Sports Betting

Alessio Pignatelli
Wenxuan Lu

1 Introduction

Sports betting has seen a surge in popularity in the United States, driven by the legalization of sports betting in an increasing number of states and the growing number of bookmakers offering lines on American sports.

In this report, we begin by exploring the fundamental terminology and concepts of sports betting. Next, we examine the business models employed by bookmakers to ensure long-term profitability. Finally, we utilize probability modeling to identify and exploit discrepancies across different bookmakers, aiming to develop winning strategies despite the safeguards implemented by bookmakers to deter consistent profits for bettors.

2 Betting Jargon

2.1 Markets

For this report, we focused on two specific markets

- Moneyline/Head-to-head
  This market, commonly referred to as "Moneyline" in the United States and "Head-to-head" (or H2H) in Europe, consists on predicting the winner in a match between two teams that does not admit ties. For example, one can bet on the winner of a basketball game (potentially after overtime), of a volleybal match, of a baseball game. Possible outcomes are
  - Home Team Win or 1
  - Away Team Win or 2
- Three-way moneyline/1X2
  This market, commonly referred to as "Three-way moneline" in the United States and "1X2" in Europe, consists of predicting the outcome of a game that allows for a tie. Examples include regular season soccer matches, rugby matches, or handball matches. Possible outcomes are
  - Home Team Win or 1
  - Tie or X
  - Away Team Win or 2

2.2 Odds

Odds are the way bookmakers represent the possible payout of a certain outcome. As we will see later, odds are closely connected to the probability (according to bookmakers) of the outcome they are associated with.

Odds can be presented in different ways, depending on the bookmakers and especially the country the bookmaker is based in. Below, we consider and explain the main three types of odds.

- American Odds As is perhaps self-explanatory, American Odds are primarily used in the United States. The American odds of an outcome are represented by an integer greater than 100 preceded by either a positive or a negative sign. The meaning varies with the sign:
  - Positive Odds
    The integer represents the potential net win when 100 units are placed at stake.
    For example: if you bet 100 on an event at +150 and you win, then you get 250 back: the original 100 plus 150 net gain.

- Negative Odds
  The integer represents the amount needed to stake for a potential win of 100.
  For example: if you bet 150 on an event at -150 and you win, then you get 250 back: the original 150 plus 100 net gain.

- European Odds
  European Odds are the most used in continental Europe. The European odds of a outcome are represented by a decimal number that is the potential total win when 1 unit is wagered. For example, if you bet 1 on an event at 1.50 and you win, you get back 1.50 (the original 1 plus 0.50 net win).

- Fractional Odds
  Fractional Odds are primarily used in the United Kingdom. The fractional odds of an outcome are represented by a fraction, where the numerator is the potential net gain and the denominator is the stake required. For example, if you bet 1 on an event at 5/1 and you win, you get back 5: the original 1 plus 5 net gain. Conversely, if you bet 1 on an event at 1/5 and you win, you get back 1.2 (the original 1 plus 0.2 net gain).

3 Implied Probabilities from Odds

The probability of a betting outcome is unknown, and different people have different estimates of the probability. The odds reflect the bookmaker's estimate of the event probability in the sense that if the bookmaker's estimate is accurate, betting according to the odds will result in neither a gain or a loss in expectation. Take the European odds as an example. For European odds x, the implied probability is $\frac{1}{x}$. For instance, the implied probability of odds 1.5 is $\frac{1}{1.5} = \frac{2}{3}$. If we bet 1 unit, then the expected return, including the original 1 unit, is $1.5 \times \frac{2}{3} + 0 \times \frac{1}{3} = 1$. We will stick with the European odds for all odds described below. For other types of odds, we can convert them to an equivalent European odds and calculate the implied probability.

It is important to understand the concepts of fair odds, subfair odds, and superfair odds. These terms help to describe whether the odds offered by a bookmaker are in the bettor's favor or the bookmaker's favor.

1. Fair Odds
   Fair odds represent the case where the implied probabilities of all possible outcomes sum up to 1. This situation is fair to both the bookmaker and the bettor. If we bet according to the implied probability, we are guaranteed with neither a gain nor a loss. For example, suppose the European odds for the winning of both teams are 2 with implied probability 50%. If we split our 1 unit and bet 0.5 on each team, we will end up with 1 unit regardless of the outcome.

2. Superfair Odds
   Superfair odds occur when the implied probabilities of all outcomes add up to less than 100%. This situation is favorable for the bettor. If we bet according to the implied probability, we are guaranteed with a gain. For example, suppose the European odds for the winning of both teams are 2.5 with implied probability 40%. If we split our 1 unit and bet 0.5 on each team, we will end up with 1.25 unit regardless of the outcome. This can arise because different bookmakers have different opinions about the event probability. When we place bets with different bookmakers, it is possible to find a combination of superfair odds. This is also referred to as arbitrage betting, where a win is guaranteed.

3. Subfair Odds
   Subfair odds occur when the implied probabilities of all outcomes add up to more than 100%. This situation is favorable for the bookmakers. If we bet according to the implied probability, we are guaranteed with a loss. For example, suppose the European odds for the winning of both teams are 1.5 with implied probability 67%. If we split our 1 unit and bet 0.5 on each team, we will end up with 0.75 unit regardless of the outcome. This is the most common scenario is real life because the bookmakers need to make a living by setting the odds to be superfair. In particular, the odds from a single bookmaker for a specific event are (almost) always subfair: the difference between the sum of the implied probabilities of all possible outcomes and 1 is called "vigorish", and is the expected gain of the bookmakers for every unit wagered by bettors.

4 Betting Strategies

The aim of this project is to explore good strategies for sports betting. A total of four different strategies are considered. For all strategies, when bookmakers offer different odds for the same outcome, if we want to bet for said outcome we do so at the highest odds (which is equivalent to the highest possible return on investment)

1. Random betting
   Random betting splits the money equally across all possible outcomes. Since at least one event occurs, this strategy ensures that theoretically, there is always some wealth remaining. It is served as a baseline comparison for all other betting strategies.

2. Bet on the most favorable only
   One popular strategy is betting on the most favorable outcome and hold back the remaining money. Theoretically, the hold-back part ensures that we always have some money to speed in the next bet.

3. Bet on the least favorable only
   Another popular strategy is betting on the least favorable outcome and hold back the remaining money. Theoretically, the hold-back part ensures that we always have some money to speed in the next bet.

4. Kelly strategy
   Kelly strategy, developed by J. L. Kelly Jr, formulates an optimal way to place bets such that the expected logarithm of wealth can be maximized. Suppose $p(s)$ is the probability for the outcome s to occur and $\alpha(s)$ is the European odds for event s. Let $a(s)$ be the fraction of money we bet on the outcome s, and b is the fraction of hold-back money. Then the Kelly strategy can be summarized by the algorithm in Figure 1. For fair and superfair odds, Kelly strategy recommends us to place bet proportional to the probability of each event to occur. For subfair odds, the strategy provides a way to select out options that are favorable to the bettor and place a fraction of money only on those favorable options.

- If $\sum \frac{1}{\alpha_s} \leq 1$, then $a(s) = p(s)$.
- If $p(s)\alpha_s < 1$ for all s, then no bets are placed.
- Otherwise:

1. Permute the indices so that $p(s)\alpha_s \geq p(s+1)\alpha_{s+1}$
2. Define $C_k = \begin{cases} \frac{1-\sum_{s=1}^{t} p(s)}{1-\sum_{s=1}^{t}\frac{1}{\alpha_s}} & (if\ k \geq 1) \\ 1 & (if\ k = 1) \end{cases}$ and $t = \min\{k \mid p(k+1)\alpha_{k+1} \leq C_k\}$
3. $b = C_t$
4. For $s = 1,2,\ldots,t$, $a(s) = p(s) - \frac{b}{\alpha_s}$; Otherwise, $a(s) = 0$.

Figure 1: Kelly Strategy

5 Applications

We assessed the betting strategies on two data sets: a volleyball data set with two options of betting and a soccer data set of three-way moneyline. For each strategy, we calculated the empirical and the expected return after betting sequentially through all games. Since the probability for each outcome is needed for the calculation, we estimated it by implied probability calculated from odds. In "Single prob", implied probability are obtained from a single odds (the highest available) for each possible outcome, while in "Mean prob"/"Median Prob" implied probability is obtained aggregating odds across bookmakers. In any case, implied probabilities are finally normalized to sum up to 1. The betting results are summarized in Figure 2 and Figure 3. Overall, Kelly strategy achieves the best outcome on superfair odds. On fair and subfair games that are in favorable to the bookmakers, none of the strategies managed to gain money due to the inaccurate estimation of the true winning probability.

5.1 Data Set I (Volleyball Games)

The first data set contains 14906 volleyball games from 1950 to 2018 across 10 bookmakers. There are two betting potions: (1) home team wins, and (2) away team wins. We first look at the betting results on superfair odds that are favorable to the bettor. Suppose we start with 1 unit of money. After betting through all the games, the Kelly strategy makes us a billionaire, regardless of the method used to estimate the implied probability. On the other hand, the other strategies do not have a sustainable gain in wealth. In reality, about 10% of the games have superfair odds (Figure 2a). To avoid profit loss, bookmakers usually impose some betting restrictions. Common rules including a minimum amount of bet of 1 dollar, a maximum amount of bet of 10,000 dollars, and a unit of 5 cents. If we start with 10 dollars and follow the betting rules, the empirical gain from the Kelly strategy becomes more realistic. As shown in Figure 2b, we gain 10,000-fold after betting through all the 14906 games over 70 years. For other strategies, we quickly lose money to an amount of fewer than the minimum amount to bet and quit the sports betting. One thing to note is that there is large discrepancy between the expected money amount and empirical money amount. This is due to the inaccuracy in probability estimation. Since the Kelly strategy is optimized with regard to the true probability of winning, the large discrepancy, especially a huge decrease in the empirical money amount compared with the expected money amount, indicates that the estimated probability does not match with the true probability. In fair and subfair games, we can earn money only if our estimation of the winning probability is closer to the truth than the bookmakers do. In the case of inaccurate probability estimation, we have no chance of gaining wealth (Figure 2c).

5.2 Data Set II (Soccer Games)

The second data set contains 479440 soccer games from 2005 to 2015 across a diverse number of bookmakers. There are three betting options: (1) home team wins, (2) away team wins, and (3) a tie occurs. We compared the betting strategies on a randomly selected 10,000 games, among which about 15% have superfair odds. The general trend is very similar to the comparison on the volleyball games (Figure 3). For games with superfair odds, the Kelly strategy is able to win money. In the setting with the same betting restrictions as stated above, there is about a 50000-fold gain in wealth if we bet according to the Kelly strategy. On fair and subfair games and for all other betting stragetigies, we failed to have a sustainable gain in money.

6 Discussion

Though the popularity of sports betting is primarily due to its entertainment component, people have explored diverse strategies to distribute their bets wisely across the betting options. Compared with random betting or betting only on the least/most favorable option, the Kelly strategy incorporates probability and has the best empirical performance. One limitation is that the Kelly strategy assumes the winning probability is known. In reality when our estimated probability is not that accurate, there can be a large difference between the expected gain and the empirical gain. Another strong assumption of the Kelly strategy is that the winning probability does not change overtime. In case where the game is dynamic and the winning probability changes from game to game, the Kelly strategy is not optimal in maximizing the total gain. Nevertheless, its good empirical performance has attracted a lot of attention and makes it a popular strategy among sports bettors.

Future developments should be in direction of achieving better estimates of the probability, as poor estimation seems the key factor preventing Kelly's strategy from being profitable even in fair or slightly subfair odds. In particular, it should be explored whether odds from certain bookmakers have better predictive power than others, and if is implied probability from odds should be aggregated weighting proportionally more the more truthworthy odds.

Strategy	Implied Prob	Expected Median Return	Expected Money Amount	Empirical Median Return	Empirical Money Amount
Kelly	Single Prob	1.005950	4.5 e+14	1.005950	4.5 e+14
Kelly	Mean Prob	1.006602	2.5 e+19	1.0043205	2.9 e+14
Kelly	Median Prob	1.057374	5.5 e+20	1.0033943	6.4 e+12
Most Favorable (betting frac: 0.5)	Single Prob	1.002975	7.1 e+7	1.145 Two Modes	2.9 e-70
Least Favorable (betting frac: 0.08)	Single Prob	1.000476	23.0	0.92 Right Skewed	8.9
Random	Single Prob	1.005950	4.5 e+14	0.925	1.1 e-23

(a) Results on superfair odds without betting restrictions

Strategy	Implied Prob	Expected Median Return	Expected Money Amount	Empirical Median Return	Empirical Money Amount
Kelly	Single Prob	1.000793	650768	1.0016535	976194
Kelly	Mean Prob	1.000721	1559826	0.9991383	957754
Kelly	Median Prob	1.000734	1617053	0.9991376	958503
Most Favorable (betting frac: 0.5)	Single Prob	1.000948	254581	1	1.96
Least Favorable (betting frac: 0.3)	Single Prob	1.001481	147739	1	2.45
Random	Single Prob	1.000786	654256	1	1.72

(b) Results on superfair odds with betting restrictions

Strategy	Implied Prob	Expected Mean Return	Expected Money Amount	Empirical Mean Return	Empirical Money Amount
Kelly	Single Prob	1	1	1	1 (didn't bet)
Kelly	Mean Prob	1.000166	9.4	0.9999781	0.28
Kelly	Median Prob	1.000174	10.4	0.9999472	0.18

(c) Results on fair and subfair odds without betting restrictions

Figure 2: Betting results on volleyball games

Strategy	Implied Prob	Expected Mean Return	Expected Money Amount	Empirical Mean Return	Empirical Money Amount
Kelly	Single Prob	1.008026	966900	1.011387	429544
Kelly	Mean Prob	1.008818	2202314	1.013908	489894
Most Favorable (betting frac: 0.5)	Single Prob	1.007039	264806	0.999855	1.90
Least Favorable (betting frac: 0.3)	Single Prob	1.006612	159567	0.9992174	2.50
Random	Single Prob	1.008651	2349244	0.9993194	1.71

(a) Results on superfair odds with betting restrictions

Strategy	Implied Prob	Expected Mean Return	Expected Money Amount	Empirical Mean Return	Empirical Money Amount
Kelly	Single Prob	0.9541276	4.2 e-176	0.9541276	4.2 e-176
Kelly	Mean Prob	0.9553148	1.3 e-171	0.9534331	1.1 e-180

(b) Results on fair and subfair odds without betting restrictions

Figure 3: Betting results on soccer games