MATH 241 Spring 2015 Homework #5

Student Jason Farkas Section B

Tuesday 17th March, 2015

Problem 1

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 $\mathbb{P} : 2^\Omega \to [0, 1]$. We will be investigating this concept here.

(a) [easy] Here is a way to produce $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ 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

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

Describe three other scenarios or devices that produce their own Ω's that also result in $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$.
ANSWER: Flipping two coins where we map outcome of landing on the same side to and landing on opposite sides as 0. Having a bag of 10 balls where 5 are blue and choosing 1 ball at random where we map other colors to 1 and the color blue balls to 0. Lastly, Where we spin a wheel of two colors that turns evenly and one color maps to 0, the other to 1.

(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.
ANASWER: All we care about is the results which pop out, we don't care how we got to the results of mapping to a 1 and mapping to a zero, at the end of the day they are the same thing. We don't care what went on, only the results of the mapping.

(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.

ANSWER: There is a sample space of all the possible price differences that could have occurred between the two days. A difference of a \$1 is one potential outcome, which was actually drawn.

Problem 2

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

(a) [easy] Draw the PMF for $X \sim \text{Bernoulli}\,(p)$.

(b) [easy] Draw the CDF for $X \sim \text{Uniform}\,(\{1, 3, 4, 9\})$.

(c) [harder] Using the r.v. from the previous question, what is $\mathbb{P}\,(X \in (3, 9))$? I am trying to trick you here.

(d) [easy] Take a r.v. X with $\text{Supp}\,[X] = [0, 1]$. Is this a "discrete r.v.?" Yes / no and explain.
ANSWER: No because it does not have a countable infinite amount in its support.

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

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

and put its PMF on the board. Write $p(x)$ for $X \sim \text{Bernoulli}\,(p)$ that is only valid for not only all values in the $\text{Supp}\,[X]$ but all values in $\mathbb{R}$. Use the indicator function and set theory notation.

$$p_{1_E}(x) = \begin{cases} \mathrm{P}(E) & \text{if } x = 1 \\ \mathrm{P}(E^c) = 1 - \mathrm{P}(E) & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

(f) [difficult] What is the parameter space of X where $X \sim \text{Bernoulli}\,(p)$ and why?
ANSWER: $X \in [0, 1]$ This is because it take a parameter of a probability which is only in between 0 and 1. Although 0 and 1 are degenerate.

(g) [difficult] Sometimes knowing the Ω matters a little bit. Let's say $X_1 \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ is generated from one coin and $X_2 \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ is generated from another coin independently tossed. Create a new r.v. $T = X_1 + X_2$. Describe its PMF using the $\sim$ notation like in the previous problem. Thus write "$T \sim$" something.
ANSWER: The number represents the number of Heads, (but could also be the number of tails.

$$T \sim \begin{cases} 0 & \text{w.p. } \frac{1}{4} \\ 1 & \text{w.p. } \frac{1}{2} \\ 2 & \text{w.p. } \frac{1}{4} \end{cases}$$

(h) [difficult] Consider the PMF we discussed for $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. Does $\int p(x)\,\mathrm{d}x = F(x) + C$ where the constant $C \in \mathbb{R}$? Explain. Think carefully about what integration really means.
ANSWER: No, it is discrete and therefore there is no curve to take the area under it.

(i) [difficult] How about the opposite? Consider the CDF we discussed for $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. Does $\mathrm{d}/\mathrm{d}x[F(x)] = p(x)$? Explain. Think carefully about what differentiation really means.
ANSWER: No, the limit from the left doesn't equal the limit from the right

Problem 3

The hypergeometric is sampling "without replacement." Imagine you have this bag of marbles with 37 marbles and 17 of them are black. We will define a "success" as drawing a black marble.

(a) [easy] Let's say you draw one marble. Call this r.v. X. Is it hypergeometric?
ANSWER: Yes.

(b) [easy] The hypergeometric distribution has three parameters. What are the parameters for X?
ANSWER: The parameters indicate the size of the sample n -which is 1 in the last example, K - the amount of successes in the total , and N the total amount in the group. There are also other potential parameters that will provide the same information .

(c) [easy] Write, but do not draw, the PDF, $p(x)$ for the r.v. X where x is the number of successes.
$p(x) = \frac{\binom{17}{x}\binom{37-17}{1-x}}{\binom{37}{1}}$

(d) [easy] What is the support of this r.v.?
ANSWER: $\text{Supp}[X] = \{0, ..., 1\}$

(e) [harder] There is another variable we learned about in class with this same support. Show that X is distributed as this type of r.v. and find its parameter(s).
ANSWER: This is a $X \sim \text{Bernoulli}\left(\frac{17}{37}\right)$

(f) [easy] Now imagine you draw 4 marbles without replacement. Call this r.v. X (and forget about the previous r.v. X from this question, parts a-e). How is X distributed? Use the notation in class and find its parameters.
ANSWER: $X \sim Hyper(4, 17, 37)$

(g) [easy] What is the support of X?
$\text{Supp}[X] = \{0, ..., 4\}$

(h) [easy] Write, but do not draw, the PMF of X.
$p(x) = \frac{\binom{17}{x}\binom{20}{4-x}}{\binom{37}{4}}$

(i) [easy] Draw the PMF of X.

(j) [easy] Draw the CDF of X.

(k) [easy] What is the probability of getting 4 successes in a row? Use the PMF.
.036

(l) [easy] What is the probability of getting 4 successes in a row? Use conditional probability. This should yield the same answer.
ANSWER: I'm not sure by what you mean to use conditional probability. What condition?

(m) [easy] Now imagine you draw 27 marbles without replacement. Call this r.v. X (and forget about the previous r.v. X). How is X distributed? Use the notation in class and find its parameters.
ANSWER: $X \sim Hyper(27, 17, 37)$

(n) [easy] What is the support of X? Why is $0 \notin \text{Supp}[X]$?
ANSWER: $\text{Supp}[X] = \{7, ..., 17\}$ It is impossible to choose a total of 27 failures out of a maximum of 20 failures. Therefore there will be a minimum of at least 7 successes chosen.

(o) [easy] Write, but do not draw, the PMF of X.
ANSWER: $p(x) = \frac{\binom{17}{x}\binom{20}{27-x}}{\binom{37}{27}}$

(p) [difficult] Find the mode of this distribution. "Mode" is defined as the most likely outcome result.
ANSWER: You are most likely to get 12 successes

Problem 4

Generally, the hypergeometric has three parameters. We will solve for its support here under several disjoint conditions and then in class we will generalize it. Call X a hypergeometric r.v. with all its parameters free - meaning they can take on any value, so please use the notation n, K, N in your answers as we did in class.

(a) [easy] Using the usual parameterization of the hypergeometric, describe the parameter space. You need to say what sets each of the parameters "lives" in.
ANSWER: $N \in \{2, 3, 4, ...\}$
$K \in \{1, 2, ..., N-1\}$
$n \in \{2, 3, ..., N-1\}$

(b) [easy] Write, but do not draw, the PMF of X.
ANSWER: $p(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$

(c) [harder] x is the free variable in $p(x)$ which you wrote in (b) and it designates the number of successes. Show that successes and failure are essentially the same thing by finding $p(n-x)$ and replacing K with $N-K$. What does this teach you?

$p(n-x) = \frac{\binom{N-K}{n-x}\binom{K}{x}}{\binom{N}{n}}$ This shows that what we call a success or failure is arbitrary, and doesn't effect the math which is based on selected a desired result from a larger group without replacement.

(d) [harder] Let's say $n \leq K$ and $n \leq N-K$. What is the support of X in this situation?
ANSWER: { 0,1,...,n }

(e) [harder] Let's say $n \leq K$ and $n > N-K$. What is the support of X in this situation?
ANSWER: { n-(N-K) ,..., n }

(f) [harder] Let's say $n > K$ and $n \leq N-K$. What is the support of X in this situation?
ANSWER: { 0,...,K }

(g) [difficult] Let's say $n > K$ and $n > N-K$. What is the support of X in this situation?
ANSWER: {n-(N-K),..., K }

(h) [E.C.] Describe the CDF of the general hypergeometric r.v.

Problem 5

We will look at hypergeometric distributions with large N. If N is really large, sampling without replacement can be approximated by sampling with replacement. In the limit, it is sampling with replacement.

(a) [easy] We will now begin deriving the binomial in pieces. Parameterize a hypergeometric by setting $K = pN$. What is the parameter space for p?
ANSWER: $X \sim Hyper(n, pN, N)$
parameter space of $p \in \{\frac{1}{2}, ..., \frac{K}{N}, ..., \frac{N-1}{N}\}$

(b) [easy] Write the PMF $p(x)$ for this r.v. using the p parameterization using x as the free variable.
ANSWER: $= \frac{\binom{N-pN}{n-x}\binom{pN}{x}}{\binom{N}{n}}$

(c) [easy] What limit do we take and why are we taking this limit?
ANSWER: We take the limit as $N->\infty$because we are looking at a really large N

(d) [easy] Rewrite the PMF without choose notation using only factorials and simplify the fraction by moving the factorial terms from denominator, $\binom{N}{n}$, to the numerator.
ANSWER: $\frac{(N-pN)!(pN)!n!(N-n)!}{(n-x)!(N-pN-n+x)!x!(pN-x)!}$

(e) [easy] Which three terms can you factor out from the limit expression? Show that they are equivalent to $\binom{n}{x}$.
ANSWER: Since n! and x! and (n-x)! are constants they are irrelevant when taking the limit as $N->\infty$ and therefore can be factored out to $\frac{n!}{(n-x)!x!} = \binom{n}{x}$