A Statistical Analysis of Flappy Bird Scores

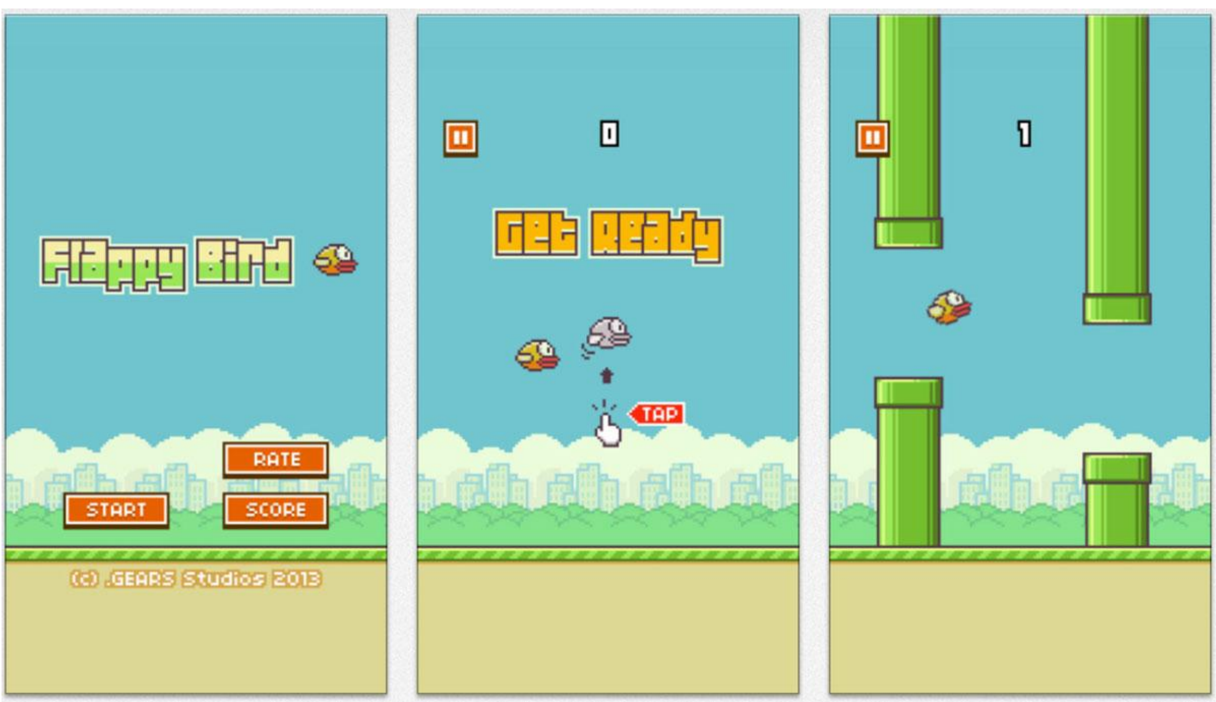
Anthony Pan and Albert Kuo

Flappy Bird is a game for Apple's iOS playable on the iPhone or iPad. The game consists of flying through a series of tubes for as long as possible until the player crashes into a tube and loses the game; the score obtained is the number of tubes that the player passes through successfully.[1] We will view each tube as a Bernoulli random variable:

- p=the probability of failing to pass through a tube
- (1-p)=the probability of not passing through one tube

If we assume that the tubes are a collection of independent and randomly distributed Bernoulli random variables—passing one tube does not reveal anything about passing any other tubes—then the game appears to take on a geometric distribution. The score of the game is the number of times that the player passes the tube, (1-p), before the first collision with a tube (p).

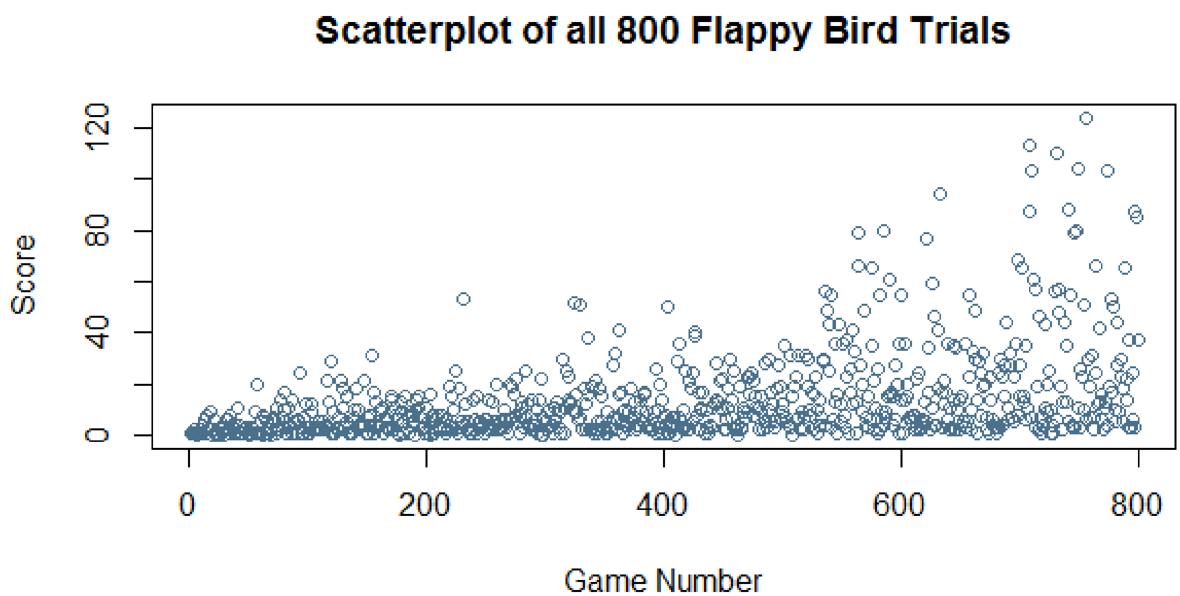
In this analysis, we will be examining the Flappy bird scores with empirically collected data to compare the distribution of scores to a geometric distribution.

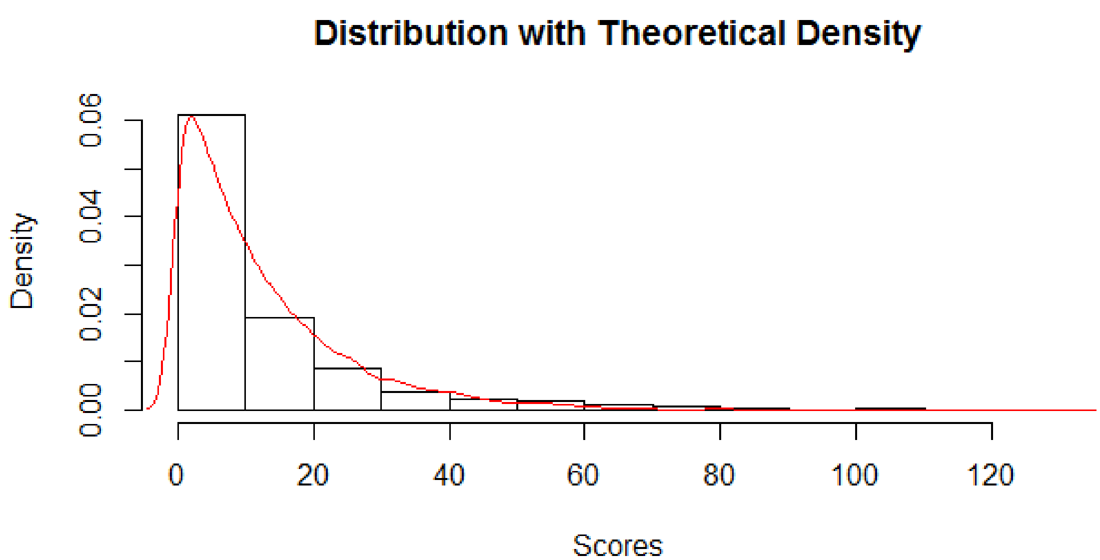


[1] A video of Flappy bird gameplay can be found at: http://www.youtube.com/watch?v=B7K4bCYV_tY

Albert's Data

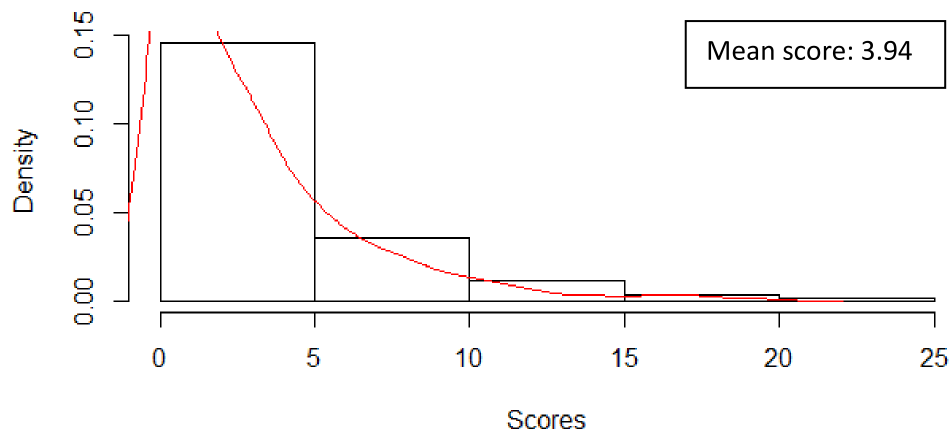
Albert has played 800 Flappy Bird games in total, and has recorded his scores for all 800 games, capturing not only trials from a seemingly geometric distribution but also his individual learning curve as time went on. As we can see by the scatterplot below, the scores have an upward trend that can be explained by the fact that each time Albert finishes a game, he becomes better due to experience. We have also plotted a histogram of scores with a theoretical geometric density to illustrate the similarity between the data set and a true geometric distribution.



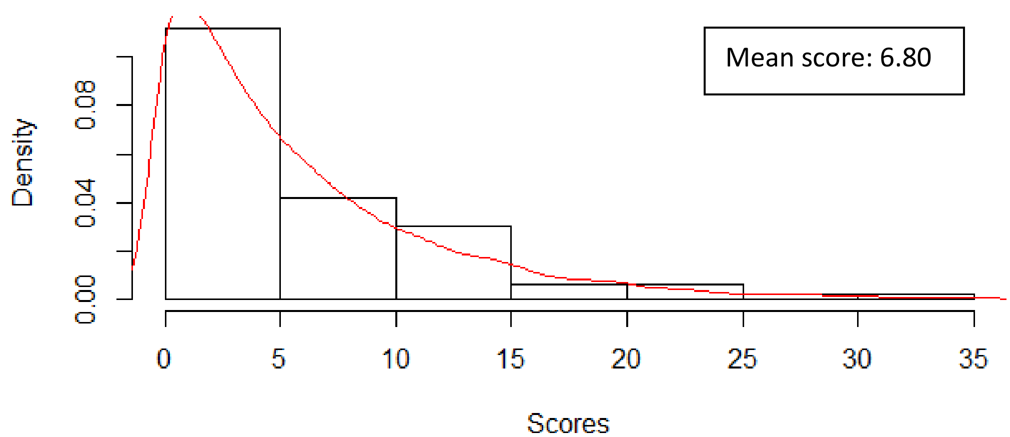


The upward trend in scores caused by the learning curve of the game suggests that the games are not independent and identically distributed random variables. To minimize the effects of the learning curve, we have decided to split the scores up into eight sections of a hundred scores each, and examine each section in comparison to a true geometric distribution. Histograms of these scores with the corresponding theoretical density of Geo($\bar{x}$) are shown below.

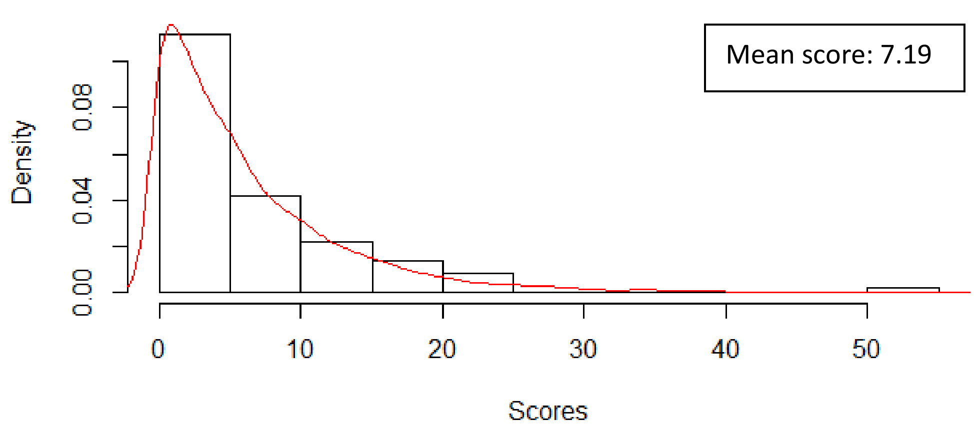
Distribution with Theoretical Density for Scores 1-100

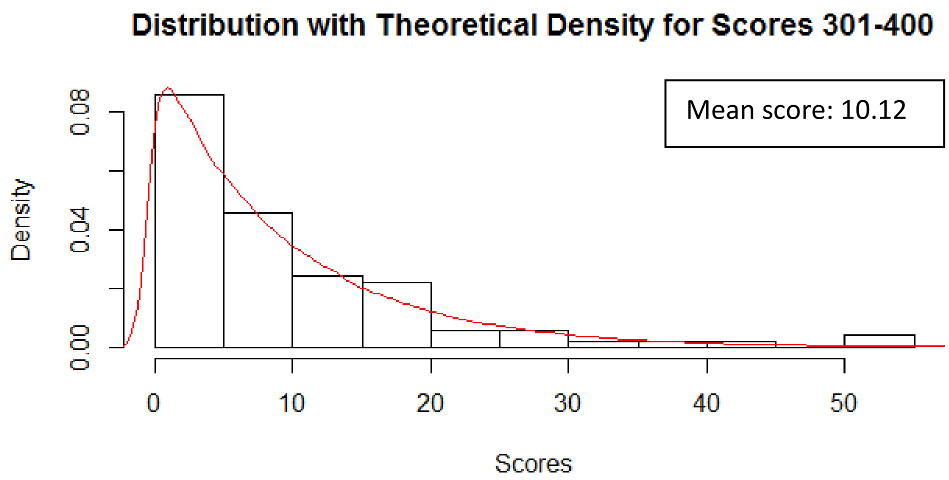


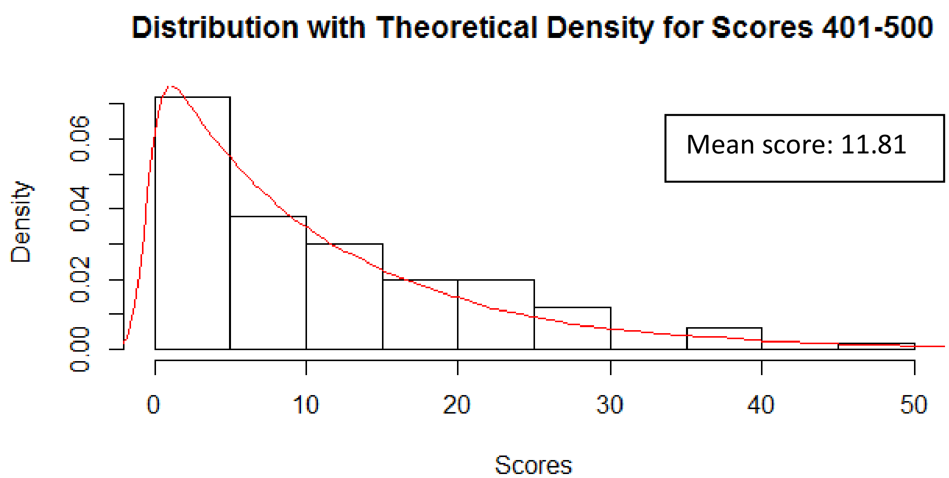
Distribution with Theoretical Density for Scores 101-200

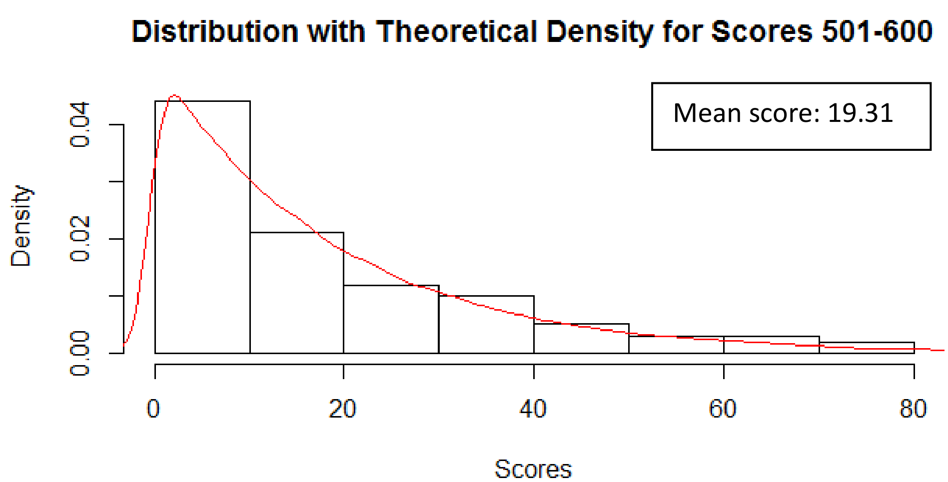


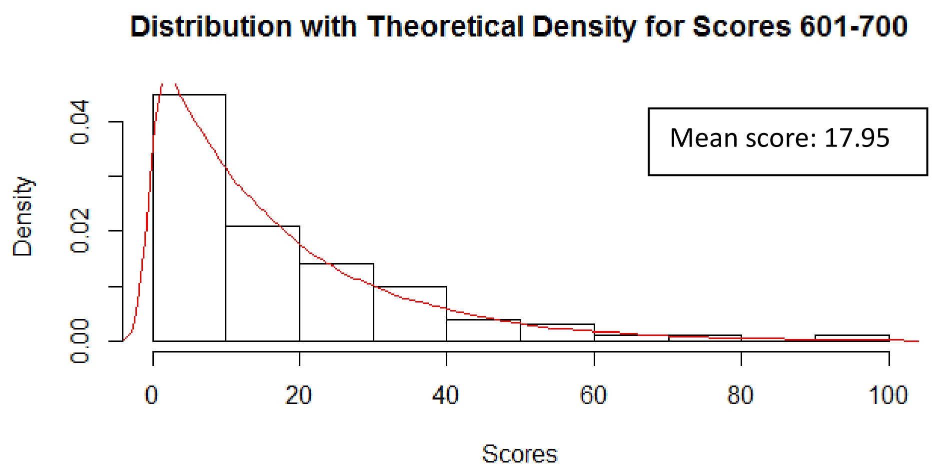
Distribution with Theoretical Density for Scores 201-300

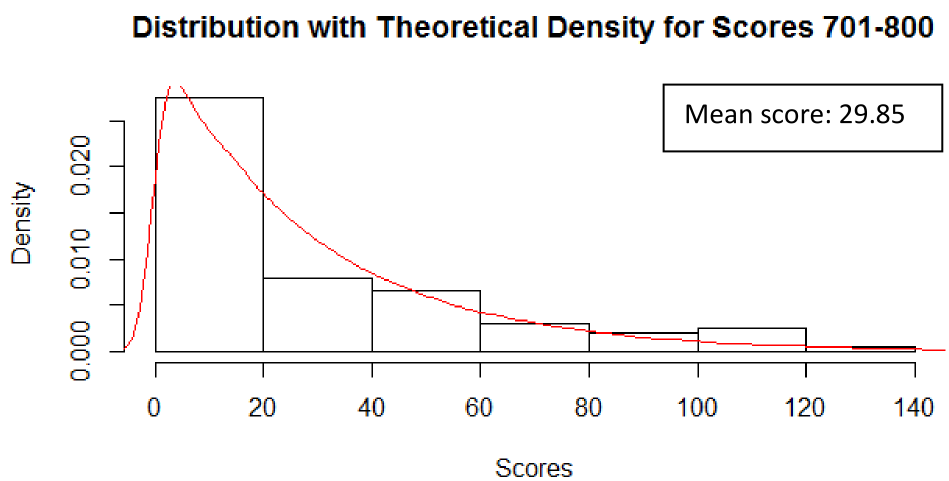



Distribution with Theoretical Density for Scores 301-400
Mean score: 10.12
Density
0.00
0.04
0.08
0
10
20
30
40
50
Scores


Distribution with Theoretical Density for Scores 401-500
Mean score: 11.81
Density
0.00
0.02
0.04
0.06
0
10
20
30
40
50
Scores


Distribution with Theoretical Density for Scores 501-600
Mean score: 19.31
Density
0.00
0.02
0.04
0
20
40
60
80
Scores





From the above graphs, we can see that a true geometric distribution with the observed average as a parameter is a pretty close fit to the data collected.

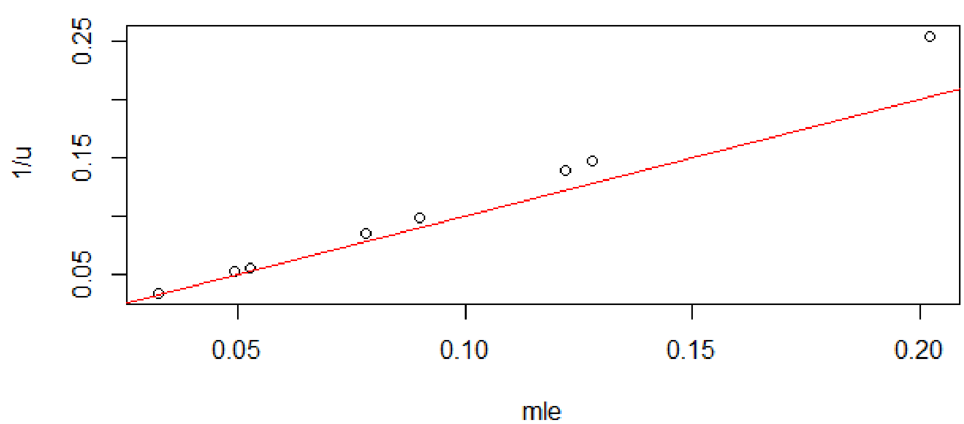
The maximum likelihood estimate (mle) of p (the probability of failing to pass through a tube) for all 800 trials is 0.0696, which means that using these 800 games as an estimator of Albert's true skill, he has an expected score of 14.37 with a 93.04% chance of passing through a tube.

In Albert's case, however, it is more interesting to look at the mle for each of the 100 trials as he gets better at the game and the mean of his score (u) increases. A table of these values is given on the next page.

Trials	Mle ($\hat{p}$)	E(X)= $\hat{u}$	$\bar{x}$	p_{obs} (1/ $\bar{x}$)	Estimated probability of flying through a tube (1- $\hat{p}$)
0-100	0.202	4.95	3.94	0.254	79.8%
101-200	0.128	7.812	6.80	0.147	87.2%
201-300	0.122	8.197	7.19	0.139	88.8%
301-400	0.0899	11.123	10.12	0.0988	91.01%
401-500	0.0781	12.804	11.81	0.0847	92.19%
501-600	0.0492	20.325	19.31	0.0518	95.08%
601-700	0.0527	18.975	17.95	0.0557	94.73%
701-800	0.0324	30.864	29.85	0.0335	96.76%

By games 701-800, Albert's sample mean score has increased to over six times the mean of his first 100 games. For the first 100 games, his sample mean score was about 5, while by the end his sample mean score was close to 31. Over the course of 800 games, Albert increased his estimated percentage of successfully flying through a tube by 16.96%, from 79.8% to 96.76%.

Plot of mle against 1/ $\bar{x}$

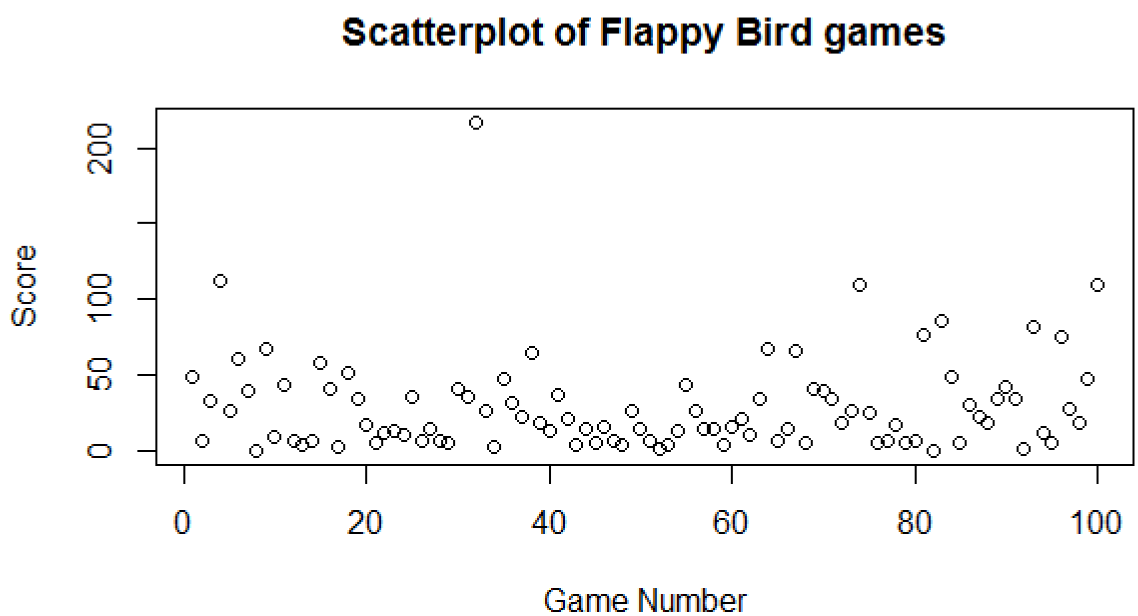


We can see that as the sample mean ($\bar{x}$) increases, 1/$\bar{x}$ gets closer to mle, which suggests that the corresponding geometric distribution becomes an increasingly better fit for the data as Albert played more games. This may be caused by a plateau in skill at the game, resulting in a more stable p value in the later trials. The realization of Albert's skill potential would imply that each game is independent from each other, while the earlier games were dependent on how much prior experience he had playing the game.

Anthony's Data

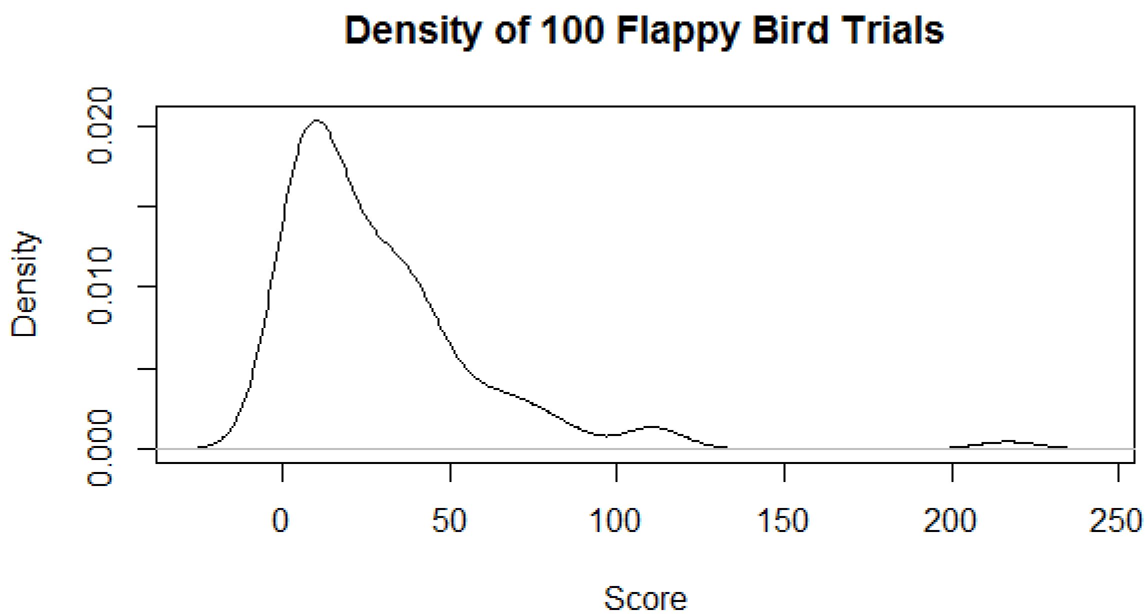
Anthony has recorded 100 trials of Flappy Bird. In contrast to Albert, when he began collecting data, he had already become familiarized with the game and had been playing for a while. We will assume that throughout these 100 trials of Flappy Bird, Anthony has already reached his skill potential, and does not get better at the game with each game played. Thus, we can assume that the games he played are independent and identically distributed.

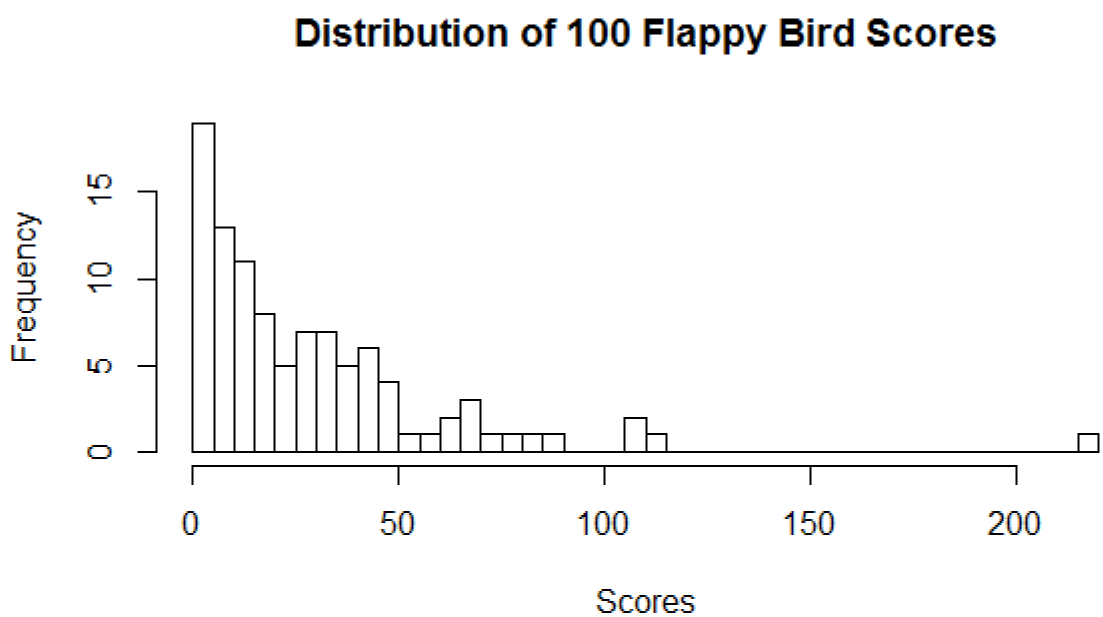
Below is a scatterplot of the 100 trials, with game number plotted against score. There does not seem to be an obvious upward trend, further supporting our assumption that Anthony has already reached his skill potential.



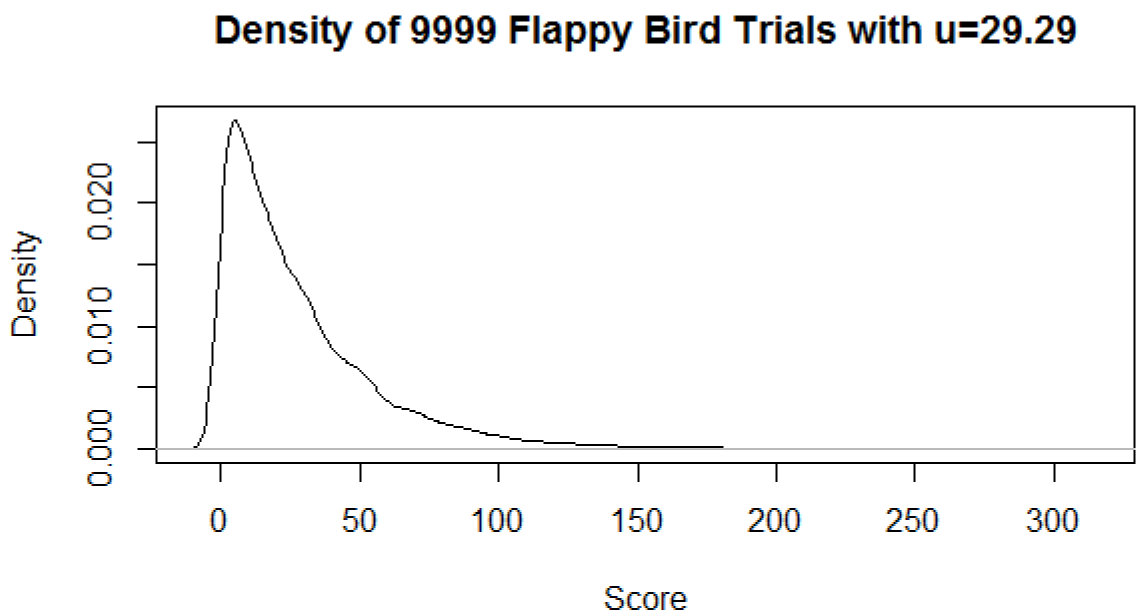
Mean:	29.29	**Min**:	0.0
Std. Dev:	31.7	**Q1**:	6.0
N:	100	**Q2**:	18.5
Skew:	2.68	**Q3**:	39.5
Kurtosis:	11.1	**Max**:	217
SE:	3.17		

Below is a histogram of the distribution and density of Anthony's 100 Flappy Bird scores with some basic information about the trials.

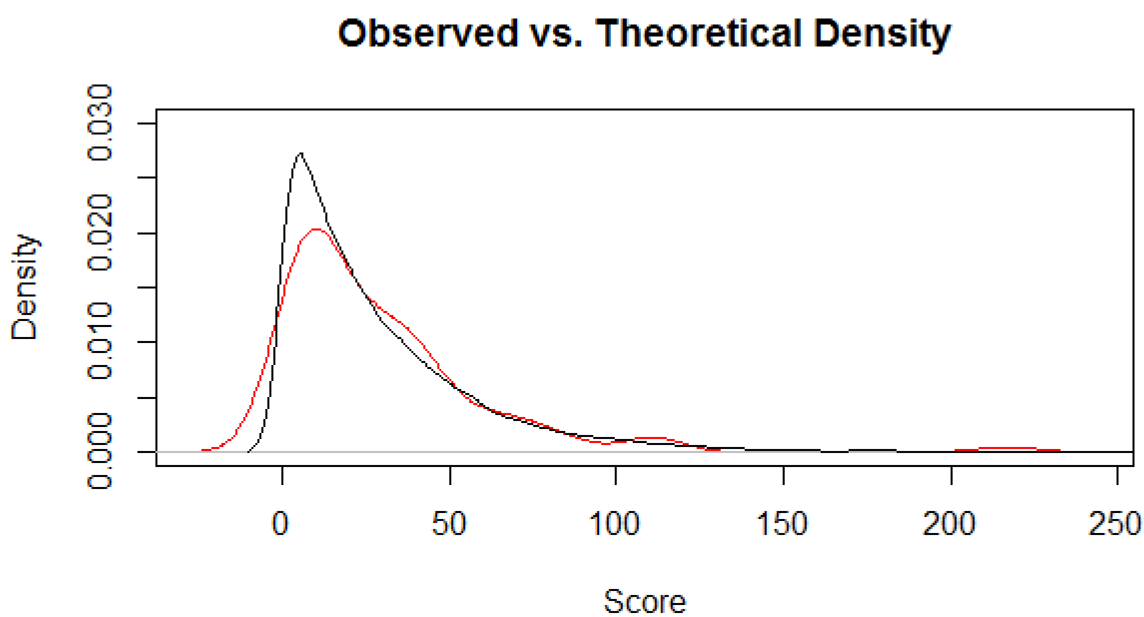




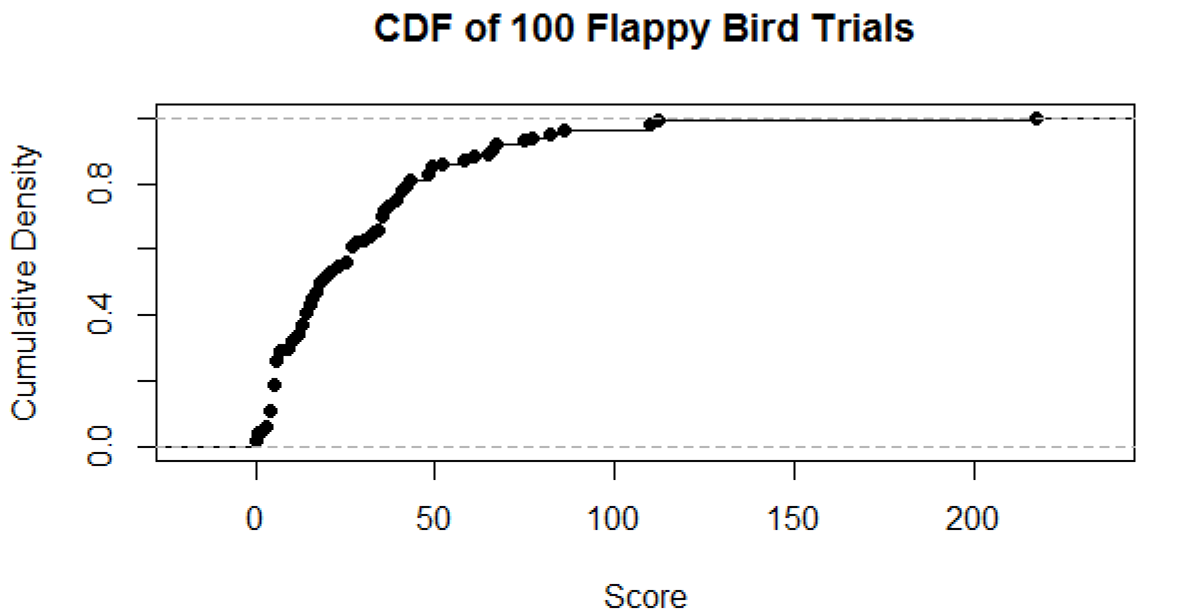
Using R, to simulate a geometric distribution with a mean=29.29, we have created a plot of the density if we had taken 9,999 trials from a true geometric distribution.

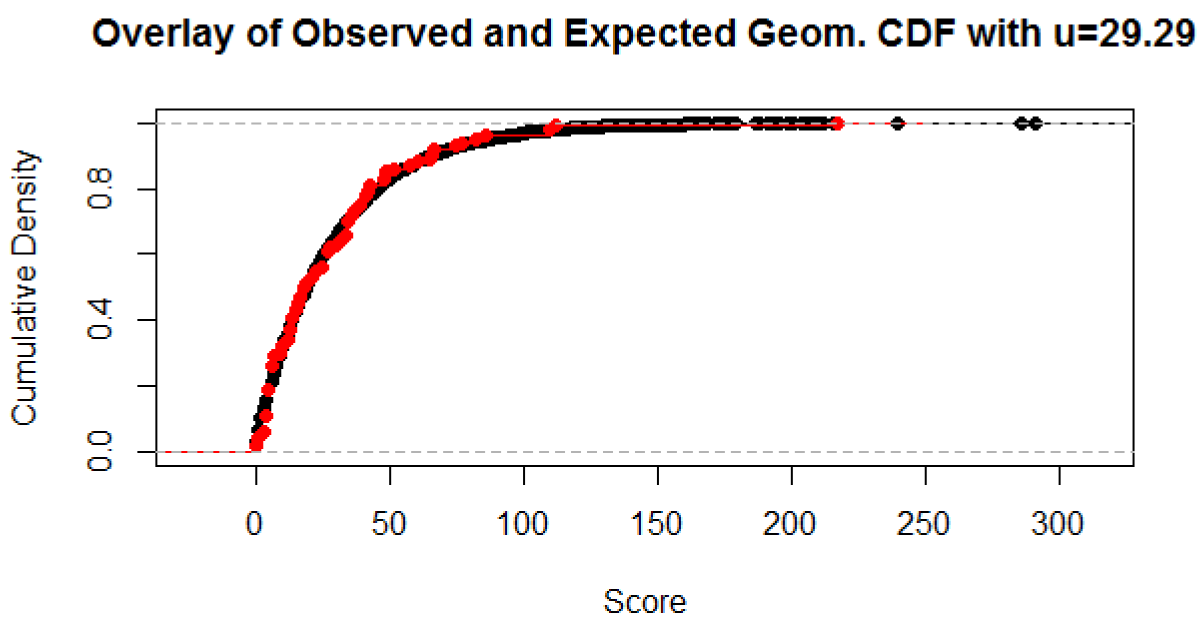


Here we have plotted the density of Anthony's collected data next to the theoretical density of a geometric distribution with mean 29.29. The theoretical density has a higher peak at the beginning, but it seems like a pretty close fit.

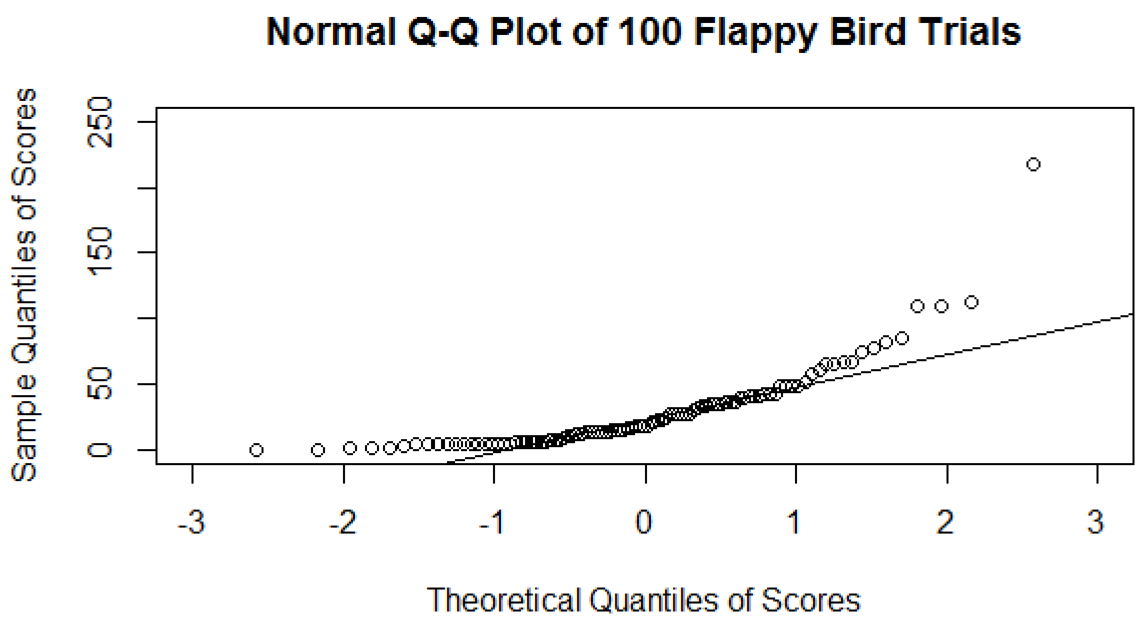


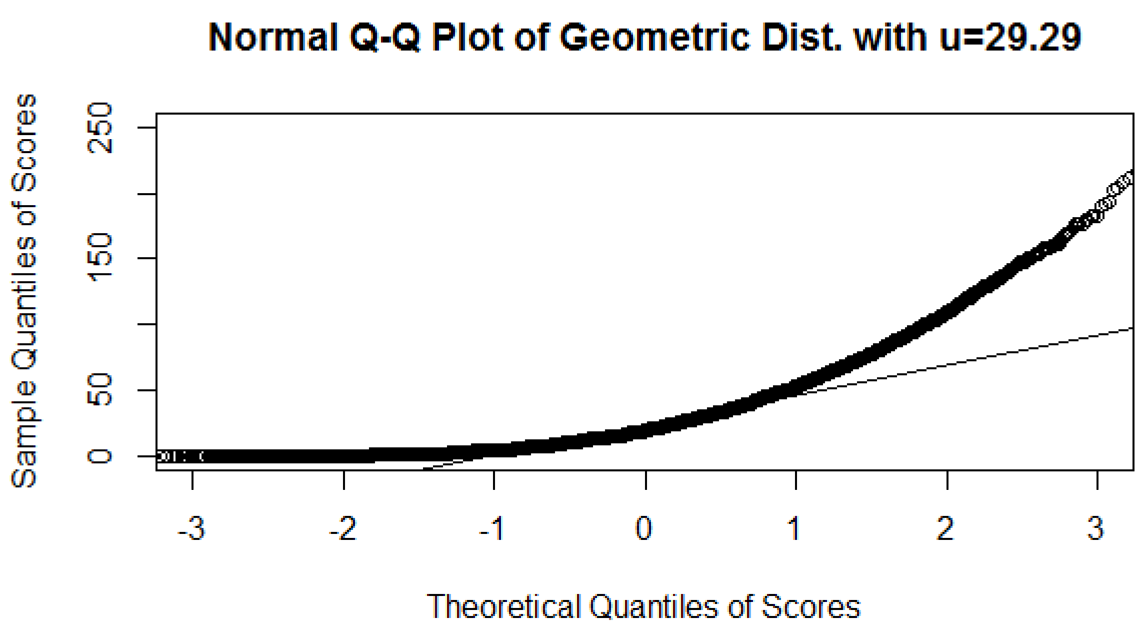
Below is a graph of the cumulative distribution of Anthony's 100 trials alone (black), and a graph of Anthony's 100 trials (red) against the expected cumulative distribution function (black) of a geometric distribution with mean 29.29.



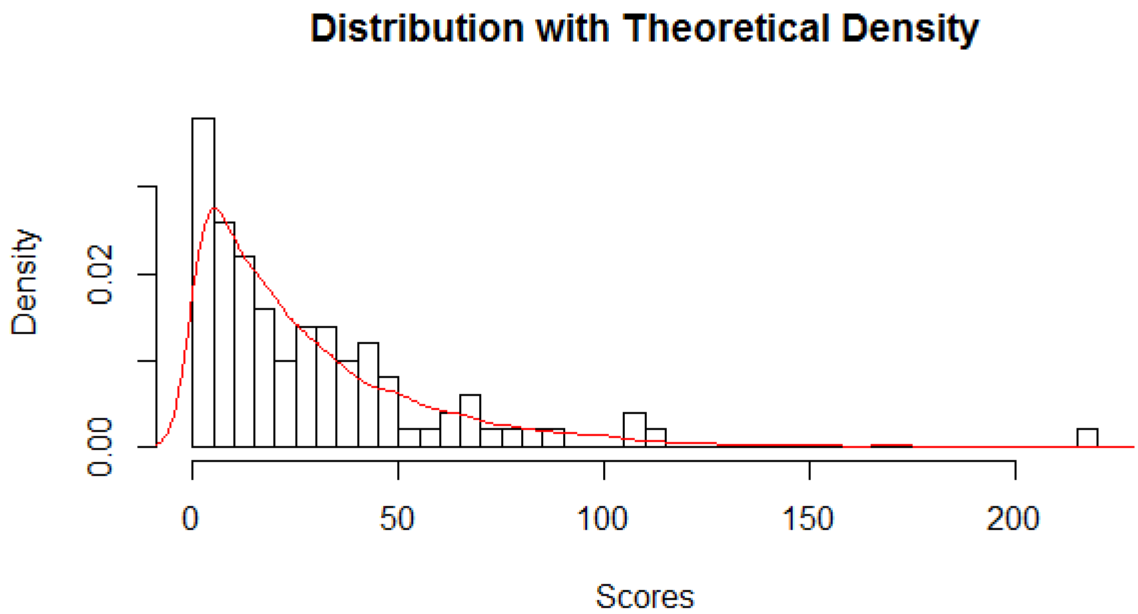


Here we have another comparison of the normal probability plots of Anthony's 100 trials along with the probability plot of a geometric distribution with u=29.29 and n=9999. Both graphs are concave up and follow a similar trend. They clearly show that the data does not follow a normal distribution and is right-skewed.

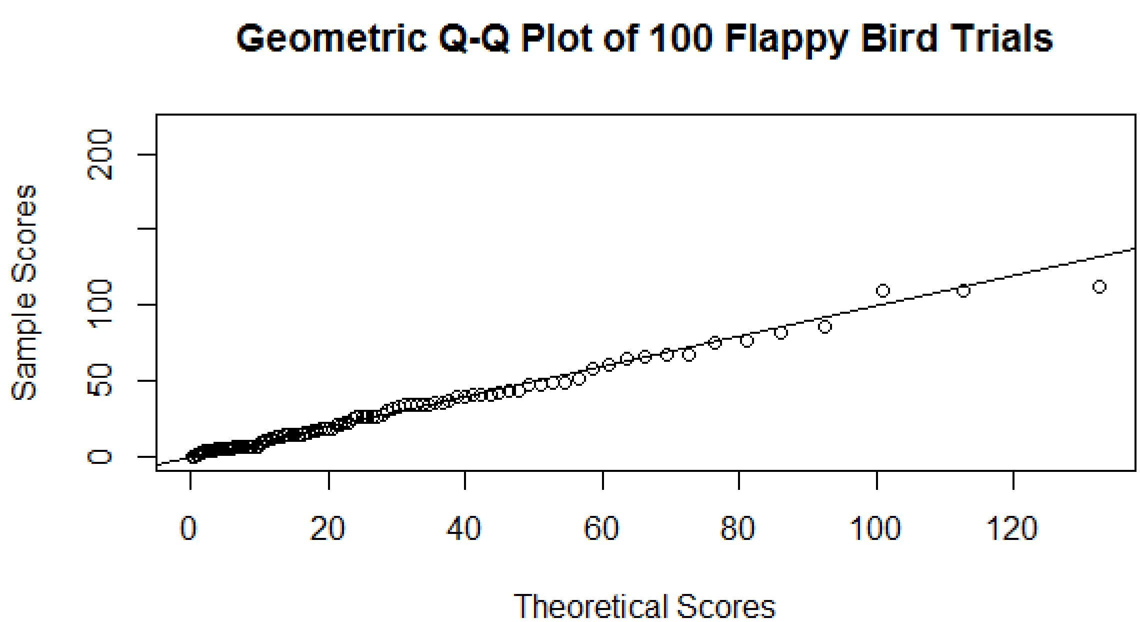




Here we have the histogram of Anthony's trials plotted against the density curve (red) of a geometric sequence with mean 29.29.



Below we've plotted the points on a geometric probability plot, with the straight line representing a true geometric distribution. Based on this probability plot and the graph comparisons above, we can say that Anthony's Flappy bird scores follow a geometric distribution very closely.



Further Analysis

Because the score distribution seems to tend toward a geometric distribution, we will make some inferences about Anthony's Flappy bird scores based on a geometric distribution.

The maximum likelihood estimate of p (the probability of not making it through one tube) for Anthony's trials is approximately .033014196. Thus, we can estimate that the probability that Anthony successfully flies through a tube while playing Flappy bird is approximately 96.70% using a geometric distribution.

We also see that Anthony has an outlying high score of 217. Assuming that Flappy bird tends toward a geometric distribution, with the p-value calculated above, the probability that Anthony scored a 217 is: $P(X=217)=(1-.033014196)^{216}(.033014196)=0.00002341$, or .00234%.

We can also calculate a 95% confidence interval for Anthony's mean score.
We have $\bar{x} \pm (1.96S)/\sqrt{n} = 29.29 \pm (1.96)(31.7)/10 = (23.0768 , 35.5032)$
Thus, we can say that we are 95% confident that the true mean of Anthony's Flappy bird scores lies between 23.0768 and 35.5032.

Conclusion:

Based on the graphs of observed versus theoretical densities, cumulative densities, and probability plots, we conclude that for the game Flappy bird, it is reasonable to assume that passing the tubes are a series of independent and identically distributed Bernoulli random variables and that the entire collection of scores tend toward a geometric distribution as a player reaches his or her skill potential.

Further Comments and Questions:

Although the geometric distribution is represented as a density curve in our plots, it is perhaps a bit misleading since it is a discrete function. This effect is more noticeable with a smaller sample of scores or a higher p value. However, even with the plot of Albert's 800 scores, we see that the density curve extends before a score of 0, which is a result of trying to fit a curve to a discrete function. Despite these drawbacks, we decided to use the curve for better clarity in our histograms.

There are also goodness of fit indicators and tests that we could have calculated to more rigorously prove that a geometric distribution is a "good" approximation of Flappy bird scores, but we haven't learned how to interpret these indicators and conduct the tests.

Another issue we faced was that when we computed the probability function for a geometric distribution, we used the inverse of the sample mean instead of the maximum likelihood estimate. We were unclear on which one would be more appropriate, but since they were fairly close to each other, our density curves would not have been significantly affected. The mle is also consistently less than the inverse of the sample mean, which suggests that our scores are less right skewed than a geometric distribution might predict.