Homework 2 (STAT 5860)

Eric Pettengill

Due by 11:59 pm, Feb. 12, 2018

Instructions:

1. Download the `Homework2.Rmd` file from the course Elearning.
2. Open `Homework2.Rmd` in RStudio.
3. Replace the *"Your Name Here"* text in the `author` with your own name.
4. Write your answer to each problem by editing `Homework2.Rmd`.
5. After you finish all the problems, click `Knit to PDF` to create a pdf file. Upload your pdf file to Homework 1 Dropbox in the course Elearning.

Set the seed number

```
set.seed(212)
```

Problem 1. The Pareto distribution has been used in economics as a model for a density function with a slowly decaying tail. The Pareto(a,b) distribution has pdf

$$f(x) = \frac{ab^a}{x^{a+1}}, \quad x \geq b$$

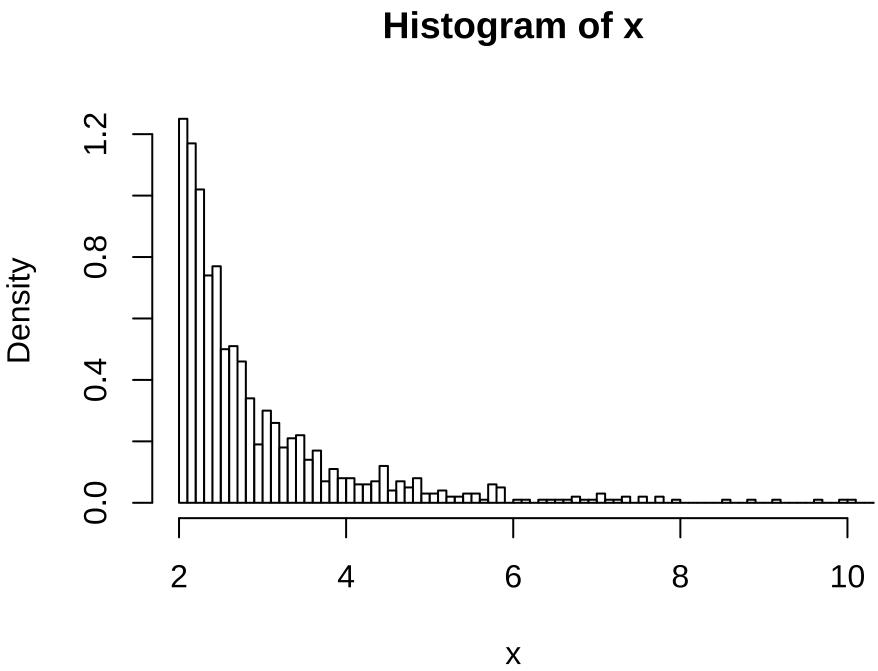
where $a > 0$ is shape paraemter and $b > 0$ is scale parameter.

(a) Use the inverse transform method to generate a random sample of size 1000 from the Pareto(3,2) distribution.

```
n <- 1000
u <- runif(n)
x <- 2/(1-u)^(1/3)
```

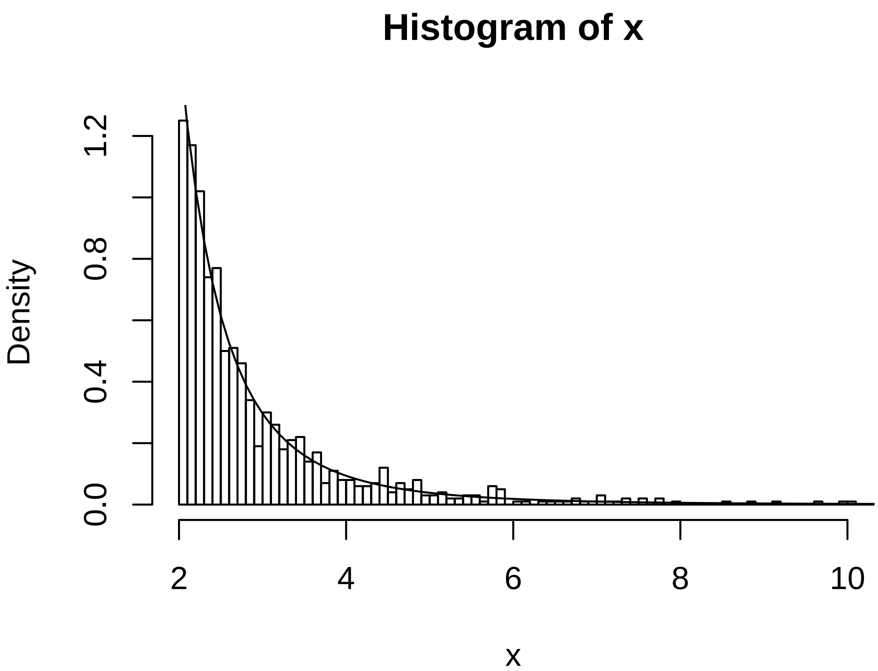
(b) Plot the density histogram of the sample.

```
hist(x, probability = TRUE, xlim = c(2,10), breaks = 500)
```



(c) Add the theoretical density curve $f(x)$ to the density histogoram.

```
hist(x, probability = TRUE, xlim = c(2,10), breaks = 500)

y <- seq(2, 30, .1)
lines(y, 24*y^(-4))
```



Problem 2. A discrete random variable X has probability mass functon (pmf)

x	0	1	2	3
$p(x)$	0.064	0.288	0.432	0.216

(a) Use the inverse transform method to generate a random sample of size 1000 from the distribution of X. Write your R code with `if else` statements. (Don't copy `while` loop code from Lecture 5 R code to solve this problem.)

```
n <- 1000
x <- rep(0, n)
u <- runif(n)

for(i in 1:n){
  if ((u[i] <= .064) & (u[i] > 0)){
    x[i] <- 0
  } else if ((u[i] <= .352) & (u[i] > .064)){
    x[i] <- 1
  } else if ((u[i] <= .784) & (u[i] > .352)){
    x[i] <- 2
  } else if ((u[i] <= 1) & (u[i] > .784)){
    x[i] <- 3
  }
}
```

(b) Calculate relative frequency for each x and compare with the theoretical probabilities.

```
for (i in 0:3) {
  print(sum(x == i)/length(x))
  }
```

```
## [1] 0.058
## [1] 0.277
## [1] 0.448
## [1] 0.217
```

Problem 3. The normal distribution is one of the most important distribution in statistics because the nature provides numerous examples of populations of measurements that, at least approximately, follow a normal distribution.

(a) Use the rejection sampling method to generate a random sample of size 1000 from the standard normal distribution. Use a standard Cauchy distribution as a proposal distribution $g(y)$. You can generate a random sample from the standard Cauchy distribution using `rcauchy()`. Moreover, you can calculate normal density and Cauchy density using `dnorm()` and `dcauchy()` respectively. Choose c such that $f(y) \leq cg(y)$ and report your choice of c.

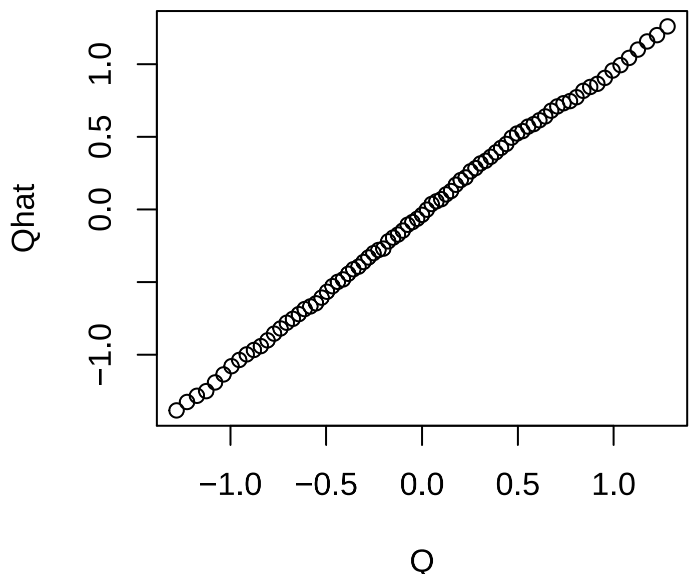
Let $c = 2$.

```
n <- 1000
k <- 0  ##ACCEPTED
j <- 0  ##ITERATIONS
x <- rep(0,n)

while (k < n){
  u <- runif(1)
  j <- j + 1
  cauch <- rcauchy(1)
```

```
  if ((dnorm(cauch)/(2*dcauchy(cauch))) > u){
    k <- k + 1
    x[k] <- cauch
  }
}
```

(b) Compute the sample percentiles and compare with the standard normal distribution percentiles.

```
p <- seq(0.1,0.9, 0.01)
Qhat <- quantile(x,p)
Q <- qnorm(p)

plot(Q,Qhat)
```



Problem 4. The random variable X has pdf

$$f(x) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}} + \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-3)^2}{2}}, \quad -\infty < x < \infty$$

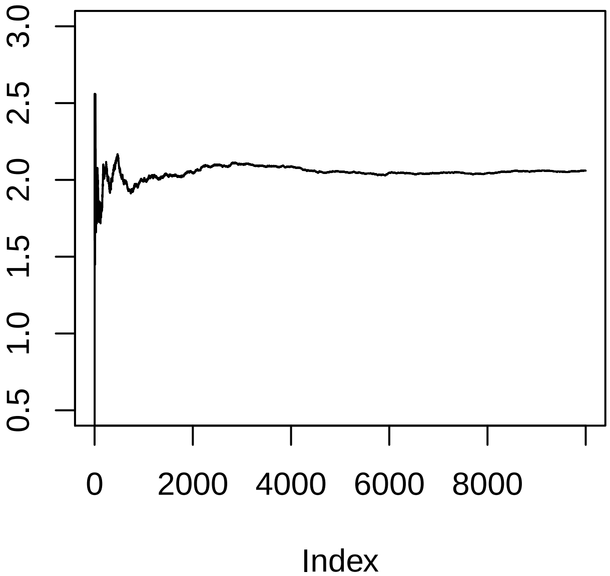
(a) Use the importance sampling method to estimate $\mathrm{E}_f(X)$.

```
n <- 10000
x <- rnorm(n, mean = 2, sd = 4)
y <- x*.5*(dnorm(x, mean = 1, sd = 1) + dnorm(x, mean = 3, sd = 1))/dnorm(x, mean = 2, sd = 4)
mean(y)
```

```
## [1] 2.059722
```

(b) Draw a plot to show the convergence of the importance sampling approximation.

```
plot(cumsum((x*.5*(dnorm(x, mean = 1, sd = 1) +
                    dnorm(x, mean = 3, sd = 1))/dnorm(x, mean = 2, sd = 4)))/(1:n),
     ylab ="",
     type = "l",
     ylim = c(0.5,3))
```


0.5
1.0
1.5
2.0
2.5
3.0
0
2000
4000
6000
8000
Index