`Project 1:`
Coin Flips according to a Categorical Distribution

Shahla Daneshmehr

February 2023

1 Abstract

Coin flipping is a fundamental cryptographic primitive that enables two distrustful and far apart parties to create a uniformly random bit [1]. When we flip a coin, there are two possible outcomes: heads or tails. The experiment has been simulated by categorical distribution with two different scenarios of different probabilities. This experiment measures one hundred times coin flips according to a categorical distribution and graphs the results.

2 Introduction

In probability theory and statistics, a categorical distribution (also called a generalized Bernoulli distribution, multinoulli distribution [2]) is a discrete probability distribution that describes the possible results of a random variable that can take on one of K possible categories, with the probability of each category separately specified. There is no innate underlying ordering of these outcomes, but numerical labels are often attached for convenience in describing the distribution, (e.g. 1 to K). The K-dimensional categorical distribution is the most general distribution over a K-way event; any other discrete distribution over a size-K sample space is a special case. The parameters specifying the probabilities of each possible outcome are constrained only by the fact that each must be in the range 0 to 1, and all must sum to 1 [3].

In fact, the categorical distribution is the generalization of the Bernoulli distribution for a categorical random variable, i.e. for a discrete variable with more than two possible outcomes, such as the roll of a die. On the other hand, the categorical distribution is a special case of the multinomial distribution, in that it gives the probabilities of potential outcomes of a single drawing rather than multiple drawings [4].

In probability theory and statistics, the Bernoulli distribution is the discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability q=1-p. Less formally, it can be thought of as a model for the set of possible outcomes of any single experiment that asks a yes/no question. Such questions lead to outcomes that are boolean-valued: a single bit whose value is success/yes/true/one with probability p and failure/no/false/zero with probability q. It can be used to represent a (possibly biased) coin toss where 1 and 0 would represent "heads" and "tails", respectively, and p would be the probability of the coin landing on heads (or vice versa where 1 would represent tails and p would be the probability of tails) [5].

This project explains two different probabilities when we flip a coin by using categorical distribution and consequently plot them in two separate graphs with green and red colors.

3 Hypotheses to Explain coin flipping

I simulate a 100 coin flips with a 50 percent probability of heads, then 100 times again with a 75 percent probability of heads. I anticipate that the outcomes will roughly match up to their probabilities. Roughly 50 of the first run should be heads, and roughly 75 of the second run should be heads.

4 Code and Experimental Simulation

Let's look at a simulated code:

```
#!/usr/bin/env python

# imports of external packages to use in our code
import sys
sys.path.append(".")

# Get our probablity config for our experiment
from Config import probs

# import our Random class from python/Random.py file
from Random import Random
random = Random()

# set up our results file for writing
result_file = open("results.txt", mode="w")

def run_experiment(prob):
    return random.Categorical(prob)

for prob in probs:
    r = []
    n_heads = 0
    for run in range(100):
        r.append(run_experiment(prob))
        n_heads += r[-1]
    result_file.write(" ".join(list(map(str, r))) + " " + str(n_heads) + "\n")
```

5 Analysis

I simulated two different scenarios (two different probabilities) and performed an analysis of the output to evaluate how well these different scenarios can be distinguished from each other. As a matter of fact, sometimes visual summaries of data can tell us more than text summaries. These plots show us one of the possible probabilities for the coin flipper with two different outlines. To acquire deep information about the plot, we need to define each axis and also different colors (red and green) to clearly understand the outcome of our simulations.

```
#!/usr/bin/env python

# imports of external packages to use in our code
import sys
sys.path.append(".")

# Get our probablity config for our experiment
from Config import probs

# imports of external packages to use in our code
import matplotlib.pyplot as plt

result_file = open("results.txt", mode="r")

results = []
n_heads = []

for line in result_file.readlines():
    line = line.rstrip().split(" ")
    r, heads = line[:-1], line[-1]
    results.append([float(result) for result in r])
    n_heads.append(float(heads))

colors = ["red", "green"]
figure, axis = plt.subplots(1,2)
for i in range(len(results)):
    print("p = ", end='')
    print(probs[i], end=': ')
    print(results[i], end="\n\n")
    axis[i].bar(["heads", "tails"], [n_heads[i], (len(results[i]) - n_heads[i])], color=colors[i
    axis[i].set_title(str(probs[i]))

plt.ylabel('number of coin flips')

plt.show()
```

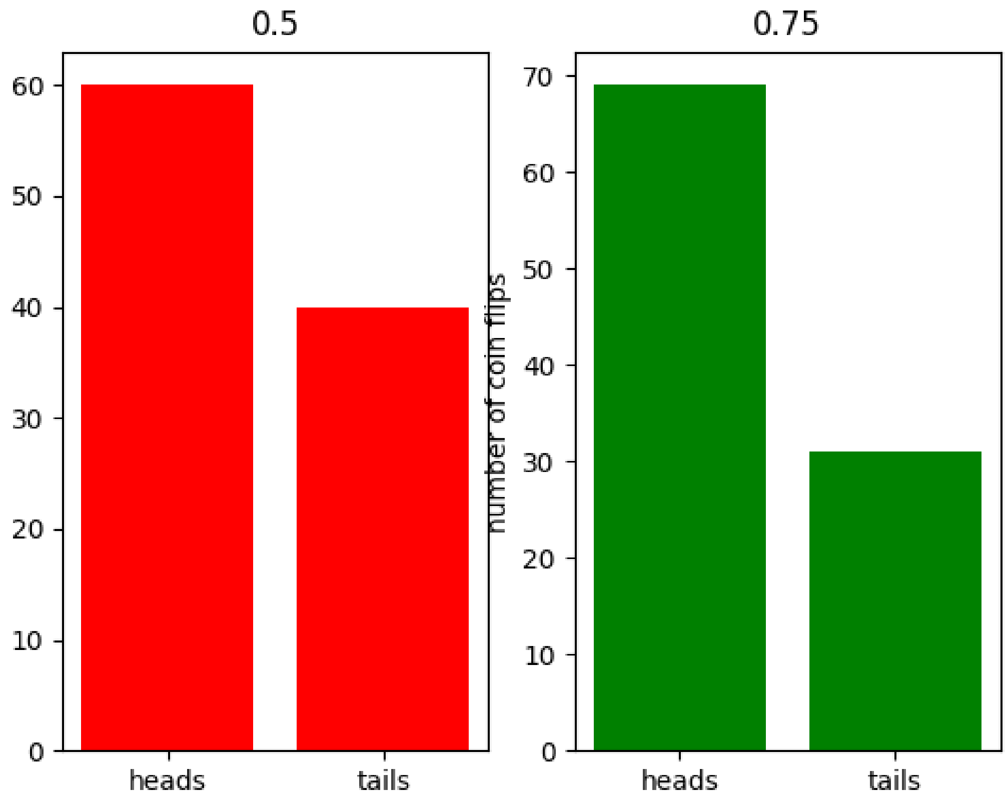

Figure 1: Simulations of coin flip probabilities

6 Conclusion

One conclusion that we can immediately draw about the probabilities of coin flips by categorical distribution in this project could be that when we try 50 percent, both columns in the plot usually appear as a very close probability. Running the program several times validated this claim. However, 75 percent probability has more variance than expected.

7 References

1. Chailloux, A. Kerenidis, I. (2009) Optimal Quantum Strong Coin Flipping, 50th Annual IEEE Symposium on Foundations of Computer Science, Atlanta, GA, USA, pp. 527-533.

2. Murphy, K. P. (2012). Machine learning: a probabilistic perspective, p. 35. MIT press.

3. Stark, H. Woods, J.W. (2002). Probability and Random Processes with Applications to Signal Processing (3rd ed.). Pearson. p. 7.

4. Albert, J. (1998). Listing All Possible Outcomes (The Sample Space). Bowling Green State University. Retrieved 2013-06-25.

5. Erhan, Çınlar (2011). Probability and stochastics. New York: Springer. p. 57.