Odds of a Perfect Bracket

Dr. Mike Johnson
Professor of Electrical and Computer Engineering, Marquette University

What are the odds of a perfect NCAA Tournament Bracket?

The short answer is that it's ***impossible to compute exactly***. The number of variables is almost countless, including seeds, win/loss records, injuries, tournament location, game time, weather, and so on. However, the odds can certainly be estimated, and we can determine an upper and lower bound. The accuracy of any estimate is determined by what variables it considers and how well the estimate accounts for those variables.

Below several possible mathematical models for estimating the odds of a perfect bracket are considered and compared.

Summary chart

Every game a coin flip: $\boldsymbol{P_{perfect}} \approx \mathbf{1\ in\ 9\ quintillion}$

Estimate odds from seed differential: $\boldsymbol{P_{perfect}} \approx \mathbf{1\ in\ 660\ billion}$

Estimate odds with historical data: $\boldsymbol{P_{perfect}} \approx \mathbf{1\ in\ 55\ billion}$

A person who can beat the odds by 5%: $\boldsymbol{P_{perfect}} \approx \mathbf{1\ in\ 3\ billion}$

A person who can beat the odds by 10%: $\boldsymbol{P_{perfect}} \approx \mathbf{1\ in\ 200\ million}$

A zero-variable model: the coin-flip

As a starting point, we can consider the mathematical model in which there is no information about what team is more likely to win any of the games. In such a case, each game can be considered a "coin flip" with a 50% probability of either team winning.

In the 64-team NCAA tournament, there are exactly 63 games (32, 16, 8, 4, 2, and 1, in each round respectively). Thus the odds of a bracket (in fact any bracket, since all brackets have the same chance) of being correct can be easily calculated as

$$(\mathbf{0.5})^{\mathbf{63}} \approx \mathbf{1\ in\ 9{,}000{,}000{,}000{,}000{,}000{,}000,\ (1\ in\ 9\ quintillion)}$$

A one-variable model: seed differential

If we were to consider any one variable, the one which likely has the most influence on the outcome would be the difference between the seeds of the two teams playing. There are several models that could be used to estimate the outcome probability for each game based on the difference in seeding between teams. We can use historical results directly, or even combine these to try and get a more accurate estimate. For example, by reviewing first round results and plotting the seed differential versus probability of winning, we find that the relationship is linear, as seen below

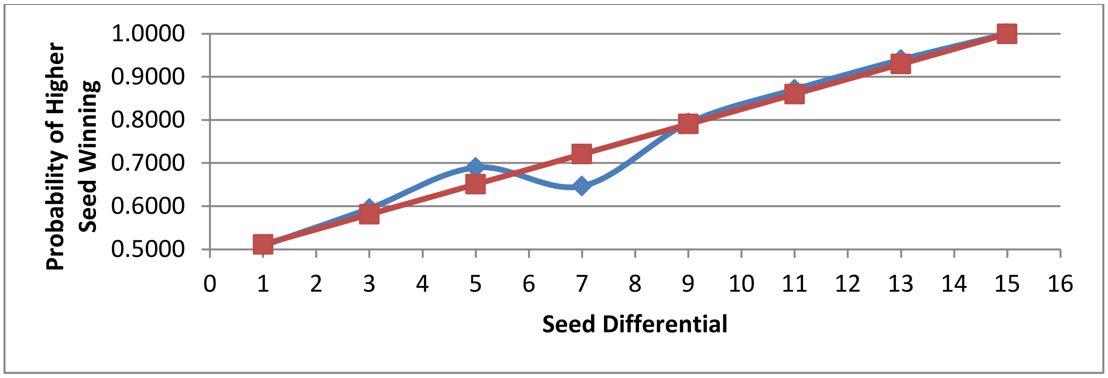

Figure 1. First-round historical probability of the higher seed winning versus the seed differential (blue) and linear approximation of estimated probability (red)

Based on this observation, we can use the above linear approximation to estimate the probability of the higher seed winning for any matchup. The odds of a perfect bracket using this approach would be

$$\prod_{game\ i=1}^{63} \boldsymbol{Pwin_i}$$

where *Pwin*$_i$ is the probability of winning each game in the bracket based solely on the seed differential. Since that is different depending on the specific bracket picks, it doesn't have a fixed value. For example, the highest probability bracket in this approach is a bracket that always picks the highest seed to win. In that case we get

$$\boldsymbol{P_{perfect} \approx 1 \text{ in } 660 \text{ billion}}$$

By comparison, if we deliberately always pick the lower seed to win, we get the probability of a perfect bracket

$$\boldsymbol{P_{perfect} \approx 1 \text{ in } (40 \times 10^{40})}$$

<u>A two-variable model: individual seed of each team</u>

The single-variable model above has some clear weaknesses. For example, it cannot tell the difference between a 1-seed playing a 4-seed, a 3-seed playing a 6-seed, or a 7-seed playing a 10-seed, since each has the same difference. If, as seems likely, these are actually different, we need to take into account both seeds. For many game situations this is not difficult, since certain seed matchups occur often. However, there are many matchups that rarely or ever occur, so there is insufficient historical data to estimate the probability of the higher seed winning. (In fact, of the 136 possible seed matchups, fully 62 of them have *never occurred at all*.)

In order to use both seeds and be able to predict the probability of winning for any possible game, we need to select a mathematical model that can handle having limited data. An example of such a model is the "nearest neighbor" model, which uses historical data that is "close" to the desired variables. A good example of how this works would be a 1-seed against a 14-seed or a 1-seed against a 15-seed. Neither of these have ever happened, so we have no data; however, 1-seeds are 19-0 against 12 seeds, 4-0 against 13 seeds, and 116-0 against 16 seeds, so we can estimate the missing probability be averaging these results, giving us the reasonable answer that the probability of the 1-seed winning these matchups is extremely high, and likely near 100%.

We can easily determine the best possible odds using historical data, since again the most likely bracket is the one that always picks the higher seed to win, and we have plenty of historical data to determine those probabilities. Using this approach we find that

$$\boldsymbol{P_{perfect} \approx \text{1 in 55 billion}}$$

A many-variable model: individual and computer-based game analysis, human prediction

If we want to consider many possible variables, a much more sophisticated approach is required. There are computer programs that can consider which teams are playing, win-loss records, injuries, and current trends, and assist human analysts in making predictions and determining the odds. (Certainly sports betting organizations are quite good at determining the odds of winning.) The actual odds of getting a perfect bracket if you consider this information varies widely depending on which teams are playing and how evenly matched the games are.

As an upper limit, though, consider that even a true sports expert could only beat the odds by a few percentage points. To imagine the chances of a perfect bracket for such an expert, we can take the historical data and increase the probabilities of the lower seed winning by a small amount to see how it impacts our chances. Doing this, we find that

Beat the odds by 5%: $\boldsymbol{P_{perfect} \approx \text{1 in 3 billion}}$

Beat the odds by 10%: $\boldsymbol{P_{perfect} \approx \text{1 in 200 million}}$

Beat the odds by 20%: $\boldsymbol{P_{perfect} \approx \text{1 in 2 million}}$

Odds of a Perfect Bracket

Appendix: Summary of historical data for the lower seed winning.

Reference web site: http://mcubed.net/ncaab/seeds.shtml

Probability of lower seed winning

1 vs. 16	1.00
1 vs. 8	0.81
1 vs. 4	0.67
1 vs. 2	0.55
2 vs. 15	0.94
2 vs. 7	0.74
2 vs. 3	0.61
3 vs. 14	0.85
3 vs. 6	0.54
4 vs. 13	0.78
4 vs. 5	0.55
5 vs. 12	0.68
6 vs. 11	0.67
7 vs. 10	0.60
8 vs 9	0.51

Probability of the most likely bracket, with all lower seeds winning is

In each of four regions:

1^{st} round: (1 d 16) x (2 d 15) x (3 d 14) x (4 d 13) x (5 d 12) x (6 d 11) x (7 d 10) x (8 d 9)
= 1 x .94 x .85 x .78 x .68 x .67 x .60 x .51 = .0869

2^{nd} round: (1 d 8) x (2 d 7) x (3 d 6) x (4 d 5)
= .81 x .74 x .54 x .55 =.178

3^{rd} round: (1 d 4) x (2 d 3)
= .67 x .61 = .41

4^{th} round: (1 d 2) = 0.55

Total: .0869 x .178 x .41 x .55 = .003488

For all four regions: .003488 x .003488 x.003488 x.003488 = 0.000000000148 (=$1.48 \text{ x } 10^{-10}$)

Semifinals: (1 d 1) x (1 d 1) = .5 x .5 = 0.25

Finals: (1 d 1) = 0.5

Total: $1.48 \text{ x } 10^{10}$ x 0.5 x 0.25 = $1.85 \text{ x } 10^{-11}$ = 1 in 55 billion