CMSC320: Homework 2 (Statistics)

James Zhang*

February 20, 2024

§1 Set Theory and Bayes Theorem

A. Recall from the Inclusion-Exclusion Principle that we have

$$A \cup B \cup C = A + B + C - A \cap B - A \cap C - B \cap C + A \cap B \cap C$$

Solving for $A \cap B \cap C$ yields

$$-(A \cap B \cap C) = A + B + C - A \cap B - A \cap C - B \cap C - A \cup B \cup C$$
$$-x = 4500 + 6000 + 5500 - 1800 - 2000 - 3000 - 10000$$
$$-x = 16000 - 10000 - 6800$$
$$x = 800 = A \cap B \cap C$$

B. Let us find the probability that a randomly selected shopper is also a reviewer. We seek $\mathbb{P}(R \mid S)$

$$\mathbb{P}(R \mid S) = \frac{\mathbb{P}(R \cap S)}{\mathbb{P}(S)}$$

by Bayes Theorem. Note that $\mathbb{P}(S) = \frac{3}{5}$ since there are 6000 shoppers out of 10000. Further, $\mathbb{P}(R \cap S) = \frac{3}{10}$ since there are 3000 people who are both shoppers and reviewers. Therefore,

$$\mathbb{P}(R|S) = \frac{3}{10} / \frac{3}{5} = \frac{1}{2}$$

C. Let us find the probability that someone is in exactly 2 categories but not all 3. Note that we were essentially given the number of people in exactly 2 categories $= 2000 + 3000 + 1800 = 6800$.

$$\mathbb{P}(\text{Exactly 2 but not all 3}) = \frac{6800}{10000} = \frac{17}{25}$$

D. This does not make sense logically because the intersection $A \cap B \cap C$ represents people who are shoppers, buyers, and reviewers. If we know that no shoppers are buyers and vice versa, then therefore then there cannot be someone who is a shopper, buyer, and reviewer. Essentially,

$$A \cap B = \emptyset \implies A \cap B \cap C = \emptyset$$

*Email: jzhang72@terpmail.umd.edu

§2 Probability Distributions

A. The probability that he aces his opponent at least once in any of the games is

$$1 - \mathbb{P}(\text{No aces}) = 1 - 0.0081 = 0.9919$$

B. Given $\mathbb{P}(Y > 119) = 0.02273$ and $\mathbb{P}(Y < 104) = 0.2335$. Taking Z-scores, we get

$$\mathbb{P}(\frac{Y-\mu}{\sigma} > \frac{119-\mu}{\sigma}) = 0.02273, \ \mathbb{P}(\frac{Y-\mu}{\sigma} < \frac{104-\mu}{\sigma}) = 0.2335$$

$$1 - \phi(\frac{119-\mu}{\sigma}) = 0.02273 \implies \phi(\frac{119-\mu}{\sigma}) = 0.97727, \phi(\frac{104-\mu}{\sigma}) = 0.2335$$

$$\frac{119-\mu}{\sigma} = 2.00037, \frac{104-\mu}{\sigma} = -0.72737$$

Now solving this sytem of equations

$$119 = 2.00037\sigma + \mu, 104 = -0.72737\sigma + \mu$$

$$15 = 2.00037\sigma + 0.72737\sigma \implies 15 = 2.72774\sigma \implies \sigma \approx 5.4991$$

$$\implies \sigma^2 \approx 30.2401$$

Plugging this back in solve for μ, we get

$$\mu \approx 107.9998 \implies Y \sim N(107.9998, 30.2401)$$

C. Using formulas for mean and variance of binomial distribution, we can get up another system of equations

$$\mathbb{E}(X) = np = 107.9998 \text{ and } \mathrm{Var}(X) = np(1-p) = 30.2401$$

By direct substitution for np into the variance formula, we get

$$1 - p = \frac{30.2401}{107.9998} \implies p = 1 - \frac{30.2401}{107.9998} \implies p \approx 0.7200$$

$$n \approx \frac{107.9998}{0.7200} = 149.9997$$

$$X \sim \text{Binomial}(149.9997, 0.72)$$

§3 Bayes Theorem

We seek $\mathbb{P}(\text{Delay} \mid \text{Predict Delay})$. Using Bayes Theorem we can find this as

$$\frac{\mathbb{P}(\text{Delay})\mathbb{P}(\text{Predict Delay}|\text{Delay})}{\mathbb{P}(\text{Delay})\mathbb{P}(\text{Predict Delay}|\text{Delay}) + \mathbb{P}(\text{No Delay})\mathbb{P}(\text{Predict Delay}|\text{No Delay})}$$

Plugging in known values, we get that

$$\mathbb{P}(\text{Delay}|\text{Predict Delay}) = \frac{0.05(0.8)}{0.05(0.8) + 0.95(0.15)} = \frac{0.04}{0.04 + 0.1425} \approx 0.2192$$

§4 Expected Value

I. Note that expectation follows linearity.

$$\mathbb{E}(X2) = 2\mathbb{E}(X) = 2(\sum_{i=1}^{6} \frac{1}{6} i) = 2(3.5) = 7$$

II. Similarly

$$\mathbb{E}(X3) = 3\mathbb{E}(X) = 3(3.5) = 10.5$$