Nicholas Carolan
email@princeton.edu
COS302-F23

Assignment #6

Due: 6:00pm Wednesday 1 November 2023

Discussants: N/A

Upload at: https://www.gradescope.com/courses/606160/assignments/3574133

Problem 1 (15pts)
Consider a sample space $\Omega = \{\text{red}, \text{green}, \text{blue}, \text{orange}, \text{yellow}\}$.

(A) What is the smallest possible valid event space $\mathcal{A}$?

(B) What is the smallest possible event space that contains the set $\{\text{blue}\}$?

(C) What is the smallest possible event space that contains both $\{\text{blue}\}$ and $\{\text{red}, \text{green}\}$?
(Hint: it has eight members.)

An event space $\mathcal{A}$ must satisfy the following:
- Contains Ω
- Be closed under complements (e.g. $\Omega^C = \emptyset$)
- Be closed under unions/intersections (e.g. $\{\text{blue}\} \cup \{\text{red}, \text{green}, \text{orange}, \text{yellow}\} = \Omega$)

(A) The smallest possible valid event space $\mathcal{A}$ is the trivial event space $\mathcal{A} = \{\emptyset, \Omega\}$.

(B) The smallest possible event space that contains the set $\{\text{blue}\}$ is $\mathcal{A} = \{\emptyset, \{\text{blue}\}, \{\text{red}, \text{green}, \text{orange}, \text{yellow}\}, \Omega\}$.

(C) the smallest possible event space that contains both $\{\text{blue}\}$ and $\{\text{red}, \text{green}\}$ is

$$\mathcal{A} = \{\emptyset, \{\text{blue}\}, \{\text{red}, \text{green}\}, \{\text{orange}, \text{yellow}\}, \{\text{blue}, \text{red}, \text{green}\}, \{\text{blue}, \text{orange}, \text{yellow}\}, \{\text{red}, \text{green}, \text{orange}, \text{yellow}\}, \Omega\}$$

.

Problem 2 (15pts)

Consider the following bivariate distribution $p(x, y)$ of two discrete random variables X and Y.

Y					
y_1	0.01	0.02	0.03	0.1	0.1
y_2	0.05	0.1	0.05	0.07	0.2
y_3	0.1	0.05	0.03	0.05	0.04
X	x_1	x_2	x_3	x_4	x_5

(A) What is the marginal distribution $p(x)$?

(B) What is the marginal distribution $p(y)$?

(C) What is the conditional distribution $p(x \mid Y = y_1)$?

(D) What is the conditional distribution $p(y \mid X = x_3)$?

(E) What is the conditional distribution $p(x \mid Y \neq y_1)$?

(A) The marginal distribution $p(x) =$

X	x_1	x_2	x_3	x_4	x_5
	0.16	0.17	0.11	0.22	0.34

(B) The marginal distribution $p(y) =$

Y	
y_1	0.26
y_2	0.47
y_3	0.27

(C) The conditional distribution $p(x \mid Y = y_1) =$

X	x_1	x_2	x_3	x_4	x_5
	$\frac{0.01}{0.26} \approx 0.04$	$\frac{0.02}{0.26} \approx 0.08$	$\frac{0.03}{0.26} \approx 0.12$	$\frac{0.1}{0.26} \approx 0.38$	$\frac{0.1}{0.26} \approx 0.38$

(D) The conditional distribution $p(y \mid X = x_3) =$

Y	
y_1	$\frac{0.03}{0.11} \approx 0.27$
y_2	$\frac{0.05}{0.11} \approx 0.45$
y_3	$\frac{0.03}{0.11} \approx 0.27$

(E) The conditional distribution $p(x \mid Y \neq y_1) =$

X	x_1	x_2	x_3	x_4	x_5
	$\frac{0.15}{0.74} \approx 0.20$	$\frac{0.15}{0.74} \approx 0.20$	$\frac{0.08}{0.74} \approx 0.11$	$\frac{0.12}{0.74} \approx 0.16$	$\frac{0.24}{0.74} \approx 0.32$

Problem 3 (20pts)

In 2014-2016, West Africa experienced a massive outbreak of Ebola. We'll concentrate on Sierra Leone and imagine a mandatory screening of every citizen. The probability of being tested positive given that the citizen has Ebola is 84%. The probability of being tested positive given that the citizen does not have Ebola is 11%. We also know that the probability of contracting Ebola for any given citizen is 0.4%. If a randomly-chosen citizen tests positive, what is that citizen's probability of actually having Ebola?

Use Bayes' theorem:

$$P(\text{Ebola} \mid \text{Positive}) = \frac{P(\text{Positive} \mid \text{Ebola})P(\text{Ebola})}{P(\text{Positive})}$$

$$P(\text{Positive}) = P(\text{Positive} \mid \text{Ebola}) * P(\text{Ebola}) + P(\text{Positive} \mid \text{No Ebola}) * P(\text{No Ebola}) = 0.84 * 0.004 + 0.11 * 0.996 = 0.11292$$

$$P(\text{Ebola} \mid \text{Positive}) = \frac{0.84 * 0.004}{0.11292} = 0.0297... \approx 2.9\%$$

Problem 4 (48pts)

In this problem you will do some mathematical calculations and also use Colab. Be sure to append your PDF and insert your link as usual.

(A) Import `numpy.random` (usually aliased to `npr`) and set the random seed.

(B) Imagine drawing 1,000 independent Bernoulli variates $X_i \in \{0, 1\}$ with the probability $p(X_i = 1) = 0.35$, and computing their sum

$$Y = \sum_{i=1}^{1000} X_i \,.$$

What are the mean and variance of Y?

(C) Generate 10,000 independent random variables Y as described above. That is, generate 10,000 sums of 1,000 independent Bernoulli variables. This is not as hard as it sounds. Use the `numpy.random.rand()` function only; do not use any functions from `scipy.stats`. Use `rand()` to generate a $10,000 \times 1,000$ matrix of independent uniform random variates in the interval $[0, 1]$, then threshold them appropriately to get 0 or 1. Finally, sum over the appropriate dimension to get 10,000 samples of Y above.

(D) Use Matplotlib to make a histogram of these samples. Use at least 100 bins so you can see the structure in the distribution. Do you recognize the shape?

(E) Compute the empirical mean and variance of the 10,000 samples you have drawn. Compare these results to your calculation from (B).

(F) Now imagine drawing 1,000 independent (continuous) variates uniformly in the interval $[-1, 1]$. What are the mean and variance of their sum?

(G) Generate 10,000 such sums using a variation of the procedure you performed for (C). As before, you'll only use `rand()` but you should scale and shift rather than threshold.

(H) Create a histogram of these sums, as in (D). Describe the shape.

(I) Compute the empirical mean and variance of the 10,000 samples and compare them to your computations in (F).

Problem 5 (2pts)

Approximately how many hours did this assignment take you to complete?

My notebook URL: `https://colab.research.google.com/drive/1OoLmLbuCK0XhbK9HlIWDHNADp46IKLzr?usp=sharing`

Changelog

- 23 October 2023 – F23 version.