DS 100/200: Principles and Techniques of Data Science

**Date: July 17, 2019** 

## Discussion #7

Name:

## **Probability**

- 1. Suppose John visits your store to buy some items. He buys toothpaste for \$2.00 with probability 0.5. He buys a toothbrush for \$1.00 with probability 0.1. Let the random variable X be the total amount John spends. Find  $\mathbb{E}[X]$ .
- 2. Suppose we have a coin that lands heads 80% of the time. Let the random variable Y be the *proportion* of times the coin lands heads out of 100 flips. What is Var[Y]?

3. Let X be a random variable with mean  $\mu = \mathbb{E}[X]$ . Using the definition  $Var(X) = \mathbb{E}[(X - \mu)^2]$ , show that for any constant c,

$$\mathbb{E}[(X-c)^2] = (\mu - c)^2 + \text{Var}(X).$$

Discussion #7

- 4. Use the above result to prove that
  - $Var(X) < \mathbb{E}[(X-c)^2]$  for any c
  - $\operatorname{Var}(X) = \mathbb{E}[X^2] \mathbb{E}[X]^2$
- 5. We roll a die 9 times and record the value of each roll on slips of paper. Then, we place all 9 slips of paper in a box:

2



The numbers in the box have the following summary statistics:

Statistic	Sum	Sum of Squares	Mean	Median
Value	27	93	3	3

For each of the following, answer the following questions: Is this value calculable from the information given? If so, either calculate it by hand or describe how you would calculate this value. If not, then suggest an estimate for the quantity. All draws are with replacement.

- (a) The expected value of a single draw from the box
- (b) The expected value of the average of nine draws from this box
- (c) The exact variance of the tickets in the box
- (d) The exact variance of the average of nine draws from the box