Replicating results of 2016 March Madness Tournament

Author: Bryan Katterman (modified from Layne Bradshaw proposal)

Problem:

Relating the outcomes of sports with games of chance still remains a complicated area of study in probability. The current pinnacle of study is the NCAA basketball tournament. Consisting of 63 games, if relegated to just pure random chance, the odds of correctly predicting the outcome of all 63 games is 1 in 9,223,372,036,854,775,808 (over nine quintillion). However, unlike the classical predictive models of chance which are based upon fairness (equal odds of outcome, example flip of a coin has equal chance of heads or tails providing coin is deemed fair), the outcome in sports matches almost never has fair odds; this brings both simplicity and complexity in creating an accurate predictive model. For factors of simplicity, as the match outcomes are rarely fare, we can apply weighting to each match which helps to narrow a successful guess of the outcome to better than chance odds in most cases. With some smart weight algorithms in place, some sports prediction groups claim to have narrowed their predictive odds to somewhere around a trillion to one for successfully predicting the outcome of the 63 game tournament (one company claims they have reduced it to about 4.3 billion to 1). This is an impressively significant improvement over pure chance odds, but still far from something anyone would consider a successful predictive model. The complexity in a weighted model is that there are an enormous number of factors that go into the weighting factors in predicting a winner. They are so complex that huge monetary prizes have been offered to anyone who can correctly guess the outcome of the tournament, or come close to it (Warren Buffet's 2014 $1-billion-dollar bracket prize being the most famous along with various prizes offered by Quicken Loans and other companies).

Approach:

In constructing the tournament prediction model, it is necessary to gather sports predictive weighting standards used for determining winning odds, 2016 tournament team statistics (prior to the start of the tournament) as well as prior tournament statistics. From there, create a stochastic model (with Monte Carlo methods) successfully implements all metrics and weighting factors, then run a large number of simulations of the tournament and compare the outcomes with the actual 2016 tournament results, as well as the model's odds compared with claimed industry standards.

Objectives:

- Gather statistics on all 64 teams for 2016 prior to start of the tournament, along with all prior year tournament statistics (seed matching outcomes, ave. injury / player loss prevalence, etc.)

    -- Research any available industry weighting standards used in establishing winning odds

- Create algorithm to simulate tournament and output results into usable statistical database- Successfully implement algorithm to simulate entire tournament

- See how many simulations on average are required to replicate actual 2016 outcome and compare to industry claimed statistics