Research Proposal

Finding schemes for arranging large tournaments based on probability models

Luhuan Wu

April 12, 2017

Motivation:

The study of probability is more fun when the concepts are incorporated into real-world examples that are of particular interests. Sports events and tournament competitions provide excellent opportunities for probability modelling. A question that interests me is "How might one arrange a tournament to choose a winner out of (say) 200 players, given a constraint on the total number of matches to be played? " (Aldous, 2017). Many factors would be taken into account in real life, such as entertainment, money, and making the top-seeded competitors most likely to win. However, in a probability research, we would try to simplify the situation by focusing on mathematical interpretations. At last, we might try to apply the scheme we would develop for large tournaments in general to certain real sports event for more fun. ("Bracketology" of NCCA has always been a trending topic!)

Method:

Main idea: The winning probability can be considered as a function of the seed and the strength, and we want the top-seeded players have more matches.

The research can be divided into stages as listed below:

- Stage 0: research on existing schemes.

  There some existing models that can be taken into consideration. For example, the most commonly used one are knock-out tournaments that have single-eliminations and double-eliminations. There are others like round-robin tournaments which are usually considered invalid given the limitation in total numbers of matches. Also, the typical designs in real-world tournaments offer a bunch of concrete solutions like bracketing in March Madness (NCAA), promotion and relegation in EPL, etc. I have collected a few of schemes in **Starting Resources** in the last part of the proposal. The resources also offer the methodology for modelling and simulation.

- Stage 1 : consider the different cases based on the information about the seed and strength prior to the tournament.

- Case 1, we have no information about the strength distribution of all teams. In this case, we can assume the strength distribution is normal distribution or other distributions.

- Case 2, we have complete strength distribution.

- Case 3, which is the tricky one that we only have partial information about the true strength of all teams. For example, the pairwise winning probability is known. Or we only know the relative rankings of competitors.

Information on strength can be used to calculate the seed order. In other cases, we might be only provided the seed order, which can be given by the experts based on various rating systems.

- Stage 2: develop frames and build probability models

  First we need to **define a criteria for "optimal scheme"**. For pairwise matches, we hope that the competitor with higher seed has more winning probability. For the whole tournament, we have similar expectation that more matches are going on between stronger competitors.

  For simplicity, we could first consider the case that the total number of the teams M are 2^N, then consider the sundry issues such as M = 2^N + n.

- Stage 3, evaluate the models under different settings via numerical simulations, and compare the results to other existing schemes.

- Stage 4(possibly/for fun), apply to a specific real world sports event.

Other idea:

The common theme in previous work on designing tournaments is based on the information about seeding or strength ranking given prior to the games. Could we develop our schemes incorporating the updating information from the games played? Consider the case when no information about the seed and strength is given before the tournaments. We could either set up preliminary competitions that offer the qualifiers, or divide the teams into several leagues and adopt the promotion and relegation scheme. Which one is more efficient to determine seeding? Can we create a dynamic winning-probability model to predict the winner?

In addition, the constraint in total number of matches is not mentioned in many existing literatures. Adding this constraint requires some optimization methods besides the probability knowledges.

Starting resources

1. More Probability Models for the NCAA Regional Basketball Tournaments
   http://sites.stat.psu.edu/~ajw13/stat497/ProbabilityModels_Schwertman_1996.pdf
2. Seed Distributions for the NCAA Men's Basketball Tournament: Why it May Not Matter Who Plays Whom*
   http://www.nessis.org/nessis11/jacobson.pdf\
3. AN ILLUSTRATION OF SOME BASIC PROBABILITY CONCEPTS: DETERMINING PROBABILITIES OF WINNING IN SINGLE ELIMINA TION TOURNAMENTS
   https://www.barton.edu/pdf/faculty-publications/bengtson-winning-probabilities-publication.pdf
4. A Simulation Model for Football Championships
   https://pdfs.semanticscholar.org/e4f2/3190679bb3c74776edb5f9304c55ba11f332.pdf
5. Bayesian locally-optimal design of knockout tournaments
   http://www.glicko.net/research/knockout.pdf
6. A Model for Intransitive Preferences
   http://www.aaai.org/ocs/index.php/WS/AAAIW14/paper/viewFile/8839/8287
7. Predicting Results of March Madness Using the Probability Self-Consistent Method
   http://article.sapub.org/10.5923.j.sports.20150504.04.html
8. Improved NCAA Basketball Tournament Modeling via Point Spread and Team Strength Information
   http://www.sph.umn.edu/faculty1/wp-content/uploads/2012/11/rr94-004.pdf