Jerry Cain and Chris Piech
CS 109

Section 5
February 16, 2021

# Section 5 Solutions

Based on the work of many CS109 instructors and course staff members.

## Overview of Section Materials

The warm-up questions provided will help students practice concepts introduced in lectures. The section problems are meant to apply these concepts in more complex scenarios similar to what you will see in problem sets and quizzes.

## 1 Warmups

### *1.1 Joint Random Variables Statistics*

**True or False?** The symbol $Cov$ is covariance, and the symbol $\rho$ is Pearson correlation. And $X \perp Y$ is just a fancy way to say that $X$ and $Y$ are independent.

| | |
|---|---|
| $X \perp Y \implies Cov(X, Y) = 0$ | $Var(X + X) = 2Var(X)$ |
| $Cov(X, Y) = 0 \implies \text{X} \perp Y$ | $X \sim \mathcal{N}(0, 1) \wedge Y \sim \mathcal{N}(0, 1) \implies \rho(X, Y) = 1$ |
| $Y = X^2 \implies \rho(X, Y) = 1$ | $Y = 3X \implies \rho(X, Y) = 3$ |

**True or False?**

| | |
|---|---|
| True | False (... $= 4Var(X)$) |
| False (antecedent necessary, not sufficient) | False (don't know how independent X & Y are) |
| False ($Y = X \implies$ ...) | False (... = 1) |

### *1.2 Conditional Expectation*

Let $X \sim Geo(p)$. Use the Law of Total Expectation to prove that $E[X] = 1/p$, by conditioning on whether the first flip is heads or tails.

$$E[X] = E[X|H]P(H) + E[X|T]P(T) = 1 \cdot p + (E[1 + X])(1 - p)$$

Solving yields $E[X] = 1/p$.

# 2 Problems

## *2.1 Random Number of Random Variables*

Let $N$ be a non-negative integer-valued random variable—that is, a random variable that takes on values in $\{0, 1, 2, \dots\}$. Let $X_1, X_2, X_3, \dots$ be an infinite sequence of independent and identically distributed random variables (independent of $N$), each with mean $\mu$, and $X = \sum_{i=1}^{N} X_i$ be the sum of the first $N$ of them.

Before doing any work, what do you *think* $E[X]$ will turn out to be? Then show it mathematically to see if your intuition is correct.

$$E[X] = E\left[\sum_{i=1}^{N} X_i\right] = \sum_n E\left[\sum_{i=1}^{N} X_i \mid N = n\right] p_N(n) = \sum_n E\left[\sum_{i=1}^{n} X_i \mid N = n\right] p_N(n)$$

$$= \sum_n E\left[\sum_{i=1}^{n} X_i\right] p_N(n) = \sum_n n\mu p_N(n) = \mu \sum_n n p_N(n) = \mu E[N]$$

Alternatively,

$$E[X] = E[E[X|N]] = E[N\mu] = \mu E[N]$$

## *2.2 Fairness in AI.*

In their 2018 paper "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," Joy Buolamwini and Timnit Gebru showed that commercial gender classifiers performed significantly worse on face images of darker-skinned females (error rates up to 34.7%) than lighter-skinned males (maximum error rate of 0.8%.) This disparity may result from training the classifiers on unbalanced datasets. To evaluate the classifiers, the authors designed their own dataset by collecting photos of national parliamentarians from three African countries and three European ones.

The probability table below shows the joint distribution of the dataset between two random variables: the demographic ($D$) of the photo subject and their country ($C$).

| **Demographic** | **South Africa** | **Senegal** | **Rwanda** | **Sweden** | **Finland** | **Iceland** |
|---|---|---|---|---|---|---|
| Darker Female | 0.12 | 0.05 | 0.04 | 0.01 | 0 | 0 |
| Darker Male | 0.15 | 0.07 | 0.02 | 0.01 | 0 | 0 |
| Lighter Female | 0.02 | 0 | 0 | 0.12 | 0.06 | 0.02 |
| Lighter Male | 0.05 | 0 | 0 | 0.14 | 0.09 | 0.03 |

a. What is the marginal probability distribution for demographic $D$? Provide your result as a mapping from values that $D$ can take to probabilities.

b. What is the conditional probability of country given that the subject is a lighter female, $P(C|D = \text{Lighter Female})$? Provide your result as a mapping from values that $C$ can take to probabilities. Is this mapping a probability distribution?

c. What is the conditional probability that the subject is from Senegal given their demographic, $P(C = \text{Senegal}|D)$? Provide your answer as a mapping from values that $D$ can take to probabilities. Is this mapping a probability distribution?

d. What are the pitfalls in using this dataset for a purpose beyond what the authors intended?

a. For each assignment to $D$, sum over all the values of $C$ that are consistent with that assignment.

$$\begin{aligned}
P(\text{Darker Female}) &= 0.12 + 0.05 + 0.04 + 0.01 + 0 + 0 = 0.22 \\
P(\text{Darker Male}) &= 0.15 + 0.07 + 0.02 + 0.01 + 0 + 0 = 0.25 \\
P(\text{Lighter Female}) &= 0.02 + 0 + 0 + 0.12 + 0.06 + 0.02 = 0.22 \\
P(\text{Lighter Male}) &= 0.05 + 0 + 0 + 0.14 + 0.09 + 0.03 = 0.31
\end{aligned}$$

b.

$$P(\text{South Africa}|\text{Lighter Female}) = \frac{P(\text{South Africa, Lighter Female})}{P(\text{Lighter Female})} = \frac{0.02}{0.22} \approx 0.09$$

Similarly,

$$\begin{aligned}
P(\text{Senegal}|\text{Lighter Female}) &= 0 \\
P(\text{Rwanda}|\text{Lighter Female}) &= 0 \\
P(\text{Sweden}|\text{Lighter Female}) &\approx 0.55 \\
P(\text{Finland}|\text{Lighter Female}) &\approx 0.27 \\
P(\text{Iceland}|\text{Lighter Female}) &\approx 0.09
\end{aligned}$$

This mapping is the conditional probability distribution $P(C|D = \text{Lighter Female})$. Its probabilities sum to 1.

c.

$$P(\text{Senegal}|\text{Darker Female}) = \frac{P(\text{Senegal, Darker Female})}{P(\text{Darker Female})} = \frac{0.05}{0.22} \approx 0.23$$

Similarly,

$$\begin{aligned}
P(\text{Senegal}|\text{Darker Male}) &\approx 0.28 \\
P(\text{Senegal}|\text{Lighter Female}) &= 0 \\
P(\text{Senegal}|\text{Lighter Male}) &= 0
\end{aligned}$$

This mapping is not a probability distribution because the conditioning event changes. We can also see that the probabilities do not sum to 1.

d. The dataset in question doesn't come close to representing the diversity of the world; it draws subjects from just six countries. Even within those countries, the dataset may be unbalanced with respect to socioeconomic and cultural groups because the subjects are all parliamentarians. For example, the dataset may underrepresent ethnic minorities within those countries.

## *2.3 Binary Trees*

Consider the following function for constructing binary trees:

```
struct Node {
    Node *left;
    Node *right;
};

Node *randomTree(float p) {
    if (randomBool(p)) {  // returns true with probability p
        Node *newNode = new Node;
        newNode->left = randomTree(p);
        newNode->right = randomTree(p);
        return newNode;
    } else {
        return nullptr;
    }
}
```

The `if` branch is taken with probability $p$ (and the `else` branch with probability $1 - p$). A tree with no nodes is represented by `nullptr`; so a tree node with no left child has `nullptr` for the `left` field (and the same for the right child).

Let $X$ be the number of nodes in a tree returned by `randomTree`. You can assume $0 < p < 0.5$. What is $E[X]$, in terms of $p$?

Let $X_1$ and $X_2$ be number of nodes the left and right calls to `randomTree`.
$E[X_1] = E[X_2] = E[X]$.

$$\begin{aligned}
E[X] &= p \cdot E[X \mid \texttt{if}] + (1-p)E[X \mid \texttt{else}] \\
&= p \cdot E[1 + X_1 + X_2] + (1-p) \cdot 0 \\
&= p \cdot (1 + E[X] + E[X]) \\
&= p + 2pE[X] \\
(1-2p)E[X] &= p \\
E[X] &= \frac{p}{1-2p}
\end{aligned}$$