Stat 450

R Assignment 3

Fall 2016

Key to HW 3 (thanks to Sierra Suhr for nice write-up!)

As discussed in class, the $UNIF(0, 1)$ distribution is very important for simulating continuous random variables with any other distribution, as long as the CDF has closed form.

Consider the following pdf that is positive on the interval [a,b] for $b > a$:

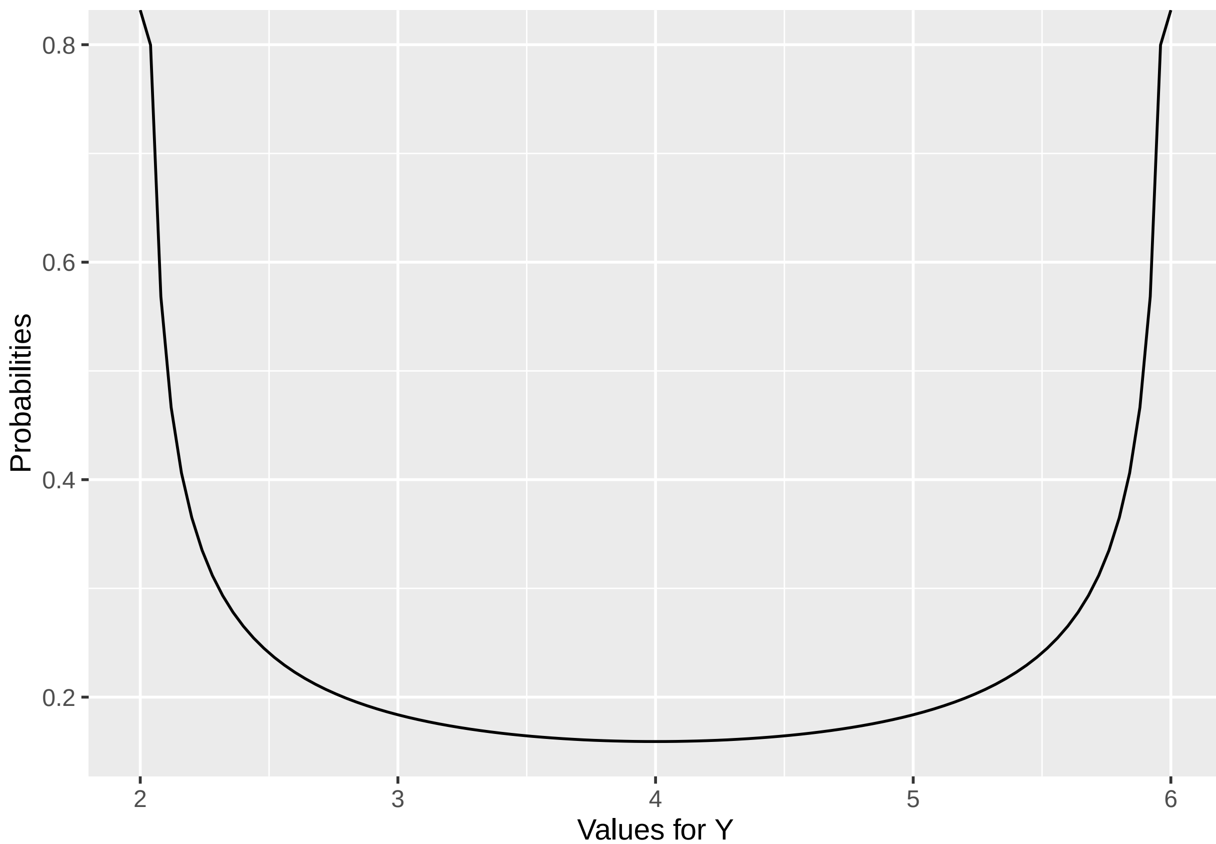
$$f(y) = \frac{1}{\pi\sqrt{(y-a)(b-y)}}, \ a \leq y \leq b$$

a. (3pts) Graph this pdf in R for $a = 2$ and $b = 6$.

```
p <- pi
pmf.fun <- function(y)
  {
  1/(p*sqrt((y-2)*(6-y)))
  }
library(ggplot2)
```

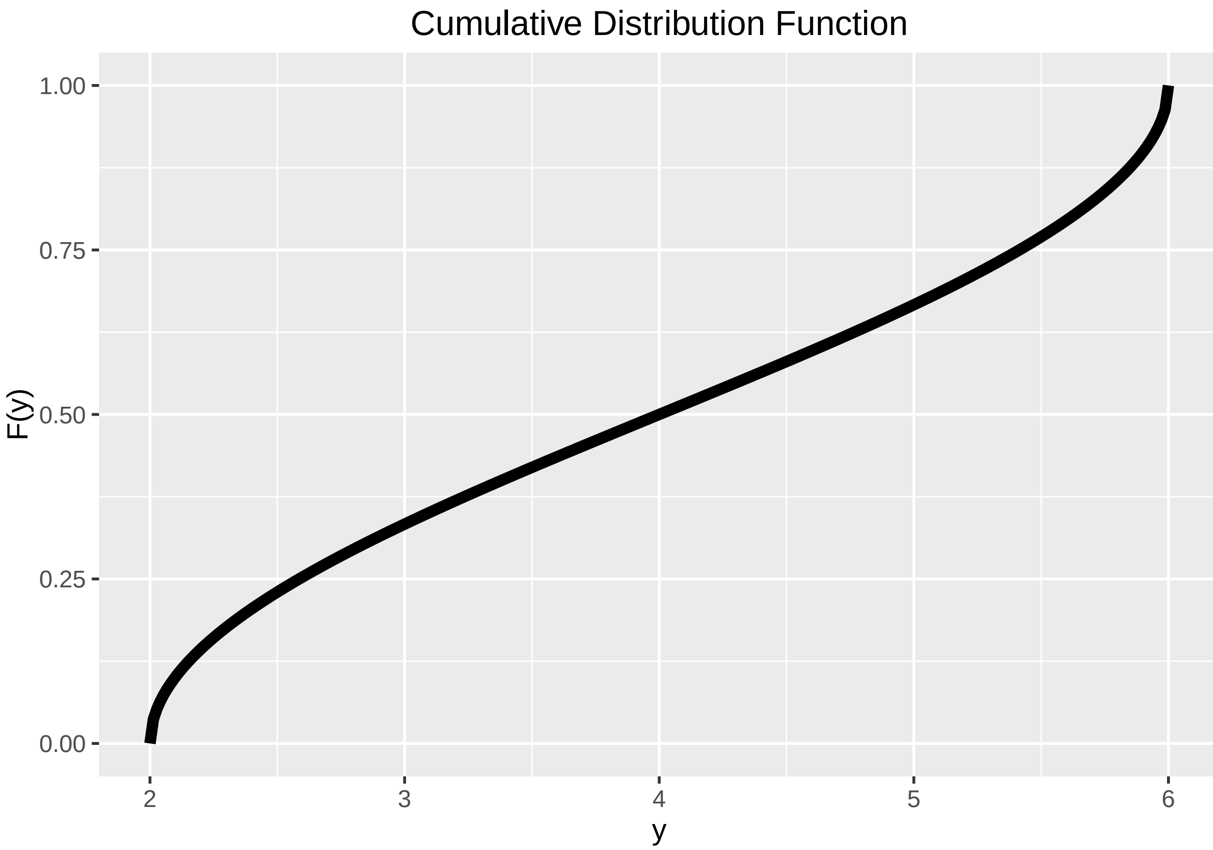
```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
ggplot(data.frame(y=c(2,6)), aes(x=y)) + stat_function(fun=pmf.fun, geom="line") +
  xlab("Values for Y") + ylab("Probabilities")
```



b. (4pts) Write a function that takes as input $\{y, a, b\}$ and returns $F(y)$. Use this function to graph the CDF, $F(y)$, in R for $a = 2$ and $b = 6$. (Hint: to find the CDF, consider the change of variable $u = \frac{y-a}{b-a}$, then the change of variable $t = \sqrt{u}$, then consider inverse trigonmetric functions. You do not need to show your work for this integral, but you do need to write code to generate $F(y)$ and plot this CDF over the relevant range of y.)

```
F.y <- function(y,a,b) {
CDF <- rep(NA,length(y))
region1 <- which(y < a)
region2 <- which(a<=y & y<=b)
region3 <- which(y>b)
CDF[region1] <- 0
CDF[region2] <- (2/pi)*asin(sqrt((y-a)/(b-a)))
CDF[region3] <- 1
return(CDF)
}
yvals <- seq(2,6,length=300)
library(ggplot2)
mydata <- data.frame(y = yvals, height= F.y(yvals,2,6))
ggplot(aes(x=y, y = height), data = mydata) + geom_line(size=2) +
ylab('F(y)') + ggtitle('Cumulative Distribution Function')
```



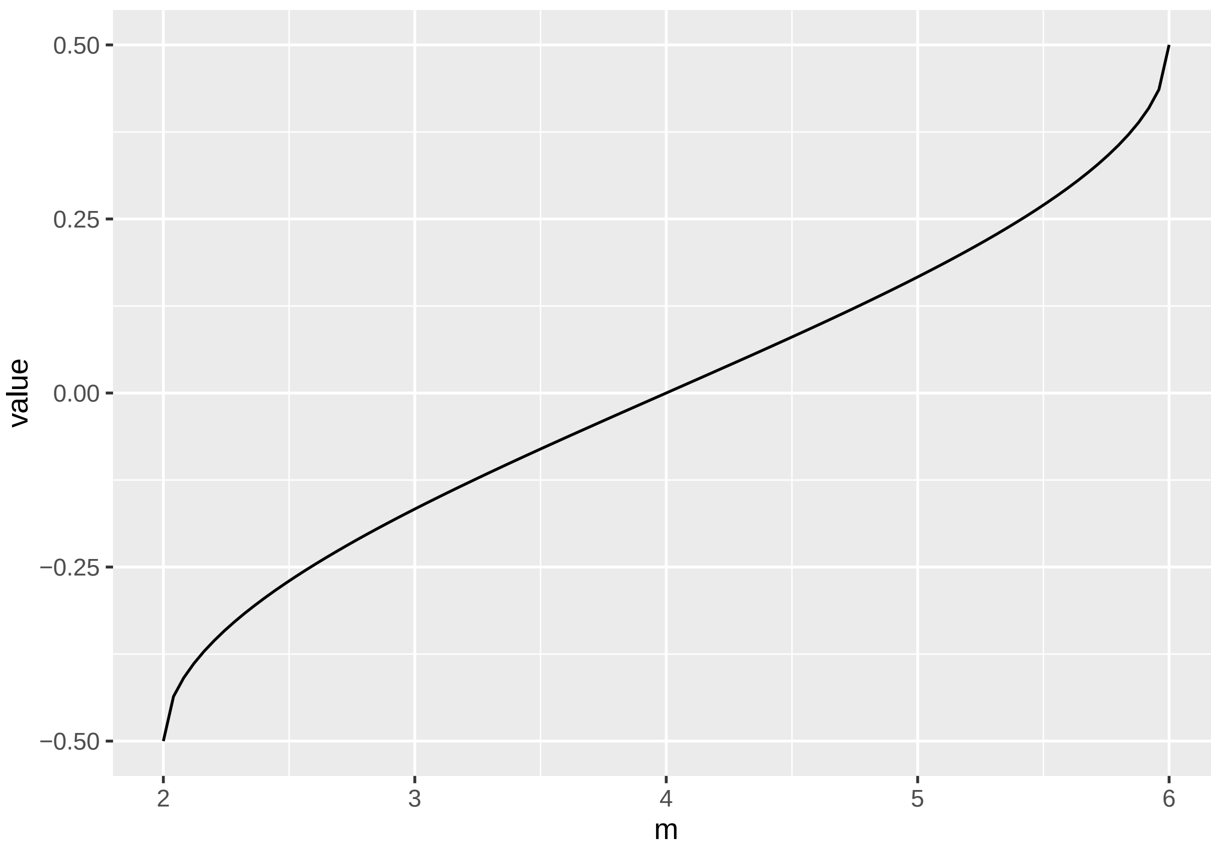
c. (1pt) Use your function to find $P(3 \leq Y \leq 5)$ for $a = 2$ and $b = 6$.

```
F.y(5,2,6)-F.y(3,2,6)
```

```
## [1] 0.3333333
```

d. (2pts) Based on your plot from c, what appears to be the median of Y? Find the exact value of the median by modifying your CDF function and using `uniroot()`. The median of Y appears to be 4.

```
integral <- function(m){
  tosolve <- F.y(m,2,6)-0.5
  return(tosolve)
}
mvals <- seq(2,6, length = 100)
newdat <- data.frame(m = mvals, value = integral(mvals))
ggplot(aes(x=m,y=value),data=newdat) + geom_line()
```

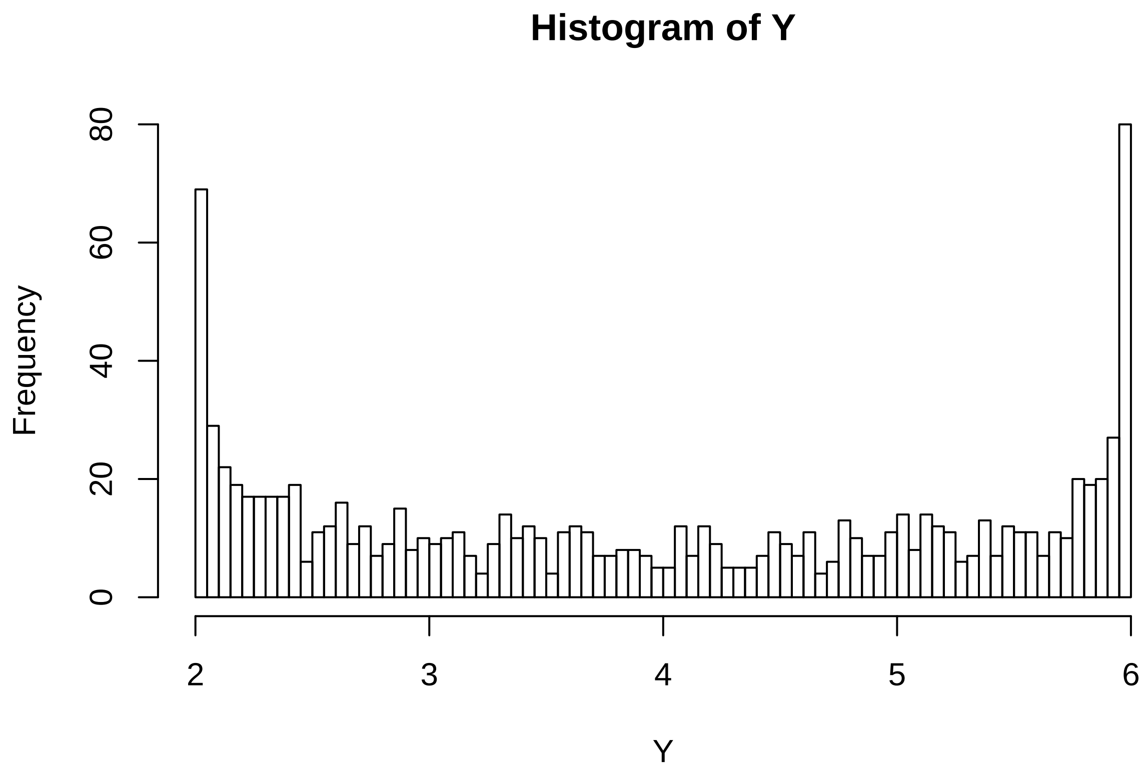



```
uniroot(integral,interval = c(2,6))$root
```

```
## [1] 4
```

e. (3pts) Now simulate 1000 realizations of Y by simulating 1000 realizations of $U \sim UNIF(0,1)$ and letting $Y = F^{-1}(u)$. Plot a histogram of these random realizations; how does its shape compare with the graph of your pdf from part a? Yes, these shapes look identical. (U-Shaped).

```
u <- runif(1000,0,1)
Y <-  4*(sin((pi*u)/2))^2+2
hist(Y,breaks=100)
```



f. (2pts) Find $\hat{P}(3 \leq Y \leq 5)$; the *empirical* (i.e., observed) proportion of your simulated data that is $\in [3, 5]$. Is it similar to your answer from c? These values are very close (0.33333)

```
count <- ifelse(Y<=5 & Y>=3, 1,0)
sum(count)/1000
```

```
## [1] 0.339
```

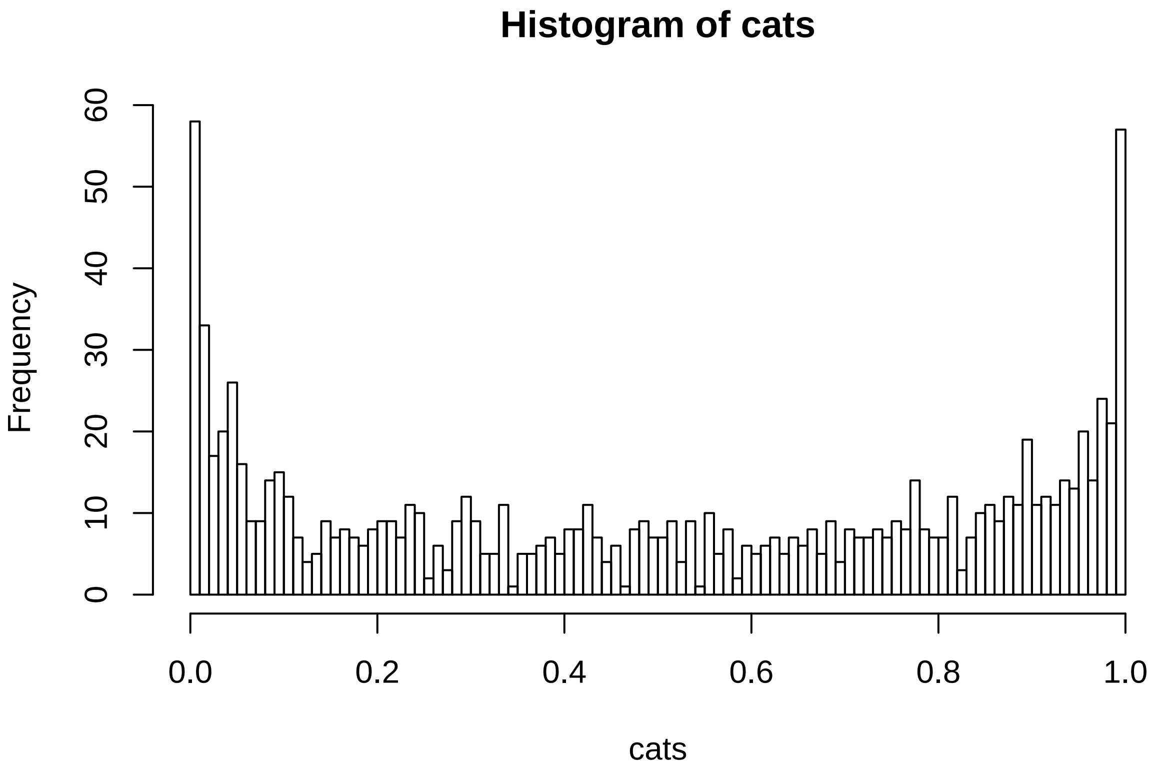
g. (2pts) Find the median of your 1000 realizations. Is it similar to your answer from d?

```
median(Y)
```

```
## [1] 3.869989
```

h. (3pts) Use `rbeta()` to simulate 1000 realizations of Y when $a = 0$ and $b = 1$, and plot a histogram of these realizations. alpha $= 0.5$ beta $= 0.5$

```
cats <- rbeta(1000,0.5,0.5)
hist(cats,breaks=100)
```


Histogram of cats
Frequency
0
10
20
30
40
50
60
0.0
0.2
0.4
0.6
0.8
1.0
cats