DS 100/200: Principles and Techniques of Data Science **Date: July 17, 2019**

Discussion #7

Name: Raguvir Kunani

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 $\mathbb{E}[X]$.

See end of worksheet

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 $\text{Var}[Y]$?

See end of worksheet

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

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

$E[(X-c)^2] = E[X^2 - 2cX + c^2]$ linearity of expectation

$= E[X^2] - 2cE[X] + E[c^2]$ linearity of expectation

$= E[X^2] - 2c\mu + c^2$ add and subtract an $E[X]^2$ term so we can factor

$= E[X^2] + E[X]^2 - 2c\mu + c^2 - E[X]^2$ rearranging terms

$= E[X]^2 - 2c\mu + c^2 + E[X^2] - E[X]^2$ def. of variance

$= \mu^2 - 2c\mu + c^2 + Var(X)$ factor $\mu^2 - 2c\mu + c^2$

$= (\mu - c)^2 + Var(X)$

4. Use the above result to prove that

   See end of worksheet

   - $\text{Var}(X) \leq \mathbb{E}[(X-c)^2]$ for any c
   - $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ plug in $c = E[X]$ into

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:

1	2	2	3	3	3	4	4	5

   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

Probability

(1) Let B be a random variable (RV) that represents the amount John spends on a toothbrush. Let P be a similar RV for toothpaste. Then $X = B + P$.

$$E[X] = E[B+P] = E[B] + E[P] = .1 + 1 = \boxed{1.1}$$

$$E[B] = (0 \times 0.9) + (1 \times 0.1) = .1$$

$$E[P] = (0 \times 0.5) + (2 \times 0.5) = 1$$

(2) Let X_i be a RV that is 1 when flip i is heads, and 0 otherwise. Then $\sum_{i=1}^{100} X_i$ is the total number of heads in 100 flips, which means Y (the proportion of heads) is $\frac{1}{100} \sum_{i=1}^{100} X_i$.

$$\begin{aligned}
\mathrm{Var}[Y] &= \mathrm{Var}\left[\frac{1}{100} \sum_{i=1}^{100} X_i\right] && \mathrm{Var}[aX] = a^2 \mathrm{Var}[X] \\
&= \left(\frac{1}{100}\right)^2 \mathrm{Var}\left[\sum_{i=1}^{100} X_i\right] && X_i \text{ are independent} \\
&= \left(\frac{1}{100}\right)^2 \sum_{i=1}^{100} \mathrm{Var}(X_i) && X_i \text{ all have same variance} \\
&= \left(\frac{1}{100}\right)^2 100 \cdot \mathrm{Var}(X_i) && X_i \text{ is a Bernoulli RV} \\
&= \frac{1}{100} \cdot \mathrm{Var}(X_i) && \text{so } \mathrm{Var}[X_i] = p(1-p) \\
&= \frac{1}{100} \cdot 0.8 \cdot 0.2 = \boxed{.0016} && = 0.8(0.2)
\end{aligned}$$

Another way to solve ②:

Let N be the number of heads in 100 flips. Then $Y = \frac{1}{100}N$ (the proportion of heads).

$$\mathrm{Var}[Y] = \mathrm{Var}\left[\tfrac{1}{100}N\right]$$

$\mathrm{Var}[aX] = a^2\,\mathrm{Var}[X]$

$$= \left(\tfrac{1}{100}\right)^2 \mathrm{Var}[N]$$

N is a Binomial RV so $\mathrm{Var}[N] = np(1-p) = 100 \cdot 0.8 \cdot 0.2$

$$= \left(\tfrac{1}{100}\right)^2 (100 \cdot 0.8 \cdot 0.2)$$

$$= \tfrac{1}{100} \cdot 0.8 \cdot 0.2 = \boxed{.0016}$$

④ Want to prove $\mathrm{Var}(X) \le E[(X-c)^2]$ for constant c

In ③ we proved $E[(X-c)^2] = (\mu - c)^2 + \mathrm{Var}(X)$.

Rewriting this, we can say $\mathrm{Var}(X) = E[(X-c)^2] - (\mu-c)^2$.

Since $(\mu - c)^2 \ge 0$, we have $\mathrm{Var}(X) \le E[(X-c)^2]$.