Module 1 Project

Analysis of a Betting strategy in Sports

Tushar Parimi

College of Professional Studies, Northeastern University

ALY 6050: Enterprise Analytics

May 29th, 2022

**Introduction**

Sports betting is a very popular thing among sport fanatics. Fans try and apply their knowledge of the sport to see financial profits by betting on teams they think will win. Although it is true, that a lot of bets or stakes are put on single match between two teams, but it is not only limited to single matches. Various stakes are put on scenarios like winner of the tournament or a series of matches in a best of example three matches etc. These attract a lot more participation and people who place bets as these scenarios might give an uneven or favorable odds to win the bet. As a single match is almost nearly a 50-50 between the two teams involved but a winner of series of matches played home and away might probably favor one of the sides more.

An average fan can make a informed decision based on the some analysis on a scenario. In this project we consider a various scenario of best of 3 and 5 matches played between two baseball teams Boston Red Sox and New York Yankees in a specific alternate order of home and away for both teams against each other.

In this scenario we will be looking to bet on Boston Red Sox to win every match they played, and the goal will be to analyze the net winnings and whether this will be favorable to us.

**Scenario**

The project basically consists of three parts. Each part of analysis will have some commonality as the we are only going to bet on Boston Red Sox win for every match. The probability that Boston Red Sox win at home is 0.59 and the probability that New York Yankees win at their home ground is 0.55.

We will receive $505 when Red Sox win a match, and we lose $525 when Red Sox do not win the match. It is to be noted that all the analysis will be based on best of n matches that means if it’s a series of 3 matches then if one team wins the first two then the third match will not happen.

This information is well visualized in the below tables in Fig1 and fig2.

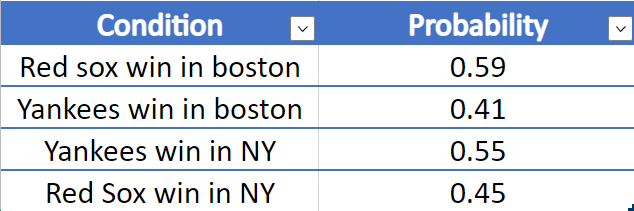


Fig1. Table shows home and away win probabilities for both the team

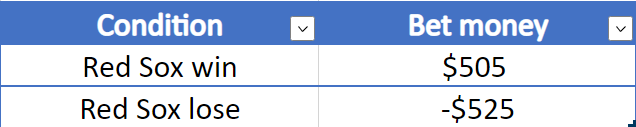


Fig2. Table shows win and lose bet money values

**Analysis**

**The first part of analysis we will go with the scenario,**

Match1 is played in Boston

Match2 is played in New York

Match3 is played in Boston

At first as our idea is to bet on Red Sox for every match we might want to know the probability that Red Sox will win the series,

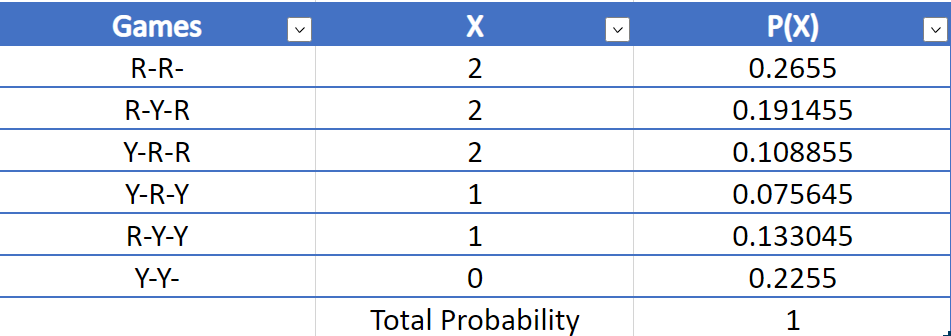


Fig3. Table showing various combination of wins and loses in the series and their probability. Here X represents no. of Red Sox wins.

From this we get,

**Probability that Red Sox win the series**= sum(P(X=2))

= 0.56581

This tells us Red Sox are more likely to win the series than Yankees. This favours us betting on Red Sox.

Next as our main goal is analysing our betting gain and loss, it makes sense to create a probability distribution of net wins from the bets.

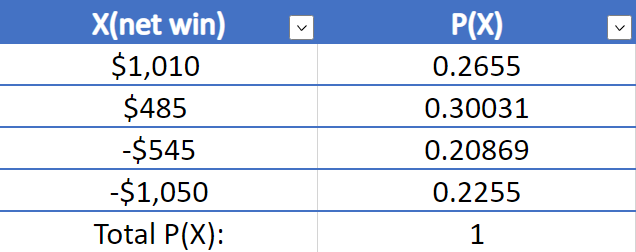


Fig4. Table represents probability distribution of a random variable representing net wins for the bets.

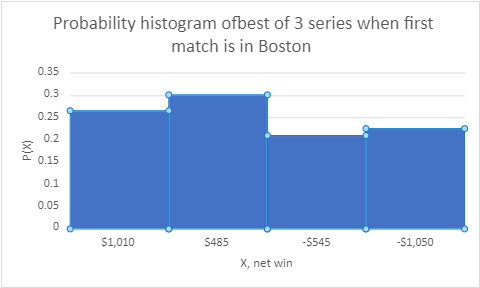


Fig5. Histogram of probability distribution

From this we calculate distribution parameters like Expectation of X (average) and standard deviation of the distribution.

**Expected mean E[X]**= Mean = 63.2943

This tells us that if a lot of these series were conducted Then the average of these series net wins will average to 63.2943. But for a single series the net wins can only take values $1010, $485, -$545 and -$1050.

**Expected Variance**= Var[X]= 648,071

**Expected Standard deviation**= SD[X]= 805.0283837

Standard deviation and Variance tells us about the variability and uncertainty of our random variable.

Next, 10000 random values of X were generated in Excel,

The observed parameters for the sample were as follows,

**Observed Mean**= 66.64

**Observed Variance**= 652209.1863

**Observed SD**= 807.5946919

The expected and observed values look some what similar, but lets estimate the 95% confidence interval for the sample,

**95% confidence interval**= (50.86173361, 82.41826639)

The expected Mean value lies within this interval. This may not be true for other samples as

confidence interval only tells us that if a lot of samples are taken then approximately 95% of them

will have expected mean within the interval (50.86173361, 82.41826639).

From this we can also say with 95% confidence that we will have positive net win by betting on the

Red Sox win.

A Chi-square test was conducted for our sample to find its goodness of fit to our net win probability distribution,

Here,

Null hypothesis: The sample comes from the net win probability distribution

Alternate hypothesis: The sample does not come from the net win probability distribution

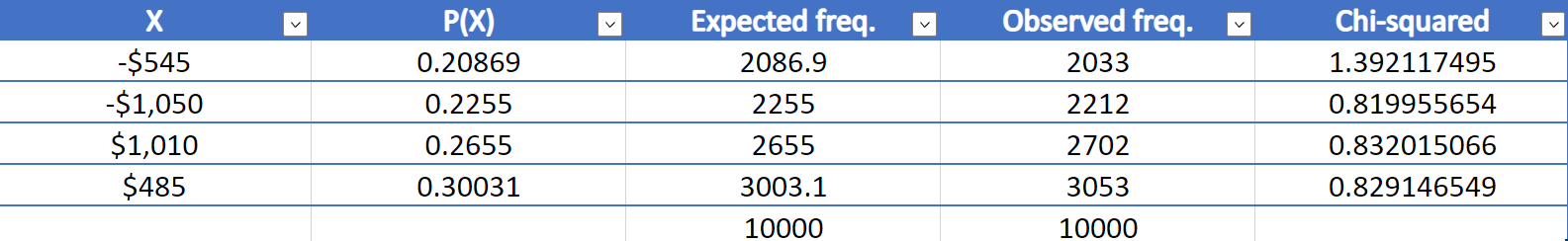


Fig6. Table showing the chi squared values for each net win case

Chi-square estimate= 1.228729922

p-value= 0.746122256

Here p-value >0.05 therefore we cannot reject Null hypothesis.

Therefore, The sample comes from the above net win probability distribution.

**In the second part of our analysis we will do everything similar to first part the home and away venue will be altered,**

Now,

Match1 is in New York

Match2 is in Boston

Match3 is in New York

Doing similar analysis as part one,

We have,

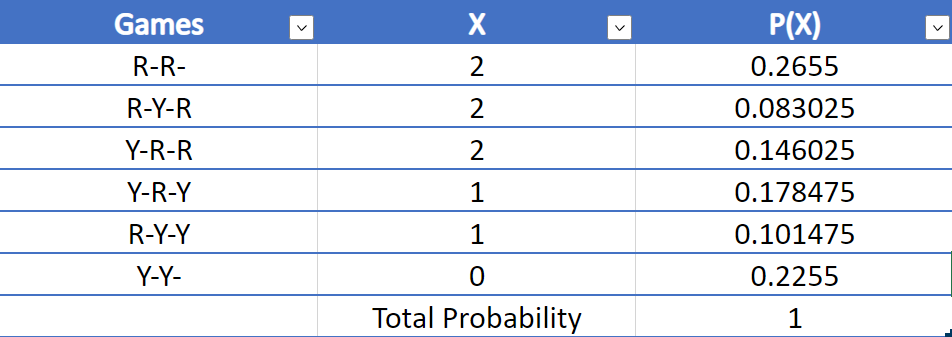


Fig7. We can see that the probabilities have changed from Fig3 as the win probabilities are different for both teams home and away.

**Probability that Red Sox will win the series**= 0.49455

This time the probability does not favour our idea of betting on Red Sox wins for every match as the probability of Red Sox winning the series is less than that of Yankees.

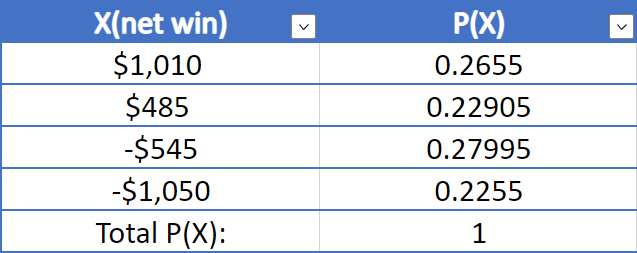


Fig8. Probability distribution of net wins when first match is played in New York

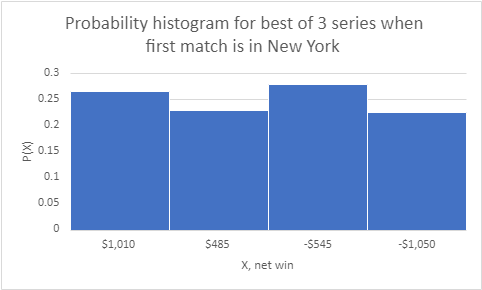


Fig9. Histogram of probability distribution

**Expected Mean**= E[X]= -10.1035

**Expected SD**= 810.1719906

Here, we got a negative average net win which does not favour our idea of betting on the Red Sox win for every match.

Similar to first part random values were generated for X,

The observed parameters from the sample were as follows,

**Observed Mean**= -13.1865

**Observed SD**= 810.6933922

**The 95% confidence interval** are found to be (-29.06557923 , 2.692579228).

The expected value of X lies in the interval.

Then Chi-square goodness of fit test was conducted,

Null hypothesis: The sample comes from the net win probability distribution

Alternate hypothesis: The sample does not come from the net win probability distribution

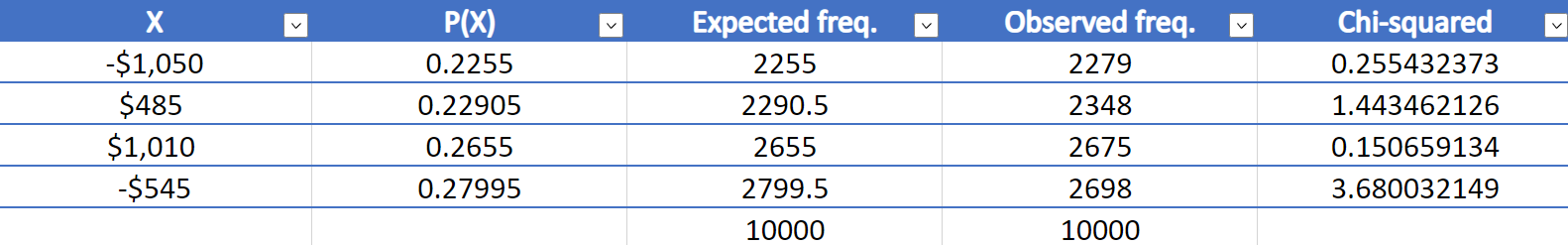


Fig10. Chi square values for the sample in 2nd part where the first games is in New York

Chi-square estimate= 5.529585781

p-value= 0.13687973

Since, p-value> 0.05, so we cannot reject Null hypothesis

Therefore, The sample represents the net win probability distribution in Fig8.

**Now proceeding to part 3 of the analysis,**

Here the series will be changed to a best of 5 series,

Match1 is in Boston

Match2 is in New York

Match3 is in Boston

Match4 is in New York

Match5 is in Boston,

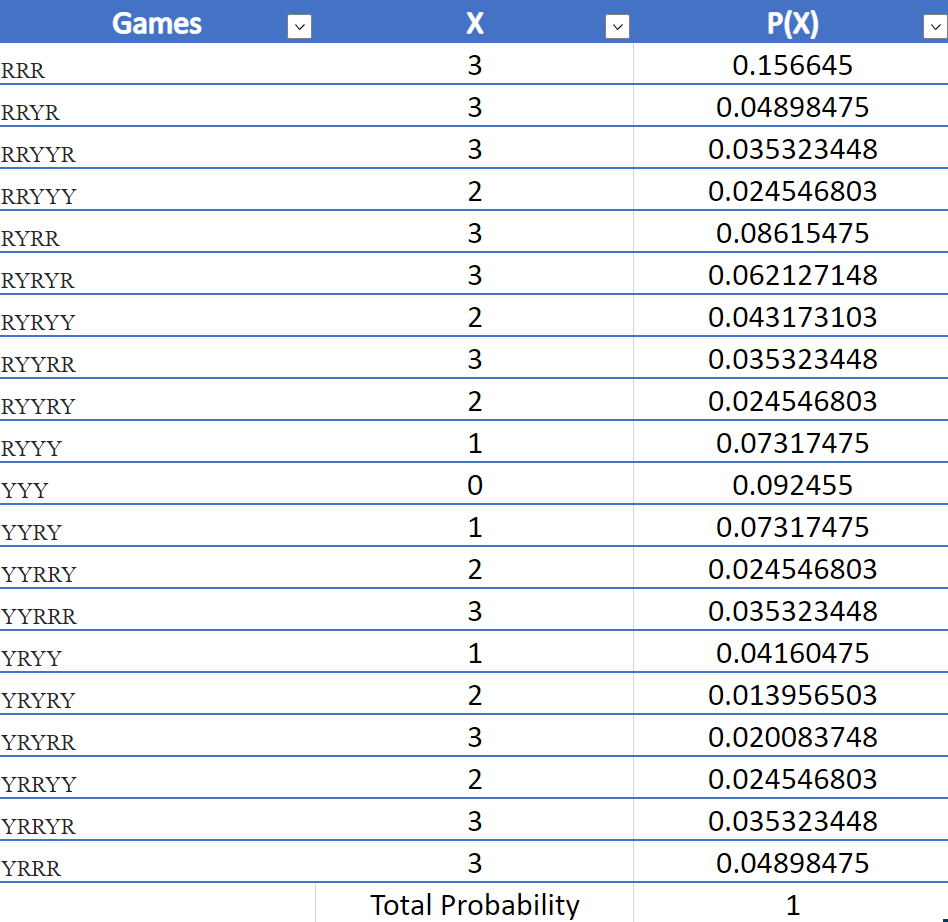


Fig11. Here we will have 20 combinations of games. Here X means no. of Red Sox wins

**Probability of Red Sox winning the series**= sum(P(X=3))

= 0.564273935

Probability distribution of net win from betting is as follows,

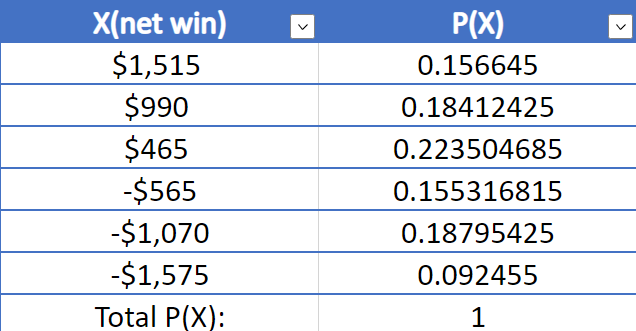


Fig12. Probability distribution of net wins for best of 5 series

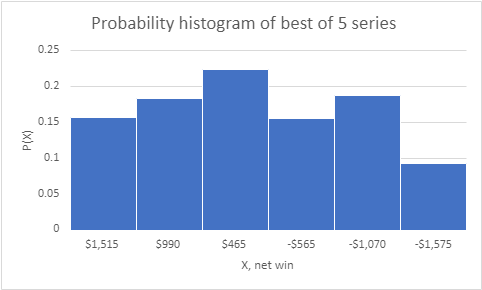


Fig13. Histogram of probability distribution

**Expected Mean**= 89.04818805

**Expected SD**= 1036.585469

In this case too 10000 random values were generated,

**Observed Mean**= 84.738

**Observed SD**= 1035.507701

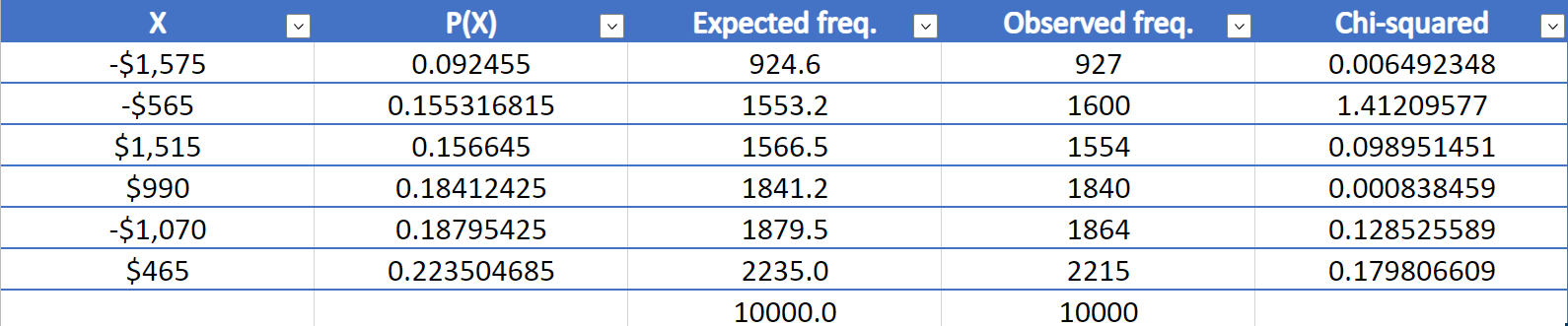
**95% confidence interval**= (64.42129815, 105.0547019)

The expected value of X lies in the interval.

The Chi-square goodness of fit test was conducted,

Null Hypothesis: The observed values come from the net wins probability distribution in Fig12.

Alternate Hypothesis: The observed values do not come from the net wins probability distribution

 Fig14. Table containing chi square values

Chi-square estimate= 1.826710225

p-value= 0.872565554

p-value not < 0.05, so cannot reject null hypothesis

Therefore the sample represents the net win probability distribution in Fig12.

**Conclusion**

From the above analysis it can be concluded that in the scenario of best of three matches where the first game is in Boston then we can probably expect a good gain from betting strategy of betting on Red Sox winning whereas when the first game is in New York then our betting strategy will not give a goo net win but more likely give us a loss. In case of best of five series we can again expect a good gain from our betting strategy of betting on Red Sox win.