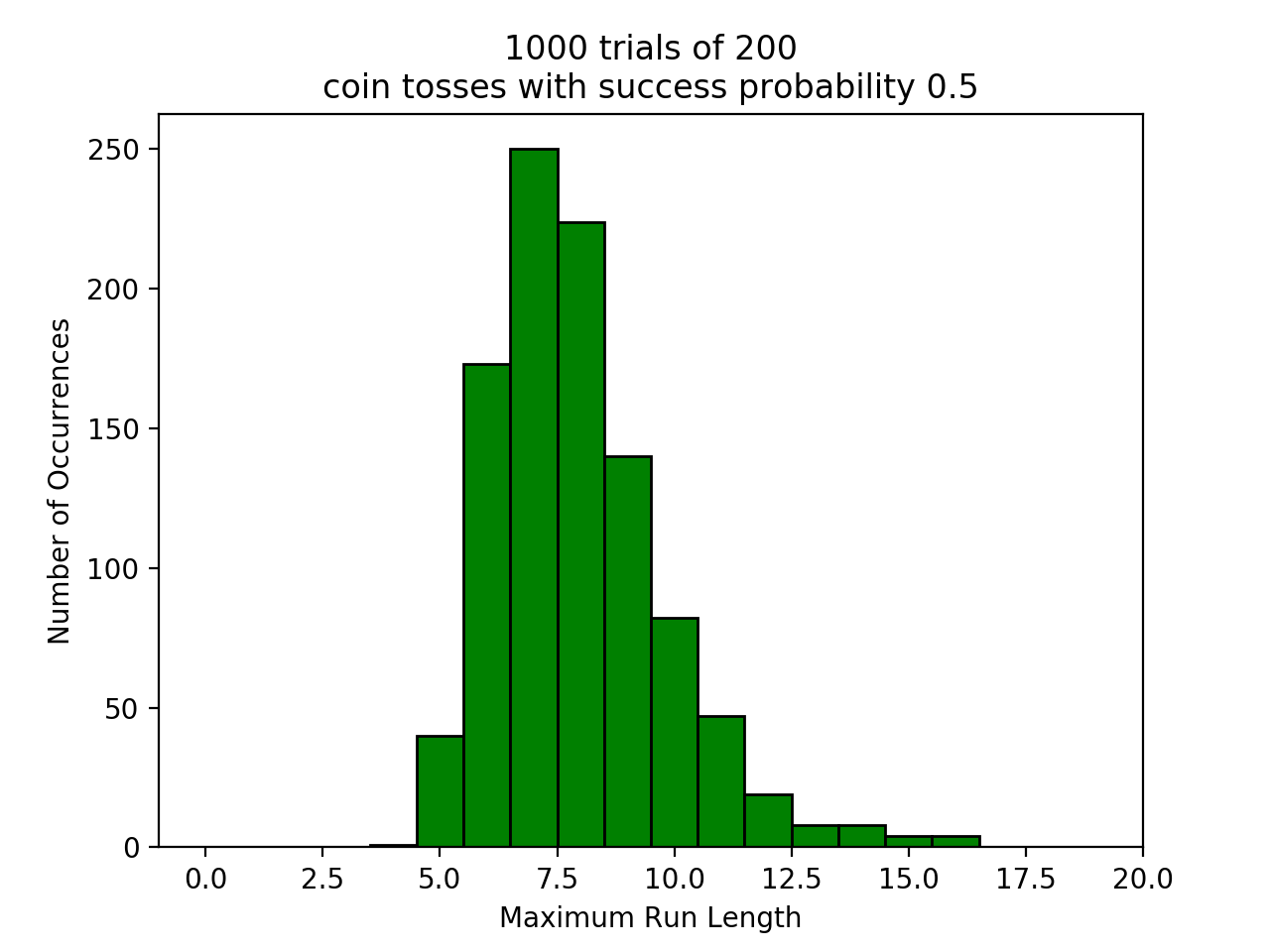
Sarah Ford

Randomness & Computation: Assignment 1 Written Responses

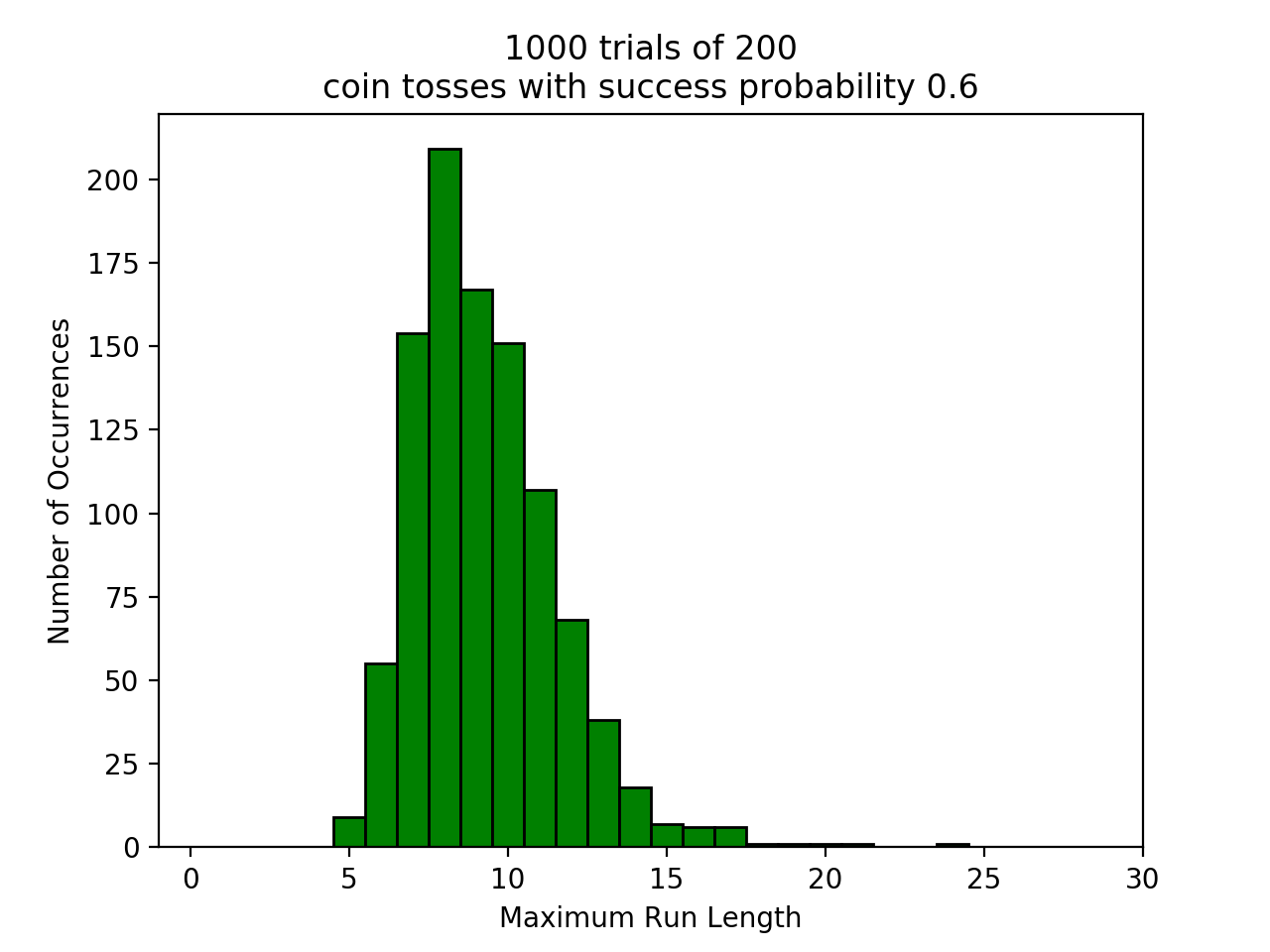
1. See the function runlengths(p,n) in the accompanying code.

2. See the function histmaxruns(p,n,numtrials,cum=False) in the accompanying code.

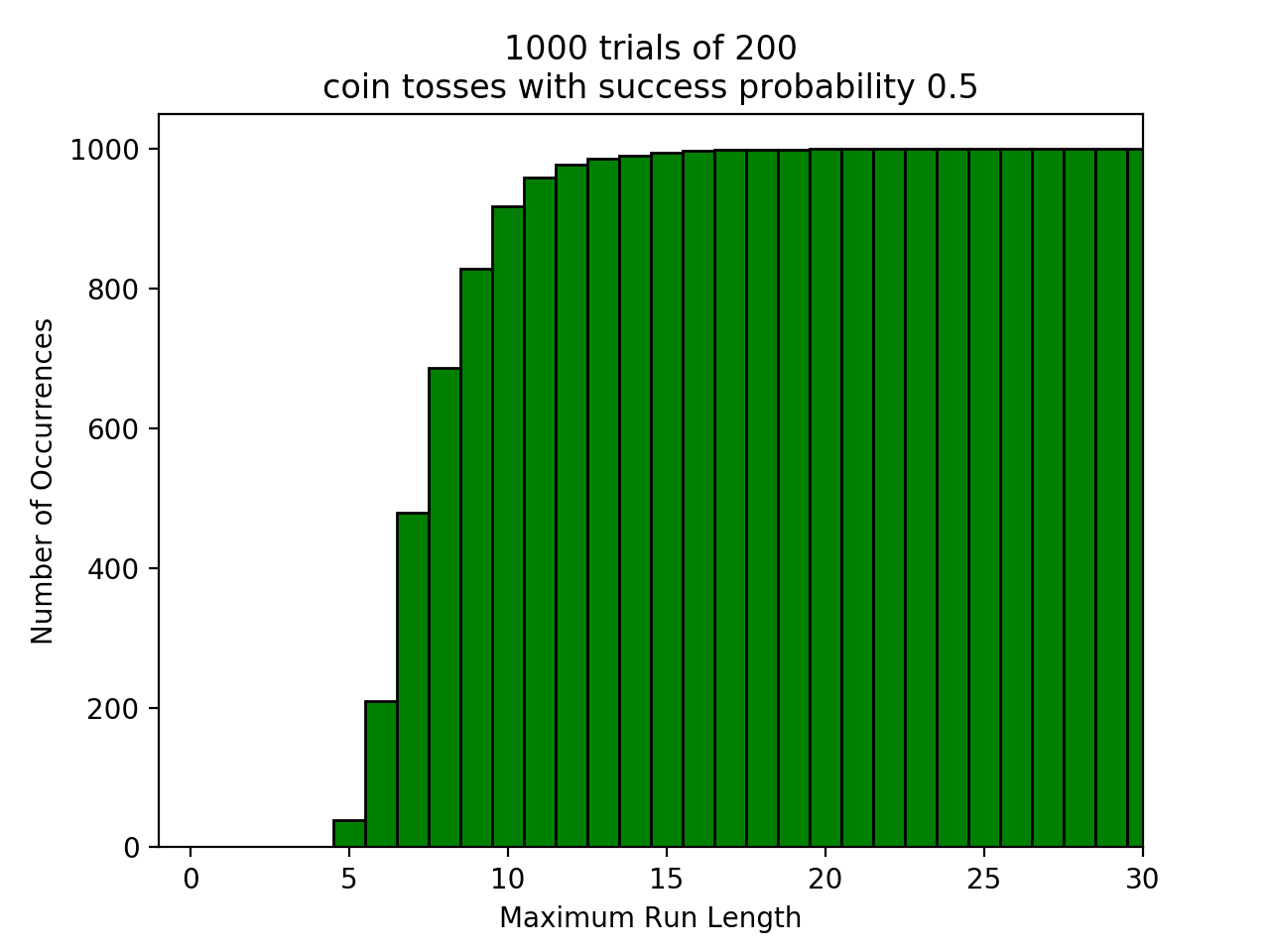
(a) After simulating 200 fair coin (p=0.5) tosses across 1,000 trials, I plotted the occurrence of each trial’s maximum run length and plotted the result on the histogram below. As evidenced by the tallest bar, the most frequently occurring run length in a sequence of 200 coin tosses was about 7 (rounded from 6.9).



(b) As in part (a) above, I again simulated 200 tosses across 1,000 trials, this time with a heads probability of 0.6 (p=0.6). The resulting histogram, plotted below, indicates that the most frequently occurring maximum run length was 8 (rounded from 7.9).

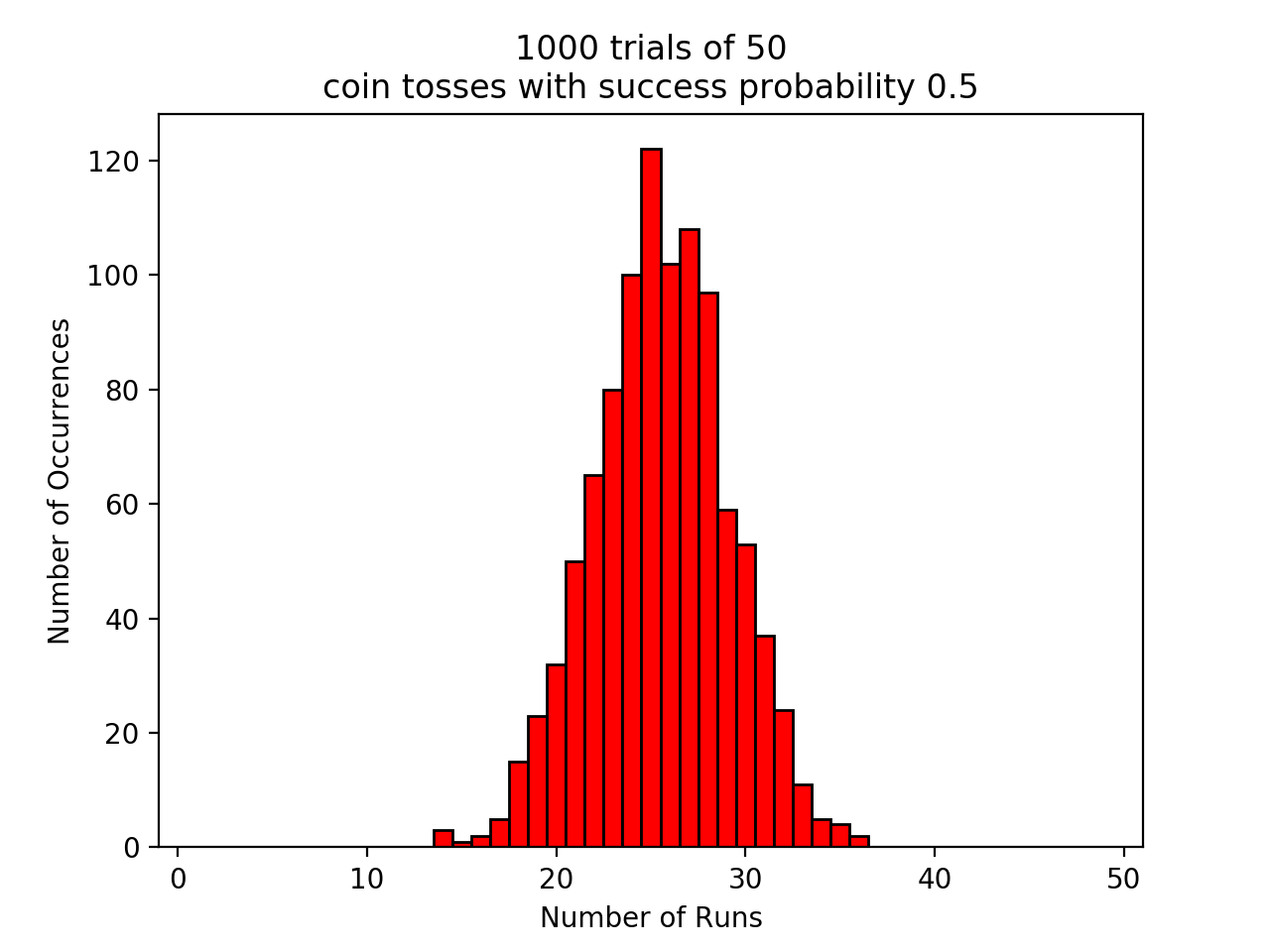


(c) In order to determine the probability that in 200 tosses of a fair coin, there is no run of length greater than 5, I simulated again 1,000 trials of 200 coin tosses and plotted the cumulative frequency histogram below. Of the 1,000 trials of 200 coin tosses simulated, the histogram indicates only about 25 occurrences (rounded from 24.7159) where the maximum run length was no greater than 5. We can therefore say that the probability that in 200 coin tosses of a fair coin, there is no length greater than 5 is equivalent to 25/1,000, or 2.5%.

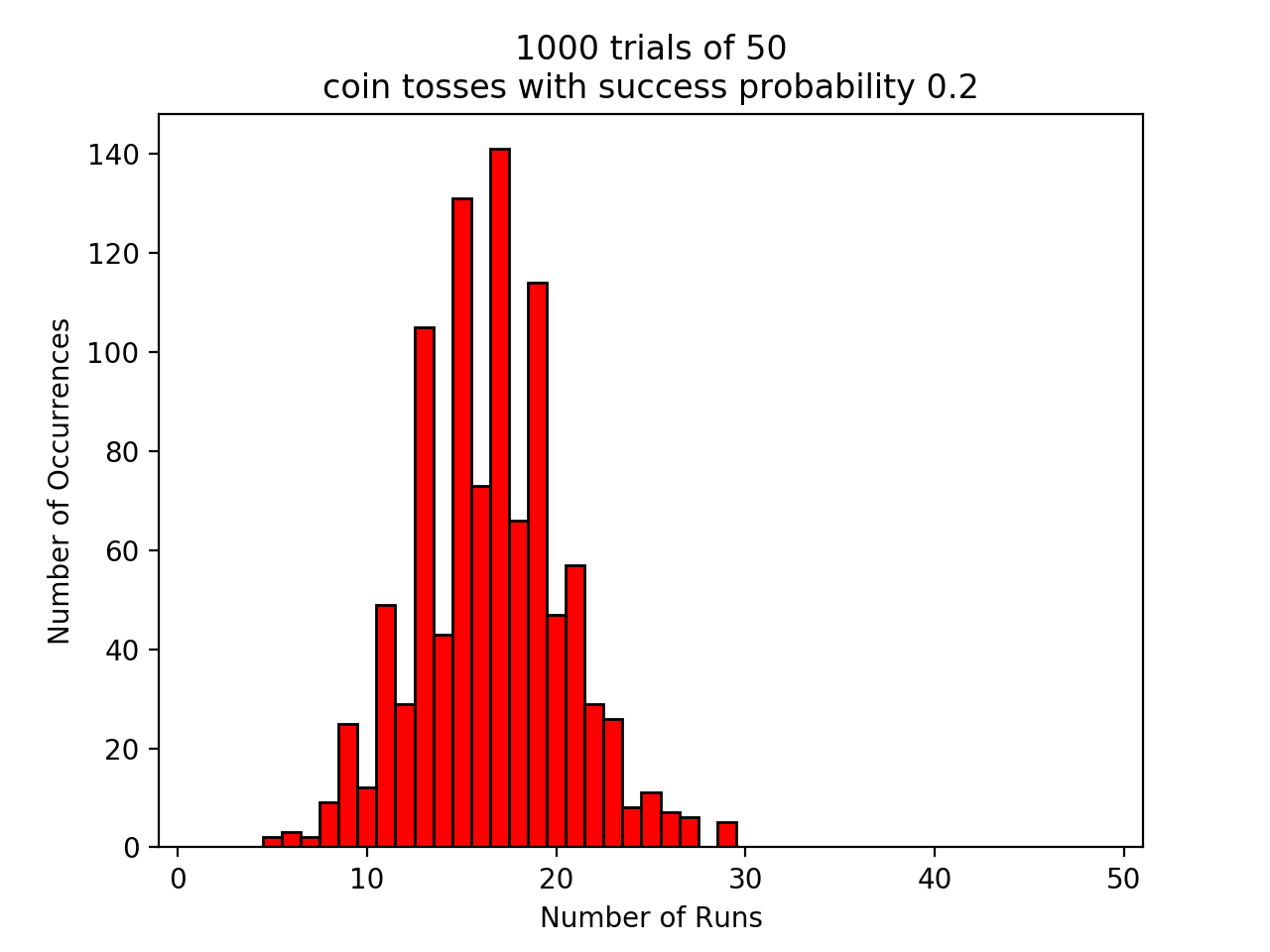


3. See the function histnumruns(p,n,numtrials,cum=False) in the accompanying code.

(a) In order to determine the most likely number of runs in 50 tosses of a fair coin, I simulated 50 tosses of a fair coin across 1,000 trials and collected the number of runs for each trial. Below, I plotted these results on a histogram. The tallest bar indicates that the most frequent number of runs (and therefore the most likely) in 50 tosses of a fair coin was about 25.

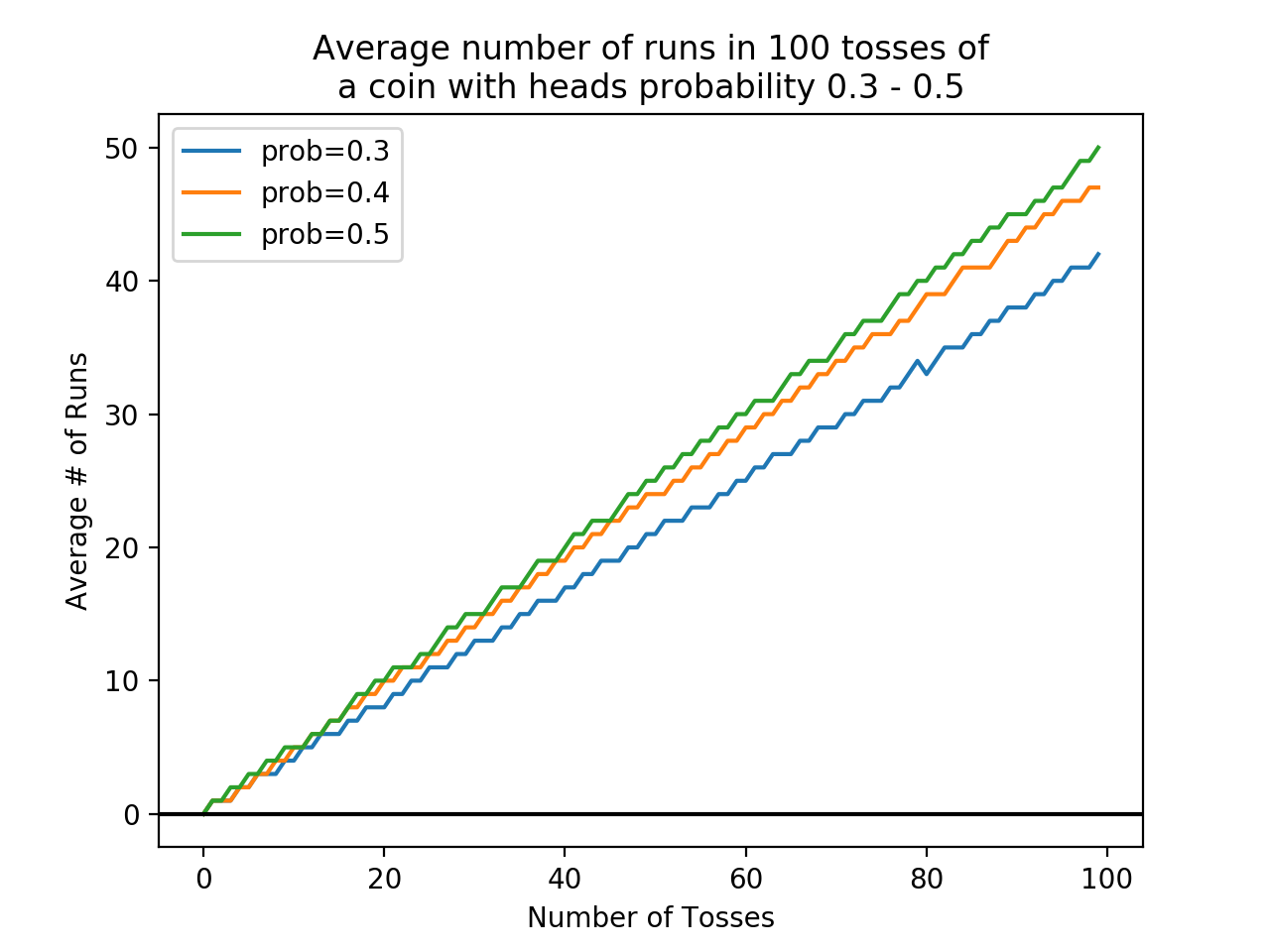


(b) When I used the above function to plot another histogram with p = 0.2, n = 50, and numtrials = 1,000, I immediately noticed the fluctuation in frequencies of number of runs within the histogram. It makes sense that, overall, there are fewer runs, since there should be more runs of Tails with the unfair coin (p=0.2). The fluctuations present below may be the result of increased random clustering of runs during trials.

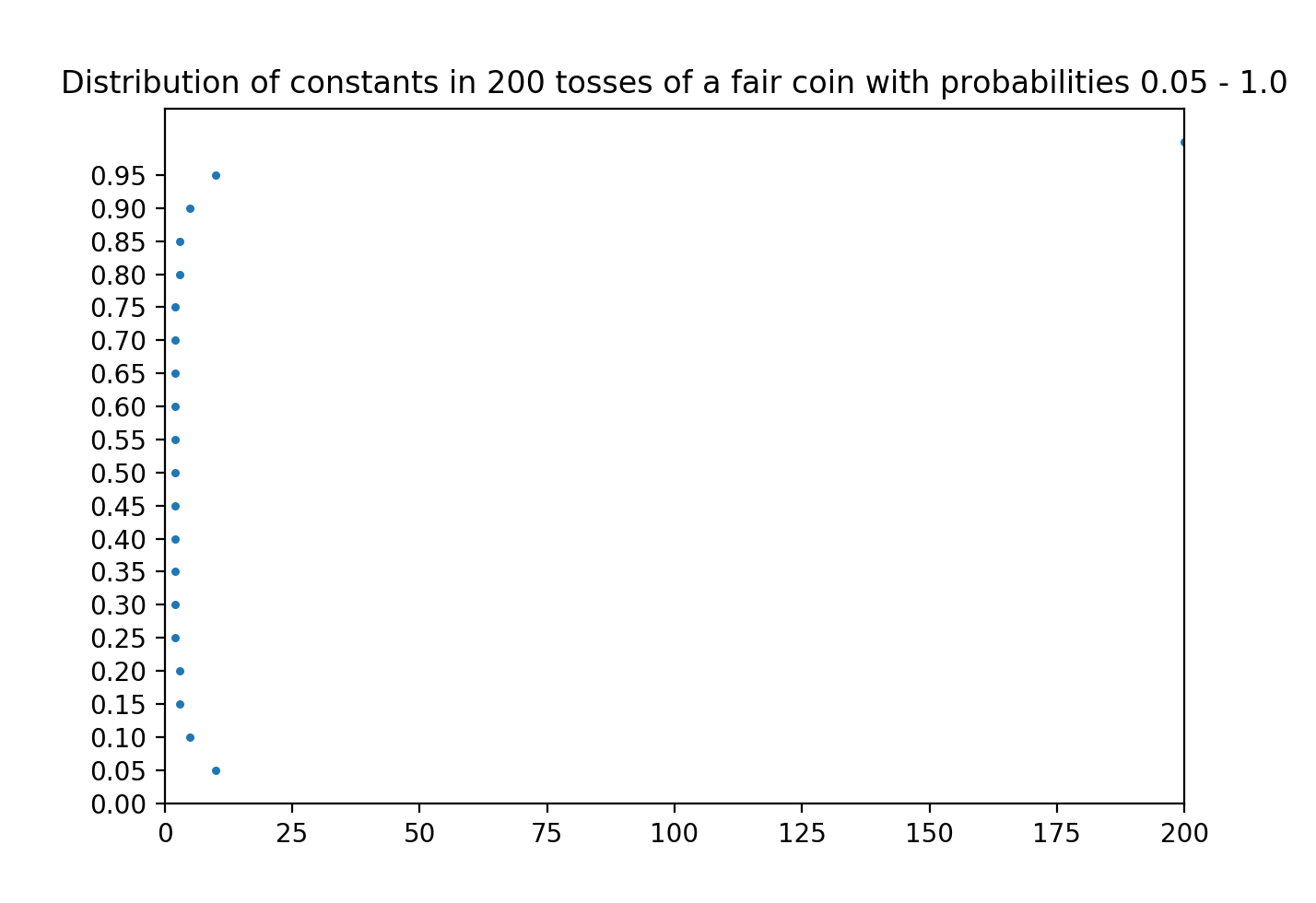


4. See the function plotnumruns() in the accompanying code.

(a) The plot below demonstrates the average number of runs within 100 tosses of a coin. The three lines represent a superimposition of 1,000 simulations of 100 tosses for coins of varying probabilities (from p= 0.3 to p=0.5).



b) The plot below illustrates the distribution of constants for 200 tosses of a coin across 1,000 simulations. Each data point represents the constant corresponding to a coin with a different probability, ranging from 0.05 to 1.0. All points are symmetrical around x = 0.5 because the more biased the coin in either direction, the less likely that the average number of runs represents exactly half of the total number of tosses. For more biased coins, the number of runs is smaller as run lengths increase, and therefore the corresponding constants must also increase with bias.



c) Of the 14 players in the Cornell Basketball Team Hot Hands study, I would argue that player number 9 is the only member of the team to deviate significantly from a coin in terms of having more streaks (and therefore fewer runs). Given our simulation above, one would expect that if a player’s shots were truly independent with 50% chance each time, then their shots should mirror the results we found with coins. Therefore, any player’s expected number of runs should be equal to the number of shots they took divided by 2 (the constant we found corresponding to a fair coin). When I calculated the absolute deviation between each player’s observed number of runs and expected number of runs, I got the array: [ 6 , 4 , 10, 2, 5, 9, 6.5, 2, 18, 1, 2, 2, 2, 0 ]. Of the absolute values in this array, the greatest deviation corresponded with the 9th player, who would have been expected to go on 50 runs but instead had 32, indicating streaks of longer runs. Thus, in accordance with our simulation conclusions above, player 9 must be the one who was found to deviate from a coin in the study.