Homework 4, DATA 556: Due Tuesday, 10/23/2018

Alexander Van Roijen

October 21, 2018

Please complete the following:

1. Problem 1 Let X be a continuous random variable with CDF F and PDF f.

   (a) Find the conditional CDF of X given $X > a$ (where a is a constant with $P(X > a) \neq 0$). That is, find $P(X \leq x|X > a)$ for all a, in terms of F.

   $$P(X \leq x|X > a) = \int_a^x f(x)dx + \int_\infty^a f(x)dx = F(x) - F(a) + F(a) = F(x) \quad (1)$$

   (b) Find the conditional PDF of X given $X > a$.

   $$\text{Take the result form above and take the derivative} \quad (2)$$

   $$f(x) = F'(x) = \frac{dF(x)}{dx} = F'(x) - F'(a) = f(x)\text{as}\frac{dF(a)}{dx} = 0 \quad (3)$$

   $$=> \text{the distributions pdf is no different} \quad (4)$$

   (c) Check that the conditional PDF from (b) is a valid PDF, by showing directly that it is non negative and integrates to 1.

2. Problem 2: A circle with a random radius $R \sim Unif(0, 1)$ is generated. Let A be its area.

   (a) Use R to simulate mean and variance of A

```
> set.seed(123)
> #this is 2a
> n=1000000
> results = runif(n,0,1)
> counter = 1
> areas=numeric(0)
> while(counter<n)
+ {
+   areas[counter] = results[counter] * results[counter] * pi
+   counter= counter + 1
+ }
```

```
> print(mean(areas))
[1] 1.045967
> print(var(areas))
[1] 0.8775327
```

(b) Find the theoretical mean and the variance of A, without first finding the CDF or PDF of A. Compare with your numerical results from (a).

$$E[A] = \pi E[r^2] \tag{5}$$

$$\text{we know r folows Unif(0,1)} => E[r^2] = Var(r) + E[r]^2 = \frac{1}{12} + \frac{1}{2}^2 = \frac{1}{3} \tag{6}$$

$$=> E[A] = \frac{\pi}{3} = 1.047198 \tag{7}$$

(c) Find the CDF and PDF of A.

$$F(x) = P(A \leq x) = P(\pi R^2 \leq x) = P(R \leq \frac{\sqrt{x}}{\sqrt{\pi}}) = F(x) = \frac{\sqrt{x}}{\sqrt{\pi}} \tag{8}$$

$$f(x) = F'(x) = \frac{1}{\pi} * \frac{d\sqrt{x}}{\sqrt{dx}} = \frac{1}{2\sqrt{x\pi}} \tag{9}$$

$$\text{with } 0 \leq x \leq \pi \tag{10}$$

3. A stick of length 1 is broken at a uniformly random point, yielding two pieces. Let X and Y be the lengths of the shorter and longer pieces, respectively, and let $R = \frac{X}{Y}$ be the ratio of the lengths of X and Y.

(a) Use simulations in R (the statistical programming language) to gain some understanding about the distribution of the random variable R. Numerically estimate the expected value of R and $1/R$.

```
> set.seed(123)
> n= 1000
> results = runif(n,0,1)
> counter = 1
> xs = numeric(0)
> ys = numeric(0)
> while(counter <= n)
```

```
+ {
+   if(results[counter]>=0.5)
+   {
+     ys[counter]=results[counter]
+     xs[counter] = 1-ys[counter]
+   }
+   else
+   {
+     xs[counter] = results[counter]
+     ys[counter] = 1 -xs[counter]
+   }
+   counter = counter + 1
+ }
> rs = xs/ys
> print(mean(rs))
[1] 0.3877773
> print(mean(1/rs))
[1] 15.82287
```

(b) Find the CDF and PDF of R.

We note that since x is exclusively smaller than y of a unit length stick, then $0 \leq X \leq 0.5$ and $0.5 \leq Y \leq 1$ and $X = 1 - Y$ and $Y = 1 - X$

$$P(R \leq r) = P(X/Y \leq r) = P(X \leq r * Y) \sim \text{DUnif}(0, \frac{0.5}{}) \tag{11}$$

(c) Find the expected value of R (if it exists).

$$E[R] = E[\frac{X}{Y}] = E[\frac{X}{1-X}] = \tag{12}$$

4. Let $U_1, ..., U_n$ be i.i.d. $Unif(0,1)$, and $X = max(U_1, ..., U_n)$.

(a) What is the PDF of X?

$$\text{CDF} = P(X \leq x) = P(U_1 \leq x, ..., U_n \leq x) = P(U_1 \leq x) * ... * P(U_n \leq x) \text{ since i.i.d} \tag{13}$$

$$=> \text{CDF} = x * x * x ... * x = x^n => \text{ PDF } = \frac{dF}{dx} = n * x^{n-1} \text{ with } 0 \leq x \leq 1 \tag{14}$$

(b) what is the $E[X]$

$$\int_{-\infty}^{\infty} x * f(x) = \int_0^1 x * f(x) = \int_0^1 x * n * x^{n-1} = \int_0^1 n * x^n = \frac{n}{n+1} * (1^n) - 0 = \frac{n}{n+1} \tag{15}$$

(c) R simulation results / approx

```
> #4c
> set.seed(123)
> n=10
> runplenty=10000
> og = runif(n,0,1)
> counter=1
> result= numeric(0)
> while(counter<=runplenty)
+ {
+   og = runif(n,0,1)
+   result[counter]=max(og)
+   counter= counter+1
+ }
> print(mean(result))
[1] 0.9091953
> print(n/(n+1))
[1] 0.9090909
```

5. (a) Find $P(X < Y)$ for X $\sim N(a, b), Y \sim N(c, d)$ with X and Y independent

$$P(X < Y) = P(X - Y < 0) \sim N(a - b, c^2 + d^2) \text{ determined via hint} \tag{16}$$

$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{x-\mu^2}{2\sigma^2}} \text{ Where } \sigma = c^2 + d^2 \text{ and } \mu = a - b \tag{17}$$

(b) R simulation

```
#5b
> set.seed(123)
> n= 1000000
> resultsx = rnorm(n,0,1)
> resultsy = rnorm(n,1,5)
> trueResults = rnorm(n,-1,sqrt(26))
> counter = 1
> numCorr=0
> numCorr2=0
> results=numeric(0)
> while(counter<=n)
+ {
+   if(resultsx[counter]<resultsy[counter])
+   {
+     numCorr = numCorr + 1
+   }
+   if(trueResults[counter]<0)
+   {
+     numCorr2 = numCorr2 + 1
+   }
+   counter= counter+1
+ }
> print(numCorr/n)
[1] 0.576969
> print(numCorr2/n)
[1] 0.577764
```

6. The heights of men in the United States are normally distributed with mean 69.1 inches and standard deviation 2.9 inches. The heights of women are normally distributed with mean 63.7 inches and standard deviation 2.7 inches. Let x be the average height of 100

randomly sampled men, and y be the average height of 100 randomly sampled women.

(a) What is the distribution of x - y?

$$\text{Similar to 5a, just with a different valued distribution} \tag{18}$$

$$x - y \sim N(69.1 - 63.7, \sqrt{2.9^2 + 2.7^2}) = N(5.4, \sqrt{15.7}) \tag{19}$$

$$E[2^X] = \sum_{n=0}^{\infty} \frac{2^n e^{-\lambda} \lambda^n}{n!} = \sum_{n=0}^{\infty} \frac{e^{-\lambda}(2 * \lambda)^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(2 * \lambda)^n}{n!} \tag{20}$$

$$\text{now let } \lambda' = 2 * \lambda => e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda')^n}{n!} \tag{21}$$

$$\text{we recognize this is the taylor expansion of the exponential function} \tag{22}$$

$$\text{we get} \sum_{n=0}^{\infty} \frac{(\lambda')^n}{n!} = e^{\lambda'} = e^{2\lambda} => E[2^X] = e^{-\lambda} e^{2\lambda} = e^{\lambda} \square \tag{23}$$

7. For X $\sim$ Geom(p), find $E[2^X]$ and $E[2^{-X}]$ (if it is finite).

$$E[2^X] = \sum_{n=1}^{\infty} 2^n (1-p)^{n-1} * p \text{ let} S = \sum_{n=1}^{\infty} 2^n (1-p)^{n-1} = 2 + 4(1-p) + ... \tag{24}$$

$$=> S * 2(1-p) = 4(1-p) + 8(1-p)^2 + .. => S - S * 2(1-p) = 2 \tag{25}$$

$$S - 2S + 2Sp = -S + 2Sp = S(-1 + 2p) = 2 => S = \frac{2}{2p-1} => E[2^X] = p * S = \frac{2p}{2p-1} \tag{26}$$

$$\text{Similarly } E[2^{-X}] = \sum_{n=1}^{\infty} 2^{-n} (1-p)^{n-1} * p \text{ we let} S = \sum_{n=1}^{\infty} 2^{-n} (1-p)^{n-1} = \sum_{n=1}^{\infty} \frac{(1-p)^{n-1}}{2^n} \tag{27}$$

$$= \frac{(1-p)^0}{2^1} + \frac{(1-p)^1}{2^2} + ... => S * \frac{1-p}{2} = \frac{(1-p)^1}{2^2} + \frac{(1-p)^2}{2^3} \tag{28}$$

$$=> S - S * \frac{1-p}{2} = S(1 - \frac{1-p}{2}) = S * \frac{1+p}{2} = \frac{1}{2} => S = \frac{1}{1+p} \tag{29}$$

$$=> E[2^{-X}] = \frac{p}{1+p} \square \tag{30}$$