Negative Binomial Distribution - World Series

This week we will be looking at the binomial distribution specifically the negative binomial distribution in the context of the world series. Much like a coin flip, the world series has two outcomes win or lose, this is where bi- two nomial-number, comes from in the name. So, before getting into the inner workings a bit, it is important to remember in probability that the sum of the each individual probability will give you the total of the series. The technical term for this in probability is this the law of total probability.

So generally when finding a probability you multiple the probability times the attempt. I know its a tired example, but back to the flipping a coin. The probability of getting heads 3 times is pretty obviously $.5^3$. However, what happens if we want to know the probability of getting 3 heads in 5 flips? Well we will use whats called a geometric distribution. This will let us figure out the possibility of getting n success in and r failures in k trials. The random letters there are just used as place holders for real numbers. In a geometric distribution you have $p^n * (1-p)^r$. This makes sense right? The coin example, barf, chance of getting 3 heads in 4 tries is the chance of 3 heads times the chance of getting one tail. Notice if we can assign a probability to a team winning the game then we can being to use some of our tools.

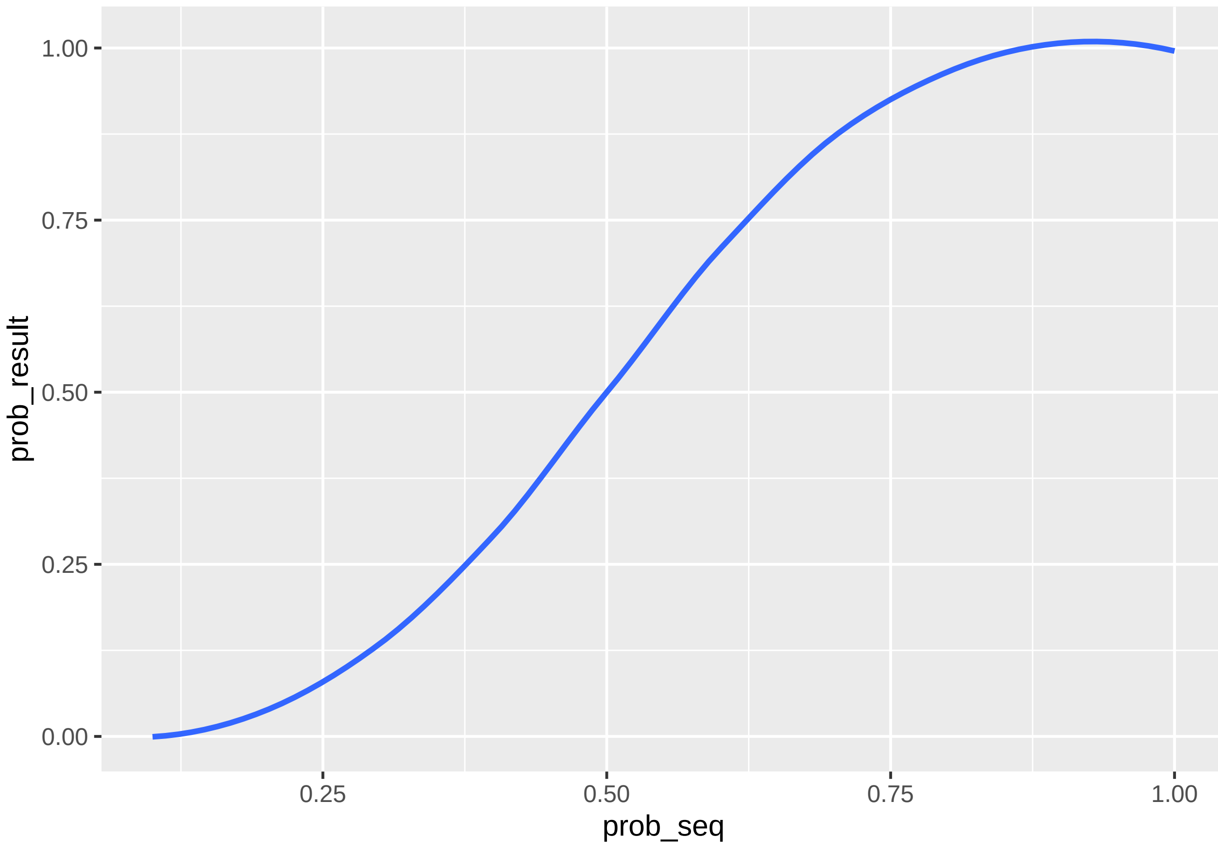
We do need a bit more because we need to find a way to find the sequence of games for playoffs. Math comes in clutch here as well! We can use the combinations formula to find all the combinations! I can give an example to help because the actual formula is a bit complex to get into here but, image we had three people in a room: me, you, and Euler. The total combinations are [me,euler], [me, you], [you, euler].

Now being the bright person that you are you realize so we just nee the combinations of 4 wins in 7 game series and the chance of getting 4 success then we just do the math! $7C4(p)^4 * p^3$ The nCr is just the fancy notation that we use to denote 7 choose 4. Unfortunately while a good idea there are two main issues here. The first is that you can win in less than seven games the second is that certain series are impossible for example win, win, lose, win, win, lose, lose because once get four wins the series is over.

We instead use whats called a negative binomial distribution. This finds the chance of the nth win in kth trial or a way in a bit more plain English what is the chance of getting your 4th win on your 4 games. Something important to note here is that in order to have your 4th win in 4 games you must of previously won all the other games. So, we can do 3C3, this idea can be generalized and we get what is mentioned above, the negative binomial distribution! We just find all the combinations that can result in our scenario time the chance of getting the scenario. So for the probability of winning the series on the 7th game would be $6C3 * p^4 * (1-p)^3$ because this is how we get our scenarios times the chance of winning that last game also finding the probability of having lost the other 3!

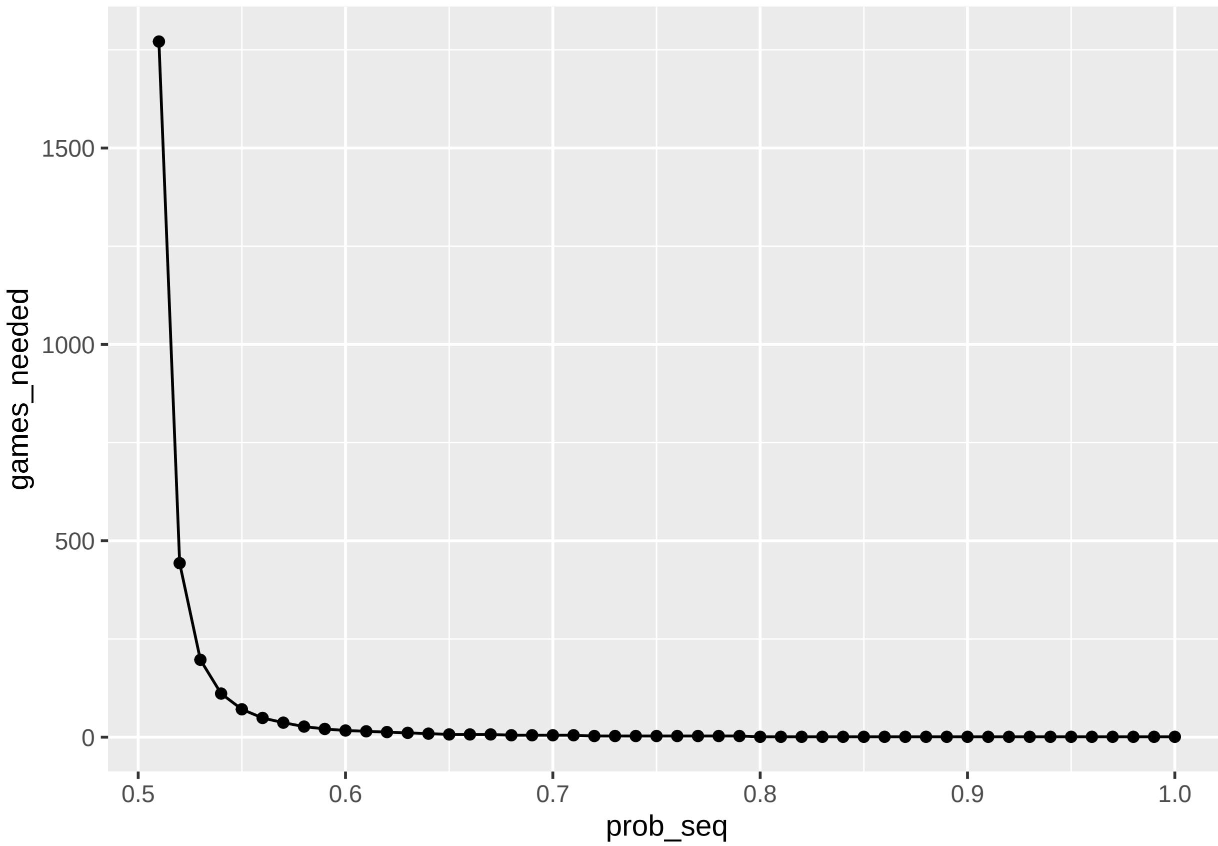
The code above uses the law of total probability and the negative binomial distribution to find the probability of the Yankees winning the world series if they have a 55% chance to win each game. The total probability can be found by adding up all the smaller ones. In this case that will be the chance of winning the 4th in 4 games, the 4th in 5 games, the 4th in 6 games and the 4th in 7 games. When you add all those up you get roughly 61%! So, if the Yankees have a 55% chance of winning each game they have a 61% chance of winning the series.

What if you want to find the probability of winning giving an arbitrary probability? Well we can basically do the same thing. However, to make the code a bit easier we use a function to give us the probability of adding all those previous games up like we did above manually. This is known in probability as a cumulative mass function. The name sounds fancy but remember cumulative just means all the previous. The graph is below!



Now, you ask what if we want to find the number of games needed for a certain probability! Well, never fear, math and code can do that as well! If you look below I have included the relevant graph. The math is the same as above we are just relaying on the computer to find the number of games until we reach a certain amount then the code stops, known as a while loop. The games of is actually 71!

However, now you think you got me. What if you wanted to know the shortest series length given some arbitrary probability? Well once again math and code prevail. We use the same tactic as above but we will put it in another piece of code called a for loop that changes in a range which will be the range of arbitrary probabilities that you want until the while loops gets the correct answer, you move along the for loop and then the whole thing starts again until it we run out of items in the for loop. The graph is below.



Lastly, lets do something really tricky what if you wanted to find the probability that the probability the braves win 55% of there games in 7 games when you know the braves either have a 55% or 45% chance to win. So in order to find this we are going to combine everything from above. The first step to recognize is that we know the braves win the series in 7 games with either a 45% chance to win or a 55% chance to win. We also know we want to find the probability of them having a 55% chance to win. Well, if you ask me the bottom sounds a bit like the law of total probability, each outcome has a 50% chance of happening and each event looks like its a negative binomial distribution of winning the 4th game on the 7th try. So for each of those that will be $6C3 * .55^4 * .45^3 * .5 + 6C3 * .55^3 * .45^4 * .5$ I know that looks complex, but remember from above this is simply all the combinations we can times the probability of those outcomes occurring. So, we have our denominator, all we need is our numerator which is simply the chance of them winning with a 55% chance at each game. Well, all that calculation is simply $6C3 * .55^4 * .45^3 * .5$ you then divide your numerator by your denominator and you get your result which in this case is roughly 55%!