Module 1 Project: Analysis of a Betting Strategy in Sports

Dhairyav Jatin Shah

College of Professional Studies, Northeastern University

Professor: Zhi(Richard) He

March 2, 2023

**Introduction**

In this problem, we are asked to analyze outcomes of 4 different series between two American NBA teams, Boston Celtics, and Miami Heat. In the first series i.e., a best of three series, the games are played in a specific order, with the first game in Miami, the second game in Boston, and the third game (if necessary) in Miami for Part 1 and vise versa in Part 2. For Part 3, the two teams would play a best of five series, where the team winning 3 first matches would win the series. The games would be played alternating between Boston and Miami, starting with Miami. For part 4, the two teams will be playing a best of 7 series, where the team winning 1st 4 games would win the series. In this series, games 2, 3, 5, and 6 are hosted by the side with home-court advantage, while games 1, 4, and 7 are hosted by the opposition. Boston has the home-court advantage over Miami. We are given the probabilities of each team winning their home game and the associated payout for betting on the Boston Celtics to win. Using this information, we are required to calculate the probability of the Boston Celtics winning the series, construct a probability distribution for the net win, calculate the expected net win and standard deviation, and estimate the expected net win using a 90% confidence interval. Finally, we will construct a frequency distribution for the random values obtained and verify how closely the distribution of these values matches the expected distribution using the Chi-square goodness of fit test.

**Part 1**

For part 1, the first game will be played in Miami and then the second game will be played in Boston, then the third game will be played in Miami if required.

1. Calculate the probability that Boston Celtics will win the series.

Using the probability table, we calculate the probability of Boston winning the series.

P(MBB) = 0.61\*0.61\*0.39 = 0.14988

P(BB) = 0.39\*0.63 = 0.24570

P(BMB) = 0.39\*0.37\*0.37 = 0.05628

P(Boston Winning) = P(MBB) + P(BB) + P(BMB) = 0.45185 i.e., 45.18%

Hence, we can say that the probability of Boston Celtics winning this series is 45.18%.

1. Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (E(X)) and the standard deviation of X.

The above table represents the probability distribution of net win (X) in the series. We know that if Boston wins one wins $100 and if Boston losses one losses $103. Using this information, we have obtained the above table. We multiply the net win with the winning probability to calculate the expected net win. We can see that the expected net win is -$11.54. Next to calculate the Std dev, we first need to find the total variance. Which is calculated by squaring the difference between the net win and sum of expected net win then diving it by the probability of that game. Our total variance obtained is **$24,835.42**. As we know std dev is the square root of the variance. Hence, we obtain the std dev as **$157.59**. The net win's standard deviation is relatively large **($157.59)**, indicating that there is a lot of variation among the possible outcomes and that the bettor's actual net win can differ significantly from the predicted figure.

1. Use Excel or R to create 2,500 random values for X. Let these random values be denoted by Y. Use these Y values to estimate your expected net win by using a 90% confidence interval. Does this confidence interval contain the E(X) in (ii)?

We use excel to generate 2,500 random values for X. The random numbers generated are from the possible net winnings i.e., **-$206, -$106, $97, $200**. The mean value we obtain from these 2500 random values is **-13.697**. The std dev obtained is **157.60**. We can observe that the std dev and mean obtained from the 2500 variables is close to the actual std dev and the expected net win we obtained in the above question. We calculated the CI and obtained **5.18 at 90% or 0.1** considering E(X). By using a 90% confidence interval, we calculate the upper and lower limits of the confidence interval **(-8.512,-18.881)**. And we can conclude that, the expected net win from (ii) lies inside the upper and lower limit bounds.

1. Construct a frequency distribution for Y. Next, use the Chi-square goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.

To construct a frequency distribution of Y, we use the formula =FREQUENCY() in excel. Once we calculate the frequency, we multiply this with the probability and calculate the expected frequency. Now we use the goodness-of-fit test to verify how closely the distribution of Y has estimated the distribution of X.

Ho = The theoretical and observed frequency is equal.

H1 = The theoretical and observed frequency are not equal.

Our p-value is **0.9**, which is greater than the confidence interval at **90% (0.1),** hence we fail reject the null hypothesis that the theoretical and observed values are equal.

1. Use your observations in parts (ii) and (iii) above to describe whether your betting strategy is favorable to you.

Based on the probability distribution we calculated earlier, we found that the expected net win for games is -$24. This means that, on average, we are expected to lose $24 over the course of three games. Using a sample of 2,500 random values for X, we estimated the expected net win using a 90% confidence interval. The resulting interval was **(-8.512,-18.881)**. This means that we can be 90% confident that our expected net win is likely to be between -**$8.512** and **-$18.881**. Based on our analysis, it seems that our betting strategy is not favorable to us. The expected net win is negative, which means that, on average, we are expected to lose money over the course of the three games. The 90% confidence interval also confirms this since the range of confidence interval is only in negative. Therefore, it appears that our betting strategy is not a good one.

**Part 2**

For part 2, the first game will be played in Boston and then the second game will be played in Miami, then the third game will be played in Boston if required.

1. Calculate the probability that Boston Celtics will win the series.

Using the probability table, we calculate the probability of Boston winning the series.

P(MBB) = 0.37\*0.39\*0.63 = 0.09091

P(BB) = 0.39\*0.63 = 0.24570

P(BMB) = 0.63\*0.61\*0.63= 0.24211

P(Boston Winning) = P(MBB) + P(BB) + P(BMB) = 0.5787 i.e., 57.87%

Hence, we can say that the probability of Boston Celtics winning this series is 45.18%. The probability of Boston winning the series increases by 12.69% if the first game is played in Boston and so on.

1. Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (E(X)) and the standard deviation of X.

The probability distribution for the series' net win (X) is shown in the accompanying table. We are aware that Boston gains $100 for each victory and loses $103 for each defeat. This data allowed us to create the table you see above. To determine the predicted net win, we multiply the net win by the winning probability. The expected net win is **$14.22**. Finding the total variance is necessary before we can calculate the standard deviation. This is determined by squaring the difference between the net victory and the total of projected net wins, then multiplying the result by the probability of the game in question. We found a total variance of **$24,534.61**. Standard deviation, as we all know, is the square root of variation. Hence, we arrive at the standard deviation of **$156.64**. The standard deviation of the net win is relatively high **($156.64)**, indicating that there is a wide range of potential outcomes and that the bettor's actual net gain can vary greatly from the expected amount.

1. Use Excel or R to create 2,500 random values for X. Let these random values be denoted by Y. Use these Y values to estimate your expected net win by using a 90% confidence interval. Does this confidence interval contain the E(X) in (ii)?

 To create 2,500 random values for X, we utilize Excel. The randomly generated numbers are taken from the potential net winnings, which are **(-$206, -$106, $97, and $200)**. We find a mean value of **16.497** using these 2500 random values. The std dev found is **155.71**. We can see that the average and standard deviation derived from the 2500 variables are rather close to the real average and projected net win we discovered in the previous question. We calculated the CI and came up with **6.10 at 90%, or 0.1**, taking E into account (X). We determine the top and lower limits of the confidence interval using a 90% confidence interval **(10.39,22.60)**. And hence, we may infer that the predicted net win from (ii) is within the bounds of the upper and lower limit.

1. Construct a frequency distribution for Y. Next, use the Chi-square goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.

 Using the formula =FREQUENCY() in Excel, we can create a frequency distribution for Y. We may then determine the expected frequency by multiplying the frequency by the likelihood. To check how accurately the distribution of Y has estimated the distribution of X, we now employ the goodness-of-fit test.

Ho = The theoretical and observed frequency is equal.

H1 = The theoretical and observed frequency are not equal.

Our p-value is **0.7**, which is greater than the confidence interval at **90% (0.1),** hence we fail reject the null hypothesis that the theoretical and observed values are equal.

1. Use your observations in parts (ii) and (iii) above to describe whether your betting strategy is favorable to you.

The predicted net win for games is -$24, according to the probability distribution we computed before. This suggests that throughout the course of three games, we would lose $24 on average. Using a sample of 2,500 random values for X, we estimated the expected net win using a 90% confidence interval. The resulting interval was **(10.39,22.60)**. This means that we can be 90% confident that our expected net win is likely to be between **10.39** and **22.60**. Our betting strategy appears to be advantageous to us, based on our study. The expectation of the net win is positive, which suggests that over the course of the three games, we should anticipate making money. This is further supported by the 90% confidence interval, which is positive in its range. We believe our betting strategy to be favorable as a result.

**Part 3**

For part 3, in this series both teams will be playing a best-of-five series where the first team to win three games will win the series. The teams will play alternate games in Boston and Miami, with the first game in Miami.

1. Calculate the probability that Boston Celtics will win the series.

From the above table, we know that if Boston wins more than 3 games to win the series. Hence, we calculate all the possibilities where Boston wins more than 3 games.

P(Boston Winning) = 0.095823 + 0.03545451 + 0.09442251 + 0.09442251 + 0.021627251 + 0.021627251 + 0.008120771 + 0.021627251 + 0.057597731 + 0.021627251 = 47.23%

Hence, we can say that the probability of Boston Celtics winning this series is 47.23%.

1. Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (E(X)) and the standard deviation of X.

In the table that follows, the probability distribution for the series' net victory (X) is displayed. We know that for every win Boston we make $100 and for every loss we lose $103.  We multiply the net win by the likelihood of winning to get the expected net win. After calculation, we observe that we will be facing a net loss of -$12.99. Before we can compute the standard deviation, we must first determine the overall variance. This is calculated by first multiplying the result by the likelihood of the game in question, then solving for the square root of the difference between the net victory and the sum of anticipated net victories. A total variance of $41,113.55 was discovered. We are all aware that the square root of variation is the standard deviation. Hence, we reach the $202.76 standard deviation. Given the wide range of possible outcomes and the relatively high ($202.76) standard deviation of the net win, it is possible for the bettor's actual net gain to differ significantly from the predicted amount.

1. Use Excel or R to create 2,500 random values for X. Let these random values be denoted by Y. Use these Y values to estimate your expected net win by using a 90% confidence interval. Does this confidence interval contain the E(X) in (ii)?

 We use Excel to generate 2,500 random values for X. The numbers that were chosen at random are **($300,$197,$94,-$309,-$209,-$109)** from the possible net winnings. Using these 2500 random values, we come up with a mean value of **-$11.65**. Standard deviation found is **202.06**. As we can see, the 2500 variables' average and standard deviation are rather like the actual average and projected net win we found in the preceding question. Taking E into consideration, we calculated the CI and arrived at 6.64 at 90%, or 0.1. (X). Using a 90% confidence interval, we determine the upper and lower boundaries of the confidence interval **(-5.00,-18.29)**. Hence, we may conclude that the expected net win from (ii) is within the upper and lower limits.

1. Construct a frequency distribution for Y. Next, use the Chi-square goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.



Using the formula =FREQUENCY() in Excel, we can create a frequency distribution for Y. We may then determine the expected frequency by multiplying the frequency by the likelihood. To check how accurately the distribution of Y has estimated the distribution of X, we now employ the goodness-of-fit test.

Ho = The theoretical and observed frequency is equal.

H1 = The theoretical and observed frequency are not equal.

Our p-value is **0.46**, which is greater than the confidence interval at **90% (0.1),** hence we fail to reject the null hypothesis that the theoretical and observed values are equal.

1. Use your observations in parts (ii) and (iii) above to describe whether your betting strategy is favorable to you.

According to the probability distribution we previously estimated, the expected net win for games is $135. This implies that we would lose, on average, $135 over the course of five games. A 90% confidence interval was used to predict the expected net win using a sample of 2,500 randomly generated values for X. The outcome was an interval of (-5.00,-18.29). In other words, we can predict with 90% certainty that our predicted net win will most certainly fall between -5.00 and -18.29. Our research suggests that our betting strategy is bad for us. We should expect to lose money throughout the course of the five games because the expectation of the net victory is negative. The 90% confidence interval, which is negative in its range, adds to this argument. We think that this makes our betting strategy unfavorable.

**Part 4**

For part 4, in this series both teams will be playing a best-of-seven series where the first team to win four games will win the series. The team with home-court advantage will host games 2,3,5 and 6 and the opposite team will host games 1,4 and 7. We assume Boston Celtics has the home-court advantage against Miami. To calculate the probability table of this series we will use R. We know that to win a best-of-seven series any team must win 4 or more games to win the series hence the number of combinations possible for each team to win more than four matches is given by.

Hence, we know that there are 35 possible combinations for each team to win a total of more than 4 games, which makes a total of 70 games played.

Table

Description automatically generated with medium confidenceGraphical user interface, text, application, email

Description automatically generated

1. Calculate the probability that Boston Celtics will win the series.

From the above table, we know that if Boston wins more than 4 games to win the series. Hence, we calculate all the possibilities where Boston wins more than 4 games.

P(Boston Winning) = 56.12%

Hence, we can say that the probability of Boston Celtics winning this series is 56.12%.

1. Construct a probability distribution for your net win (X) in the series. Calculate your expected net win (E(X)) and the standard deviation of X.

In the table that follows, the probability distribution for the series' net victory (X) is displayed. We know that for every win Boston we make $100 and for every loss we lose $103.  We multiply the net win by the likelihood of winning to get the expected net win. After calculation, we observe that we will be facing a net loss of $32.64. Before we can compute the standard deviation, we must first determine the overall variance. This is calculated by first multiplying the result by the likelihood of the game in question, then solving for the square root of the difference between the net victory and the sum of expected net victories. A total variance of $56,677.33 was discovered. We are all aware that the square root of variation is the standard deviation. Hence, we reach the $238.00 standard deviation. Given the wide range of possible outcomes and the relatively high ($238.00) standard deviation of the net win, it is possible for the bettor's actual net gain to differ significantly from the predicted amount.

1. Use Excel or R to create 2,500 random values for X. Let these random values be denoted by Y. Use these Y values to estimate your expected net win by using a 90% confidence interval. Does this confidence interval contain the E(X) in (ii)?

 We use Excel to generate 2,500 random values for X. The numbers that were chosen at random are **($400,$297,$194,$91,-$412,-$312,-$212,-$112)** from the possible net winnings. Using these 2500 random values, we come up with a mean value of **$32.64**. Standard deviation found is **238.07**. As we can see, the 2500 variables' average and standard deviation are rather like the actual average and projected net win we found in the preceding question. Taking E(X) into consideration, we calculated the CI and arrived at **7.83 at 90%, or 0.1. (X).** Using a 90% confidence interval, we determine the upper and lower boundaries of the confidence interval **(23.466,39.135)**. Hence, we may conclude that the expected net win from (ii) is within the upper and lower limits.

1. Construct a frequency distribution for Y. Next, use the Chi-square goodness of fit test to verify how closely the distribution of Y has estimated the distribution of X.



Using the formula =FREQUENCY() in Excel, we can create a frequency distribution for Y. We may then determine the expected frequency by multiplying the frequency by the likelihood. To check how accurately the distribution of Y has estimated the distribution of X, we now employ the goodness-of-fit test.

Ho = The theoretical and observed frequency is equal.

H1 = The theoretical and observed frequency are not equal.

Our p-value is **0.104**, which is greater than the confidence interval at **90% (0.1),** hence we fail to reject the null hypothesis that the theoretical and observed values are equal.

1. Use your observations in parts (ii) and (iii) above to describe whether your betting strategy is favorable to you.

According to the probability distribution we previously estimated, the expected net win for games is -$672. This implies that we would lose, on average, -$672 over the course of seven games. A 90% confidence interval was used to predict the expected net win using a sample of 2,500 randomly generated values for X. The outcome was an interval of **(23.466,39.135)**. In other words, we can predict with 90% certainty that our predicted net win will most certainly fall between **23.46** and **39.135**. Our research suggests that our betting strategy is favorable for us. We should expect to win money throughout the course of the seven games because the expectation of the net victory is positive. The 90% confidence interval, which is positive in its range, adds to this argument. We think that this makes our betting strategy favorable.

**Reference**

1. Thankur.M(n.d.) *Probability Distribution Formula*[https://www.educba.com/probability-distribution-formula**/**](https://www.educba.com/probability-distribution-formula/)
2. *Lesson 1-5 — Examples of Computation with Probability Distribution* <https://northeastern.instructure.com/courses/131431/pages/lesson-1-5-examples-of-computation-with-probability-distributions?module_item_id=8233526>
3. Zach(Apr 2020) *How to Perform a Chi-Square Goodness of Fit Test in Excel* <https://www.statology.org/chi-square-goodness-of-fit-test-excel/>