HW5 Solutions

Dr. Nick

3/4/2020

4.4.6

a. The median $\tilde{x}$ can be computed

$$F(x) = \int_0^x 2tdt = x^2, \qquad \tilde{x} = \sqrt{2}/2.$$

Then,

$$P(X_{(1)} > \sqrt{2}/2) = \left(1 - F\left(\frac{\sqrt{2}}{2}\right)\right)^3 = (1 - (1/2))^3 = 1/8$$

b. The joint density is given by (see textbook)

$$g(y_1, y_2, y_3) = 48 y_1 y_2 y_3.$$

Then, the joint of (Y_1, Y_2) can be found by integrating over y_1 to get

$$g(y_2, y_3) = 24 y_2^3 y_3.$$

Next, compute several expectations:

$$E(Y_2 Y_3) = \int_0^1 \int_0^{y_3} 24 y_2^4 y_3^2 dy_2 dy_3 = 3/5$$

$$g(y_2) = \int_{y_2}^1 24 y_2^3 y_3 dy_3 = 12 y_2^3 (1 - y_2^2)$$

$$E(Y_2) = 24/35$$

$$E(Y_3) = 6/7$$

$$\sigma_{Y_2} = 0.1726149$$

$$\sigma_{Y_3} = 0.123718$$

$$Corr(Y_2, Y_3) = 0.5733819$$

4.4.7

The formulas in the text for continuous distributions make use of differentiation of CDFs and use the fact that $P(X < c) = P(X \leq c)$ when X is continuous. For discrete r.v.'s we have to be more careful about inequalities, and pmfs are not exactly analogous to pdfs so the formulas have a different interpretation.

We can use a counting argument to argue that the given function is the pmf of Y_1. First,

$$P(Y_1 = 1) = P(\text{not all are bigger than 1}) = 1 - (5/6)^5$$

Next,

$$P(Y_1 = 2) = P(\text{not all are bigger than 2 and none are 1})$$

$$= P(\text{none are 1}) - P(\text{none are 1 and all are bigger than 2})$$

$$= (5/6)^5 - (4/6)^5$$

This argument can be repeated for $P(Y_1 = 3)$ and so on…and these probabilities match the pmf given.

4.4.8

Use Formula 4.4.3:

$$5!(1 - e^{-y_2})(e^{-y_2} - e^{-y_4})e^{-y_4}e^{-y_2}e^{-y_4}$$

Put $y_2 = z_1$ and $y_4 = z_1 + z_2$ and note the absolute value of the determinant of the Jacobian is simply 1:

$$= 5!(1 - e^{-z_1})e^{-4z_1}e^{-2z_2}(1 - e^{-z_2})$$

which factors into a function of z_1 and a function of z_2. Therefore, the two are independent.

4.7.6

```
data<-matrix(c(15,25,32,17,11,9,18,29,28,16),5,2)
chisq.test(data)
```

```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 6.4019, df = 4, p-value = 0.1711
```

```
expected.counts<-t(outer(colSums(data), rowSums(data))/sum(data))
test.stat<-sum(((data-expected.counts)^2)/expected.counts)
test.stat
```

```
## [1] 6.401891
```

```
qchisq(.95,4)
```

```
## [1] 9.487729
```

```
pchisq(test.stat,4)
```

```
## [1] 0.828922
```

Do not reject the null hypothesis of homogeneity.

4.8.2

```
U <- runif(10000)
mc.est <- mean(1/(1+U))
est.se <- sd(1/(1+U))/sqrt(10000)
abs.error<-abs(mc.est - log(2))
abs.error
```

```
## [1] 0.001581317
```

```
perc.error <- abs.error/log(2)
perc.error
```

```
## [1] 0.002281358
```

```
c(mc.est - log(2) - 1.96*est.se, mc.est - log(2) + 1.96*est.se)
```

```
## [1] -0.001150324  0.004312958
```

4.8.18

a. To use the inverse CDF method first find the CDF

$$F(x) = \int_0^x \beta t^{\beta-1} dt = x^\beta.$$

Then, we find the inverse:

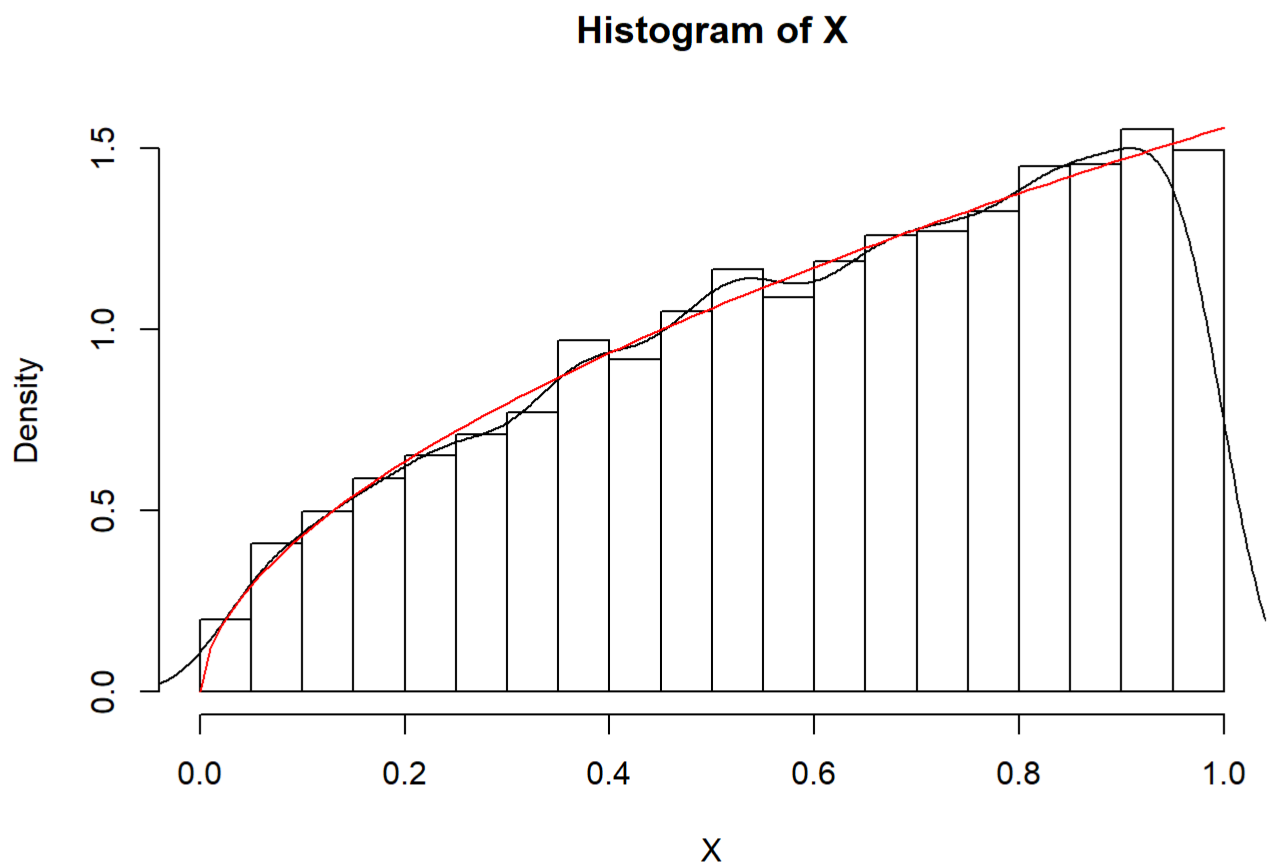
$$y = x^\beta \Rightarrow x = y^{1/\beta}.$$

Therefore, to general X we generate $U \sim Unif(0,1)$ and assign $X = U^{1/\beta}$. For a simulation of this see the R code below (this was optional):

```
beta<- rexp(1)+1
beta
```

```
## [1] 1.556455
```

```
U <- runif(10000)
X <- U^{1/beta}
hist(X, freq = FALSE)
lines(density(X))
curve(beta*x^(beta-1), 0,1,add = TRUE, col = 'red')
```



b)Notice that $\beta \exp(-(-x+1))$ dominates $\beta x^{\beta-1}$ when $\beta \geq 2$. This first function is a constant multiple of a shifted, reflected, and restricted exponential distribution with CDF $\frac{e^{x-1}-e^{-1}}{1-e^{-1}}$ for $x \in (0,1)$ so we can obtain samples from the inverse CDF method.

For $1 < \beta \leq 2$ use a uniform distribution for the proposal distribution g.

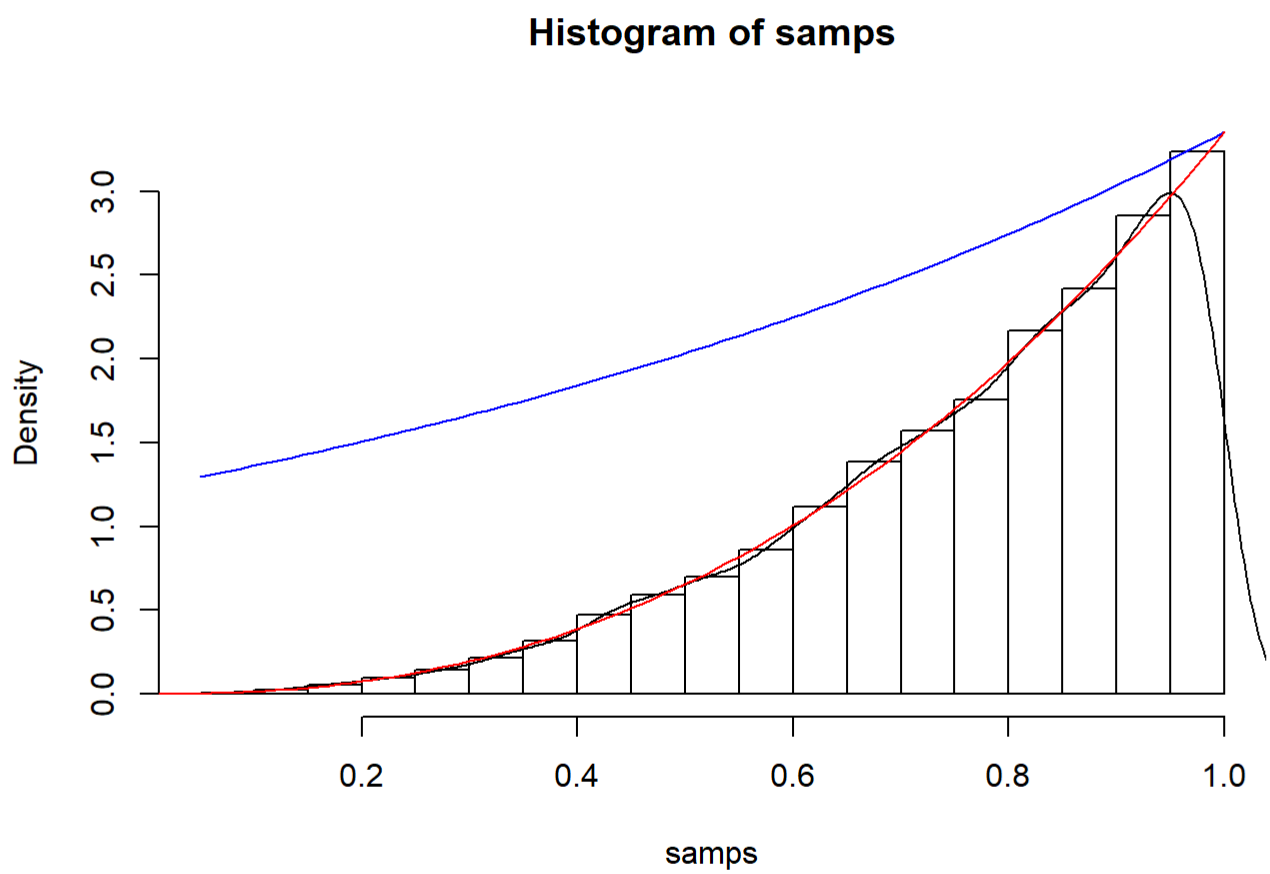
```
beta<- rexp(1)+1  # just picking a beta at random, it has to be at least 1
beta
```

```
## [1] 3.353966
```

```
desired.samps <- 10000
num.samps <- 0
num.tries<-0
samps<-rep(NA, desired.samps)
if(beta>2){
while(num.samps < 10000){
 x <- log(runif(1)*(1-exp(-1))+exp(-1))+1 #reflected and shifted exponential
  u <- runif(1)
  if(u<=((beta*(x^(beta-1)))/(beta*(exp(x-1)/(1-exp(-1)))))) {
    num.samps <- num.samps+1
    samps[num.samps]<-x
  }
  num.tries <- num.tries+1
}

hist(samps, freq=FALSE, ylim = c(0,beta))
lines(density(samps))
curve(beta*exp(-(1-x)), add=TRUE, col='blue')
curve(beta*x^(beta-1), 0,1,add = TRUE, col = 'red')
print(num.samps/num.tries)
print(1/beta)
} else {
while(num.samps < 10000){
 x <- runif(1) #reflected and shifted exponential
  u <- runif(1)
  if(u<=((beta*(x^(beta-1)))/(beta))) {
    num.samps <- num.samps+1
    samps[num.samps]<-x
  }
  num.tries <- num.tries+1
}

hist(samps, freq=FALSE, ylim = c(0,beta))
lines(density(samps))
abline(a=beta, b=0, col="blue")
curve(beta*x^(beta-1), 0,1,add = TRUE, col = 'red')
print(num.samps/num.tries)
print(1/beta)
}
```




```
## [1] 0.3019324
## [1] 0.2981545
```

4.8.21

This problem requires a good deal of geometry. Consider the joint distribution of (U, V). Since each is independent $Unif(-1, 1)$ they are jointly independent coordinates on the square in the Euclidean plane with vertices $(\pm 1, \pm 1)$. And, $W = U^2 + V^2$ is the square of the hypotenuse of the triangle formed by $(0, 0)$, $(U, 0)$, and (U, V). If we consider only points (U, V) with $W < 1$ then we have the circle of radius 1 which is inscribed in the above square. Points (U, V) such that $W < 1$ must be uniformly distributed on the circle of radius 1 with center at the origin. Next turn your attention to X_1 and X_2. We have

$$X_1 = \frac{U}{\sqrt{W}}\sqrt{-2\log W}, \quad \text{and} \quad X_2 = \frac{V}{\sqrt{W}}\sqrt{-2\log W}.$$

Thinking back to the triangle formed by $(0, 0)$, $(U, 0)$, and (U, V) notice that $\frac{U}{\sqrt{W}} = \cos\theta$ and $\frac{V}{\sqrt{W}} = \sin\theta$ where θ is the angle btween the hypotense and the x-axis. Given W which is the square of the radius (hypotenuse) we actually do not know anything about θ. Hence, θ is independent from W which means $\frac{V}{\sqrt{W}}$ and $\frac{U}{\sqrt{W}}$ are independent of W. What is the distribution of W, the square of the radius? Well, the probability $W < w$ is simply the area of the circle with radius $\sqrt{w}$ divided by the area of the circle of radius 1, which is $\frac{\pi w}{\pi} = w$.

Therefore, W is uniformly distributed on $(0,1)$ because its CDF matches that of a Uniform $(0,1)$ CDF. What is the distribution of θ? Since the points (U,V) are uniform on the disk, it's clear that the angle θ is uniformly distributed on $(0,2\pi)$ with density $1/(2\pi)$. Then, write X_1 and X_2 as

$$X_1 = \cos\theta\sqrt{-2\log W}, \quad \text{and} \quad X_2 = \sin\theta\sqrt{-2\log W}.$$

We can solve these equations for θ and W noting that $\cos\arctan x = \frac{1}{\sqrt{1+x^2}}$ so that

$$\theta = \frac{x_1}{\sqrt{x_1^2+x_2^2}}, \quad \text{and} \quad W = e^{-\frac{1}{2}(x_1^2+x_2^2)}.$$

Then, use the transformation method:

$$\frac{\partial W}{\partial x_1} = -x_1 e^{-\frac{1}{2}(x_1^2+x_2^2)}, \quad \text{and} \quad \frac{\partial W}{\partial x_2} = -x_2 e^{-\frac{1}{2}(x_1^2+x_2^2)}$$
$$\frac{\partial \theta}{\partial x_1} = -\frac{x_2}{x_1^2+x_2^2}, \quad \text{and} \quad \frac{\partial \theta}{\partial x_2} = \frac{x_1}{x_1^2+x_2^2}$$

using the fact that $\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$. The determinant of the Jacobian (in absolute value) is $e^{-\frac{1}{2}(x_1^2+x_2^2)}$ and the density of θ is $\frac{1}{2\pi}$ so the joint density of (X_1, X_2) is the product of standard normal densities, as claimed.