Computational Statistics Homework 1

Liam Dillingham

February 22, 2019

1 Question 1

The function $p(x) = sin(x)$ is a density for $x \in (0, \pi/2)$.

1.1 Describe an inverse CDF transform method to sample random variables with this density. Plot the histogram and the true density for visual verification.

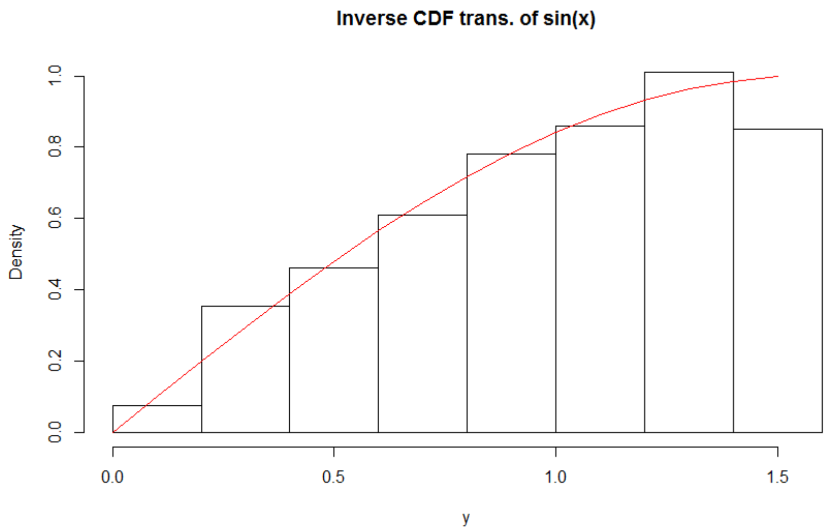
if the PDF is $p(x) = sin(x)$ from $(0, \pi/2)$, then we write the CDF as:

$$\int_0^{\pi/2} sin(x)dx$$

so the result of the integral should be $-cos(x)$. Now we need the inverse of $cos(x)$, which is $arccos(x)$. so our inverse CDF transform should be: $-arccos(x)$, but it resulted in the incorrect output. We can see the code for my inverse transform, and how I drew the graph below:

```
isin_rvgen = function(n=1000) {
  u = runif(n)
  x = acos(1-u)
  return(x)
}

x = seq(0, pi/2, 0.1)
fx = sin(x)
y = isin_rvgen()
hist(y, freq = F, main = "Inverse CDF trans. of sin(x)")
lines(x, fx,  col="red")
```

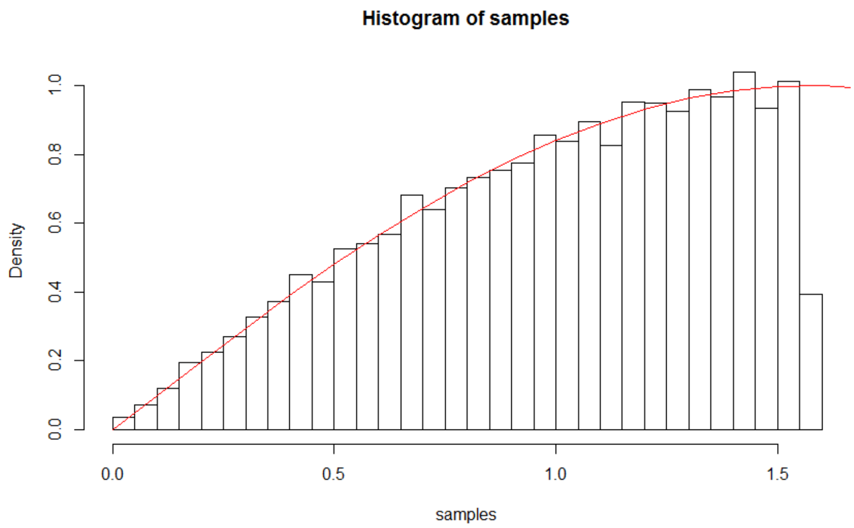


The red line is the graph of $sin(x)$ from $(0, \pi/2)$. and the histogram is contains the samples from our inverse CDF, $arccos(x)$.

1.2 Set up a rejection sampling method to sample from $p(x)$ using a proposal density $g(x)$. Plot the histogram and the true density for visual verification

Hint: You can take $g(x)$ as the uniform density on the interval $(0, \pi/2)$.

```
arfunc = function(n=1000) {
  rv = rep(0,n)
  a = 0
  for(i in 1:n) {
    x = runif(1)
    if(x > -acos(x)) {
      rv[i] = x
      a = a + 1
    }
  }
  cat("Efficiency is ", a/n)
  plot(rv)
  title("accept reject plot")
}
arfunc()
```



2 Question 2

Determine a method to draw samples from the distribution with PDF: $f(x) \propto exp(-x^4/12)$, for $x \in R$. Turn in derivation and code. Plot histogram and true density for visual verification.

We can relate this density to its general form, where $f(x) \propto exp(-kx^n)$. In our case, $k = 1/12$, and $n = 4$. Let $Y = X^4$. Note $x > 0$ so $y > 0$. We want to perform a change of variables so we can have our density in terms of y.

$\frac{dy}{dx} = 4x^3$. From the change of variables theorem, $f_y(y) = f_x(x) \mid \frac{dy}{dx} \mid$. Thus, $f_y(y) = exp(-x^4/12)\frac{1}{4}x^{-3}$. Now we need to swap out x for y. Observe the relationship: $y = x^4$ Which is change to $x = y^{1/4}$. By subsitution, we have

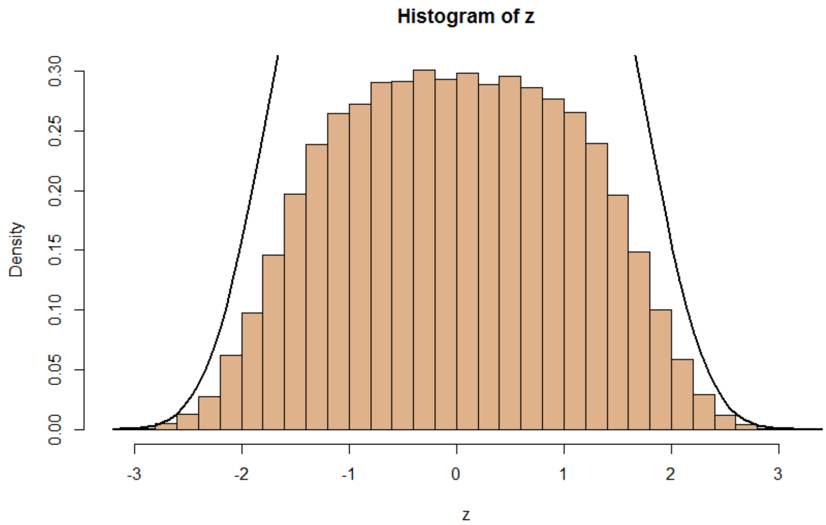
$$f_y(y) = exp(-y/12)\frac{1}{4}y^{-3/4}$$

Note that this is similar to the Gamma density:

$$f(u) = \frac{\lambda^\alpha e^{-\lambda u} u^{\alpha-1}}{\Gamma(\alpha)}, u \geq 0$$

However, we are sampling over all of R, so we need to do a coin-flip on every sample, and decide whether the sample must be negative or positive.

```
z = rgamma(1e5, shape = (1/4), rate = (1/12))
z = z^(1/4)
f = runif(1e5)
for(i in 1:1e5) {
  if(f[i] < 0.5) {
    z[i] = z[i] * -1
  }
}
hist(z, breaks = 30, freq = F, col = rgb(0.75,0.4,0.1,0.5)) # z is your sample
lambda = 1/12; alpha = 1/4
target <- function(x){exp(-lambda*x^(1/alpha))/
     integrate(function(x) exp(-lambda*x^(1/alpha)),0,Inf)$value}
curve(target,lwd=2,add=T)
```



3 Question 3

Suppose that $X \in R^{nxn}$ is a random matrix with independent $\mathcal{N}(0,1)$ entries. Let ℓ_1 be the smallest Eigenvalue of $S_n = X^T X/n$ and let $Y = n\ell_1$. Edelman showed that as $n \to \infty$, the PDF of Y approaches

$$f(y) = \frac{1+\sqrt{y}}{2\sqrt{y}} e^{-(y/2+\sqrt{y})},\ 0 < y < \infty$$

3.1 Develop a method to sample Y from the density $f(y)$ given in the equation above. Show derivation and code. Plot histogram and true density for visual verification.

I don't have a theoretical explanation for how I got the results that I did. I'm not very familiar with all of the density functions so I did a lot of searching until I found your email about truncated distributions. I read the documentation on the "truncdist" package and tried it out. After some tinkering I managed to get a function that covered our original $f(y)$. I am not sure about the quality of the sampler at the extreme tails, but this could be improved by modifying the parameters underlying the truncated distribution (such as the variance in my underlying normal dist.) but I did not have enough time to learn the documentation.

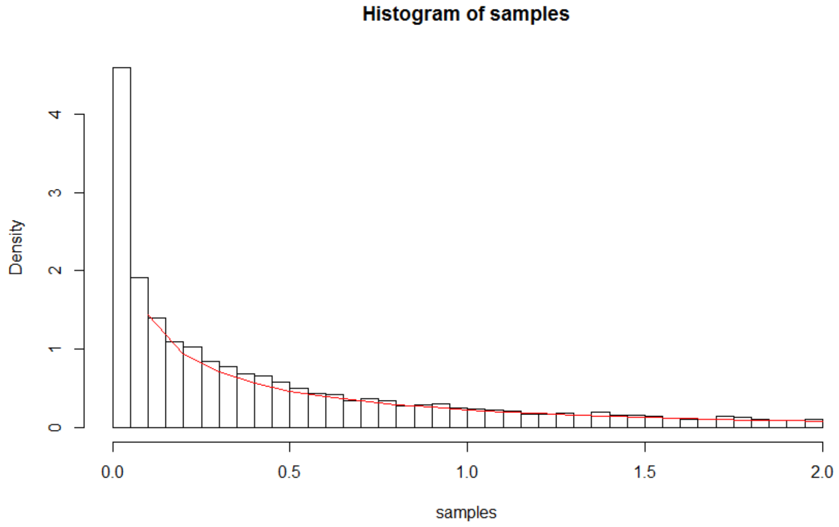
```
target = function(y) { ((1+sqrt(y)) / (2 * sqrt(y))) * exp(-1 * (y/2 + sqrt(y))) }

arfunc = function(M) {
  while(TRUE) {
    # value between [0,1]
    u = runif(1)

    # value from envelope dist
    x = rtrunc(1, spec = 'norm', a=0, b=2)
    if(u <= target(x) / dtrunc(x, spec = 'norm', a=0, b=2) * M) {
      return(x)
    }
  }
}

n = 10000
samples = rep(0,n)
for(i in 1:n) {
  samples[i] = arfunc(0.05)
}

plot(samples)
hist(samples, freq = F, breaks = 30)
lines(seq(0, 2, 0.1), target(seq(0, 2, 0.1)), col="red")
```



(As a side note, I can email my source code if you want)

3.2 Test method by estimating $E(log(Y))$ by simple Monte Carlo, and giving a 99% confidence interval. Edelman found that the answer was roughly -1.68788.