Homework 2, DATA 556: Due Tuesday, 10/09/2018

Alexander Van Roijen

October 28, 2018

Please complete the following:

1. Let X be a random variable with CDF F, and $Y = \mu + \sigma X$, where μ and σ are real numbers with $\sigma > 0$. Find the CDF of Y, in terms of F.

$$P(X \leq x) = F(x) \tag{1}$$
$$P(Y \leq y) = P(\mu + \sigma X \leq y) \tag{2}$$
$$= P(\sigma X \leq y - \mu) \tag{3}$$
$$= P(X \leq \frac{y-\mu}{\sigma}) \tag{4}$$
$$= F(\frac{y-\mu}{\sigma}) \tag{5}$$

2. (a) Show that p(n) $= (\frac{1}{2})^{n+1}$, for $n = 0, 1, 2, ...$ is a valid PMF for a discrete random variable.

   i. Showing p(n) $> 0 \ \forall n \geq 0$ by induction

$$\text{base case, n=0:} \tag{6}$$
$$p(0) = (\frac{1}{2})^{0+1} = 0.5 > 0 \tag{7}$$
$$\text{Now assume this is true for all 0 ... n, NTS for n+1} \tag{8}$$
$$p(n+1) = (\frac{1}{2})^{n+1+1} = (\frac{1}{2})^{n+1} * (\frac{1}{2})^{1} = p(n) * p(1) > 0 \text{ by the induction hypothesis} \tag{9}$$

   ii. $\sum_{n=0}^{\infty} p(n) = 1$

   This is trivial, as a well known geometric series that sums to one. Thus this property is also satisfied and the problem is done □

(b) Find the CDF of a random variable with PMF from (a)

$$F(x) = p(X \leq x) = p(X = 0) + p(X = 1) + p(X = 2)... + P(X = x) \tag{10}$$

$$\text{This is due to the discrete nature of } p(n) \tag{11}$$

$$p(X = x) = (\frac{1}{2})^{n+1} \tag{12}$$

$$=> F(x) = \begin{cases} 0 & x < 0 \\ (\frac{1}{2})^{x+1} + F(x-1) & x \geq 0 \end{cases} \tag{13}$$

$$\text{Further, this sum will look eerily like the geometric series shown above} \tag{14}$$

3. Let X, Y and Z be discrete random variables such that X and Y have the same conditional distribution given Z, i.e., for all a and z we have $P(X = a|Z = z) = P(Y = a|Z = z)$ Show that X and Y have the same distribution (unconditionally, not just when given Z)

$$\text{The law of total probability states} P(X) = \sum_{i=0}^{\infty} P(X|Z = z_i) * P(Z = z_i) \tag{15}$$

$$\text{We NTS P(X) = P(Y)} \tag{16}$$

$$P(X = x|Z = z) = P(Y = x|Z = z) \forall x, z \in S \tag{17}$$

$$=> \sum_{i=0}^{\infty} P(X = x|Z = z_i) = \sum_{i=0}^{\infty} P(Y = x|Z = z_i) \tag{18}$$

$$=> \sum_{i=0}^{\infty} P(X = x|Z = z_i) * P(Z = z_i) = \sum_{i=0}^{\infty} P(Y = x|Z = z_i) * P(Z = z_i) \tag{19}$$

$$\text{the above is true as we are simply multiplying each factor of the sum by the same value on both sides} \tag{20}$$

$$=> \text{Thus, by the definition of the law of total probablity we have } P(X) = P(Y) \Box \tag{21}$$

4. (a) Let X ∼ DUnif(C), and B be a nonempty subset of C. Find the conditional distribution

of X, given that X is in B.

$$\text{we have } B \subset C \text{ and X} \sim \text{DUnif(C)} => P(X = x) = \frac{1}{|C|} \tag{22}$$

$$P(X = x|x \in B) = \frac{P(X = x \cap x \in B)}{P(x \in B)} \text{ (Definition of the conditional)} \tag{23}$$

$$\text{we further know the following} \tag{24}$$

$$P(x \in B) = \sum_{j=0}^{\infty} P(X = x_j|x \in B) * P(x_j) \text{ (law of total probability)} \tag{25}$$

$$\text{we know} P(X = x_j) = \frac{1}{|C|} \text{ and } \sum_{j=0}^{\infty} P(X = x_j|x \in B) = |B| \text{ as } B \subset C \text{ and } B \neq \emptyset \tag{26}$$

$$=> P(X = x|x \in B) = \frac{P(X = x \cap x \in B)}{\frac{|B|}{|C|}} = \frac{(\frac{1}{|C|})}{\frac{|B|}{|C|}} = \frac{1}{|B|} \tag{27}$$

$$=> P(X = x|x \in B) \sim \text{DUnif(B)} \tag{28}$$

Since we know that $x \in B$, then we do not need to concern ourselves of any values outside of the space of $B \subset C$

Further, since X follows a uniform distribution within the space of C and $B \subset C$, then similarly, X has equal chances of taking values on within B

Thus, $X|B \sim DUnif(B)$

(b) If X follows HGeom(w, b, n), what is the distribution of $n - X$?

$$XHGeom(w, b, n) => P(X = k) = \frac{\binom{w}{k} * \binom{b}{n-k}}{\binom{n}{k}} \tag{29}$$

$$P(n - X = k) = P(X = n - k) = \frac{\binom{w}{n-k} * \binom{b}{n-(n-k)}}{\binom{n}{n-k}} = \frac{\binom{w}{n-k} * \binom{b}{k}}{\binom{n}{k}} \tag{30}$$

$$\text{*This is due to the symmetric property of n choose k and n choose n - k} \tag{31}$$

$$=> n - X \sim \text{HGeom(b,w,n)} \tag{32}$$

5. A copy machine is used to make n pages of copies per day. The machine has two trays in which paper gets loaded, and each page is taken randomly and independently from one of the other trays. At the beginning of the day, the trays are refilled so that they each have m pages. Using simulations in R, find the smallest value of m for which there is at least a

95 percent chance that both trays have enough paper on a particular day, for n = 10, n = 100, n = 1000, and n = 10000.

```
n=c(10,100,1000,10000)

findMinM2 = function(n,p){
m=(n/2)
numTries=10000
while (TRUE) {
numCorrect=0
results = rbinom(numTries,n,0.5)
l = ((m-results)>=0)
r = ((m - (n-results))>=0)
numCorrectvec=l&r
numCorrect=length(numCorrectvec[numCorrectvec=="TRUE"])
#print(numCorrect)
sum = numCorrect/numTries
if(sum>=0.95)
{
return(m)
}
else
{
m= m+1
}
}

for (ns in n)
{
print(findMinM2(ns,0.5))
```

```
}
> [1] 8
> [1] 60
> [1] 531
> [1] 5097
```

it is worth noting that this is easily checked using pbinom. This code can be found on my github!