A Demonstration of the Difference Between Binomial and Hypergeometric in Statistics

Yvanny Chang
Prospect High School
San Jose, CA 95129, USA
yvanny.chang@gmail.com

Abstract

This STEM project is to use JAVA random simulation to differentiate statistics between Binomial (with replacement) and Hypergeometric (without replacement). The author used the Powerball lottery probability to determine the Jackpot probability by matching the numbers of five balls and the Mega number. The Mega number of the red ball can be duplicated the same number of any white ball picked already. There is about 12% higher probability for Jackpot if the Mega number cannot duplicate the ball of any white ball picked. The author derived the JAVA script to simulate two probability scenarios. Binomial Approximation of Hypergeometric Probability by adjusting the sample size in the pool was also studied. When the pool size is larger, the probability difference between whether the Mega number can duplicate the white ball number becomes smaller. The ratio of two probabilities will approach 1 when the pool size becomes larger. JAVA script can demonstrate this Probability Approximation pattern. This STEM project is successful since the author can integrate learning across Statistics, and JAVA as a Data Science approach.

Keywords
Binomial, hypergeometric, Java, powerball, probability

1. Introduction
In statistics, binomial distribution allows duplicates, while hypergeometric distribution cannot create duplicates. The aim of the project is to find examples of binomial and hypergeometric distribution in real life. The example used was the Powerball lottery to calculate the differences between the two distributions. On January 13, 2016, Powerball created the largest jackpot in U.S. history of 1.5 billion as shown in Figure 1. The rule of matching with the red ball is following the binomial distribution, and the white ball is following hypergeometric distribution.

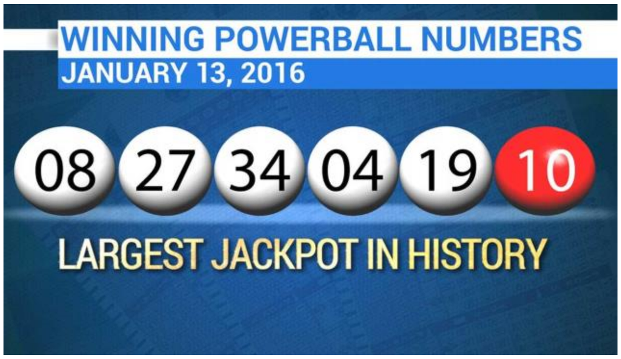

Figure 1. Powerball Jackpot

2. Objective

The purpose was to compare the difference between binomial and hypergeometric distribution. To do this, java was used to calculate the probability of matching, using both binomial and hypergeometric. In addition, different probabilities of matching were also tested. There are two main objectives in this project. The first is to match the winning probability by doing the calculations by hand. The second objective was to test multiple scenarios using Java.

3. Method

Java was used to calculate the difference. It was used for many reasons, but the main reason is that Java can be used to calculate matching probabilities very quickly. After the code for calculating the probability of matching is complete, all that had to be done was change two variables. (Refer to **Figure 3**) By using java, we can easily calculate differences in short amounts of time just by changing two variables. We used X and Y in the java program to easily calculate different variables.

3.1 Hand Calculation

For X number of white balls, Y number of red balls, the probability of the following winning chance will be studied:

- 5+1: Match 5 white balls and 1 red ball
- 5+0: Match 5 white balls and no red ball
- 4+1: Match 4 white balls and 1 red ball
- 4+0: Match 4 white balls and no red ball

Taken the example of "5+1" case, the total space= C(X, 5) * C(Y, 1). The probability of matching all 5 white balls is C(5,5)=1; the probability of matching 1 red ball is C(1,1)=1. Thus, the total probability of 5+1 is **1** in C(X, 5) * C(Y, 1).

For another example of "4+0" case, the total space= C(X, 5) * C(Y, 1). The probability of matching 5 white balls = C(5,4) * C(X-5,1). The probability of matching 0 red ball=C (Y-1,1)=Y-1. So, the total probability of "4+0" is 1 in C(X,5)*Y/(C(5,4)*C(Y-5,1))/(Y-1).

By applying the same method, we can calculate for any matching scenario by hand. However, it is not the most efficient method.

3.2 Java Program

Instead of calculating everything by hand, writing a Java program to calculate the matching probabilities of different matching scenarios makes it much more efficient.

```
public static double FiveAndOne(int X, int Y){
        return combination(X,5)*Y;
}

public static double FourAndZero(int X, int Y){
        return combination(X,5)*(Y/(Y-1))/(combination(5,4)*combination(Y-5,1));
}
```

This part of the Java program calculates the matching scenarios in the Powerball Lottery.

This is how the Java program works. Our program allows to easily switch between binomial (y=x) and hypergeometric (y=x-5).

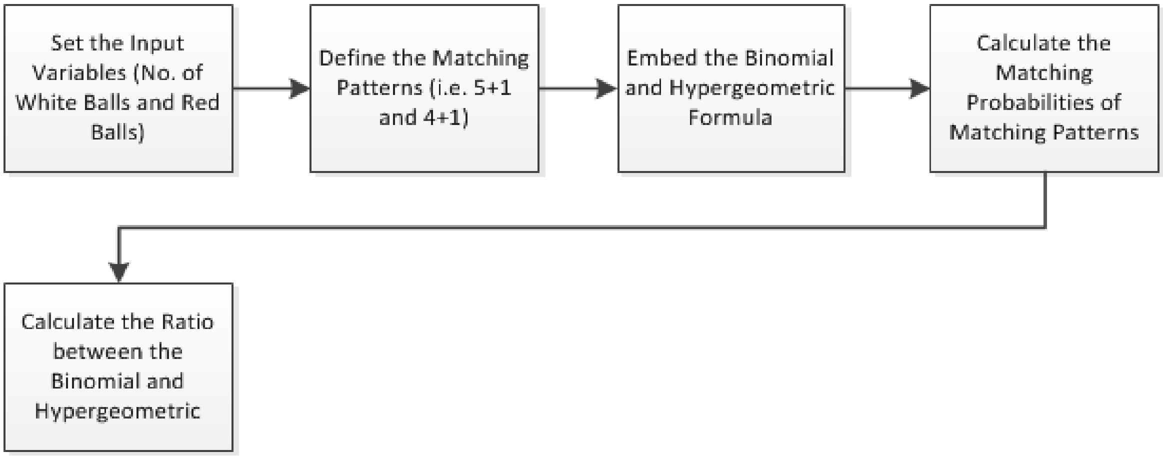

Figure 2. JAVA flowchart.

4. Result

Figure 3 shows that the probability of matching calculated using Java successfully matches the official results. To determine which is better, the ratio between probability of matching with binomial and hypergeometric was calculated. These are the matching probabilities of the winning Powerball combinations. The greater the chances of winning, the smaller the prize. Our results from the Java program matched all numbers perfectly.

Figure 4 shows the ratio between the probability of matching with binomial and hypergeometric shows how much higher the probability of matching with hypergeometric as opposed to binomial. For each X, there are two Y values. The Y values correspond to binomial and hypergeometric distribution. The ratio between binomial and hypergeometric distribution for the jackpot (5+1) and 4+1 case shows that as pool size decreases, probability of matching with hypergeometric increases from 9-20%. At certain points, binomial and hypergeometric have the same chance of matching as well as hypergeometric having a guaranteed chance of matching. Our hand calculations of the matching probability also matched our java calculations.

Figure 5 plots our ratios between binomial and hypergeometric distributions. The X-axis is the number white balls and the Y-axis is the ratio. Trend shows that as the X-axis increases the ratio of matching decreases, approaching 1. That means for larger numbers, we can use hypergeometric to approximate binomial results.

NUMBERS MATCHED	LIKELIHOOD
5 white and red	1 in 292,201,338
5 white	1 in 11,688,053.52
4 white and red	1 in 913,129.18
4 white	1 in 36,525.17
3 white and red	1 in 14,494.11
3 white	1 in 579.76
2 white and red	1 in 701.33
1 white and red	1 in 91.98
Red	1 in 38.32
Nothing	1 in 1.04

Figure 3. Likelihood of winning

X	Y	5+1	4+1
59	59	1.09	1.09
	54		
55	55	1.10	1.10
	50		
50	50	1.11	1.11
	45		
45	45	1.13	1.13
	40		
40	40	1.14	1.14
	35		
35	35	1.17	1.17
	30		
30	30	1.20	1.20
	25		

Figure 4. Ratio of binomial and hypergeometric

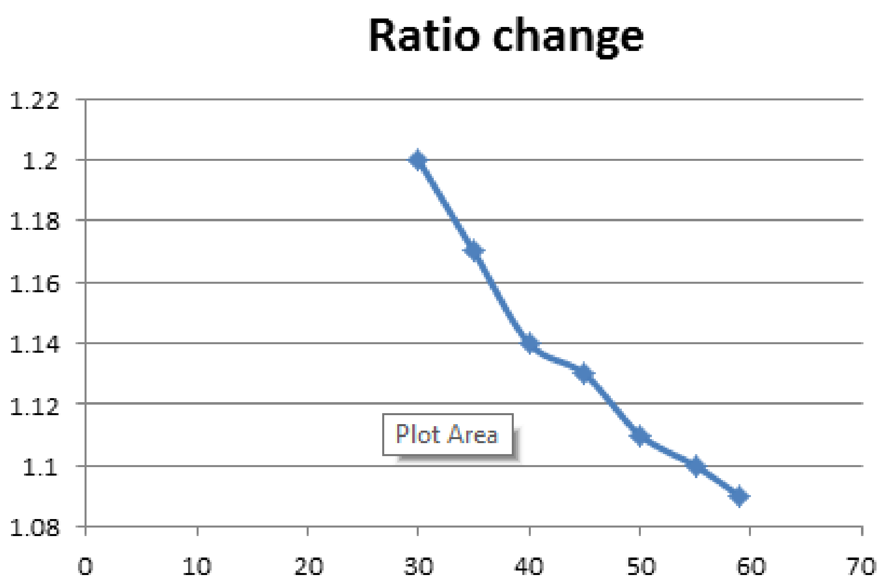

Figure 5. Ratio change in probability of matching.

5. Conclusion

This project used JAVA random simulation to define the difference between Binomial (with replacement) and Hypergeometric (without replacement). The Powerball lottery was used to determine difference between the two distributions. A Java program was used to simulate two probability scenarios. The calculations of matching probability done by hand also matched the calculations done with java. Different scenarios were also studied by varying pool size. There is a 9-20% higher probability of matching if using hypergeometric distribution. When the pool size is larger, the probability difference between binomial and hypergeometric distribution becomes smaller. The ratio of two probabilities will approach 1 when the pool size becomes larger. This STEM project is successful since the author can integrate learning across Statistics, and JAVA as a Data Science approach.

References

Feller, William. On the normal approximation to the binomial distribution. *Selected Papers I. Springer International Publishing*, pp 655-665 2015

Loertscher, Simon, Ellen V. Muir, and Peter G. Taylor. A general non-central hypergeometric distribution. *Communications in Statistics-Theory and Methods 46.9* pp 4579-4598. 2017

Kiersz, Andy. We Did the Math for the $450 Million Powerball Jackpot and Concluded It's Not worth Buying a Ticket. *Business Insider. Business Insider, Inc*, 05 Jan. 2016. Web. 13 Nov. 2016.

V. Ariyabuddhiphongs, Journal of Gambling Studies, March 2011, Volume 27, Issue 1, pp 15–33

Welte JW, Tidwell MC, Barnes GM, Hoffman JH, Wieczorek WF, J Gambl Stud. 32(2):379-90. doi: 10.1007/s10899-015-9551-0, 2016

Douglas C. Montgomery and George C. Runger. Play Overview | What's My Game Today? Powerball Has Changed. N.p., n.d. Web. 13 *Applied Statistics and Probability for Engineers* pp 26-27, pp 93-94, 2016

Biographies

Yvanny Chang is an incoming freshman to Prospect High School (as of 2017). He is 14 years old and this is his first paper. Yvanny Chang went to Easterbrook Discovery School from kindergarten to eighth grade. In his spare time, he likes to read, listen to music, and hang out with his friends.