Fundamentals of Data Analysis and Decision Science
Duke University

EGR 238L/MATH 238L
Spring 2022

Probability
Lab Assignment

(10) 1. Consider the experiment of flipping one fair coin whose two sides we will call heads (H) and tails (T).

(a) What is the sample space Ω?

$S = \{H, T\}$

(b) What is the probability of each outcome in Ω?

50% each

(c) Take a coin (say, a US quarter) out of your pocket, and flip it 10 times. How many times does it came up H, and how many does times it came up T? Do this entire thing again a few (at least 3) more times.

1st trial: 6 H, 4 T   2nd: 5 H, 5 T   3rd: 4 H, 6 T   4th: 7 H, 3 T

(d) Read the abstract[1] of the paper "Dynamical Bias in the Coin Toss" by Persi Diaconis, Susan Holmes, and Richard Montgomery[2] In a few sentences discuss what this might mean for the distinction between theoretical models and actual data in real life.

Theoretically it should be a 50/50 chance of flipping a coin, but in actuality it is just a physics problem which obeys these mechanics, and the probability is higher for the side which ~~it~~ it started on.

[1]Read the rest if you like as well!
[2]https://statweb.stanford.edu/ susan/papers/headswithJ.pdf

(10) 2. Consider the experiment of flipping two fair coins, one with our left hand (call this the "first coin") and one with our right hand (call this the "second coin").

(a) What is the sample space Ω?

$S = \{HH, HT, TH, TT\}$

(b) Let A be the event that "the first coin came up H and the second coin came up T." Compute $P(A)$.

$$\frac{C = \{HT\}}{S = \{HH, HT, TH, TT\}} = \frac{1}{4} = 25\%$$

(c) Let B be the event that "exactly one H came up." Compute $P(B)$.

$$\frac{C = \{HT, TH\}}{S = \{HH, HT, TH, TT\}} = \frac{2}{4} = \frac{1}{2} = 50\%$$

(10) 3. Suppose instead we flip seven fair coins.

(a) Let C be the event that "the first coin came up H and all other coins came up T'." Compute $P(C)$.

$C = \{HTTTTTT\} = \frac{1}{2^7} = \frac{1}{128}$

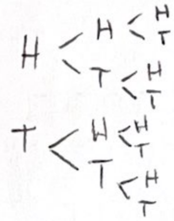
$2^7 = 128$

(b) Let D be the event that "exactly one H came up." Compute $P(D)$.

$C = \{HTTTTTT, THTTTTT, TTHTTTT, TTTHTTT, TTTTHTT, TTTTTHT, TTTTTTH\}$

$\frac{7}{2^7} = \boxed{\frac{7}{128}}$

(30) 4. Consider the experiment of flipping three fair coins. Let X be the discrete random variable that counts how many heads came up in the experiment.



HHH
HHT
HTH
HTT
THH
THT
TTH
TTT

(a) What are the possible values that X can take?

$x = \{0, 1, 2, 3\}$

(b) For each of these values x_i, what is $P(X = x_i)$?

$0 = \frac{1}{8}$

$1 = \frac{3}{8}$

$2 = 3/8$

$3 = 1/8$

(c) What is the expected value $E[X]$?

$(0)\frac{1}{8} + (1)\frac{3}{8} + 2(3/8) + (3)1/8 = \frac{12}{8} = \boxed{1.5}$

(d) What is the variance $Var[X]$? What is the standard deviation?

$E[(X-\mu)^2] = E(x^2) - [E(x)]^2$

$n = 3 \quad p = .5 \qquad np = 1.5$

$Var[x] = np(1-p) = 1.5(1-.5) = \boxed{.75}$

(e) What is the probability that a specific run of the experiment will result in $X = E[X]$?

0, X cannot equal 1.5

(f) Take three coins out of your pocket, and run this experiment yourself a large number of times. Each time you run it, note the value of X. Visualize your results in the form of a histogram.

Results (10 trials)

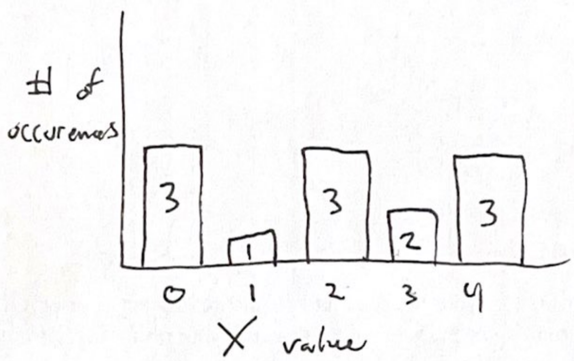
1
2
2
2
1
3
0
1
2
2

(20) 5. Consider a fair six-sided die, with faces numbered 1 through 6. Let's play the following game which will produce a discrete random variable X. We roll this fair die four times. If we never see a 1, then $X = 0$. Otherwise, let X be the roll on which we *first* see a 1. For example, if we roll $(2,1,3,1)$, then $X = 2$. If we roll $(6,4,2,1)$, then $X = 4$, and if we roll $(3,6,2,2)$, then $X = 0$.

(a) Play this game a large number of times and sketch (to scale) a histogram of the observed X-values.

10 trials

(3,2,5,4) (4,3,1,1)
(1,5,1,4) (6,2,2,3)
(6,1,2,3) (5,4,3,1)
(2,2,3,5) (4,1,2,3)
(5,1,5,6) (5,4,6,1)



(b) Compute the theoretical expected value $E[X]$.

$X=0 = \left(\frac{5}{6}\right)^4 = 48\%$

$X=1 = \frac{1}{6} = 16.7\%$

$X=2 = \frac{5}{6}\cdot\frac{1}{6} = 13.9\%$

$X=3 = \frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6} = 11.7\%$

$X=4 = \frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6} = 9.6\%$

$E(X) = .167 + .278 + .351 + .384 = \boxed{1.18}$

(30) 6. Now let's make the last problem a little bit harder (warning: you will need to use your Calculus skills, specifically series and sequences, a little bit here!). The game is as follows. We roll a fair six-sided die over and over again, until we see our first 1. Then we stop. Let Y be the roll on which we see this 1.

Compute the expected value $E[Y]$.

# of rolls

1: 1/6

2: $\frac{5}{6} \times \frac{1}{6}$

⋮

$$E(Y) = \sum_{y_i} y_i \cdot P(Y = y_i)$$

1: 1/6

2: $5/6 \cdot \frac{1}{6}$

3: $5/6 \cdot \frac{5}{6} \cdot \frac{1}{6}$

4: $5/6 \cdot 5/6 \cdot 5/6 \cdot 1/6$

⋮

$$E(Y) = \sum_{y_i} y_i P(Y = y_i)$$

$$E(Y) = \sum_{y_i} Y_i \cdot \left(\left((i-1) \cdot \frac{5}{6}\right) \cdot \frac{1}{6}\right)$$

(30) 7. Now you're going to play a game called Solo Pig[3]. The rules are as follows. You start with zero points. At each turn of the game, you have one of two options, Stop or Roll. If you Stop, the game ends and you keep the points you have. If you Roll, then one of two things happens. If you roll a 1, the game ends and you finish with zero points. If you roll something other than a 1, then you add the number of dots rolled to your score, and you get another turn. The game continues until you roll a 1 or you choose Stop. For example, a series of rolls $(5,3,3,4,1)$ would give me zero points, but a series of rolls $(5,3,3,4)$ followed by a Stop choice would leave me with 15 total points.

Play this game a few times yourselves, experimenting with different strategies. Write down the results. Then, using your answer to the previous question, come up with what you think is the optimal strategy for Solo Pig[4]

wrote python code to simulate game,

4, 2, 2, 3, 1 = 0

5, 1 = 0

3, 3, 4, Stop = 10

5, 6, Stop = 11

2, 2, 1, = 0

4, 3, 5, 2, 3, 1 = 0

The optimal Strategy is to play until youre around 10-15 points, and then Stop before risking a 1 being rolled.

[3]There's a more common, and more complicated dice game called Pig, which I encourage you to look up.
[4]That is, at what total score is it a good idea to Stop rather than to Roll?

(10) 8. Provide two different possible sets of 5 observations that have a mean equal to 20, yet a mean of 20 is a potentially misleading summary statistic.

When a dataset is skewed it can possibly be misleading to use mean. An example of this is Ages of people using SnapChat, with sample {14, 16, 15, 15, 40} the median is 15 years old, but the average is 20, if you throw out the outlier, you can move the average to 15.

In millions, the value of an artists paintings, valued at {5, 5, 5, 5, 80} the 80 million painting was very special and owned by a celebrity, but the average value is not 20 because of skewed data.

9. Suppose there are 40 people in the classroom across the hall, and you know that the mean age of the people in that room is 20 years old and half are wearing blue shirts while half are wearing white shirts. What can you say about a "typical" occupant of that classroom? What can you say about the population of people in that classroom as a whole?

- An occupant of the room is wearing either blue of white
- the combined age of everyone is 800 years
- 20 people wearing blue, 20 wearing white