Margaret Connor
Winter 2019 | CSS 490

Assignment 1
I/ The math needed to answer the following questions is simple. Unless explicitly requested, you aren't required to show your calculations, but it's a good idea anyway. That way, if you understand the concept, but make an arithmetic error, it's possible I can give partial credit.

1. [12 points] We have a large jar filled with balls of three colors, red, white, and black, which we will draw from the jar one at a time. There are equal numbers of each color in the jar, so the chances of drawing any color is 1/3. Now consider an experiment (random process) in which three balls are drawn consecutively from the jar. We record the sequence of colors as the outcome of the experiment. For example, "RRW" means the first ball was red, the second ball was red, and the third ball was white.
    a. Is the probability space for this process discrete or continuous?
    The probability space is discrete
    b. How many possible outcomes are there?
    There are 27 possible outcomes
    c. Enumerate the outcomes.
    The enumerated outcomes are in the table below

RWR	RWW	RGW	RRR	RRW	RRB	RBR	RBW	RBB
WWR	WWW	WWB	WRR	WRW	WRB	WBR	WBW	WBB
BWR	BWW	BWB	BRR	BRW	BRB	BBR	BBW	BBB

    d. What is the probability of each outcome?
    The probability of each outcome is 1/27, assuming that the color balls are replaced to make their numbers equal to other colors before each ball is drawn.
    e. What is the probability of the event "one ball of each color was drawn", where the order in which the colors were drawn does not matter?
    p(RWB) + p(RBW) + p(WBR) + p(WRB) + p(BRW) + p(BWR) = 2/9

2. [14 points] A marketing survey looked at the preferences of hot drink size among 1275 random customers of a coffee shop chain. The survey was also interested in whether the customer's gender affects their preference. The results of the survey were used to estimate the probabilities in this joint probability distribution:

	Tall (T)	Grande (G)	Venti (V)

Female (F)	0.12	0.24	0.06
Male (M)	0.08	0.38	0.12

a. What is p (M, T), the joint probability that a customer in the survey was both male and prefers tall drinks?
   Male and Tall = p(M, T) = 0.08
b. What is p (F), the marginal probability that a customer in the survey was female?
   p(F) = p(F, T) + p(F, G) + p(F, V) = 0.12 + 0.24 + 0.06 = 0.42
c. What is p(G), the marginal probability that a customer in the survey prefers Grande drinks?
   p(G) = p(M, G) + p(F, G) = 0.62
d. What is p (V | M), the conditional probability, given a customer in the survey was male, that he prefers venti drinks?
   p(V | M) = p(V, M) / p(M) (0.08 + 0.38 + 0.12) = .206895517
e. What is p (F | V), the conditional probability, given a customer in the survey prefers venti drinks, that the customer was female?
   p(F | V) = p(F, V) / p(V) = 0.06 / (0.06 + 0.12) = 0.333333333 = 1/3
f. There are two random variables in this situation, drink size and gender. Are they independent or dependent? Explain how you arrived at the answer and show your calculations.
   They are dependent of one another because the all genders do not have identical preference mix. Female surveyors prefer tall drinks over venti male surveyors do not.

3. [12 points] What is the expected value of a single roll of a fair six-sided die? Show your calculations.
   The expected probability of rolling a 5 is 1 / 6 = 0.1667
   6 possible outcomes, Ω = 1,2,3,4,5,6
   p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = ⅙
   p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = 1
   Expected value = $\sum_{x}^{|\Omega|} \quad (f(x) \cdot p(x))=1 \cdot \frac{1}{6}+2 \cdot \frac{1}{6}+3 \cdot \frac{1}{6}+4 \cdot \frac{1}{6}+5 \cdot \frac{1}{6}+6 \cdot \frac{1}{6}=3.5$
4. [14 points] The following application of Bayes rule often occurs in actual medical practice. Suppose you have tested positive for a disease. What is the probability you actually have the disease? It depends on the sensitivity and specificity of the test, and on the prevalence (prior probability) of the disease.
   We'll denote:
      a positive test as: Test = pos
      a negative test as: Test = neg
      presence of disease as: Disease = true
      absence of disease as: Disease = false
   We know from clinical studies done on the test before FDA approval that the sensitivity and specificity of the test are:

p (Test = pos | Disease = true) = 0.95 (true positive rate, or sensitivity)
p (Test = neg | Disease = false) = 0.90 (true negative rate, or specificity)

From which we can also deduce:

p (Test = neg | Disease = true) = 0.05 (false negative rate)
p (Test = pos | Disease = false) = 0.10 (false positive rate)

We also know from public health surveys that the disease is relatively rare. The prevalence in the general population is:

p (Disease = true) = 0.01

From which we can deduce:

p (Disease = false) = 0.99

a. Use Bayes rule to calculate p (Disease = true | Test = pos), i.e. the probability you actually have the disease, given the test was positive.

   p(true|pos) = p(pos|true) * p(true) / p(pos)
   p(true|pos) = 0.95 *0.01 / p(pos)

   Solve for p(pos)

   p(pos|true) = 0.95 = p(pos, true)/p(true)
   p(neg|false) = 0.90 = p(neg, false)/p(false)
   p(neg|true) = 0.90 = p(neg,true)/p(false)
   p(pos|false) = 0.10 = p(pos,false)/p(false)

   p(true) = 0.01
   p(false) = 0.99

   p(x) = (x, y) + p(x, !y)
   p(x,y) = p(x|y)*p(y)
   p(x|y) = p(x,y)/p(y)

   p(pos) = p(pos|false) + (p(true)*p(pos|true)) = 0.10 + 0.01*0.95 = 0.1095
   p(true|pos) = 0.95 *0.01 / 0.1095 = 0.0867579909 = 8.67%

b. Calculate the ratio p (Disease = true | Test = pos) / p (Disease = true). [In Bayesian statistics, a ratio like this is interpreted as the effect of new evidence on our beliefs about a probability. In this case, we are concerned with the probability of disease, and the new evidence is the test result.]

   p(true|pos)/p(true) = 0.0867579909 /.01 = 0.8675799087 = 86.75799087%

5. [14 points] Calculate the indicated products of row vector v and matrix A. You can report your results as simple row-by-row listings of elements, using tabs or spaces to separate the elements, as done below for matrix A.

Commented [1]: p(B| A) = p(A |B) * p(B) / p(A) where p(A) = p(A, B) + p(A, not B)

v = [5 3 1] A= 0 1 2
2 1 1

AT = 0 2
1 1
2 1

a. v · AT

5 14

b. A · AT

5 3
3 6

c. AT · A

4 2 2
2 2 3
2 3 5

d. A · vT

5 14

e. Explain why the answers to a) and d) are similar, using the transposition rule.
Transpose rule is defined as $(AB)^T = B^T A^T$, then we define transposition.
Using this definition and the transposition rule $(v \cdot A^T)^T = v^T \cdot A$
This property makes $v \cdot A^t$very similar to $A^t \cdot v$

6. [12 points] What is the cosine of the angle between vectors u and v? Show how you set up the solution, and your actual calculations.
u = [4 2 0 -1] v = [3 -3 2 0]
Using the dot product alternative form $x \cdot y = ||x||||y||cos(\theta)$, so $(x \cdot y)/(||x||||y||) = cos(\theta)$

$$[4\,2\,0\,-1] \cdot [3\,-3\,2\,0]/||[4\,2\,0\,-1]|| \cdot ||[3\,-3\,2\,0]|| = cos(\theta)$$

$$[4\,2\,0\,-1] \cdot [3\,-3\,2\,0] = 6$$

$$||[4\,2\,0\,-1]|| = \sqrt{4^2 + 2^2 + 0^2 - 1^2} = \sqrt{21}$$

$$||[3\,-3\,2\,0]|| = \sqrt{3^2 - 3^2 + 2^2 + 0} = \sqrt{22}$$

$$[4\,2\,0\,-1] \cdot [3\,-3\,2\,0]/||[4\,2\,0\,-1]|| \cdot ||[3\,-3\,2\,0]||=6/(\sqrt{462}) = cos(\theta)$$

$\theta = 73.79080203$ degrese

II/ [2 points each] For each example below, state whether the feature is

I. Numerical: discrete or continuous
II. Categorical or nominal: ordinal

If you're doubtful about any part of an answer, add a few words of explanation.

A. Brightness outdoors as measured by a light meter.
   Numerical continuous
B. Brightness outdoors as indicated by people's subjective judgments (e.g. "it's pretty bright", "hard to see").
   Ordinal
C. Angles as measured in degrees between 0 and 360.
   Numerical continuous
D. Telephone area code.
   Numerical discrete
E. Bronze, silver, and gold medals as awarded at the Olympics.
   Ordinal
F. Height above sea level in meters.
   Numerical continuous
G. Number of patients in a hospital.
   Numerical discrete
H. Ability to pass light in terms of the following values: opaque, translucent, transparent.
   Ordinal
I. Military rank.
   Ordinal
J. Density of a substance in grams per cubic centimeter.
   Numerical continuous
K. Coat check number. (When you attend an event, you often can give your coat to someone who, in turn, gives you a number to use to claim your coat when leaving.)
   Numerical discrete