

Module 4 Peer Review Assignment

Problem 1

A continuous random variable with cumulative distribution function F has the median value m such that $F(m) = 0.5$. That is, a random variable is just as likely to be larger than its median as it is to be smaller. A continuous random variable with density f has the mode value x for which $f(x)$ attains its maximum. For each of the following three random variables,

(i) state the density function,

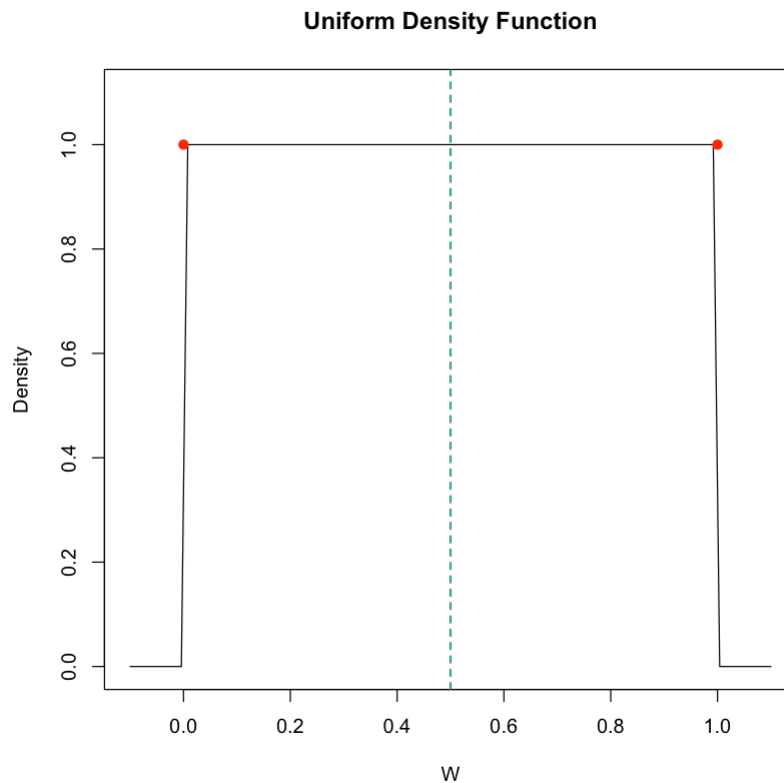
(ii) compute the median, mode and mean for the random variable, and

(iii) Provide at least one graph for the density function using values of the parameter(s) that you select. Indicate the median, mode, and mean values on your graph. (The purpose of this problem is to see the relative locations of the median, mode, and mean for the different random variables).

a) W which is uniformly distributed over the interval $[a, b]$, for some value $a, b \in \mathbb{R}$.

```
In [1]: a <- 0
b <- 1

curve(dunif(x, min = a, max = b), from = a-0.1, to = b+0.1, ylim = c(0, 1),
      ylab = "Density", xlab = "W", main = "Uniform Density Function")
abline(v = (a + b)/2, col = "blue", lty = 2)
points(c(a, b), dunif(c(a, b), min = a, max = b), col = "red", pch = 1)
abline(v = (a + b)/2, col = "green", lty = 2)
```



(i) The density function of W is $f_W(w) = 1/(b - a)$ for $a \leq w \leq b$ and $f_W(w) = 0$ otherwise

(ii)

- the mean is $\frac{1}{2}(a + b)$
- the median is $\frac{1}{2}(a + b)$
- mode is $w \in (a, b)$

b) X which is normal with parameters μ and σ^2 , for some value $\mu, \sigma^2 \in \mathbb{R}$.

```

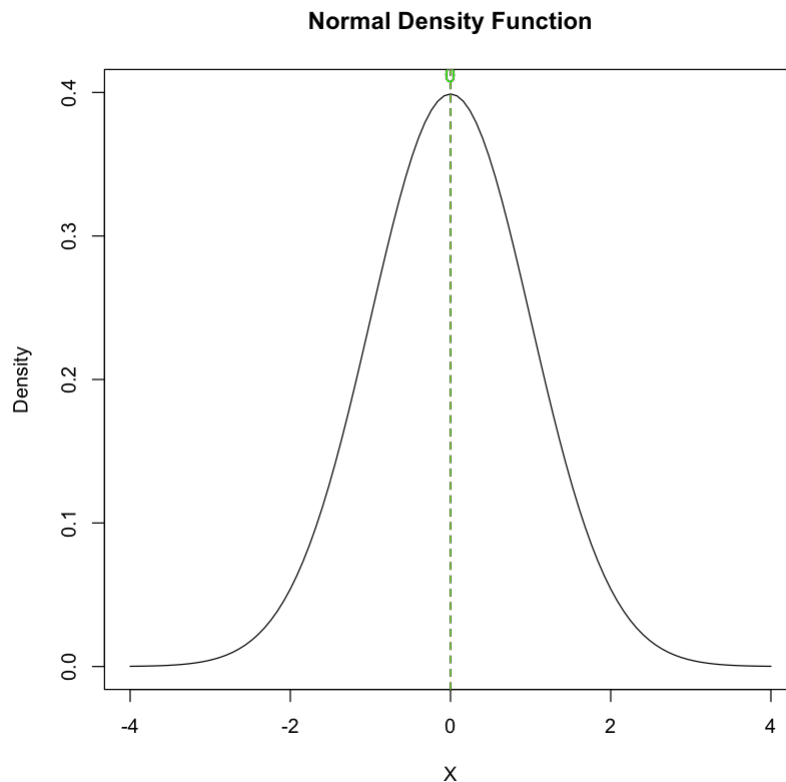
In [2]: mu <- 0
        sigma <- 1

        f <- function(x) {
          1/(sigma * sqrt(2*pi)) * exp(-(x-mu)^2/(2*sigma^2))
        }

        median <- mu
        mode <- mu
        mean <- mu

        plot(
          f, xlim = c(mu-4*sigma, mu+4*sigma), ylim = c(0, 0.4),
          ylab = "Density", xlab = "X", main = "Normal Density Function"
        )
        abline(v = median, col = "blue", lty = 2)
        abline(v = mode, col = "red", lty = 2)
        abline(v = mean, col = "green", lty = 2)
        text(median, f(median), round(median, 2), pos = 3, col = "blue")
        text(mode, f(mode), round(mode, 2), pos = 3, col = "red")
        text(mean, f(mean), round(mean, 2), pos = 3, col = "green")

```



(i) ?

(ii)

- mean μ
- median μ

- mode μ

c) Y which is exponential with rate $\lambda \in \mathbb{R}$.

```
In [5]: library(ggplot2)

# Set rate parameter
lambda <- 0.5

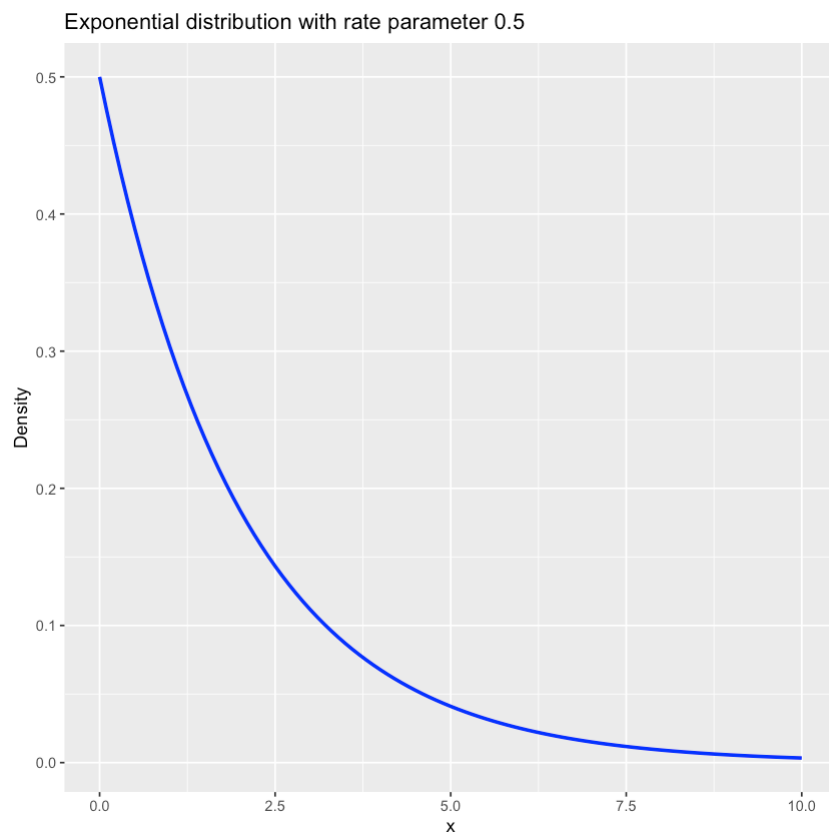
# Create function for exponential density
f <- function(x) lambda * exp(-lambda * x)

# Create x values
x <- seq(0, 10, by = 0.1)

# Create y values
y <- f(x)

# Create data frame
df <- data.frame(x, y)

# Create plot
ggplot(df, aes(x = x, y = y)) +
  geom_line(color = "blue", size = 1) +
  ggtitle(paste("Exponential distribution with rate parameter", lambda)) +
  xlab("x") + ylab("Density")
```



(i) $f_Y(y) = \lambda e^{-\lambda y}$ for $y \geq 0$ and $f_Y(y) = 0$ for $y < 0$

(ii)

- mean $1/\lambda$
- median $(\ln 2)/\lambda$
- mode λ

Problem 2

For this problem, we're going to visualize what's happening when we go between different normal distributions.

Part A)

Draw at least 10000 samples from the standard normal distribution $N(0, 1)$ and store the results. Make a density histogram of these samples. Set the x -limits for your plot to $[-10, 10]$ and your y -limits to $[0, 0.5]$ so we can compare with the plots we'll generate in **Parts B-D**.

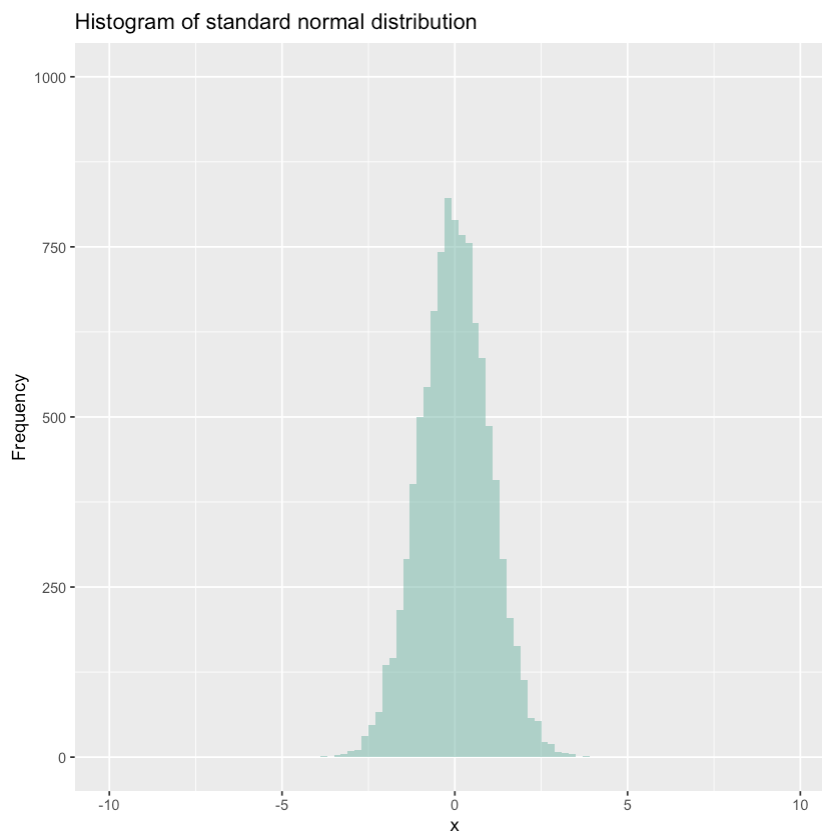
```
In [7]: library(ggplot2)

# Set seed for reproducibility
set.seed(123)

# Generate 10000 samples from  $N(0,1)$ 
samples <- rnorm(10000)

# Create histogram plot
ggplot(data.frame(samples), aes(x = samples)) +
  geom_histogram(binwidth = 0.2, fill = "#69b3a2", alpha = 0.5) +
  ggtitle("Histogram of standard normal distribution") +
  xlim(-10, 10) + ylim(0, 1000) +
  xlab("x") + ylab("Frequency")
```

Warning message:
 “Removed 2 rows containing missing values (`geom_bar()`).”



Part b) Now generate 10000 samples from a $N(2, 3)$ distribution and plot a histogram of the results, with the same x -limits and y -limits. Does the histogram make sense based on the changes to parameters?

Note: Be careful with the parameters for `rnorm`. It may help to check the documentation.

```
In [8]: library(ggplot2)

# Set seed for reproducibility
set.seed(123)

# Generate 10000 samples from N(2,3)
samples <- rnorm(10000, mean = 2, sd = 3)

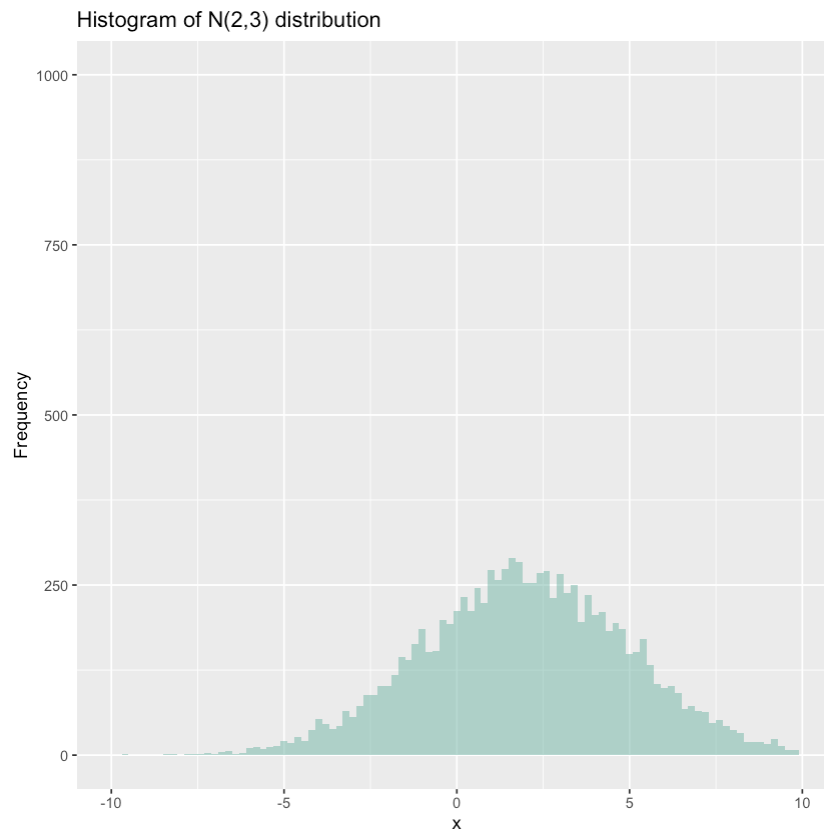
# Create histogram plot
ggplot(data.frame(samples), aes(x = samples)) +
  geom_histogram(binwidth = 0.2, fill = "#69b3a2", alpha = 0.5) +
  ggtitle("Histogram of N(2,3) distribution") +
  xlim(-10, 10) + ylim(0, 1000) +
  xlab("x") + ylab("Frequency")
```

Warning message:

"Removed 42 rows containing non-finite values (`stat_bin()`)."

Warning message:

"Removed 2 rows containing missing values (`geom_bar()`)."



The histogram makes sense based on the changes to the parameters. The mean of the distribution has shifted to 2, so the histogram will be shifted to the right compared to the standard normal distribution. The standard deviation of the distribution has increased to 3, therefore the histogram will be wider compared to the standard normal distribution. The shape of the histogram will still resemble a bell curve.

Part c)

Suppose we are only able to sample from the standard normal distribution $N(0, 1)$. Could we take those samples and perform a simple transformation so that they're samples from $N(2, 3)$? Try this, and plot another histogram of the transformed data, again with the same axes. Does your histogram based of the transformed data look like the histogram from **Part B**?

```
In [9]: library(ggplot2)

# Set seed for reproducibility
set.seed(123)

# Generate 10000 samples from N(0,1)
samples_std <- rnorm(10000)

# Transform the samples to N(2,3)
samples <- 2 + 3 * samples_std

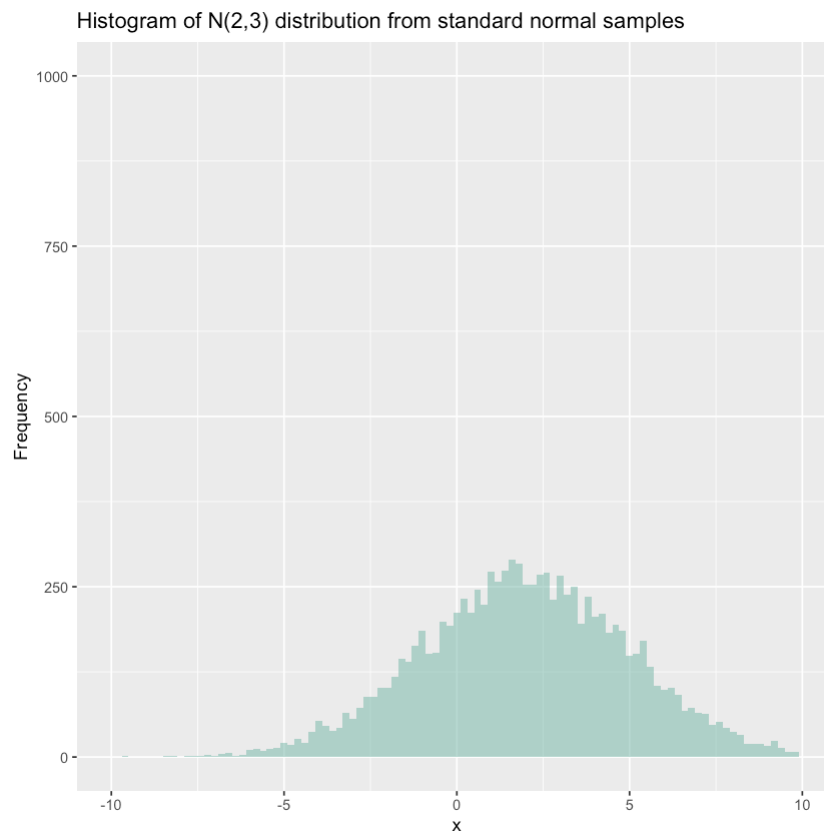
# Create histogram plot
ggplot(data.frame(samples), aes(x = samples)) +
  geom_histogram(binwidth = 0.2, fill = "#69b3a2", alpha = 0.5) +
  ggtitle("Histogram of N(2,3) distribution from standard normal sampl") +
  xlim(-10, 10) + ylim(0, 1000) +
  xlab("x") + ylab("Frequency")
```

Warning message:

“Removed 42 rows containing non-finite values (`stat_bin()`).”

Warning message:

“Removed 2 rows containing missing values (`geom_bar()`).”



Yes, we can use a simple transformation to convert samples from the standard normal distribution $N(0, 1)$ to samples from the $N(2, 3)$ distribution. Specifically, if we take a sample x from $N(0, 1)$, then the transformed sample $y = 2 + 3x$ will be a sample from $N(2, 3)$.

Part d)

But can you go back the other way? Take the $N(2, 3)$ samples from **Part B** and transform them into samples from $N(0, 1)$? Try a few transformations and make a density histogram of your transformed data. Does it look like the plot of $N(0, 1)$ data from **Part A**?

```

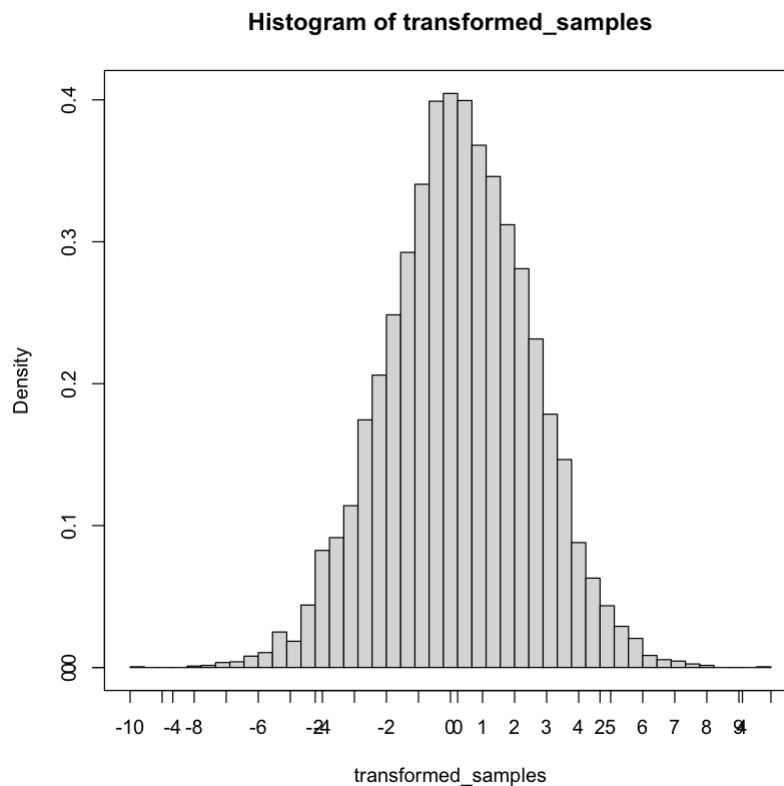
In [2]: # Generate 10000 samples from a normal distribution with mean 3 and st
samples <- rnorm(10000, mean = 3, sd = 4)

# Transform the samples to samples from the standard normal distributi
transformed_samples <- (samples - 3) / 4

# Create a histogram of the transformed samples
hist(transformed_samples, breaks = 50, freq = FALSE)

# Set x-limits and y-limits
xlim <- c(-10, 10)
ylim <- c(0, 0.5)
plot.window(xlim = xlim, ylim = ylim)
axis(1, xlim[1]:xlim[2])
axis(2, ylim[1]:ylim[2])
box()

```



Yes, transforming samples from the $N(3,4)$ distribution to samples from the standard normal distribution $N(0,1)$

In []:

