# Homework 3, DATA 556: Due Tuesday, 10/16/2018

**Alexander Van Roijen**

October 12, 2018

Please complete the following:

1. Problem 1

   (a) Find the mean and the variance of a Discrete Uniform random variable on 1, 2, . . . , n.

   Let $X \sim$ DUnif

$$P(X = x) = \frac{1}{n} \text{ by the def of uniform distrib} \tag{1}$$

$$E[X] = \sum_{j=0}^{n} P(X = j) * j = \sum_{j=0}^{n} \frac{1}{n} * j = \frac{1}{n} * \sum_{j=0}^{n} j = \frac{1}{n} * \frac{n(n+1)}{2} = \frac{n+1}{2} \tag{2}$$

   It is a similar proof with a different subsitution for the expected value of X squared

$$\tag{3}$$

$$E[X^2] = \sum_{j=0}^{n} P(X = j) * (j^2) \text{ LOTUS} \tag{4}$$

$$= \sum_{j=0}^{n} \frac{1}{n} * (j^2) = \frac{1}{n} * \sum_{j=0}^{n} (j^2) = \frac{1}{n} * \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6} \tag{5}$$

$$E[X^2] - (E[X])^2 = \frac{(n+1)(2n+1)}{6} - (\frac{n+1}{2})^2 \tag{6}$$

$$\frac{4 * (n+1)(2n+1)}{24} - (\frac{6 * (n+1)^2}{24}) = \frac{n^2 - 1}{12} \tag{7}$$

   (b) Let X be a discrete random variable that satisfies the properties listed in the homework. Find E[X]

$$\text{We have a symmetry such that } P(X = x) = P(X = -x) \tag{8}$$

$$E[X] = \sum_{x=-n}^{n} P(X = x) * x \tag{9}$$

$$= P(X = -n) * -n + P(X = -(n-1)) * (-(n-1)) + ... + P(X = n) * n \tag{10}$$

$$= -1 * (P(X = -n) * n - P(X = n) * n) + ...P(X = 0) * 0 \tag{11}$$

$$= -1 * (P(X = n) * n - P(X = n) * n) + ...P(X = 0) * 0 = 0 + 0 + ... + 0 \tag{12}$$

$$\text{due to the symmetry of the probs} \tag{13}$$

$$=> E[X] = 0 \tag{14}$$

2. We have X with PMF $P(X = k) = \frac{-1*p^k}{log(1-p)*k}$ for k = 1, 2, . . . .. Here p is a parameter with 0 ¡ p ¡ 1. Find the mean and the variance of X

First, note that the value $\frac{-1}{log(1-p)}$ is a positive constant

$$E[X] = \sum_{k=1}^{\infty} P(X=k) * k = \sum_{k=1}^{\infty} \frac{-1 * p^k}{log(1-p) * k} * k = \sum_{x=1}^{\infty} \frac{-1 * p^k}{log(1-p)} \tag{15}$$

$$= \frac{-1}{log(1-p)} \sum_{x=1}^{\infty} p^k \tag{16}$$

$$=> \sum_{x=1}^{\infty} p^k = S => S * p = \sum_{x=2}^{\infty} p^k => S - Sp = p => S = \frac{p}{(1-p)} \tag{17}$$

$$=> E[X] = \frac{-1}{log(1-p)} * \frac{p}{(1-p)} \text{similarly, the } E[X^2] \text{is proved in a similar manner except} \tag{18}$$

$$S - Sp = \sum_{x=1}^{\infty} p^k \tag{19}$$

$$\text{which we already proved so by substituting we get} \tag{20}$$

$$S - Sp = \frac{p}{(1-p)} => S = \frac{p}{(1-p)^2} \tag{21}$$

$$=> E[X^2] = \frac{-1}{log(1-p)} * \frac{p}{(1-p)^2} \tag{22}$$

$$=> Var(X) = E[X^2] - E[X]^2 = \frac{p(p - log(1-p))}{log(1-p)^2 * (1-p)^2} \tag{23}$$

3. (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

$$\text{base case, n=0:} \tag{24}$$

$$p(0) = (\frac{1}{2})^{0+1} = 0.5 > 0 \tag{25}$$

$$\text{Now assume this is true for all 0 ... n, NTS for n+1} \tag{26}$$

$$p(n+1) = (\frac{1}{2})^{n+1+1} = (\frac{1}{2})^{n+1} * (\frac{1}{2})^1 = p(n) * p(1) > 0 \text{ by the induction hypothesis} \tag{27}$$

   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)

$$F(x) = p(X \leq x) = p(X = 0) + p(X = 1) + p(X = 2)... + P(X = x) \tag{28}$$

$$\text{This is due to the discrete nature of } p(n) \tag{29}$$

$$p(X = x) = (\frac{1}{2})^{n+1} \tag{30}$$

$$=> F(x) = \begin{cases} 0 & x < 0 \\ (\frac{1}{2})^{x+1} + F(x-1) & x \geq 0 \end{cases} \tag{31}$$

$$\text{Further, this sum will look eerily like the geometric series shown above} \tag{32}$$

4. 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)

$$\text{The law of total probability states} P(X) = \sum_{i=0}^{\infty} P(X|Z = z_i) * P(Z = z_i) \tag{33}$$

$$\text{We NTS P(X) = P(Y)} \tag{34}$$

$$P(X = x|Z = z) = P(Y = x|Z = z) \forall x, z \in S \tag{35}$$

$$=> \sum_{i=0}^{\infty} P(X = x|Z = z_i) = \sum_{i=0}^{\infty} P(Y = x|Z = z_i) \tag{36}$$

$$=> \sum_{i=0}^{\infty} P(X = x|Z = z_i) * P(Z = z_i) = \sum_{i=0}^{\infty} P(Y = x|Z = z_i) * P(Z = z_i) \tag{37}$$

the above is true as we are simply multiplying each factor of the sum by the same value on both sides (38)

$$=> \text{Thus, by the definition of the law of total probablity we have } P(X) = P(Y) \square \tag{39}$$

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!