Omega 50 (2015) 141–148

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Modeling the winning seed distribution of the NCAA Division I men's basketball tournament

Arash Khatibi [a,*], Douglas M. King [a], Sheldon H. Jacobson [b]

[a] Department of Industrial and Enterprise Systems Engineering, University of Illinois, Urbana, IL 61801, United States
[b] Department of Computer Science, University of Illinois, Urbana, IL 61801, United States

ARTICLE INFO

Article history:
Received 24 November 2013
Accepted 11 August 2014
This manuscript was processed by Associate Editor Rosenberger
Available online 19 August 2014

Keywords:
Exponential distribution
Markov model
NCAA basketball tournament

ABSTRACT

The National Collegiate Athletic Association's (NCAA) men's Division I college basketball tournament is an annual competition that draws widespread attention in the United States. Predicting the winner of each game is a popular activity undertaken by numerous websites, fans, and more recently, academic researchers. This paper analyzes the 29 tournaments from 1985 to 2013, and presents two models to capture the winning seed distribution (i.e., a probability distribution modeling the winners of each round). The Exponential Model uses the exponential random variable to model the waiting time between a seed's successive winnings in a round. The Markov Model uses Markov chains to estimate the winning seed distributions by considering a seed's total number of winnings in previous tournaments. The proposed models allow one to estimate the likelihoods of different seed combinations by applying the estimated winning seed distributions, which accurately summarize aggregate performance of the seeds. Moreover, the proposed models show that the winning rate of seeds is not a monotonically decreasing function of the seed number. Results of the proposed models are validated using a chi-squared goodness of fit test and compared to the frequency of observed events.

1. Introduction

Sports events attract the attention of millions of people around the world. International competitions such as the Olympic Games and the FIFA soccer World Cup draw the attention of an enormous number of people from many countries. In the United States, events like the National Collegiate Athletic Association (NCAA) Division I men's basketball tournament (hereafter referred to as the *NCAA tournament*) are the focus of significant media and fan interest. Predicting the winners of sports competitions is a popular activity by both pundits and fans. Results of games are also the subject of billions of dollars of betting each year [9]. The importance and popularity of sports predictions have inspired academic researchers to study the tournaments results and design mathematical models to estimate the games outcomes. For the NCAA tournament, methods to assign teams to regions, assess the field, and predict the winners have been widely investigated [3,7,8,11].

The NCAA tournament is an annual single-elimination tournament. The tournament's history and current structure are described by Jacobson and King [5]. The tournament's current format with 64 teams playing in six rounds began in 1985 (hereafter referred to as the *modern era*). Beginning in 2001, the number of participants increased to 65, and further increased to 68 teams in 2011. Four of these 68 teams (termed the First Four) are eliminated in play-off games before the formal start of the tournament (called round 1). This paper ignores these games and focuses on six rounds starting with 64 teams (called rounds 2, 3, 4, 5, 6, and 7). Competitions are held in four regions (typically labeled Midwest, West, South, and East), each having 16 teams that are assigned seed numbers from 1 (strongest) to 16 (weakest). Round 2 games are determined based on the seed numbers: the seed number n ($=1, 2,\ldots,8$) plays against the seed number $17-n$ in each of the four regions. Also, each seed's route to later rounds is known. For example, in each region the winners between seeds 1 and 16 and seeds 8 and 9 in round 2 play against each other in round 3. Fig. 1 shows each seed's route to later rounds in each region. Round 3 winners are collectively called the Sweet Sixteen, while the winners in round 4, which are the eight remaining teams, are called the Elite Eight. Teams appearing in round 6 are the Final Four, comprising the regional champions. This progression means that competitions continue in each region until the end of round 5, where the four regional champions are determined. These four teams play in the national semifinals (round 6) to determine the final two teams that play in the national championship game (round 7).

* Corresponding author. Tel.: +1 2177215484.
E-mail addresses: khatibi2@illinois.edu (A. Khatibi), dmking@illinois.edu (D.M. King), shj@illinois.edu (S.H. Jacobson).

http://dx.doi.org/10.1016/j.omega.2014.08.004
0305-0483/

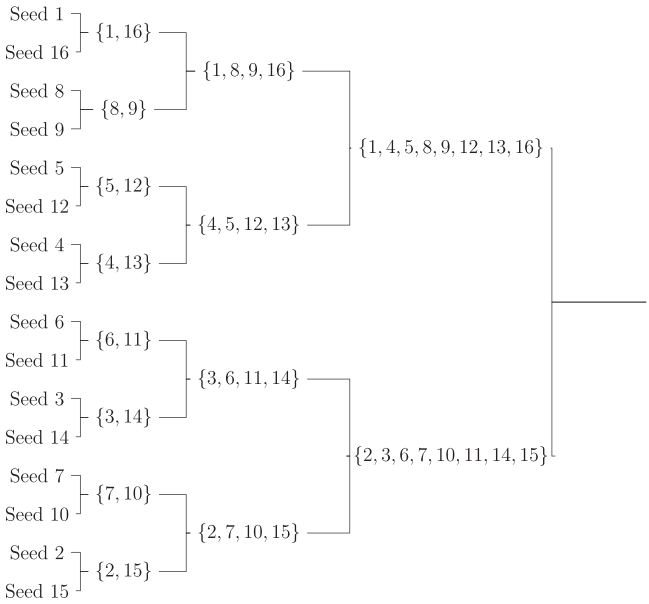

Fig. 1. Tournament bracket for a single region, including the set of seeds that can win each game.

Teams participating in the NCAA tournament are assigned a seed number based on several factors, including their seasonal performance [1]. Hence, teams' seed numbers are approximate metrics for comparing their relative strength. Teams with numerically smaller seed numbers are referred to as higher seeds. For example, in a game between seeds 3 and 8, seed 3 is the higher seed. Upsets (where the higher seeded team loses) occur frequently, which makes the task of predicting the winners difficult.

While there is considerable interest in predicting the results of individual games, the task of predicting a single winner of each game must consider the uncertainty inherent in game outcomes. Sports fans and analysts interested in making such predictions must first hold a deep understanding of these uncertainties, which they then translate into their predictions. Hence, the current study focuses on estimating probability distributions for the winning seeds in each game in each round, with the aim of providing a deeper understanding of the uncertain performances of seeds throughout the tournament. Sports fans and analysts can then apply these distributions in many different ways, such as choosing game winners randomly according to these distributions, or comparing different sets of proposed winners in a round by computing the relative likelihood of each set according to the estimated seed distributions. Therefore, the key result of the current study is not the ability of the proposed models to accurately predict individual game winners, but their ability to estimate seed distributions that accurately summarize aggregate performance of the seeds.

Jacobson et al. [6] propose a truncated geometric random variable for modeling the winning seed distributions. The potential winning seeds of each game are defined by sets for each round. For each set, each seed's winning probability defines a Bernoulli random variable. Jacobson et al. [6] consider the necessary and sufficient conditions on this sequence of Bernoulli trials such that the number of trials until the first success follows a geometric distribution [10]. Comparing their proposed method with the data for 26 tournaments (from 1985 to 2010), they conclude that the (truncated) geometric distribution provides a good fit to seeds' historical winning probabilities in later rounds, particularly round 5 and beyond. Although academic researchers have put more effort on designing mathematical models to predict the results of sports games in recent years, there are few models available for the NCAA tournament [6]. Furthermore, the existing models suffer from a lack of accuracy in predicting the results.

This paper proposes two models to estimate the winning seed distribution for each round of the NCAA tournament: the Exponential Model and the Markov Model. The Exponential Model estimates the winning seed distribution by modeling the winning frequency of each seed in a given round. It defines an exponential random variable for estimating the time between each seed's consecutive winnings in each round. The Exponential Model uses the 29 tournaments from 1985 to 2013 as the training data set for parameter estimation. Therefore, it does not consider any chance for occurrence of events that have not happened during the tournaments from 1985 to 2013. To overcome this limitation, a Markov Model is designed that uses a seed's performance in prior rounds to estimate its winning rate in later rounds. While the Exponential Model estimates the winning seed distribution in round j ($=2, 3, 4, 5, 6, 7$) by considering each seed's performance in round j, the Markov Model uses the total number of games that a seed has won in rounds 2, 3,…,j as a measure of its performance. Comparing the results with the frequency of observed events during the modern era tournaments suggests a good fit for both the Exponential and Markov Models. Moreover, it shows that the seeds' estimated winning rates in each round do not monotonically decrease as the seed number increases. Note that both the Exponential and Markov Models incorporate the potential upsets by considering the seeds performance in 29 modern era tournaments. Moreover, the parameters of the Markov model are estimated using the data of 29 modern era tournaments, which can be considered as the training data set for both proposed models.

The paper is organized as follows. Section 2 describes the Exponential Model that uses the exponential distribution for modeling the winning seed distribution in each round. Section 3 discusses the Markov Model proposed for estimating the probability of rare events. Section 4 uses the X^2 goodness of fit test to compare the estimated winning seed distribution by the Exponential and the Markov Models with the data of the 29 modern era tournaments (1985–2013) and the Geometric Model [6]. Section 5 summarizes the results and provides the concluding remarks and several future research directions.

2. The Exponential Model

The exponential distribution represents a family of continuous probability distributions. It describes the time between events in a Poisson process, a stochastic process that counts the number of random events occurred in an interval. The exponential distribution is the continuous analog of the geometric distribution in that they are memoryless and they are used to model waiting time situations. As discussed in the previous section, Jacobson et al. [6] propose a truncated geometric random variable to model the distribution of seeds that win in the last three rounds of the tournament. In this paper, the continuous counterpart of the geometric distribution is used to estimate the winning seed distribution in all rounds. The Exponential Model estimates the winning probability of a seed in a given round by modeling how often the seed wins in that round.

The average number of tournaments between consecutive events, such as the championship of a team with seed number 2, gives a measure for computing its occurrence frequency over a given number of tournaments. The Exponential Model assumes that a seed's winning in a round is a Poisson process, which occurs continuously and independently at a constant rate, and defines an exponential random variable for modeling the time between the seed's successive winnings in that round. The mean of the

exponential random variable is computed using the maximum likelihood estimation (MLE) on the results of modern era tournaments (1985–2013). Assuming an exponential random variable for modeling each seed's winning in each round, the MLE estimates the mean of the defined random variable by calculating the average interarrival time of each event (seed's winning in the given round) in the 29 modern era tournaments. If seed number i has appeared in the$j$th round n times during the previous 29 tournaments, its average interarrival time is $29/n$. Seed i's winning in the $(j-1)$th round is modeled as a Poisson process with rate $\lambda_{i,j-1}=n/29$ per tournament. Hence, an exponential random variable with mean $29/n$ is associated with seed i for computing its winning probability in round $j-1$. For example, in round 2, seed number 2 teams have played 116 times against seed number 15 teams and won 109 of these games. An exponential random variable with mean $29/109=0.266$ is associated with seed 2 teams to estimate their success rates in round 2.

Table 1 shows the seeds' winning rates in each round. Note that there are four teams of each seed playing in each tournament. Hence, the defined interarrival time is not real-time, but rather, a measure to compute the average number of teams of a given seed that advance to a specific round in each tournament. For example, seed 1's winning rate in round 2 is 4, which means that in each tournament seed number 1 teams win all their 4 games against seed 16 in round 2.

Table 2 reports seeds' winning probabilities in each round based on the Exponential Model. In a game between seeds i and j, seed i's winning probability is the probability that its associated exponential random variable, which models the waiting time to seed i's next win, is smaller than seed j's. In other words, the winner is the one whose Poisson event (winning) arrives first. For example, if $f_{i,j}$ denotes the exponential random variable associated with seed i in round j, and $\lambda_{i,j}$ represents its rate, the probability that seed 2 defeats seed 15 in round 2 is given by

$$p(f_{2,2}<f_{15,2})=\frac{\lambda_{2,2}}{\lambda_{2,2}+\lambda_{15,2}}=\frac{3.76}{3.76+0.24}=0.94 \quad (1)$$

The same method is used to estimate seeds' winning probabilities in other rounds. First, sets of possible winners are defined such that the winners in a round come from distinct sets. For example, $S=\{2,7,10,15\}$ is a set of possible winners in round 3 since only one of the four seeds $(2,7,10,15)$ can reach round 4 in each region. A seed number 2 team wins in round 3 (and advances to round 4) if its Poisson event arrives first, that is if seed 2's arrival time is the smallest. Hence, the probability that a seed 2 team wins in round 3 is given by

$$p(f_{2,3}=\min_{i\in S}(f_{i,3}))=\frac{\lambda_{2,3}}{\lambda_{2,3}+\lambda_{7,3}+\lambda_{10,3}+\lambda_{15,3}}$$
$$=\frac{2.59}{2.59+0.66+0.72+0.03}=0.65 \quad (2)$$

Randomness inherent in games' outcomes and the frequency of stronger opponents playing weaker opponents provide the opportunity for upsets to occur. The Exponential Model uses the seeds' appearance rate in each round to estimate the winning seed distribution in future tournaments, and hence, it takes potential upsets into account. For example, since seed 9 plays against seed 1 (the strongest seed) in round 3, its winning rate is less than seed 10, which plays against the winner of seeds 2 and 15 (0.72 per tournament for a seed 10 versus 0.17 per tournament for a seed 9). Hence, the estimated winning probabilities of seeds in each round incorporate the seeds' performance in previous tournaments and do not monotonically decrease as the seed number increases.

There are events with zero rate in Table 1, which represent the seeds that have not appeared in some rounds in the modern era tournaments. For example, seed 5 teams have advanced to the national final game three times. Although they have not won any of these three finals, assigning zero probability of success in round 7 to seed 5 teams (as shown in Table 2) indicates that such an outcome is impossible. However, any seed could conceivably win in any round in future tournaments. Therefore, assigning a zero success probability to seeds based on their performance history in a round, which is a result of using the maximum likelihood estimator on a small data set (29 modern era tournaments), does not reflect an accurate estimate. Note that in 2013, a seed 15 team reached round 4, and a seed 9 team reached round 6. These two events had not happened in any of the tournaments from 1985 to 2012, indicating the limitation of the Exponential Model in estimating the probability of rare events. One method to solve this problem is to estimate the mean of the defined exponential random variables using estimators such as Bayesian inference. Since this method works on the basis of the Exponential Model, a new model is proposed that incorporates a seed's winnings in prior round to estimate its winning probability in a given round. This model, referred to as the Markov Model, defines a Markov chain for each set of possible winners in a given round, and estimates a non-zero winning probability for all seeds that have won at least a single game during the 29 modern era tournaments. The next section discusses the Markov Model to estimate the rates of historically unobserved events.

Table 1
Winning rate of each seed in different rounds, $\lambda_{i,j}$ (per tournament).

Seed number (i)	Round (j)					
	2	3	4	5	6	7
1	4.00	3.48	2.76	1.62	0.93	0.62
2	3.76	2.59	1.86	0.86	0.41	0.14
3	3.41	2.07	1.03	0.48	0.31	0.14
4	3.14	1.76	0.62	0.45	0.10	0.03
5	2.59	1.34	0.28	0.21	0.10	0
6	2.66	1.34	0.45	0.10	0.07	0.03
7	2.41	0.66	0.24	0	0	0
8	1.93	0.34	0.24	0.14	0.07	0.03
9	2.07	0.17	0.07	0.03	0	0
10	1.59	0.72	0.24	0	0	0
11	1.34	0.52	0.17	0.10	0	0
12	1.41	0.69	0.03	0	0	0
13	0.86	0.21	0	0	0	0
14	0.59	0.07	0	0	0	0
15	0.24	0.03	0	0	0	0
16	0	0	0	0	0	0

Table 2
Winning probability of each seed in each round based on the Exponential Model.

Seed number (i)	Round (j)					
	2	3	4	5	6	7
1	1.00	0.87	0.69	0.41	0.47	0.62
2	0.94	0.65	0.47	0.22	0.21	0.14
3	0.85	0.52	0.26	0.12	0.16	0.14
4	0.78	0.44	0.16	0.11	0.05	0.03
5	0.65	0.34	0.07	0.05	0.05	0
6	0.66	0.34	0.11	0.03	0.03	0.03
7	0.60	0.16	0.06	0	0	0
8	0.48	0.09	0.06	0.03	0.03	0.03
9	0.52	0.04	0.02	0.01	0	0
10	0.40	0.18	0.06	0	0	0
11	0.34	0.13	0.04	0.03	0	0
12	0.35	0.17	0.01	0	0	0
13	0.22	0.05	0	0	0	0
14	0.15	0.02	0	0	0	0
15	0.06	0.01	0	0	0	0
16	0	0	0	0	0	0

3. The Markov model

This section describes a method to estimate the probability of rare events that have not happened in the modern era tournaments. The Markov Model defines a Markov chain for each set of possible winners in a given round, with nodes representing the seed numbers. Then, it computes the stationary values for the Markov chain when the transitions between nodes are defined based on the seeds' total number of wins.

The Markov Model considers seeds' performance in earlier rounds of the tournament to estimate probabilities of historically unobserved events. It defines sets of possible winners for each round. For example, there are four sets of possible winners in round 3: (1, 8, 9, 16), (2, 7, 10, 15), (3, 6, 11, 14), (4, 5, 12, 13). From each region and each set, one seed reaches round 4. Two sets of eight teams exist for round 4: (1, 4, 5, 8, 9, 12, 13, 16) and (2, 3, 6, 7, 10, 11, 14, 15). The winners of these two sets play against each other in round 5. Therefore, winners of round 5 can be any of the seeds (1, 2, 3,…,16), and the single set of round 5 (and also rounds 6 and 7) contains all 16 seeds. A Markov chain, which is represented by a directed graph, is defined for each set, with states (nodes) representing the seeds. In the Markov chain defined for round k ($=3, 4, 5, 6, 7$), there is a transition from state i to state j with weight m (i.e., an edge from node i to node j) if m is the total number of games that a seed number j team has defeated a seed number i team in rounds 2, 3,…,k. The transition weights are normalized such that the matrix representing the transition probabilities becomes a stochastic matrix (i.e., the sum of each row is equal to one). Since normalizing the winning probabilities by this method limits the stationary values to 0.5 or less (because no state can transition back to itself), it underestimates the success probability of stronger seeds. To solve this problem, a transition from state i to itself is defined whose weight w is the total number of wins of seed i in rounds 2, 3,…,k. The long-run proportion of the time that the process is in each state (seed) is the Markov chain's stationary vector or dominant eigenvector. These stationary values are the seeds' estimated winning probabilities in the respective set, and are used to compute the seed's expected winning frequency in a particular round. Since the Markov Model uses a seed's total number of wins in prior rounds to estimate its winning probability in later rounds, it estimates a non-zero winning rate for a seed if it has won at least a single game during the modern era tournaments. Note that no seed has won all its games in rounds 3–7, and hence, the dangling node problem in the field of web-page ranking does not occur (where the dangling node problem happens when a user arrives at a web-page that does not link out to another web-page, resulting in an absorbing state in the Markov chain [4]).

The Markov Model uses a seed's total number of wins to estimate its success probability. Since a seed 16 team has not won any games from 1985 to 2013, the Markov Model estimates a zero winning probability for seed 16 in all rounds. A seed 16's winning rate can be estimated by assuming that seed 16 has won a single game in round 2 of the 29 modern era tournaments and using the Markov Model. Using this method, the winning probability of seed 16 versus seed 1 in round 2 is estimated to be less than 0.01. Since the winning probability of seed 16 is even smaller in later rounds, this paper focuses on the rare events for seeds that have won at least a single game in the modern era tournaments and assumes a zero winning probability for seed 16 in all rounds.

The Markov Model is now explained in more detail. For rounds $k=2, 3, 4, 5$, there are 2^{5-k} sets of 2^{k-1} seeds. For set n in round k, there is a $2^{k-1} \times 2^{k-1}$ matrix, $M_{k,n}$. For rounds $k=6, 7$, there is a single set of 16 teams. The (i, j)th entry of the matrix is the number of games that the corresponding seed of row j defeated the corresponding seed of row i in rounds 2 through k in the 29 modern era tournaments when the sum of the row elements is normalized to one. The (i, i)th entry of the matrix is the total number of wins of the corresponding seed of row i in rounds 2 through k when the sum of the row elements is normalized to one.

Solving $\pi^{k,n} = \pi^{k,n} M_{k,n}$ with $\sum_j \pi_j^{k,n} = 1$ (where $\pi^{k,n}$ denotes the stationary values of $M_{k,n}$), and multiplying the stationary values by the total number of potential games of the given seed in round k of 29 tournaments results in the expected number of times that the given seed would advance to round $k+1$. Note that the number of potential games in rounds $k=2, 3, 4, 5$ is 116 since there are 4 regions and 29 years worth of tournaments. The number of potential games in rounds $k=6, 7$ is 58 and 29, respectively. Hence, sum of the expected number of times winning in round k is equal to $2^{5-k} \times 116$. Note also that $\sum_{j,n} \pi_j^{k,n}$ is equal to the number of seeds per region in round k.

3.1. Results of each round

Sets of possible winners in round 3 are (1, 8, 9, 16), (2, 7, 10, 15), (3, 6, 11, 14), (4, 5, 12, 13), which result in 4×4 matrices $M_{3,1}$, $M_{3,2}$, $M_{3,3}$, and $M_{3,4}$, respectively (see the Appendix). For example, the first, second, third, and the fourth rows of $M_{3,2}$ correspond to seeds 2, 7, 10, and 15, respectively. A seed 2 team has lost 17 out of 67 games to a seed 7 team, 17 out of 42 games to a seed 10 team, and 7 out of 116 games to a team of seed 15 in rounds 2 and 3. The first row entries of $M_{3,2}$ correspond to 184, 17, 17, and 7 when the sum of them is normalized to one. The same procedure produces the other rows entries.

In round 4, Markov chains are defined for the sets (1, 4, 5, 8, 9, 12, 13, 16) and (2, 3, 6, 7, 10, 11, 14, 15). For this case, sets of eight teams result in 8×8 matrices $M_{4,1}$ and $M_{4,2}$ (see the Appendix). In rounds 5, 6, and 7, a Markov chain is defined for sets of 16 teams since from round 5 it is possible for all seeds to play each other. For these rounds, the single set of 16 teams results in 16×16 matrices $M_{5,1}, M_{6,1}$, and $M_{7,1}$ (see the Appendix). The ith row of these matrices corresponds to seed number i.

Stationary values of the matrices provide the seeds' expected number of wins in each round. The stationary values and the expected number of times that each seed wins in rounds 3, 4, 5, 6, and 7 are shown in Tables 3 and 4.

3.2. The rare events

The Markov Model is designed to address the problem of the Exponential Model in estimating the frequency of rare events.

Table 3
Stationary values (π_j^k) of the seeds in rounds 3–7.

Seed (j)	Round (k)				
	3	4	5	6	7
1	0.8140	0.5783	0.2990	0.3025	0.3202
2	0.5831	0.3881	0.1906	0.1932	0.1856
3	0.4656	0.2079	0.1049	0.1163	0.1149
4	0.4068	0.1523	0.0868	0.0789	0.0781
5	0.3505	0.1027	0.0575	0.0534	0.0513
6	0.3403	0.1284	0.0613	0.0635	0.0628
7	0.2280	0.1116	0.0429	0.0416	0.0390
8	0.1015	0.0585	0.0292	0.0284	0.0296
9	0.0845	0.0455	0.0219	0.0206	0.0210
10	0.1682	0.0848	0.0337	0.0326	0.0306
11	0.1470	0.0527	0.0272	0.0274	0.0267
12	0.1685	0.0414	0.0222	0.0199	0.0191
13	0.0742	0.0213	0.0113	0.0099	0.0097
14	0.0471	0.0162	0.0074	0.0078	0.0076
15	0.0207	0.0103	0.0043	0.0041	0.0038
16	0	0	0	0	0

Table 4
Expected number of times winning for each seed in rounds 3–7 in 29 tournaments.

Seed (j)	Round (k)				
	3	4	5	6	7
1	94.42	67.08	34.68	17.54	9.28
2	67.64	45.02	22.11	11.20	5.38
3	54.01	24.11	12.17	6.75	3.33
4	47.19	17.67	10.06	4.57	2.27
5	40.66	11.91	6.69	3.10	1.49
6	39.47	14.90	7.11	3.68	1.82
7	26.45	12.95	4.98	2.41	1.13
8	11.78	6.79	3.38	1.65	0.86
9	9.80	5.28	2.53	1.19	0.61
10	19.51	9.83	3.91	1.89	0.89
11	17.06	6.12	3.15	1.59	0.77
12	19.55	4.80	2.57	1.16	0.55
13	8.60	2.47	1.32	0.59	0.28
14	5.46	1.88	0.86	0.45	0.22
15	2.40	1.19	0.49	0.24	0.11
16	0	0	0	0	0

Seeds 13, 14, 15, and 16 have not won any games in rounds 4, 5, 6, and 7 of the 29 modern era tournaments. Therefore, their winning probabilities in these rounds are zero based on the Exponential Model. However, the Markov Model uses the performance history of these seeds in previous rounds to estimate their winning rate in a given round. For example, the expected appearance of 0.49 for a seed 15 team in round 6 means that a seed 15 is expected to reach round 6 one time every 29/0.49=59 tournaments.

Seeds 7, 10, and 12 teams' winning in round 5 and beyond, seeds 9 and 11 winning in rounds 6 and 7, and seed 5's winning in round 7 are the other rare events that have not happened in the 29 modern era tournaments. The stationary values of the Markov chains estimate the frequency of rare events (Tables 3 and 4). For example, a seed 15 team is estimated to appear in the national championship game approximately one time each 121 tournaments since its winning expected frequency in round 6 is 0.24 in 29 tournaments. The average number of championships of seed 15 teams is estimated to be approximately one time in 264 tournaments. Note that the Markov Model estimates a non-zero winning rate for all seeds (except seed 16) since they have won at least a single game in the 29 modern era tournaments. Moreover, the Markov Model shows that the winning rate is not a monotone function of the seed number.

4. Validating the results

This section evaluates the Exponential and Markov Models. The X^2 goodness of fit test is used to assess how well the proposed models fit the observed values in the past 29 tournaments from 1985 to 2013. The Exponential Model estimates the winning seed distribution in the NCAA tournament in each round by defining exponential random variables for each seed. The Markov Model is designed to estimate the probability of rare events, which have not happened in the 29 modern era tournaments.

The X^2 test statistic is used as a measure to compare the performance of the Exponential and the Markov Models with the Geometric Model [6] for all rounds. The X^2 test statistic sums the square of the difference between the expected frequency of each seeds appearance in a given round and its actual appearance in the previous 29 tournaments, normalized by the actual number of appearances. Table 5 shows the X^2 test statistic and the p-values for the Exponential and the Markov Models and compares them with the Geometric Model [6] for rounds 3–7, whose results have been updated to include data from all of the 29 modern era tournaments [2]. The expected appearance frequency of each seed in different rounds is estimated using the three models and compared to the frequency of observed events in the 29 modern era tournaments. Note that the number of degrees of freedom in calculating the p-values is the total number of seeds minus the number of sets in each round.

Table 5
X^2 test statistic and p-value for different rounds.

Round	Geometric model		Exponential model		Markov model	
	X^2	p-value	X^2	p-value	X^2	p-value
3	30.35	<0.01	<0.01	>0.99	11.14	0.52
4	14.64	0.40	<0.01	>0.99	21.50	0.09
5	24.67	0.05	<0.01	>0.99	23.51	0.07
6	7.96	0.93	<0.01	>0.99	16.80	0.33
7	13.51	0.56	<0.01	>0.99	15.82	0.39

Table 6
Probability of round 6 seeding for Geometric (Geo), Exponential (Exp), and Markov models.

Year (s) occurred	Seeds	Probability			Expected waiting time		
		Geo	Exp	Markov	Geo	Exp	Markov
1985	1, 1, 2, 8	0.0055	0.0148	0.0060	181	68	167
1986	1,1, 2, 11	0.0013	0.0106	0.0056	757	95	180
1987, 1988	1, 1, 2, 6	0.0145	0.0106	0.0125	69	95	80
1989, 1998, 2003	1, 2, 3, 3	0.0233	0.0150	0.0075	43	66	133
1990	1, 3, 4, 4	0.0055	0.0074	0.0028	181	135	353
1991, 2001, 2009	1, 1, 2, 3	0.0625	0.0508	0.0214	16	20	47
1992	1, 2, 4, 6	0.0069	0.0059	0.0073	145	170	137
1993	1, 1, 1, 2	0.0526	0.0571	0.0204	19	18	49
1994, 2004	1, 2, 2, 3	0.0385	0.0270	0.0137	26	37	73
1995, 2012	1, 2, 2, 4	0.0233	0.0253	0.0113	43	40	88
1996, 2005	1, 1, 4, 5	0.0089	0.0116	0.0054	112	86	187
1997, 1999	1, 1, 1, 4	0.0200	0.0300	0.0093	50	33	108
2000	1, 5, 8, 8	0.0001	0.0003	0.0002	21180	3199	5685
2002	1, 1, 2, 5	0.0233	0.0222	0.0118	43	45	85
2006	2, 3, 4, 11	0.0002	0.0017	0.0011	4089	574	883
2007	1, 1, 2, 2	0.0500	0.0455	0.0195	20	22	51
2008	1, 1, 1, 1	0.0208	0.0269	0.0080	48	37	125
2010	1, 2, 5, 5	0.0034	0.0029	0.0023	292	347	442
2011	3, 4, 8, 11	0.0001	0.0003	0.0002	70990	3527	5762
2013	1, 4, 4, 9	0.0003	0.0005	0.0006	3159	2168	1689

Since the same set of data is used to both estimate parameters and evaluate how well each model fits the observed data, a model with a large number of parameters will tend to exhibit a stronger fit with observed data than a model with fewer parameters. For example, the Exponential Model defines one exponential random variable (with its own corresponding rate parameter) for each seed in each round, producing a strong fit to the observed data and leading to p-values higher than 0.99 for all rounds. The Markov Model shows a stronger fit than the Geometric Model in rounds 3 and 5. While the Geometric Model does not show a good fit in round 3, the p-values of the Markov Model show that it provides a good fit to the results of all rounds. The applicability of the Markov Model in all rounds is its main strength compared to the Geometric Model.

Table 6 reports the occurrence probability of different seed combinations in round 6 for the Geometric, Exponential and Markov Models. It also shows the expected waiting time (measured in number of tournaments) for each of these seed combinations to happen. This table includes all round 6 seed combinations that occurred in the 29 modern era tournaments. For example, the seed combination (1, 1, 2, 3) has reached round 6 three times in the modern era tournaments. Therefore, its expected frequency is

Table 7
Number of 1 seed teams in round 6 for Geometric (Geo), Exponential (Exp), and Markov models.

Scenario	Probability			Expected occurrence			Times occurred
	Geo	Exp	Markov	Geo	Exp	Markov	
Exactly zero	0.158	0.121	0.241	4.6	3.5	7.0	2
Exactly one	0.370	0.337	0.412	10.7	9.8	11.9	12
Exactly two	0.326	0.351	0.264	9.4	10.2	7.6	11
Exactly three	0.127	0.163	0.075	3.7	4.7	2.2	3
Exactly four	0.019	0.028	0.008	0.5	0.8	0.2	1

computed by multiplying the winning probabilities of seeds 1, 1, 2, and 3 in round 5 and multiplying the result by the number of possible permutations of these seeds. This results in an expected occurrence frequency of once every 20 tournaments by the Exponential Model and once every 47 tournaments by the Markov Model. The minimum occurrence probability among the sets that have occurred during the 29 modern era tournaments belongs to the seed combination (3, 4, 8, 11) that happened in 2011.

Table 7 shows the probability and the expected number of seed number 1 teams in round 6 estimated by the Geometric Model (Geo), Exponential Model (Exp), and the Markov Model. The most probable combination estimated by the Markov Model is one seed number 1 team reaching round 6, which occurred 12 times in the 29 modern era tournaments. The Geometric, Exponential, and Markov models estimate the expected frequency of exactly one seed number 1 team in round 6 to be 10.7, 9.8, and 11.9 in 29 tournaments, respectively. Compared to the Exponential Model, the Markov Model underestimates the performance of stronger seeds. As shown in the table, the Markov Model estimates larger expected occurrence for a Final Four combination including no seed number 1 teams and smaller expected occurrence for a Final Four including four seed 1 teams (compared to the other two models). The intuitive explanation is that the Markov Model decreases the estimated winning probabilities of stronger seeds to compute a non-zero value for rare events (since the seeds' winning probabilities in a set sum to one). The Markov Model defines the links weights based on a seed's total number of wins in rounds up to the given round. Since the number of games in later rounds tends to be smaller, the (stronger seeds') smaller number of wins in later rounds is dominated by larger number of wins in the beginning rounds. Therefore, the estimated winning rates of stronger seeds by the Markov Model tend to be smaller than the Exponential Model. Note that the X^2 goodness of fit test statistic is 2.53, 1.86, 8.58 for the Geometric, Exponential, and Markov Models, respectively. The corresponding p-values are 0.64, 0.76, and 0.07 for the Geometric, Exponential, and Markov Models, respectively.

4.1. 2014 tournament

The results of tournaments from 1985 to 2013 are used as the training data set for the Exponential and Markov Models to estimate the seeds performance in the 2014 tournament. The Final Four seeds in the 2014 tournament are (1, 2, 7, 8) and a seed 7 won the championship for the first time in the modern era tournaments. This Final Four seed combination and the championship of a seed 7 are rare events whose probabilities are estimated to be zero by the Exponential Model. The probability of a Final Four seed combination of (1, 2, 7, 8) is 0.0007 and 0.0016 by the Geometric and Markov Models, respectively. The Geometric Model estimates a probability of 0.0065 for the championship of a seed 7, which is expected to occur once each 153 tournaments. The Markov Model estimates a winning probability of 0.039 in the final game for seed 7, which is expected to occur once every 25 tournaments. Hence, the Markov Model estimates a larger probability than the other two models for these events happened in 2014.

5. Conclusions and future work

Significant media attention and widespread international interest in sports events and predicting their results have motivated academic researchers to study sport analytics and propose mathematical models to estimate the games outcomes. This paper analyzes the NCAA Division I men's basketball tournament and proposes two models to estimate the winning seed distribution in each round. The Exponential Model assumes that a seed's winnings in a round is a Poisson process and defines an exponential random variable to model the time between each seed's consecutive winnings in each round. The winning probability of a seed in a given round is estimated by comparing its associated exponential random variable with other seeds' that it may confront in that round.

Since the Exponential Model uses maximum likelihood to estimate the mean of the exponential random variable, it assumes no chance for events that have not happened in the modern era tournaments (referred to as rare events). The Markov Model estimates the probability of rare events by defining a Markov chain for each set of distinct possible winners in a round and computing the stationary values of the Markov chain. Both the Exponential and Markov models estimate the potential upsets by using seeds' performance history in modern era tournaments to compute the likelihood of an event. They show that the winning frequency of different seeds in each round is not a monotone decreasing function of the seed number.

A X^2 goodness of fit test statistic is used to validate the models. While the Exponential Model shows very strong fit to the frequency of observed events, it cannot predict historically unobserved events. The Markov Model estimates a seed's winning probability in a given round by using the seed's total number of winnings in rounds up to that round and is capable of estimating rare events. The proposed models can be used by sports fans to assess the seeds' performance in each round and compute the likelihood of different seed combinations. They can also be used for other single-elimination competitions with changes to the definition of a seed.

There are several future research directions. Estimating the models' parameters of each region based on the results of its own games can improve the accuracy of the results. Moreover, since the most recent tournaments are better indicators of teams strength, larger weights can be assigned to the results of the most recent tournaments in parameter estimation. Finally, incorporating other features such as the games venues may also improve the results.

Acknowledgement

The computational results were obtained with support from the Simulation and Optimization Laboratory in the Department of Computer Science at the University of Illinois at Urbana-Champaign. This material is based upon work supported in part by (while the third author served at) the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Government, or the National Science Foundation.

Appendix

Matrices $M_{3,1}$, $M_{3,2}$, $M_{3,3}$, $M_{3,4}$ are the transition matrices of round 3. Matrices $M_{4,1}$ and $M_{4,2}$ are the transition matrices of round 4, and matrices $M_{5,1}$, $M_{6,1}$, and $M_{7,1}$ are the transition matrices of rounds 5, 6, and 7, respectively. Note that the results are rounded to four digits after the decimal point for round 3, and to three digits after the decimal point for the other rounds:

$$M_{3,1} = \begin{array}{c} \text{seed} \\ 1 \\ 8 \\ 9 \\ 16 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 8 & 9 & 16 \end{array} \\ \begin{pmatrix} 0.9342 & 0.0439 & 0.0219 & 0 \\ 0.2674 & 0.3837 & 0.3488 & 0 \\ 0.3125 & 0.3182 & 0.3693 & 0 \\ 1.0000 & 0 & 0 & 0 \end{pmatrix} \end{array}, \quad M_{3,2} = \begin{array}{c} \text{seed} \\ 2 \\ 7 \\ 10 \\ 15 \end{array} \begin{array}{c} \begin{array}{cccc} 2 & 7 & 10 & 15 \end{array} \\ \begin{pmatrix} 0.8178 & 0.0756 & 0.0756 & 0.0311 \\ 0.2688 & 0.4785 & 0.2473 & 0.0054 \\ 0.1582 & 0.4430 & 0.3987 & 0 \\ 0.8862 & 0.0163 & 0.0325 & 0.0650 \end{pmatrix} \end{array},$$

$$M_{3,3} = \begin{array}{c} \text{seed} \\ 3 \\ 6 \\ 11 \\ 14 \end{array} \begin{array}{c} \begin{array}{cccc} 3 & 6 & 11 & 14 \end{array} \\ \begin{pmatrix} 0.7395 & 0.1256 & 0.0558 & 0.0791 \\ 0.1865 & 0.6010 & 0.2021 & 0.0104 \\ 0.1548 & 0.4968 & 0.3484 & 0 \\ 0.7444 & 0.0902 & 0.0226 & 0.1429 \end{pmatrix} \end{array}, \quad M_{3,4} = \begin{array}{c} \text{seed} \\ 4 \\ 5 \\ 12 \\ 13 \end{array} \begin{array}{c} \begin{array}{cccc} 4 & 5 & 12 & 13 \end{array} \\ \begin{pmatrix} 0.6860 & 0.1353 & 0.0580 & 0.1208 \\ 0.1728 & 0.5969 & 0.2147 & 0.0157 \\ 0.1146 & 0.4777 & 0.3885 & 0.0191 \\ 0.6454 & 0.0780 & 0.0567 & 0.2199 \end{pmatrix} \end{array}.$$

$$M_{4,1} = \begin{array}{c} \text{seed} \\ 1 \\ 4 \\ 5 \\ 8 \\ 9 \\ 12 \\ 13 \\ 16 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 4 & 5 & 8 & 9 & 12 & 13 & 16 \end{array} \\ \begin{pmatrix} 0.891 & 0.043 & 0.021 & 0.030 & 0.015 & 0 & 0 & 0 \\ 0.112 & 0.620 & 0.108 & 0.015 & 0 & 0.046 & 0.097 & 0 \\ 0.122 & 0.143 & 0.530 & 0.009 & 0.004 & 0.178 & 0.013 & 0 \\ 0.253 & 0.011 & 0 & 0.401 & 0.330 & 0.005 & 0 & 0 \\ 0.304 & 0.011 & 0.005 & 0.309 & 0.370 & 0 & 0 & 0 \\ 0.107 & 0.102 & 0.424 & 0 & 0 & 0.350 & 0.017 & 0 \\ 0.027 & 0.619 & 0.075 & 0.007 & 0.007 & 0.054 & 0.211 & 0 \\ 1.000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

$$M_{4,2} = \begin{array}{c} \text{seed} \\ 2 \\ 3 \\ 6 \\ 7 \\ 10 \\ 11 \\ 14 \\ 15 \end{array} \begin{array}{c} \begin{array}{cccccccc} 2 & 3 & 6 & 7 & 10 & 11 & 14 & 15 \end{array} \\ \begin{pmatrix} 0.793 & 0.047 & 0.020 & 0.057 & 0.057 & 0.003 & 0 & 0.023 \\ 0.084 & 0.687 & 0.098 & 0.011 & 0.014 & 0.044 & 0.062 & 0 \\ 0.090 & 0.155 & 0.556 & 0.013 & 0.009 & 0.168 & 0.009 & 0 \\ 0.244 & 0.029 & 0.015 & 0.468 & 0.224 & 0.015 & 0 & 0.005 \\ 0.140 & 0.050 & 0.022 & 0.391 & 0.391 & 0.006 & 0 & 0 \\ 0.059 & 0.141 & 0.453 & 0 & 0 & 0.347 & 0 & 0 \\ 0 & 0.733 & 0.089 & 0.007 & 0.007 & 0.022 & 0.141 & 0 \\ 0.879 & 0.008 & 0 & 0.016 & 0.032 & 0 & 0 & 0.064 \end{pmatrix} \end{array}$$

$$M_{5,1} = \begin{array}{c} \text{seed} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \end{array} \\ \begin{pmatrix} 0.833 & 0.048 & 0.019 & 0.034 & 0.017 & 0.005 & 0 & 0.024 & 0.012 & 0 & 0.007 & 0 & 0 & 0 & 0 & 0 \\ 0.054 & 0.745 & 0.037 & 0.011 & 0.008 & 0.017 & 0.048 & 0.006 & 0.003 & 0.048 & 0.003 & 0 & 0 & 0 & 0.020 & 0 \\ 0.039 & 0.076 & 0.664 & 0.010 & 0.003 & 0.089 & 0.010 & 0 & 0 & 0.013 & 0.039 & 0 & 0 & 0.056 & 0 & 0 \\ 0.105 & 0.007 & 0.007 & 0.627 & 0.101 & 0.004 & 0 & 0.014 & 0 & 0 & 0 & 0.043 & 0.091 & 0 & 0 & 0 \\ 0.118 & 0 & 0.008 & 0.139 & 0.538 & 0 & 0 & 0.008 & 0.004 & 0 & 0 & 0.172 & 0.013 & 0 & 0 & 0 \\ 0.024 & 0.086 & 0.147 & 0.008 & 0.004 & 0.539 & 0.012 & 0.004 & 0 & 0.008 & 0.159 & 0 & 0 & 0.008 & 0 & 0 \\ 0.019 & 0.236 & 0.028 & 0.009 & 0 & 0.014 & 0.453 & 0.005 & 0 & 0.217 & 0.014 & 0 & 0 & 0 & 0.005 & 0 \\ 0.243 & 0.011 & 0.005 & 0.011 & 0 & 0 & 0 & 0.407 & 0.317 & 0 & 0 & 0.005 & 0 & 0 & 0 & 0 \\ 0.300 & 0 & 0.005 & 0.011 & 0.005 & 0 & 0 & 0.306 & 0.372 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.021 & 0.132 & 0.047 & 0.010 & 0.005 & 0.021 & 0.368 & 0 & 0 & 0.389 & 0.005 & 0 & 0 & 0 & 0 & 0 \\ 0.011 & 0.057 & 0.137 & 0 & 0 & 0.440 & 0 & 0 & 0 & 0 & 0.354 & 0 & 0 & 0 & 0 & 0 \\ 0.107 & 0.006 & 0 & 0.101 & 0.421 & 0 & 0 & 0 & 0 & 0 & 0 & 0.348 & 0.017 & 0 & 0 & 0 \\ 0.027 & 0 & 0 & 0.619 & 0.075 & 0 & 0 & 0.007 & 0.007 & 0 & 0 & 0.054 & 0.211 & 0 & 0 & 0 \\ 0 & 0 & 0.733 & 0 & 0 & 0.089 & 0.007 & 0 & 0 & 0.007 & 0.022 & 0 & 0 & 0.141 & 0 & 0 \\ 0 & 0.879 & 0.008 & 0 & 0 & 0 & 0.016 & 0 & 0 & 0.032 & 0 & 0 & 0 & 0 & 0.064 & 0 \\ 1.000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$M_{6,1} =$

seed	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.826	0.053	0.027	0.033	0.016	0.004	0	0.022	0.011	0	0.007	0	0	0	0	0
2	0.066	0.731	0.040	0.011	0.011	0.019	0.045	0.008	0.003	0.045	0.003	0	0	0	0.019	0
3	0.044	0.082	0.663	0.009	0.003	0.085	0.009	0	0	0.013	0.038	0	0	0.053	0	0
4	0.118	0.010	0.014	0.611	0.097	0.007	0	0.014	0	0	0	0.042	0.087	0	0	0
5	0.119	0	0.008	0.140	0.539	0	0	0.008	0.004	0	0	0.169	0.012	0	0	0
6	0.024	0.089	0.145	0.008	0.004	0.540	0.012	0.004	0	0.008	0.157	0	0	0.008	0	0
7	0.019	0.236	0.028	0.009	0	0.014	0.459	0.005	0	0.217	0.014	0	0	0	0.005	0
8	0.243	0.010	0.005	0.010	0.005	0	0	0.409	0.311	0	0	0.005	0	0	0	0
9	0.304	0	0.005	0.011	0.005	0	0	0.304	0.370	0	0	0	0	0	0	0
10	0.021	0.132	0.047	0.010	0.005	0.021	0.368	0	0	0.389	0.005	0	0	0	0	0
11	0.011	0.062	0.140	0	0	0.433	0	0.006	0	0	0.348	0	0	0	0	0
12	0.107	0.006	0	0.101	0.421	0	0	0	0	0	0	0.348	0.017	0	0	0
13	0.027	0	0	0.619	0.075	0	0	0.007	0.007	0	0	0.054	0.211	0	0	0
14	0	0	0.733	0	0	0.089	0.007	0	0	0.007	0.022	0	0	0.141	0	0
15	0	0.879	0.008	0	0	0	0.016	0	0	0.032	0	0	0	0	0.064	0
16	1.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$M_{7,1} =$

seed	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.826	0.053	0.025	0.034	0.015	0.006	0	0.023	0.011	0	0.006	0	0	0	0	0
2	0.080	0.719	0.044	0.010	0.010	0.018	0.044	0.008	0.003	0.044	0.003	0	0	0	0.018	0
3	0.046	0.089	0.659	0.009	0.003	0.083	0.009	0	0	0.013	0.037	0	0	0.052	0	0
4	0.124	0.010	0.014	0.608	0.096	0.007	0	0.014	0	0	0	0.041	0.086	0	0	0
5	0.130	0	0.008	0.138	0.532	0	0	0.008	0.004	0	0	0.167	0.012	0	0	0
6	0.028	0.088	0.144	0.008	0.004	0.540	0.012	0.004	0	0.008	0.156	0	0	0.008	0	0
7	0.019	0.236	0.028	0.009	0	0.014	0.453	0.005	0	0.217	0.014	0	0	0	0.005	0
8	0.241	0.010	0.010	0.010	0.005	0	0	0.410	0.308	0	0	0.005	0	0	0	0
9	0.304	0	0.005	0.011	0.005	0	0	0.304	0.370	0	0	0	0	0	0	0
10	0.021	0.132	0.047	0.010	0.005	0.021	0.368	0	0	0.389	0.005	0	0	0	0	0
11	0.011	0.062	0.140	0	0	0.433	0	0.006	0	0	0.348	0	0	0	0	0
12	0.107	0.006	0	0.101	0.421	0	0	0	0	0	0	0.348	0.017	0	0	0
13	0.027	0	0	0.619	0.075	0	0	0.007	0.007	0	0	0.054	0.211	0	0	0
14	0	0	0.733	0	0	0.089	0.007	0	0	0.007	0.022	0	0	0.141	0	0
15	0	0.879	0.008	0	0	0	0.016	0	0	0.032	0	0	0	0	0.064	0
16	1.000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

References

[1] Baumann R, Matheson VA, Howe CA. Anomalies in tournament design: the madness of March Madness. Quantitative Analysis in Sports 2010;6(2) (Article 4).

[2] BracketOdds. Seed distributions for March Madness 2013: A tool for bracketologists; 2014 (Last accessed on 21.04.14).

[3] Fearnhead P, Taylor BM. Calculating strength of schedule, and choosing teams for March Madness. American Statistician 2010;64(2):108–15.

[4] Ipsen ICF, Selee TM. Pagerank computation, with special attention to dangling nodes. SIAM Journal on Matrix Analysis and Applications 2007;29(4):1281–96.

[5] Jacobson S, King DM. Seeding in the NCAA men's basketball tournament: when is a higher seed better?. Gambling Business and Economics 2009;3(2):63–87.

[6] Jacobson SH, Nikolaev AG, King DM, Lee AJ. Seed distributions for the NCAA men's basketball tournament. Omega 2011;39(6):719–24.

[7] Koenker R, Bassett JGW. March Madness quantile regression bracketology, and the Hayek Hypothesis. Business and Economic Statistics 2010;28(1):26–35.

[8] Kvam P, Sokol J. A logistic regression/Markov chain model for NCAA basketball. Naval Research Logistics 2006;53(8):788–803.

[9] Metrick A. March Madness? Strategic behavior in NCAA basketball tournament betting pools Economic Behavior and Organization 1995;30:159–72.

[10] Shishebor Z, Towhidi M. On the generalization of negative binomial distribution. Statistics and Probability Letters 2004;66:127–33.

[11] Smith JC, Fraticelli BMP, Rainwater C. A bracket assignment problem for the NCAA men's basketball tournament. International Transaction in Operational Research 2006;13(3):253–71.