Applying Betting Theory to March Madness

Nicolas Lepore

April 30, 2018

Abstract

The purpose of this paper is to find the most effective betting strategy to apply to the NCAA Tournament, also known as March Madness. The problem faced is that each year, many teams defy their odds and win games that weren't predicted, gaining or losing casino gamblers large sums of money in the process. Some gamblers go for the unfavored teams to capitalize on great rewards, while others choose the safe winner and seek marginal profit. An attempt is made to make sense of the randomness in a strategy that returns positively despite the erratic nature of the games.

My method was simulating last year's March Madness in MATLAB, and employing each strategy with a separate bankroll to see how they would fare overall. Each strategy worked by taking in a probability and a bankroll, and made gains multiplied by a casino-like approximation that used the probability to produce betting odds. The result was plotted on a graph of the logarithm of money to compare the performances of each.

The results showed that the Kelly Criterion double bet strategy, a method in which more certain games receive larger bets than uncertain ones and bets were made on both sides of the game, was the most effective in generating profit. The Risk strategies, the ones that bet on the unfavored team, fared poorly, and all Safe strategies fared well enough to bring back positive returns. The Safe Risk method, in which each favored team bet received 90% of the strategy's bankroll, was the most divergent, since it either gained or lost a sizable portion on each bet.

In conclusion, there are better and worse ways to bet with March Madness, and the probabilities associated with each game hold more truth than most gamblers in Las Vegas would like to believe.

1 Introduction

Every year, March Madness seems to become less and less predictable. The games do not abide by the probabilities assigned to them, and this can especially be a headache (or a wonder) for those betting on the games in Las Vegas. The goal is to find the most effective algorithmic method of betting in the games, and other scholars' work was used to guide the study.

This paper explores betting theory as described by William Poundstone's *Fortune's Formula* [Pou05] and J. L. Kelly, Jr.'s *A New Interpretation of Information Rule* [JK56]. The problem Kelly addresses is assigning value to the error of an informational channel as described by Shannon in his paper, *A Mathematical Theory of Communication*. By losing or gaining money based on a lack of information, Kelly uses this error and relates it to betting probabilities, and from there devises a method to make a large return from the betting. With a perfect channel, i.e. 100% probability of a team winning, and 1 to 1 odds, a person can make a maximum 2^{63} times their initial bet (since usually a person can make 2^{N}, and March Madness has 63 games total). Since that isn't possible otherwise, a realistic version will be presented.

From these, the Kelly Criterion for betting will be used, and it will be applied to games that have transpired. For this specific instance, it is assumed that there is no *track take*, meaning that no money will be taken as a cut for the casino, since it simplifies our scenario and allows for further analysis into each method. Without a *track take* the Kelly bet will ignore posted odds and only act on the probabilities.

I will be using two instances of the Kelly Criterion, one in which the strategy bets for one side, and one in which it bets for both sides. I will call these Kelly Single and Kelly Double.

The rest of my methods were generated on my own, by mathematical or trial and error means.

Some terms used in this paper are apart of a betting language that may be confusing to a layperson. The term *bankroll* refers to the funds available to a gambler, including all the money the gambler has won and excluding the money he has lost. So when betting a percentage of the bankroll, it means that the bet is a portion of the pot available to the gambler. The word *odds* refers to a simplified betting language meant to represent probability in a way that quantifies the return of the investment as well. So betting at 2:1 odds means that if the bet is for the favored side, if the favored team wins the bettor will earn half of his betting amount plus the initial bet is returned to him, whereas if he bet on the unfavored team and they win, he makes double his bet and his bet is returned to him. If he loses the bet, then he loses only the amount he bet on the game. The discrepancy between the probability and odds is known as the *edge*.

2 Foundation

To lay the foundation for this task, each of the following was done...

1. Researched and read about Kelly's Criterion in *Fortune's Formula* and *A New Interpretation of Information Rule*.
2. Downloaded probability data from FiveThirtyEight [JBS18], an open website that offers statistical data on everything from sports to elections.
3. Cut out excess data, formatted, and sent the data to MATLAB.
4. Devised a method to match certain competitors against each other and compare their probabilities of winning, since the data did not already display match-ups.
5. Identified the winners in each round.
6. Created a casino-like simulator of betting odds, taking in a probability and returning the closest equivalent in betting terms. The casino makes no money from the betting, and therefore no *track take* exists in this scenario. The probabilities are simply rounded to the closest odds. The odds were generated somewhat arbitrarily by not allowing the odds to become larger than having a 5 on one side of the odds (1:5 or 5:1), but the casino included some outliers as well. Data on odds is not uniform, so an attempt was made at an original scheme for odds mapping.

3 Simulation

3.1 Assumptions and Limitations

- Each bet was simulated by betting each event sequentially, although many of the rounds are happening simultaneously. The bankroll would look different if simulated in a real casino, since some bankrolls are not updated by the time the next bet is made (games aren't done before next round of betting starts).
- Only took into account the FiveThirtyEight data, although other probabilities exist.
- Could not identically replicate how a casino translates probability to odds.
- Used one instance of the March Madness tournament.
- Most savvy gamblers would never bet half of their bankroll on a single game, and so they diversify, similar to investors on the stock market. Some of the strategies bet somewhat unrealistically large amounts.

3.2 Strategies

First multiple betting strategies were created that take in probability and a bankroll, and return a suitable bet.

	$p > 50$	$p < 50$
Safe Bet	Bet 10% For	Bet 10% Against
Safe Long	Bet 25% For	Bet 25% Against
Risk Bet	Bet 10% Against	Bet 10% For
Risk Long	Bet 25% Against	Bet 25% For
Safe Risk	Bet 90% For	Bet 90% Against
	$p > 75$	$p < 75$
Very Safe	Bet 10% For	Bet 0
Very Risk	Bet 10% Against	Bet 0

Gradual Safe: Gradually bet more if the percentage was more certain, and bet less if the game could go either way. At above 90%, it bets 0.5 of the bankroll, at between 80% and 90% it bets 0.4 of the bankroll, and so on. Bets against the same way.

Kelly Single Bet: [JK56] Using this Kelly Criterion equation

$$(1 + l) = 2q$$

where:

- l is the fraction of the current bankroll to bet on the favored team
- q is the probability of winning, $q > 0.5$.

As an example, if a gamble has a 60% chance of winning ($p = 0.60$, $q = 0.40$), and the gambler receives 1-to-1 odds on a winning bet, then the gambler should bet 20% of their bankroll on the favored team.

Kelly Double Bet: However, where there is no *track take*, the Kelly Criterion calls for a bet to the other side as well. Quite simply, the strategy bets the probability of each event happening as a fraction of the bankroll. Thus:

$$l_1 = q * b, l_2 = p * b$$

where:

- b is the bankroll
- l_1 is the fraction of the bankroll to bet on the favored team
- l_2 is the fraction of the bankroll to bet on the unfavored team
- q is the probability of winning
- p is the probability of losing, which is $1 - q$.

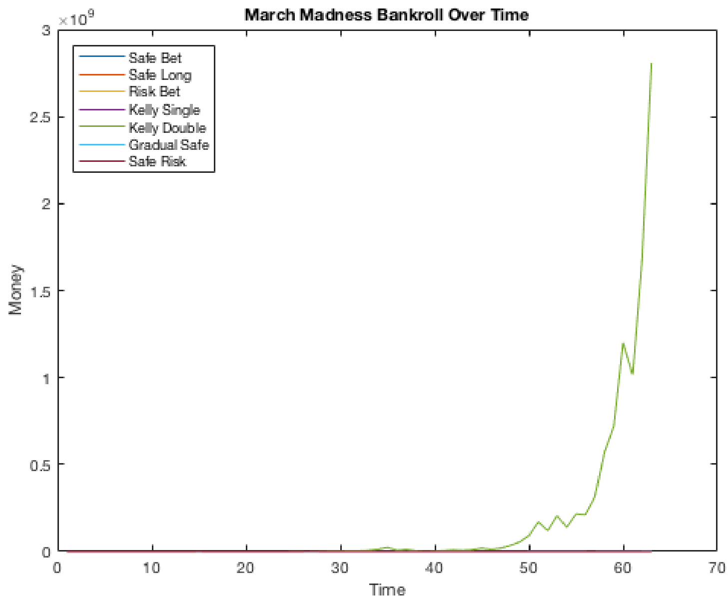
3.3 MATLAB

To use the data from FiveThirtyEight, each round of the tournament was sectioned. For each game, every strategy was implemented. If the bet was correct, the bet was multiplied by the odds determined by the "casino", and returned it to the strategy's bankroll. If the bet was incorrect, the bet amount was deducted from the strategy's bankroll. The most a strategy could lose was its bet amount, and the gains were determined by how risky the bet was.

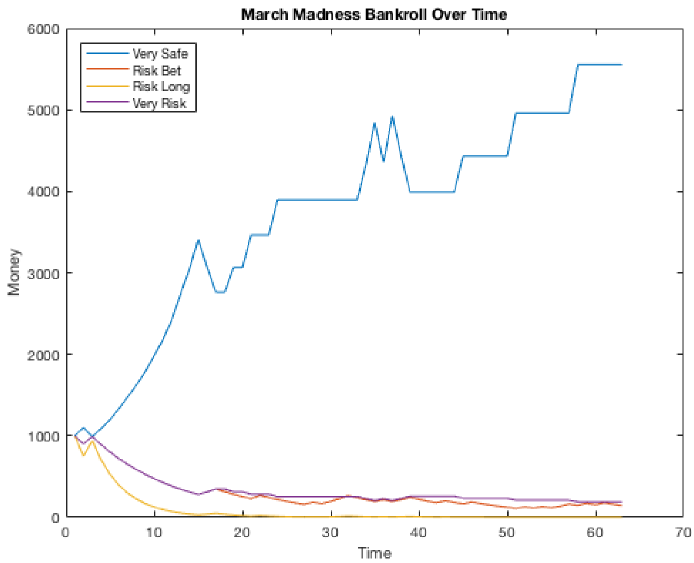
- `kmadness.m` is the main driver that culminates each programs, and it is where all of the betting occurs.
- `oddscalc.m` houses the approximation for the casino.
- `vscreate.m` houses the system that matches each competitor, done by taking a probability and finding the other probability that both add up to 1 (100%).
- `teamofid.m` looks up the team name associated with the team id.
- `safebet.m`, `safelong.m`, `risk.m`, `risklong.m`, `saferisk.m`, `vsafe.m`, `vrisk.m`, `gradsafe.m`, `kellybet.m`, and `kellybet2.m` are the individual strategies described above.

4 Results

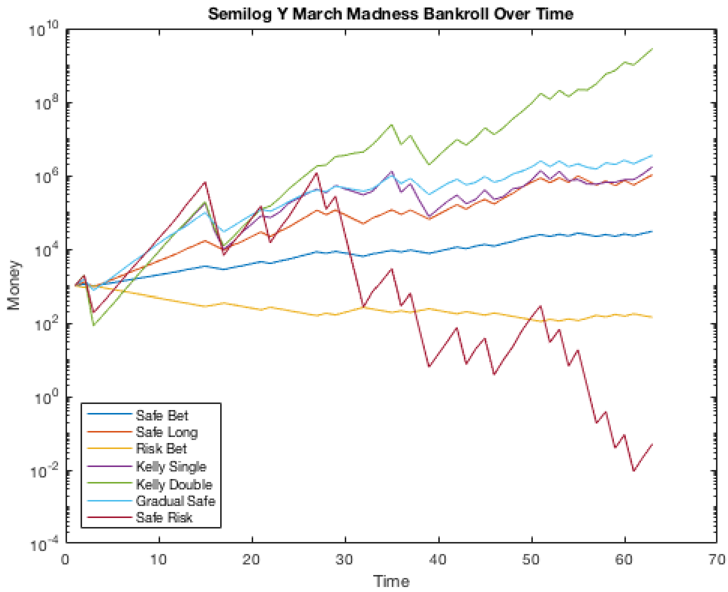
To accurately digest the data, strategies were occasionally omitted that made the others unrecognizable. First, the data altogether, but without Very Safe, Very Risk, and Long Risk since they are negligible.



Next, the data with Safe Bet, Long Safe, Safe Risk, Gradual Risk, Kelly Single, and Kelly Double omitted.



Now, plotted with the logarithm of money (y-axis) and all strategies except Very Safe, Very Risk, and Long Risk.



5 Discussion

5.1 Safe Risk

The Safe Risk strategy was the most erratic, for clear reasons. For every wager won, the strategy captured almost all possible gain. However, for every wager lost, the strategy lost almost all of the gains it made. Safe Risk traversed across the entire semilog y plot, but this behavior gives us some observations into the data set.

In the first round of the tournament, Safe Risk achieved the highest bankroll of all the strategies, simply because the first round is characterized by strong teams going up against weak ones. There were upsets clearly shown in the data, but overall the trend showed favoritism. However, as the betting saw closer games, these all-in bets resulted in too many losses to make up for, and the Safe Risk strategy ended up in the worst position. Safe Risk experienced both the highest and the lowest bankroll, so another possible optimal method is to bet 90% in each first round match-up, and not bet for the rest of the tournament.

5.2 Risk Strategies

During March Madness, some bettors believe that it is most profitable to bet for every unfavored team. By always betting for the lower odds, the one game that ends up an upset could pay off all loses and then some. However, from the simulation, we see that the risky situations lose too often to be a viable option. Maybe March Madness statistics and odds have some truth to them after all?

5.3 Gradual Safe

Gradual Safe takes into account the risk of the scenario, and bets more on events that are less risky, betting less on events that are more risky. In this way, the Gradual Safe is most similar to human behavior, where a regular gambler would bet a lot on most certain gambles and a little when they are afraid to lose. However, in a March Madness with more upsets, the strategy could pan out very differently.

5.4 Kelly Single Bet

The Kelly Single performed very similarly to the Gradual Safe strategy, but acted with more precision since it was dictated by an algorithm instead of probability thresholds. However, due to the nature of March Madness and the randomness of games, it didn't perform quite as well as

Gradual Safe. Either Gradual Safe or Kelly Single would be good strategies to work off of since they will produce similar results.

5.5 Kelly Double Bet

The Kelly Double proved to be the best by a sizable margin. The strategy is inherently aggressive, especially because it consistently bets on both sides. Not only that, it bets the entire bankroll on every single game. By simply betting the probability as a proportion of the available bankroll, it bets similarly to Gradual Safe and Kelly Single but even simpler than the two, and also covers itself for when the bet fails by betting for the unfavored side, and sometimes even earns more in this technique.

If the casino did take a cut of the bets, as real world casinos do, this method would be less desirable since each bet on both sides would lose money due to transaction costs. That is why, when a *track take* exists, people usually only bet for one side.

6 Conclusion

For betting in March Madness, multiple strategies could land you with a positive return. The strategies Gradual Safe, Safe Long, Safe Bet, and the Kelly Single/Double all performed well enough to bring in revenue. The Safe Risk strategy also had the potential to make the most, but because of its risky nature, it also had the ability to lose the most at one time.

The Safe strategies, the ones that bet on the favored team, all seemed to be viable options in the tournament. The Risk strategies did not perform nearly as well, since March Madness probabilities turn out to be more accurate than anticipated, at least in this year. After all, Villanova was the favored team, and they won the title again this year without hardship.

For future March Madness tournaments, Kelly Double will prove to be the best with the least risk associated with it. Real-life gamblers have learned this from the trade on their own, but the data backs up the strategy. With Kelly's Criterion implemented in this way, even if the posted odds are as accurate as the probabilities, the strategy can generate revenue.

To further research, new strategies would be tested, and factors would be taken into account such as which team is the home team, how much momentum does each team have (how many games did they win in a row before this upcoming game), and how did each team fare in the last March Madness tournament. Multiple sets of March Madness data would prove different results, and a new method would figure out how to bet simultaneously instead of all at once.

March Madness is mad, but there turns out to be a lot of truth to the numbers after all.

References

[JBS18] Rachael Dottle Jay Boice and Nate Silver. 2018 March Madness Predictions. *FiveThirtyEight*, 2018.

[JK56] Jr. J.L. Kelly. A New Interpretation of Information Rate. *ATT*, 1956.

[Pou05] William Poundstone. *Fortune's Formula: The Untold Story of the Scientific Betting System That Beat the Casinos and Wall Street.* 2005.