MATH 241 Spring 2015 Homework #4

Professor Adam Kapelner

Due 5PM outside my office, Tuesday, Mar 3, 2015

(this document last updated Tuesday 24th February, 2015 at 8:29pm)

Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out". Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read the section on independence in and conditional probability in Chapter 3. Chapter references are from the 7th edition of Ross.

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]", purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

This homework is worth 100 points but the point distribution will not be determined until after the due date. Late homework will be penalized 10 points per day.

15 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com` (make sure you upload both the hwxx.tex and the preamble.tex file). If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. **Handing it in without this printout is NO LONGER ACCEPTABLE.** Keep this page printed for your records. Write your name and section below where section A is if you're registered for the 9:15AM–10:30AM lecture and section B is if you're in the 12:15PM-1:30PM lecture.

NAME: _____ SECTION (A or B): _____

Conditional Probability We will solve more problems using conditional probability.

Problem 1

New York is a "concrete jungle where dreams are made of." To this extent, a young upstart tries to drum up business in three of the tallest buildings in the city. Below from left to right are pictured One World Trade Center (104 floors), the Empire State Building (103 floors) and the Bank of America Tower (55 floors).

Consider the case where this person enters one of the three buildings randomly and goes to a random floor.

(a) [easy] Draw a probability tree of this random event. Use "..." notation so your trees don't take up the whole page.

(b) [easy] Are the building selection and floor selection *independent* (*i.e. informationally irrelevant*)? Justify your answer using the definition of statistical independence.

(c) [easy] What is the probability of the businessman winding up on floor 23 of One World Trade Center on a given day?

(d) [harder] What is the probability of the businessman winding up on floor 23 of any building on a given day?

(e) [difficult] If the businessman is on floor 50, what is the probability he is in the Bank of America Tower?

(f) [difficult] What are the *odds* he winds up on any floor between 1 through 55? Interpret "odds" to mean what it usually means: *odds against.*

(g) [E.C.] In one week, the businessman was on floor 12, 15 18, 32 and 59. What is the probability he visited One World Trade Center for more than one of those days?

Problem 2

Assume that the overall probability of contracting breast cancer in a 45 year old American woman is 0.1% on average (or one in a thousand). A typical diagnostic test is a mammographic scane. Assume also that a mammograph scan reading is 80% *sensitive* on average and 95% *specific* on average. Here, "sensitive" means among patients with cancer, the probability that the test is positive and "specific" means among patients without cancer, the probability that the test is negative.

(a) [easy] Denote cancer as C and no cancer as C^C and mammography positive as T and mammography negative as T^C. What is $\mathbb{P}(C)$, $\mathbb{P}(T \mid C)$ and $\mathbb{P}(T^C \mid C^C)$? These are readable from the problem statement above. You must use this notation going forward to get full credit.

(b) [easy] Now solve for $\mathbb{P}(C^C)$, $\mathbb{P}(T^C \mid C)$ and $\mathbb{P}(T \mid C^C)$ using the complement rule.

(c) [easy] Draw a tree with two branches: C vs. C^C and then draw a second set of branches for T vs. T^C (four branches). Mark all four conditional probabilities in this tree's configuration and all four marginal probabilities on the right. Check your answers by assuring that these four marginal probabilities form a partition of $\mathbb{P}(\Omega) = 1$.

(d) [easy] Draw the "inverted" tree. It has two branches: T vs. T^C and then a second set of branches for C vs. C^C (four branches). Mark all four conditional probabilities in this tree's configuration and all four marginal probabilities on the right. Check your answers by assuring that these four marginal probabilities form a partition of $\mathbb{P}(\Omega) = 1$.

(e) [harder] What is $\mathbb{P}(T)$? Use the law of total probability here and explain what the law is and how exactly you're using it to solve this problem.

(f) [difficult] What does $\mathbb{P}(T)$ mean? Answer *in English*.

(g) [harder] Now the money question: if a woman is scanned and tests positive, what is the probability she has cancer? Use the notation I have provided and answer as a *percentage* so it is more viscerally interpretable to you. Do not be alarmed if the answer surprises you.

(h) [harder] You may have done the previous question over and over and gotten frustrated. Your answer is probably correct though. Can you explain why it's so low? Comment on the usefulness of mammography given the post test probability of cancer which you computed.

(i) [harder] If a woman is scanned and tests positive, what is the probability she does *not* have cancer?

(j) [harder] If a woman is scanned and tests negative, what is the probability she does *not* have cancer?

(k) [harder] What is the ratio of $\frac{\mathbb{P}(C \mid T)}{\mathbb{P}(C \mid T^C)}$? What does this ratio mean? What does your answer suggest? Is it possible these scans aren't such a terrible diagnostic tool after all?

Problem 3

We will follow up here with questions on the Monte Hall game.

(a) [harder] In class, we used Bayes Theorem (the law of total probability with many applications of Bayes Rule) to show that if you pick door 1 and door 2 opens, then the probability that the car is in door 3 is 2/3. Repeat that calculation here instead with the following situation: you initially selected door 1, door 3 opens, find the probability the car is in door 2. Hint: it is the same thing as the class notes, but with notation switched around.

(b) [harder] Now imagine a variant of the game is played in the following way: there are four doors, you pick one and the game show host opens up two doors to reveal two goats. You now have the option to keep the door you selected initially or switch to the other door that remains closed. What is the probability of winning if you switch? You can use Bayes Theorem as in (a) or draw a tree like we did in class.

(c) [E.C.] Imagine the variant where there are now n doors. You choose 1 and the game show host opens up $n-2$ doors to reveal $n-2$ goats. You have the option to keep the door you selected initially or switch to the other closed door. What is the probability of winning if you switch?

Trees   We will solve problems using trees and introduce the Bernoulli random variable.

Problem 4

You play a game with your friend. You both roll a die. Whoever rolls higher wins. If you roll the same number, you tie.

(a) [easy] What is the probability you tie?

(b) [easy] What is the probability you win? Draw a tree to figure this out. The first branch is the numerical value of your roll, the second branch is the numerical value of your friend's roll. Then you mark on the right whether you Win (W), Tie (T) or Lose (L).

(c) [harder] Imagine upon ties, the game continues: you both roll again. You play until someone has a higher roll than the other. What is the probability you win this game? Use the algebraic trick we talked about in class.

(d) [E.C.] Consider the same game as described above with one rule change: your friend automatically wins if you both tie on rolling a 1 and a 1. What are fair odds (against) on this variant of the game?

Random Varibles We now begin question about the second unit of this class: r.v.'s.

Problem 5

In class we spoke about how random variables map outcomes from the sample space to a number *i.e.* $X : \Omega \to \mathbb{R}$. That is they are set functions, just like the probability function which is $\mathbb{P} : \Omega \to [0, 1]$. We will be investigating this concept here.
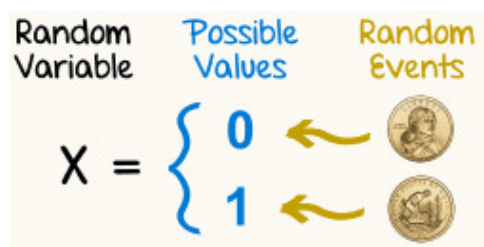
(a) [easy] Here is a way to produce $X \sim$ Bernoulli $\left(\frac{1}{2}\right)$ using the Ω from a roll of a die. Map outcomes 1,2,3 to 0 and outcomes 4,5,6 to 1. This works because

$$\mathbb{P}\left(X = 0\right) = \mathbb{P}\left(\{\omega : X(\omega) = 0\}\right) = \mathbb{P}\left(\{1\} \cup \{2\} \cup \{3\}\right) = 1/2 \text{ and}$$
$$\mathbb{P}\left(X = 1\right) = \mathbb{P}\left(\{\omega : X(\omega) = 1\}\right) = \mathbb{P}\left(\{4\} \cup \{5\} \cup \{6\}\right) = 1/2.$$

Describe three other scenarios or devices that produce their own Ω's that also result in $X \sim \text{Bernoulli}\left(\frac{1}{2}\right).$

(b) [harder] We talked about in class how the sample space no longer needs to be considered once the random variable is described. Why? Use your answer to (a) to inspire this answer. Write it *in English* below.

(c) [difficult] Back to philosophy... Let's say X models the price difference that IBM stock moves in one day of trading. For instance, if the stock closed yesterday at \$56.24 and today it closed at \$57.24, the random variable would be \$1 for today. According to our definition of a random variable, there is a sample space with outcomes being drawn ($\omega \in \Omega$) that is "controlling" the value of X. Describe it the best you can *in English*. There are no right or wrong answers here, but your answer must be coherent and demonstrate you understand the question.

We will now study probability mass functions (PMF's) denoted as $f(x)$ and cumulative distribution functions (CDF's) denoted as $F(X)$ and review the r.v.'s we did in class.

(a) [easy] Draw / graph the PMF for $X \sim \text{Bernoulli}(0.2)$. What is $\mathbb{P}(X = 1)$?

(b) [easy] Take a r.v. X with $\text{Supp}[X] = [0, 1]$. Is this a "discrete r.v.?" Yes / no and explain.

(c) [difficult] In class we defined the Bernoulli r.v. as:

$$X \sim \begin{cases} 1 & \text{w.p.} \quad p \\ 0 & \text{w.p.} \quad 1 - p \end{cases}$$

but we did not put its PMF on the board. Write $p(x)$ for $X \sim \text{Bernoulli}(p)$ that is only valid for values in the $\text{Supp}[X]$.

(d) [difficult] What is the parameter space of X where $X \sim \text{Bernoulli}(p)$ and why?