

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

Alexander Van Roijen

October 14, 2018

Please complete the following:

1. Problem 1

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

Let $X \sim \text{DUnif}$

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

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

It is a similar proof with a different substitution for the expected value of X squared (3)

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

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

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

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

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

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

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

$$= P(X = -n) * -n + P(X = -(n-1)) * (-(n-1)) + \dots + P(X = n) * n \quad (10)$$

$$= -1 * (P(X = -n) * n - P(X = n) * n) + \dots P(X = 0) * 0 \quad (11)$$

$$= -1 * (P(X = n) * n - P(X = n) * n) + \dots P(X = 0) * 0 = 0 + 0 + \dots + 0 \quad (12)$$

$$\text{due to the symmetry of the probs} \quad (13)$$

$$\Rightarrow E[X] = 0 \quad (14)$$

2. We have X with PMF $P(X = k) = \frac{-1 * p^k}{\log(1-p) * k}$ for $k = 1, 2, \dots$. 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

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

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

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

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

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

$$\text{which we already proved so by substituting we get} \quad (20)$$

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

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

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

3. (a) Use LOTUS to show that for $X \sim \text{Pois}(\lambda)$ and any function $g(\cdot)$, $E[Xg(X)] = \lambda E[g(X+1)]$

$$E[Xg(X)] = \sum_{x=0}^{\infty} x * g(x) \frac{e^{-\lambda} * \lambda^x}{x!} = \sum_{x=1}^{\infty} g(x) * \lambda \frac{e^{-\lambda} * \lambda^{x-1}}{(x-1)!} \quad (24)$$

$$\text{now let there be a substitution such that } j = x-1 \quad (25)$$

$$= \sum_{j=0}^{\infty} g(j+1) * \lambda \frac{e^{-\lambda} * \lambda^j}{j!} = \lambda * \sum_{j=0}^{\infty} g(j+1) \frac{e^{-\lambda} * \lambda^j}{j!} \quad (26)$$

$$\text{considering } g(x) \text{ is any valid function, we conclude} \quad (27)$$

$$E[Xg(X)] = \lambda * E[g(X+1)] \quad (28)$$

We bring this back to this form using LOTUS and our lack of knowledge on $g(x)$ (29)

(b) Find the third moment $E(X^3)$ for $X \sim \text{Pois}(\lambda)$

$$\text{let } g(x) = x^2 \Rightarrow E(Xg(X)) = E[g(X+1)] = \lambda * E[(X+1)^2] \quad (30)$$

$$= \lambda * E[X^2 + 2X + 1] = \lambda * (E[X^2] + 2E[X] + 1) \quad (31)$$

using the known properties of the variance and mean of the uniform dist we get
(32)

$$= \lambda * (\lambda^2 + \lambda + 2\lambda + 1) = \lambda^3 + 3\lambda^2 + \lambda \quad (33)$$

4. Show that for any events $A_1 \dots A_n$, $P(A_1 \cap \dots \cap A_n) \geq \sum_{j=1}^n P(A_j) - n + 1$

lets first begin by studying this same situation for indicator random variables for these events.

One more thing to keep in mind is that we know

$$0 \leq P(X = x) \leq 1 \forall x \text{ and similarly } I(A) = 1 \text{ or } I(A) = 0 \text{ first we show } I(A_1 \cap \dots \cap A_n) \geq \sum_{j=1}^n I(A_j) - n + 1$$

$$\text{Case 1: } I(A_1 \cap \dots \cap A_n) = 0 \quad (34)$$

$$\Rightarrow \sum_{j=1}^n I(A_j) \leq (n-1) \quad (35)$$

$$\text{this is self explanatory as at most } n-1 \text{ indicator terms are } 1 \quad (36)$$

$$\Rightarrow \sum_{j=1}^n I(A_j) - n + 1 \leq I(A_1 \cap \dots \cap A_n) \quad (37)$$

$$\text{Case 2: } I(A_1 \cap \dots \cap A_n) = 1 \quad (38)$$

$$\Rightarrow \sum_{j=1}^n I(A_j) = n \quad (39)$$

$$\text{this is self explanatory as all } n \text{ indicator variables must be } 1 \quad (40)$$

$$\Rightarrow \sum_{j=1}^n I(A_j) - n + 1 = n - n + 1 = 1 \leq I(A_1 \cap \dots \cap A_n) = 1 \quad (41)$$

$$\text{Now we look at the broader case of } P(A_1 \cap \dots \cap A_n) \geq \sum_{j=1}^n P(A_j) - n + 1 \quad (42)$$

$$\text{we know that the total law of probability follows that} \quad (43)$$

$$\text{can we say sum of all } A \text{ probs must be less or equal to } 1? \quad (44)$$

$$\text{or can we say that the expected value of indicator} = P \text{ implies its true?} \quad (45)$$

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!