

Midterm 01 - STAT440

Joseph Sepich (jps6444)

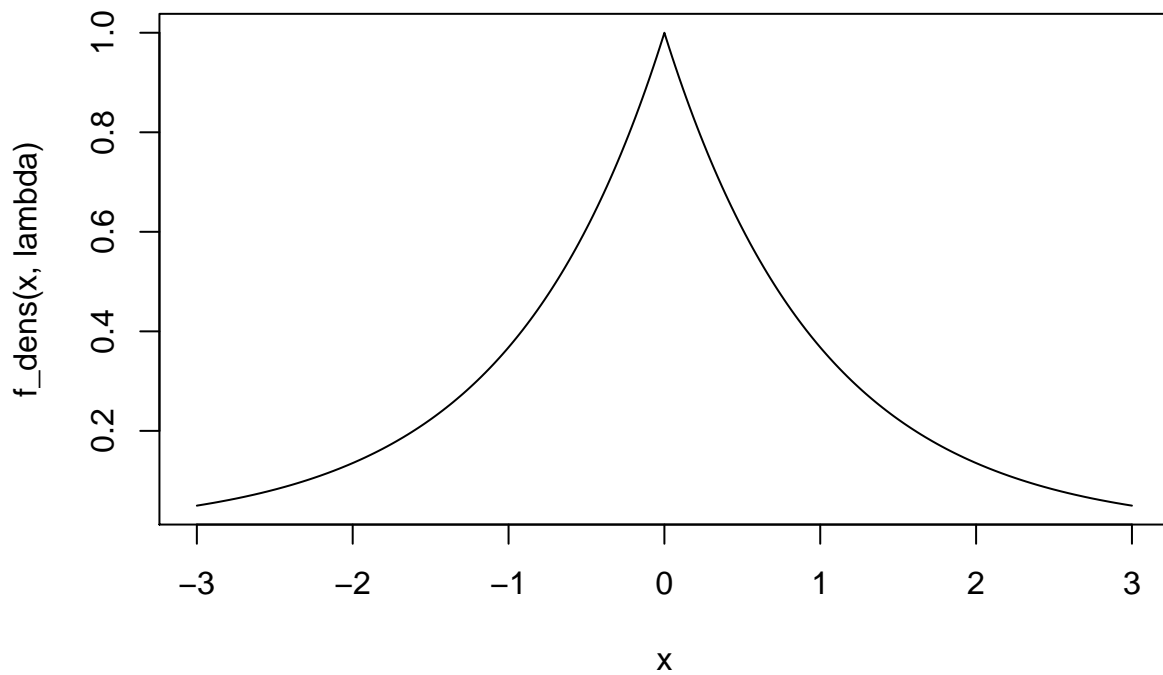
09/21/2020

Problem 1 - Inverse Transform

$$f(x) = K e^{-\lambda|x|}, x \in R, \lambda > 0$$

Part a

```
f_dens <- function(x, lambda) {  
  exp(-1 * lambda * abs(x))  
}  
  
lambda <- 1  
x <- seq(-3, 3, 0.01)  
  
plot(x, f_dens(x, lambda), "l")
```



The density is essentially an exponential density function that is mirrored over the y-axis. Logically this makes sense, since the exponent of e includes $|x|$.

Part b

As I said in the previous part the function $f(x)$ is essentially mirrored over the y-axis. This implies that the same area should lie under the curve on each side, so if we “chop it in half” along the y-axis, then we only have to evaluate the integral from 0 to ∞ (or $-\infty$ to 0) and double the result to get the value of $\int_{-\infty}^{\infty} f(x)dx$.

Part c

To get K we can evaluate the integral in the previous part and set it equal to 1. Note that we can ignore the absolute value function around x when only evaluating positive numbers.

$$2 \int_0^{\infty} f(x)dx = 2 \int_0^{\infty} K e^{-\lambda x} dx = \frac{-2K}{\lambda} (e^{-\lambda x})|_0^{\infty} = \frac{-2K}{\lambda} (0 - 1) = \frac{2K}{\lambda}$$

$$\frac{2K}{\lambda} = 1$$

$$2K = \lambda$$

$$K = \frac{\lambda}{2}$$

This value makes sense when you consider the regular exponential distribution pdf would have $K = \lambda$, so since we are going to have that distribution mirrored on both sides, we then only want half the area, so divide that value by two to get our actual value of $K = \frac{\lambda}{2}$.

Part d

To compute the CDF let's first only consider the case where $x \geq 0$. Since we know our function is symmetric over the y-axis we can say that any value greater than 0 is half the area (of 1) plus whatever is under the curve past 0:

$$F(a) = 0.5 + \int_0^a \frac{\lambda}{2} e^{-\lambda x} dx = 0.5 - \frac{1}{2} (e^{-\lambda x})|_0^a = 0.5 - 0.5(e^{-\lambda a} - 1) = 1 - \frac{e^{-\lambda a}}{2}$$

So for positive x we get $F(x) = 1 - \frac{e^{-\lambda x}}{2}$, but what about negative x values. Since the density is symmetric we know that $F(x) = 1 - F(-x)$, so we can say for $x < 0$ we get:

$$F(a) = 1 - (0.5 + \int_0^a \frac{\lambda}{2} e^{-\lambda x} dx) = 0.5 - \int_0^a \frac{\lambda}{2} e^{-\lambda x} dx$$

Basically we start at $\frac{1}{2}$ and if x is positive we find the area to add to the right, and if it is negative, then we find the area we should subtract to the left (which is the same value just + or -).

$$F(x) = 0.5 + \text{sign}(x) \left(0.5 - \frac{e^{-\lambda|x|}}{2} \right)$$

Part e

We can clearly see the inflection point at $y = \frac{1}{2}$ below, which will likely be helpful in part f.

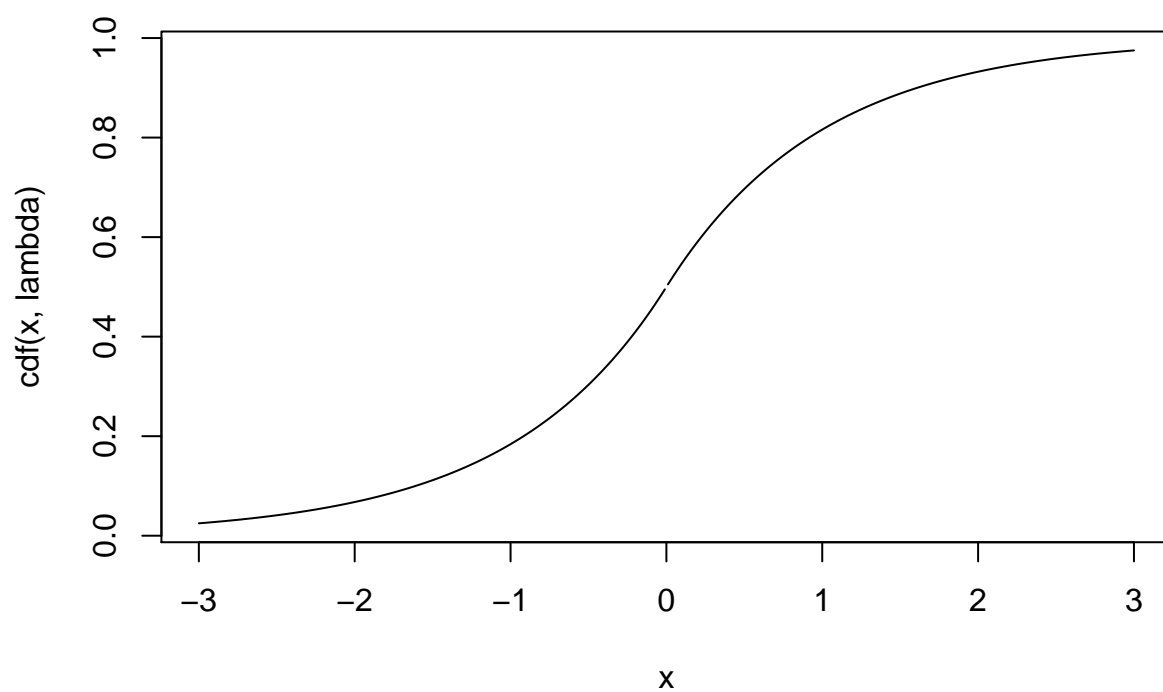
```

cdf <- function(x, lambda) {
  sign_x <- x / abs(x)
  0.5 + sign_x * (0.5 - 0.5 * exp(-1 * lambda * abs(x)))
}

lambda <- 1
x <- seq(-3, 3, 0.01)

plot(x, cdf(x, lambda), "l")

```



Part f

$$0.5 + \text{sign}(x)\left(0.5 - \frac{e^{-\lambda|x|}}{2}\right) = y$$

$$\text{sign}(x)\left(\frac{1}{2} - \frac{e^{-\lambda|x|}}{2}\right) = y - \frac{1}{2}$$

Now we know from the CDF, plotted above, that the sign of x is associated with values of y . Let's first look at the region $y < \frac{1}{2}$:

$$\frac{e^{-\lambda|x|}}{2} - \frac{1}{2} = y - \frac{1}{2}$$

$$\frac{e^{-\lambda|x|}}{2} = y$$

$$-\lambda|x| = \ln(2y)$$

$$|x| = \frac{\ln(2y)}{-\lambda}$$

Recall that is the region where x should be negative. Since $\frac{\ln(2y)}{-\lambda}$ is a positive value on the domain of $0 \leq y < \frac{1}{2}$ we must multiply by negative one and we get:

$$x = -1 * \frac{\ln(2y)}{-\lambda}$$

So when $y < \frac{1}{2}$ we get $F^{-1}(y) = -1 * \frac{\ln(2y)}{\lambda}$. How about $y \geq \frac{1}{2}$:

$$\frac{1}{2} - \frac{e^{-\lambda|x|}}{2} = y - \frac{1}{2}$$

$$-\frac{e^{-\lambda|x|}}{2} = y - 1$$

$$\frac{e^{-\lambda|x|}}{2} = 1 - y$$

$$-\lambda|x| = \ln(2(1 - y))$$

Again following the logic above, we know x must be positive, and for the domain of y , the function is already positive. This gives us.

$$x = \frac{\ln(2(1 - y))}{-\lambda}$$

$$F^{-1}(y) = \frac{\ln(2(1 - y))}{-\lambda}$$

Part g

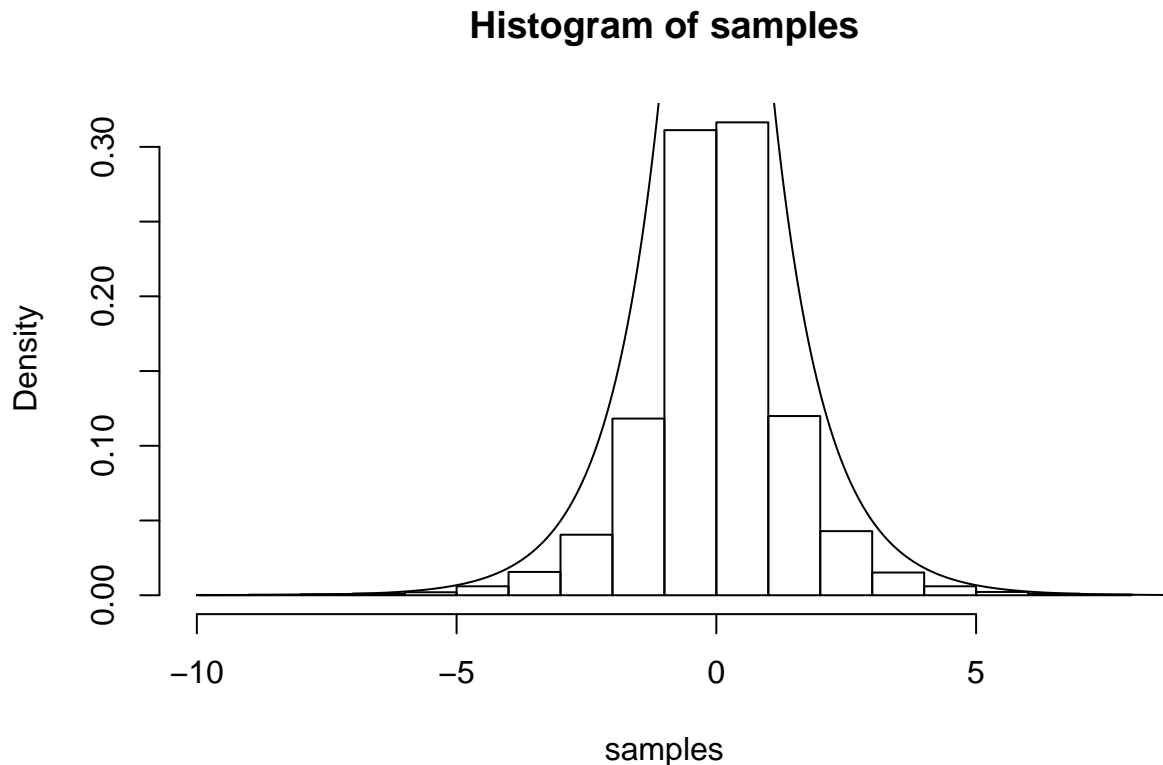
```
n <- 10000

cdf_inverse <- function(y, lambda){
  y_val <- y
  y[y >= 0.5] <- 1 - y[y >= 0.5]
  out <- log(2 * y) / (-1 * lambda)
  out[y_val < 0.5] <- -1 * out[y_val < 0.5]
  out
}

lambda <- 1
samples <- runif(n)
samples <- cdf_inverse(samples, lambda)

hist(samples, freq=FALSE)

x <- seq(-10, 10, 0.01)
lines(x, f_dens(x, lambda))
```



Problem 2 - Rejection Sampling

Here we will discuss rejection sampling to draw samples from the density from the previous question, which is a mirrored exponential distribution.

Part a

Say we consider a uniform distribution as our candidate $g(x) = 1, x \in (0, 1)$. This would not be a desirable candidate density purely for the fact that you can only sample from the values of x between 0 and 1. While you could modify this slightly to include $(-1, 1)$, there is no way to sample from the entire support of $f(x)$ with the uniform distribution. It should also be noted that the scaling factor c here is equal to the exponential parameter λ divided by two, so for distributions with large λ this candidate density does not scale well (grows linearly).

Part b

Using an exponential distribution $g(x) = \lambda e^{-\lambda x}, x \in [0, \infty)$ we would have a fantastic candidate density. Since this matches the density $f(x)$, besides being 2 times larger, you can actually adjust your sampling technique to not reject any. What you could do is sample from this candidate density, and if you were supposed to reject, you can instead accept for a value of $-x$. This works, because the function is exactly twice as large, so the region above x in the candidate density is actually the region above both x and $-x$ in the density $f(x)$.

Part c

A standard normal distribution would not be a good density to use to sample from this exponential function. Since the tails of the standard normal decay exponentially faster, and the supports are all real numbers, there is no scaling factor c that would enable the tails to be large enough to cover the tails of $f(x)$.

Part d

A cauchy distribution centered around 0 with rate 1 would be an okay candidate density function.

$$\begin{aligned}h(x) &= \frac{f(x)}{g(x)} = \lambda/2e^{-\lambda x}(\pi(1+x^2)) \\h'(x) &= \pi\lambda xe^{-\lambda x} - \frac{\pi}{2}\lambda^2 e^{-\lambda x}(1+x^2) \\ \frac{\pi}{2}\lambda^2 e^{-\lambda x}(1+x^2) &= \pi\lambda xe^{-\lambda x} \\ \frac{1}{2}\lambda(1+x^2) &= x \\ 0 &= \frac{x^2\lambda}{2} - x + \frac{\lambda}{2} \\ x &= \frac{-(-1) \pm \sqrt{1-4(\lambda/2)^2}}{\lambda} = \frac{1 \pm \sqrt{1-\lambda^2}}{\lambda}\end{aligned}$$

Since $\lambda > 0$ the x for the max value of c here is imaginary, so once again we follow the same logic from part c. There is no scaling factor large enough to enable rejection sampling, because the tail decay of the cauchy distribution is too fast compared to that of the exponential distribution.

Problem 3 - Importance Sampling

Part a

```
H <- function(z, beta, FUN=function(x) 1) {
  vals <- dnorm(z) / ((1 - pnorm(beta)) * FUN(z))
  vals[z < beta] <- 0
  vals
}
```

Part b

Recall that in importance sampling we are sampling from our candidate and finding the expected value of $\frac{h(x)f(x)}{g(x)}$ where x is the value sampled from $g(x)$, $h(x) = x$, and $f(x) = f_{Z|Z>\beta}(x)$.

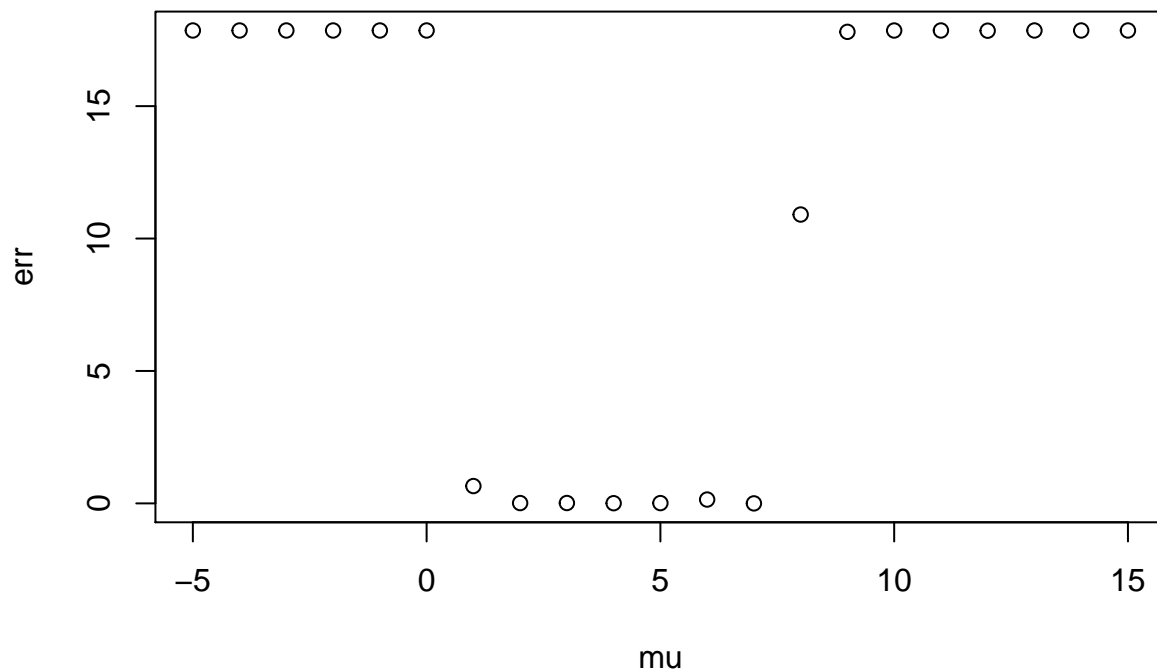
```
sample_H <- function(n, beta, mu, sigma=1) {
  # sample from g(x)
  candidate_samples <- rnorm(n, mu, sigma)
  # get h(x) * f(x) / g(x)
  transform <- candidate_samples * H(candidate_samples, beta, function(x) dnorm(x, mu, sigma))
  # estimate theta
  return(sum(transform) / n)
}
```

Part c

```
beta <- 4
n <- 10000
mu <- c(-5:15)
err <- vector("numeric", length(mu))
theta_actual <- exp(-1 * beta^2 / 2) / (sqrt(2 * pi)*(1-pnorm(beta)))

for (i in 1:length(mu)) {
  theta_est <- sample_H(n, beta, mu[i])
  err[i] <- (theta_est - theta_actual)^2
}

plot(mu, err)
```



Part d

Means (μ) of 2 to 7 have the best estimates when $\beta = 4$. This means that drawing samples from a normal distribution whose center was near our value β had the best estimates. This happened because these candidate densities had the most samples near/just over the value of β . Since our pdf starts and decays past β these samples drawn (just near β) were the most “important” or gained the highest weight compared to samples drawn that were under β or far too large. Samples that were less than β actually would gain zero weight and those that were too large would have a very small weight due to the decaying tail.