Probability Project

Jack Doughty

April 2022

1 Mario Kart Custom Tracks

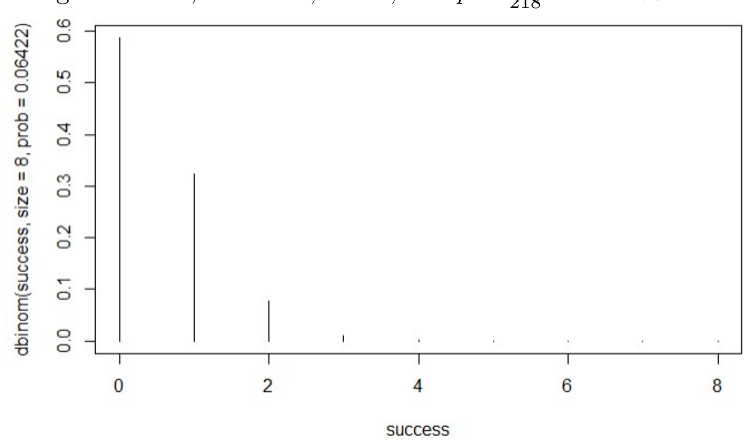
Mario Kart Wii CTGP has 218 custom tracks, of which 14 were made by the author Sniki (https://www.youtube.com/channel/UCEhmfD0ms0YFPmQsckPyn6Q). For our problems, we are going to let

1.1 Binomial Distribution and its PMF

The PMF for any binomial distribution is as follows:

$$P(x) = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, 2...n$$

Using this PMF, let $n = 8$, $x = 2$, and $p = \frac{14}{218} = 6.422\%$.



1.2 Calculating specific values

1.2.1 Mean

For the mean we know for a binomial distribution it is simply $n*p = 8*0.06422 =$ **0.5138** which means we would estimate 0.5138 tracks made by Sniki would be picked on average every time 8 tracks are picked in the same manner.

1.2.2 Standard Deviation

Our standard deviation is $\sqrt{n * p * (1-p)} = \sqrt{8 * 0.06422 * 0.93578} = \sqrt{0.481} =$ **0.693** which implies that we can be expected to vary with 0.693 of our expected mean 0.5138 (which makes a lot of sense considering the majority of our expected outcomes are 0 or 1 tracks being selected).

1.2.3 Median

By the definition of median, we know that 1/2 of all possible scenarios need to be less than or equal to our median value and the other half needs to be greater than or equal to. However to the best of my understanding median values for binomial distributions must be integers. Referencing our earlier graph, it's evident most of our scenarios result in 0 or 1 track by Sniki being played. While we could solve this with using the equations:

$$P(x \geq a) = 0.5 \; P(x \leq a) = 0.5$$

An easier solution would be to just utilize R and the *qbinom* function with $p = 0.5$. Using this, we get a median value of **0**.

1.2.4 75th Percentile

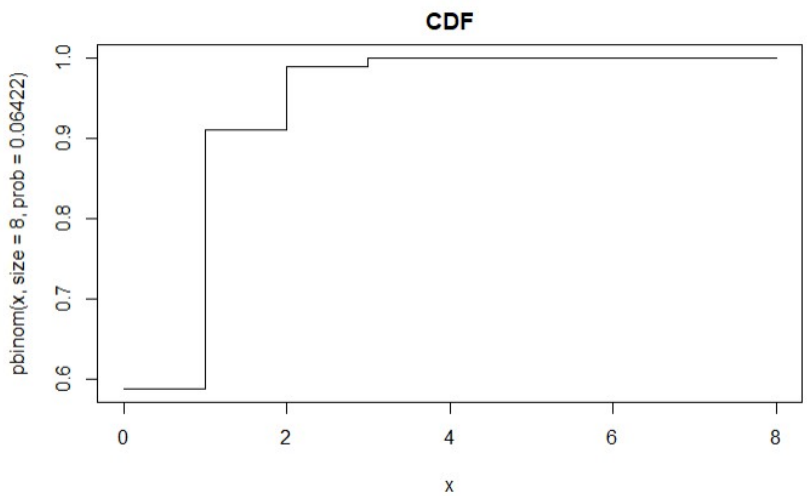
Similar to the median, we can use R and the *qbinom* function with $p = 0.75$. This time however we get a value of **1**, which makes a lot of sense when referencing our earlier graph as roughly 75% of the time we play 1 or less Sniki tracks.

1.3 Our Cumulative Distribution Function

The general form for the binary distribution CDF is as follows:

$$F(x; n, p) = P(X \leq x) = \sum_{i=0}^{x} f(i; n, p)$$

Using this and substituting in our proper values we get a graph here for our CDF:



1.4 Scenarios

1. In a tournament competition, there are 8 races played. Assuming tracks are selected at random with equal probability and they can be re-picked. We can rewrite 2 scenarios mathematically as follows:

   a. $P(X = 2)$. This is the probability that in our 8 race tournament exactly 2 tracks made by Sniki are played. Using our PMF with $x = 2$ we can solve this.

$$P(x) = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$P(2) = \binom{8}{2} 0.06422^2 (1 - 0.06422)^6$$
$$P(2) = 28 * 0.06422 * 0.6284$$
$$P(2) = 0.07754 \text{ or } \mathbf{7.754\%}$$

   b. $P(X \geq 1)$. This is the probability that in our 8 race tournament there is at least 1 track made by Sniki played.

   The easiest way to solve this would be to calculate the probability that 0 tracks made by Sniki are played, and subtract that value from 1. This is because the only scenarios where there are not at least 1 track made by Sniki are played is when 0 are made.

$$P(x) = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$P(0) = \binom{8}{0} 0.06422^0 (1 - 0.06422)^{8-0}$$
$$P(0) = 1 * 1 * 0.9358^8$$
$$P(0) = 0.5880$$
$$P(x \geq 1) = 1 - P(0)$$

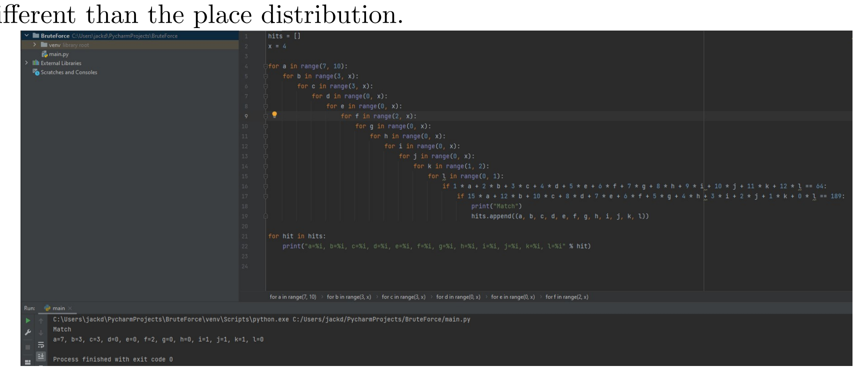
$$P(x \geq 1) = 1 - 0.5880$$
$$P(x \geq 1) = 0.4120 \text{ or } \mathbf{41.20\%}$$

2 DS Wario Stadium

My favorite track in Mario Kart Wii, and definitely my best Track is DS Wario Stadium. I have played the track 18 times in the lounge format (12 player events), which means our outcomes are going to be the final position of the race which ranges 1st through 12th. While I cannot seem to find the exact distribution of my results in these 18 races, the API says I have an average score of 10.50 and an average place of 3.56. We want to try and find the probability of me getting a certain place or better each race. Lets make a PMF for the probability of me getting 1st at least 2 times in 4 races.

2.1 PMF

Since we have no named distribution here, it will require a bit more work in order for us to create a PMF. We can start with using our average place values. With this, we can use $E|X| = \sum_x xp(x)$ and substitute in to get $3.56 = \sum_x xp(x)$. We can use this and brute force calculate all the possible values for each of our races, along with using the average score since the score distribution is significantly different than the place distribution.

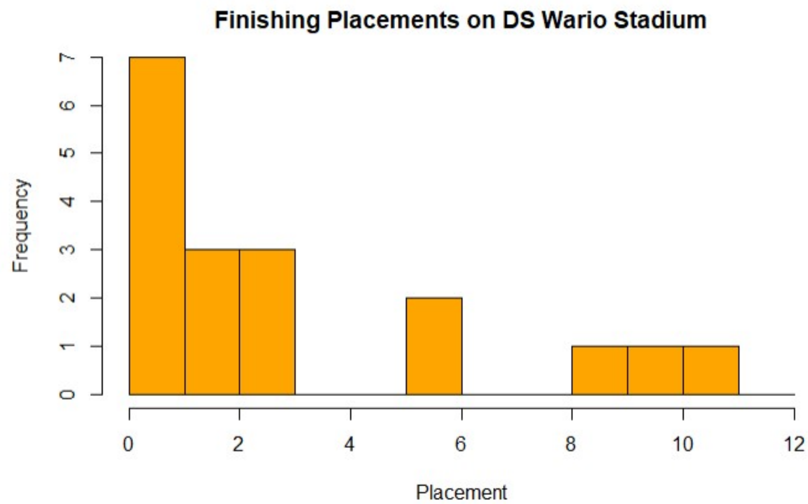


As you can see here we found our distribution of races to be as follows:

1st: 7
2nd: 3
3rd: 3
4th: 0
5th: 0
6th: 2
7th: 0
8th: 0
9th: 1
10th: 1

11th: 1
12th: 0

From here we are able to make a figure of our distribution:



If we want to determine the probability of me getting 1st at least 2 times in 4 races, we can use the binomial PMF with this unnamed distribution.

$$P(x) = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, 2...n$$
$$P(2) = \binom{4}{2} 0.3889^2 (1 - 0.3889)^2$$
$$P(2) = 6 * 0.1512 * 0.3735$$
$$P(2) = 0.3388 \text{ or } \mathbf{33.88\%}$$

2.2 Calculating specific values

2.2.1 Mean

This was already supplied and used to calculate the data set, our mean is **3.56**.

2.3 Standard Deviation

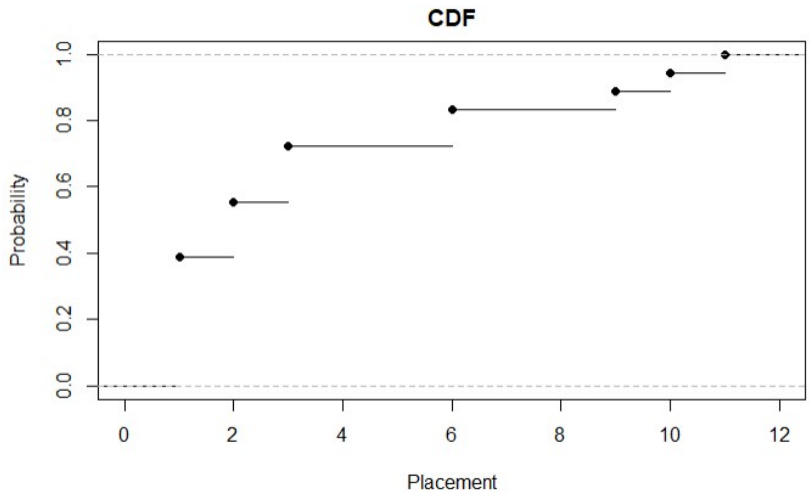
We can calculate this with the dataset we have in r, which comes out to **3.365** as our standard deviation.

2.4 Median and 75th Percentile

For both of these we can use the graph produced and our data set. For the median we know it lies between our 9th and 10th value, which happens to give us **2**. As for our 75th percentile it's quite obviously 1 because we would say 1 is the 100th percentile, not 12. 1 makes up over 38% of our data, so therefore it is also the 75th percentile.

2.5 Our Cumulative Distribution Function

Fortunately since we already have our data in r, we just can use the *ecdf* function and plot it to get the graph you can see here.



2.6 Scenarios

1. Assume I play DS Wario Stadium 12 times in a row.

   a. $P(X = 2forx > 8)$. What are the odds that I get a spot below 8th exactly 2 races?

$$P(x) = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$P(2) = \binom{12}{2} 0.1765^2 (1 - 0.1765)^{10}$$
$$P(2) = 66 * 0.03115 * 0.1435$$
$$P(2) = 0.2950 \text{ or } \mathbf{29.50\%}$$

   b. $P(X = 10, x < 3)$. What are the odds I get 3rd or better all but 2 races?

$$P(x) = f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$P(10) = \binom{12}{10} 0.765^1 0(1 - 0.765)^2$$
$$P(10) = 66 * 0.0684 * 0.0553$$
$$P(10) = .2499 \text{ or } \mathbf{24.99\%}$$

3 R Work

If you would rather have the actual file I would be more than willing to send it, but hopefully this screenshot will suffice.

Probability Project.Rmd

````
---
title: "Probability Project"
output: html_document
---

```{r}
success <- 0:8
plot(success, dbinom(success, size=8, prob=.06422),type='h')
median(dbinom(success, size=8, prob=.06422))
qbinom(p = 0.5, size = 8, prob = .06422)
qbinom(p = 0.75, size = 8, prob = .06422)
```
```{r}
placements <- c(7, 3, 3, 0, 0, 2, 0, 0, 1, 1, 1, 0)
actplac <- c(1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 6, 6, 9, 10, 11)
hist(actplac, main = "Finishing Placements on DS Wario Stadium", col = "orange", xlab = "Placement", breaks=0:12)
```
```{r}
sd(actplac)
median(actplac)
plot(ecdf(actplac), main = "CDF", ylab = "Probability", xlab = "Placement")
```


````