P_M4_1

October 26, 2021

1 Module 4 Peer Review Assignment

2 Problem 1

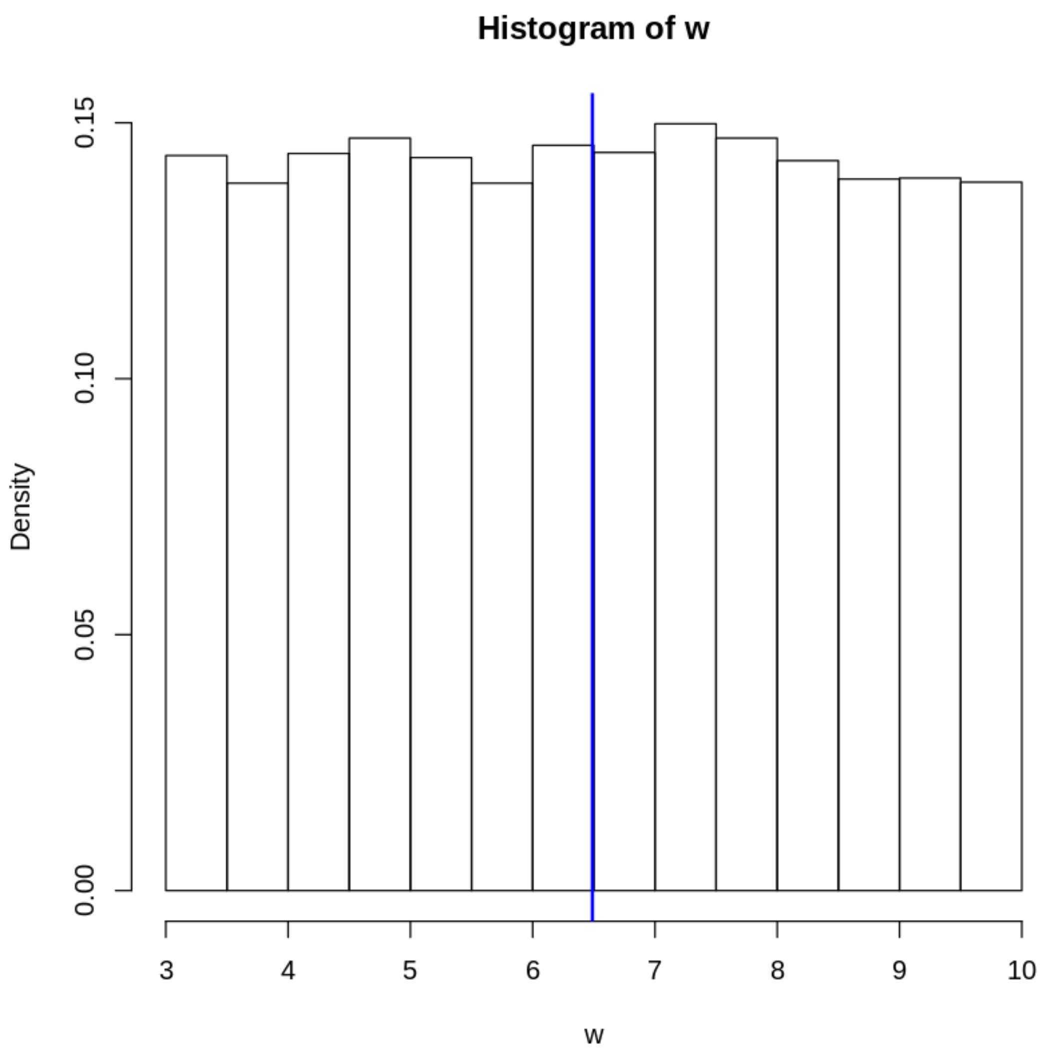
A continuous random variable with cumulative distribution function F has the median value m such that $F(m) = 0.5$. That is, a random variable is just as likely to be larger than its median as it is to be smaller. A continuous random variable with density f has the mode value x for which $f(x)$ attains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

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

[63]:
```
Mode <- function(u) {
  u_vals <- unique(u)
  u_vals[which.max(tabulate(match(u, u_vals)))]
}
```

[64]:
```
# Your Code Here
set.seed(13544)
w = runif(10000, 3, 10)
median = round(median(w), 4)
mean = round(mean(w), 4)
print(paste0("The median of the distribution, represented by the blue line, is:␣
↪", median))
print(paste0("The mean of the distribution, also represented by the blue line,␣
↪is: ", mean))
print('The mode for the uniform distribution is all values in the range [a,b].␣
↪In this case, it is all values within [3,10]')
hist(w, prob=TRUE)
abline(v=mean, col="blue", lwd="2")
```

```
[1] "The median of the distribution, represented by the blue line, is: 6.5021"
[1] "The mean of the distribution, also represented by the blue line, is:
6.4883"
[1] "The mode for the uniform distribution is all values in the range [ab]. In
this case, it is all values within [3,10]"
```



i) f(w) $= \frac{1}{a-b}, a \leq w \leq b$

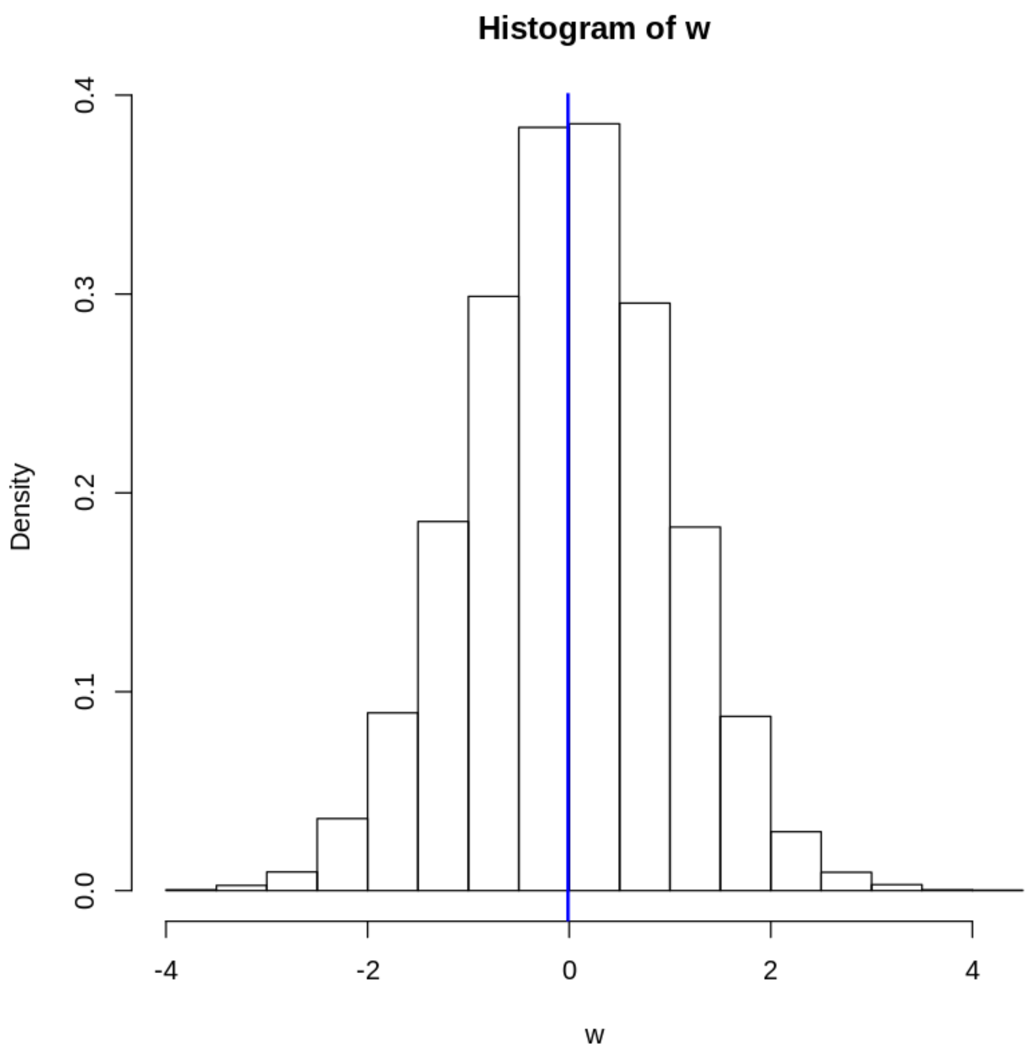
ii) Please see the print statements above. The median and mode are the same since a uniform distribution will have equal density for all w. There is no skewness in the graph. Also, all values within [a,b], in this case [3,10], are the mode since all values have equal density. The mode was not plotted as it would cover the whole plot. The blue line represents both the median and the mean.

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

[65]:
```
# Your Code Here
set.seed(13545)
w = rnorm(10000, 0, 1)
median = round(median(w), 4)
```

```
mean = round(mean(w), 4)
mode = round(Mode(w), 4)
print(paste0("The median of the distribution, represented byt the blue line, is:
 ↪ ", median))
print(paste0("The mean of the distribution, also represented by the blue line,␣
 ↪is: ", mean))
print(paste0("The mode of the distribution, also represented by the blue line,␣
 ↪is: ", mode))
hist(w, prob=TRUE)
abline(v=mean, col="blue", lwd="2")
```

```
[1] "The median of the distribution, represented byt the blue line, is: -0.0075"
[1] "The mean of the distribution, also represented by the blue line, is:
-0.0116"
[1] "The mode of the distribution, also represented by the blue line, is:
0.7085"
```



i) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

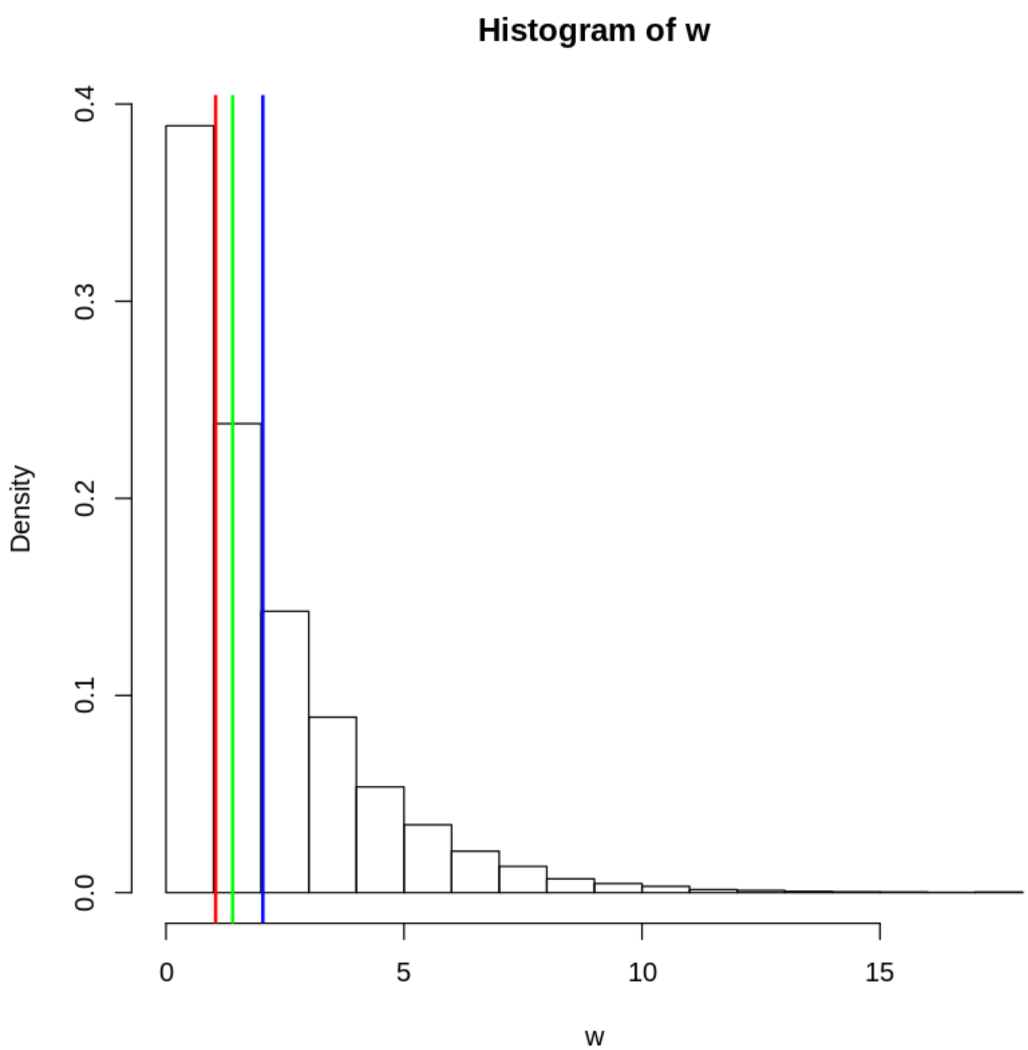
ii) The mode, median and mean are all the same in this distribution. The highest peak is around 0 (representing a mode of 0, an equal amount of density is to the left and right of 0 (therefore stating the median is 0). The blue line represents the mean, median and mode.

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

[66]:
```
set.seed(13545)
lambda = 5
w = rexp(10000, 0.5)
median = round(median(w), 4)
mean = round(mean(w), 4)
mode = round(Mode(w), 4)
```

```
print(paste0("The median of the distribution, represented byt the green line,␣
 ↪is: ", median))
print(paste0("The mean of the distribution, also represented by the blue line,␣
 ↪is: ", mean))
print(paste0("The mode of the distribution, also represented by the red line,␣
 ↪is: ", mode))
hist(w, prob=TRUE)
abline(v=mean, col="blue", lwd="2")
abline(v=mode, col="red", lwd="2")
abline(v=median, col="green", lwd="2")
```

```
[1] "The median of the distribution, represented byt the green line, is: 1.3998"
[1] "The mean of the distribution, also represented by the blue line, is:
2.0347"
[1] "The mode of the distribution, also represented by the red line, is: 1.0428"
```



i) $f(x) = \lambda e^{-\lambda x}$

ii) The mean, median and mode are stated above. the mean is represented by the blue line, the mode is represented by the red and the green line represents the mode. It makes sense the mean is around 2 since we know the mean of an exponential distribution is 1/lambda and lambda is 0.5 in this case.

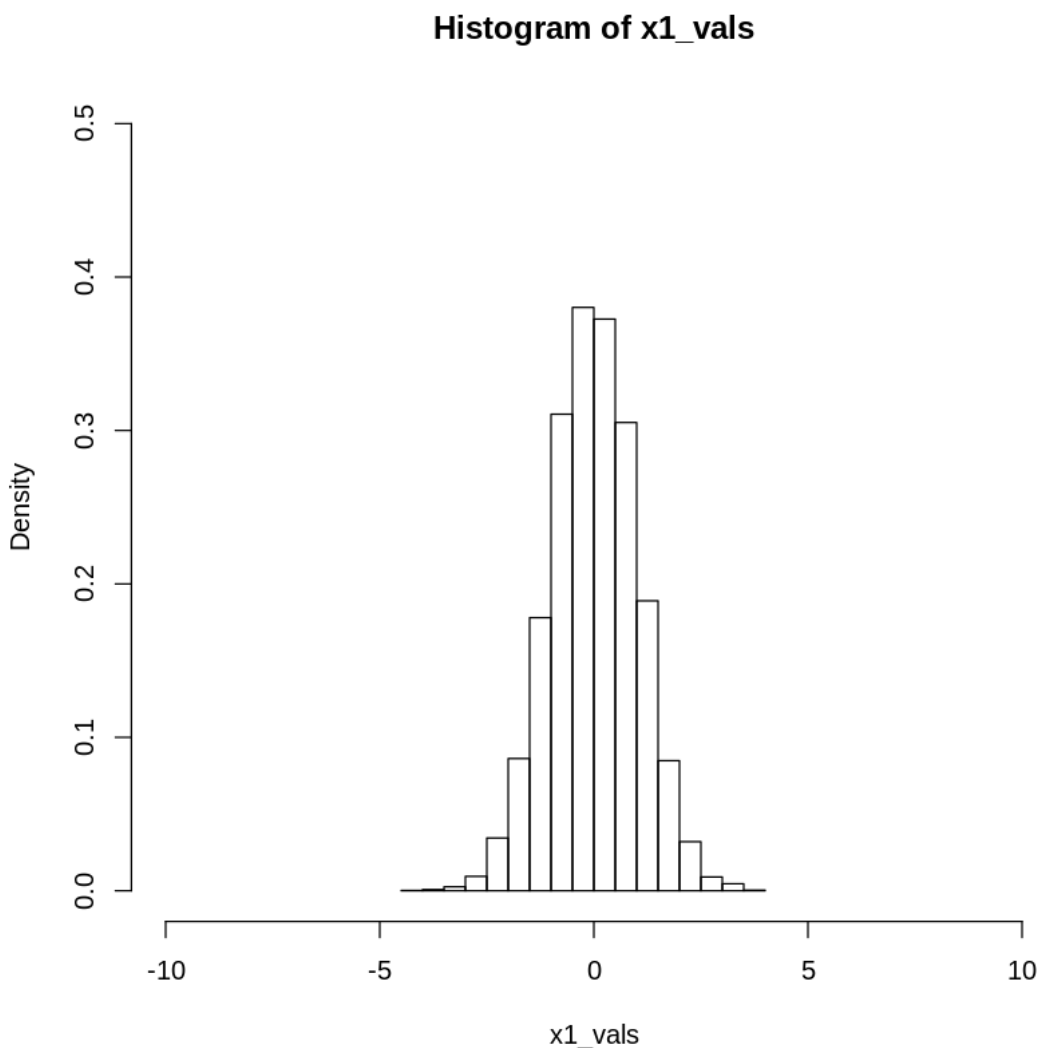
3 Problem 2

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**.

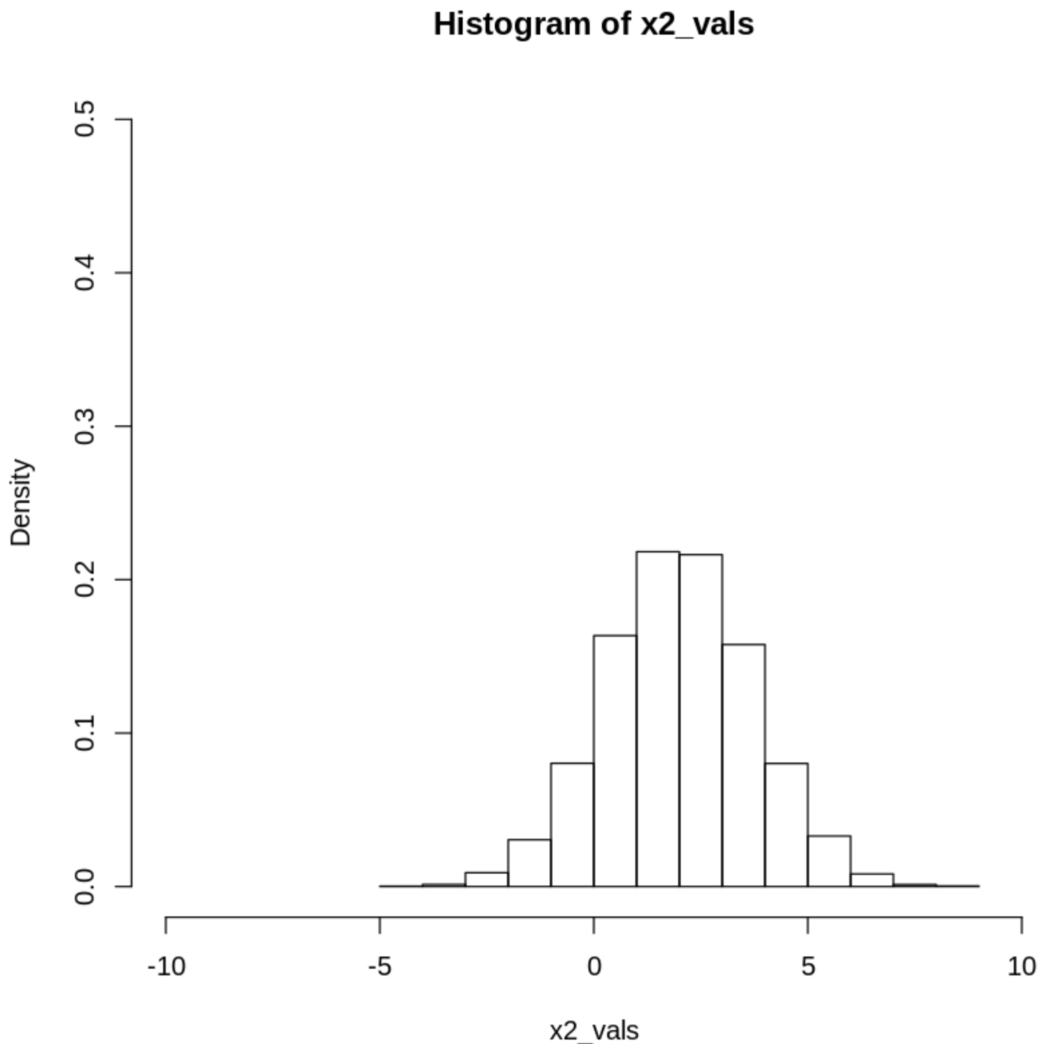
[67]:
```
x1_vals = rnorm(10000, 0, 1)
hist(x1_vals, prob=TRUE, xlim=c(-10,10), ylim=c(0,0.5))
```



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.

[68]:
```
set.seed(23422)
x2_vals = rnorm(10000, 2, sqrt(3))
hist(x2_vals, prob=TRUE, xlim=c(-10,10), ylim=c(0,0.5))
```

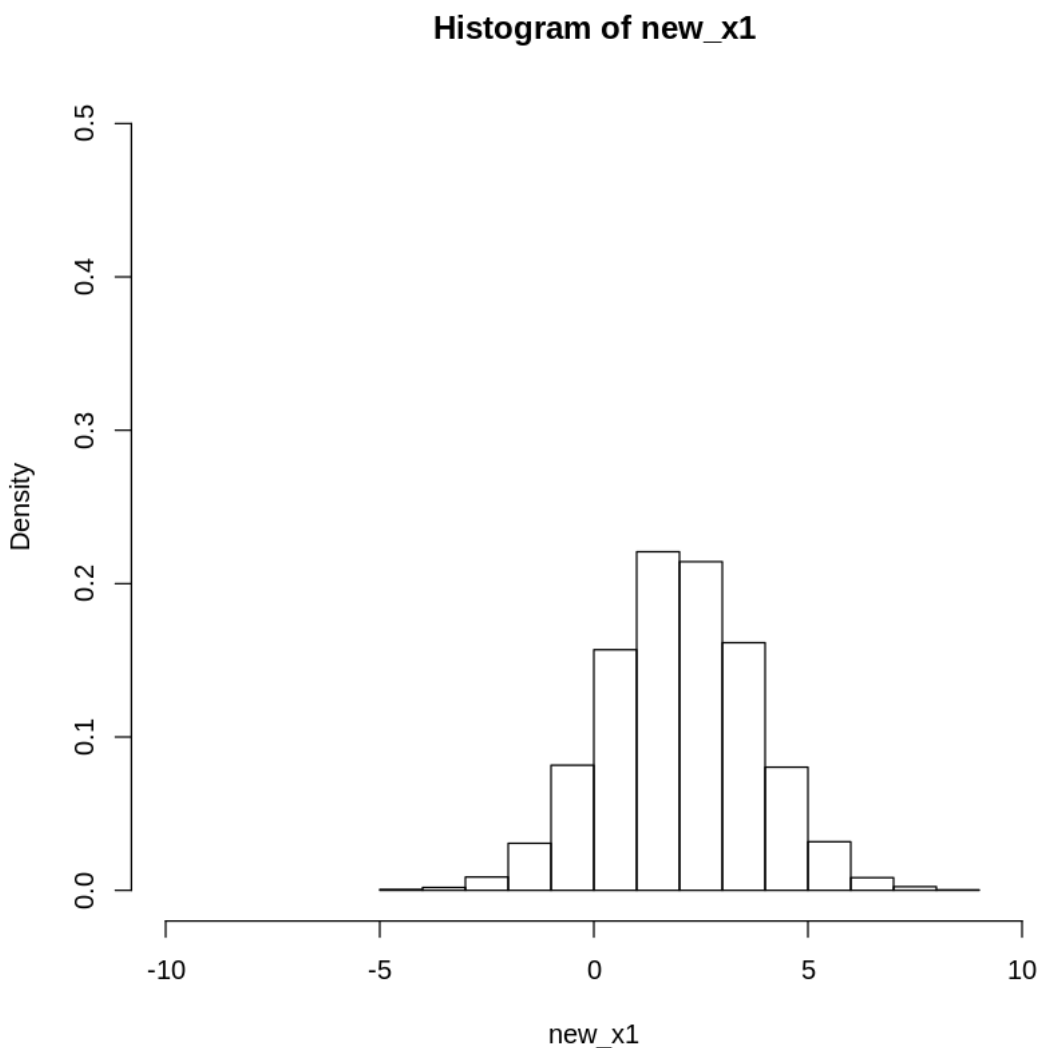


Yes, this makes sense. First, there is no skewness in the histogram (symmetrical); this is due to the distribution staying the same with many trials. Also, the mean and median seem to be around 3. We can tell because most of the density lies between 0 and 4 (via the peak of the histogram). Compared to the previous histogram, this is due to a shift from 0 to 3 of the mean. Lastly, the density of histogram is more spread out due to the tranformation from 1 to 3.

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**?

[69]:
```
new_x1 =  sqrt(3) * x1_vals + 2
hist(new_x1, prob=TRUE, xlim=c(-10,10), ylim=c(0,0.5))
```



We can transform the distribution to $N(2,3)$ using by calculating $\sqrt{3}X+2$ where $X \sim N(0,1)$. This wil shift the the graph to the right 2 values and increase the spread of the density plot so it matches the normal distribution. We can show this with the follwing:
Letting $X \sim N(2,3)$

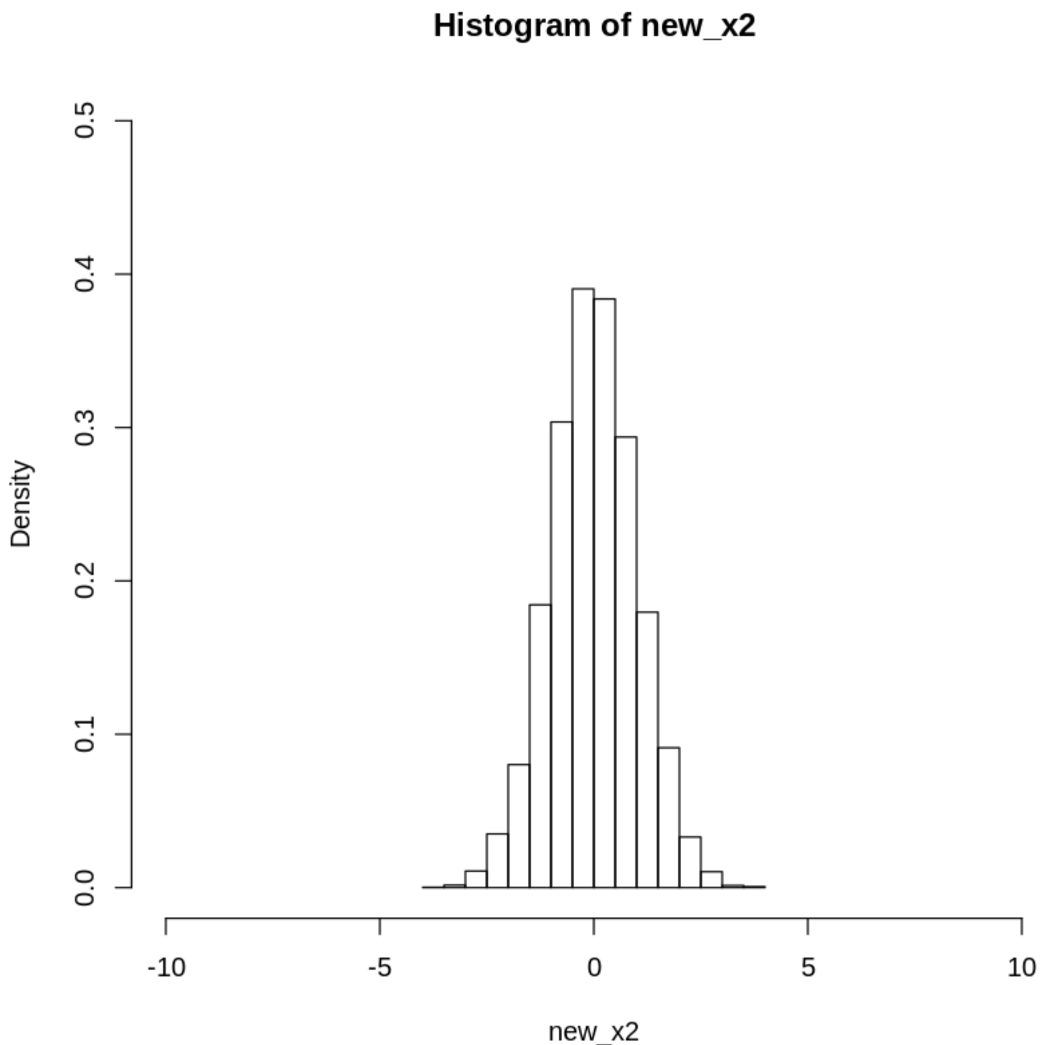
$$\frac{x-2}{\sqrt{3}} = N(0,1) \Rightarrow x-2 = \sqrt{3}N(0,1) \Rightarrow x = \sqrt{3}N(0,1)+2 \Rightarrow N(2,3) = \sqrt{3}\cdot x+2$$

Where $X \sim N(0,1)$. As we can see, the graphs are very similar.

Part d)

But can you go back the other way? Take the $N(3,4)$ 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**?

[70]:
```
# Your Code Here
new_x2 =  (x2_vals - 2)/sqrt(3)
hist(new_x2, prob=TRUE, xlim=c(-10,10), ylim=c(0,0.5))
```

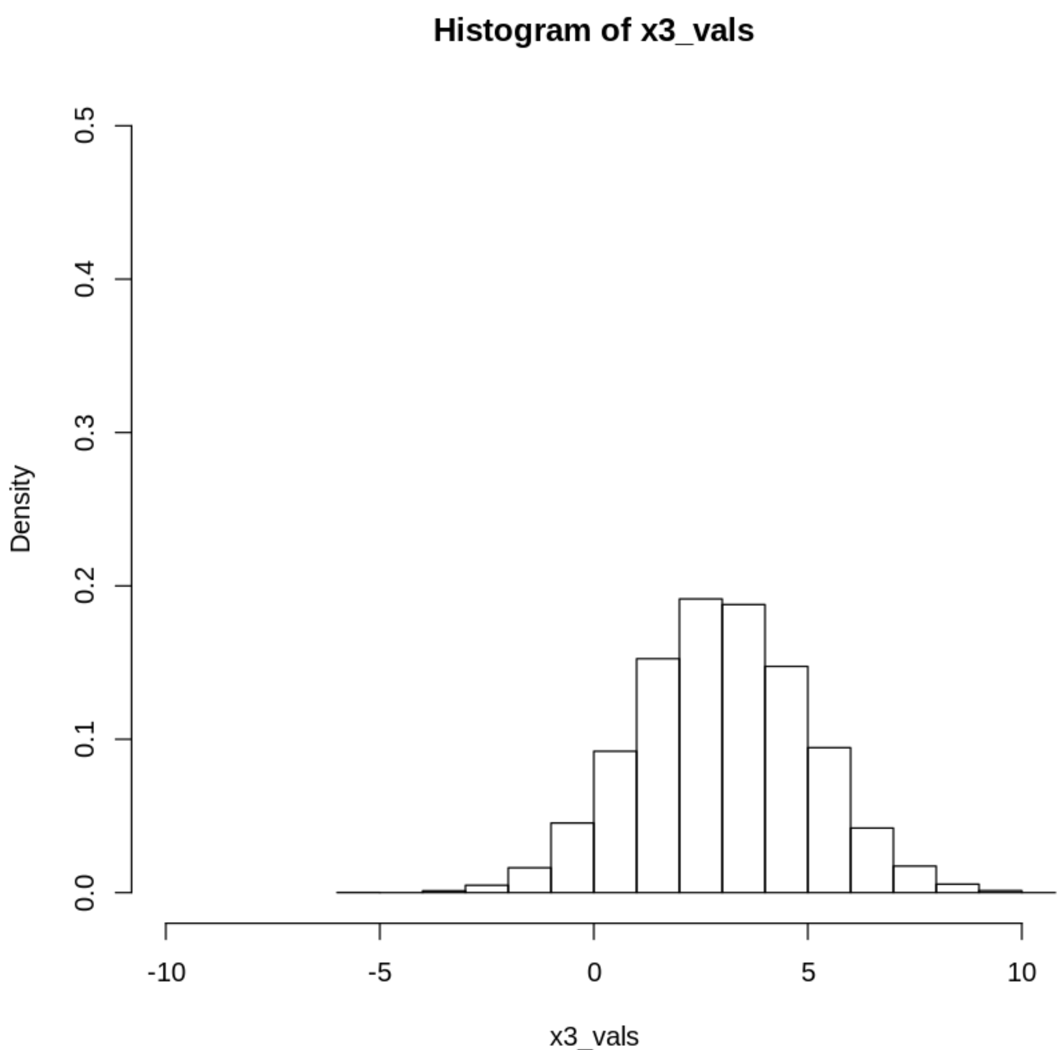


First, the dsitribution in part B was N(2,3), not N(3,4). and we can transform the values to standard normal by using the standard normal scaler. for this distribution we get :

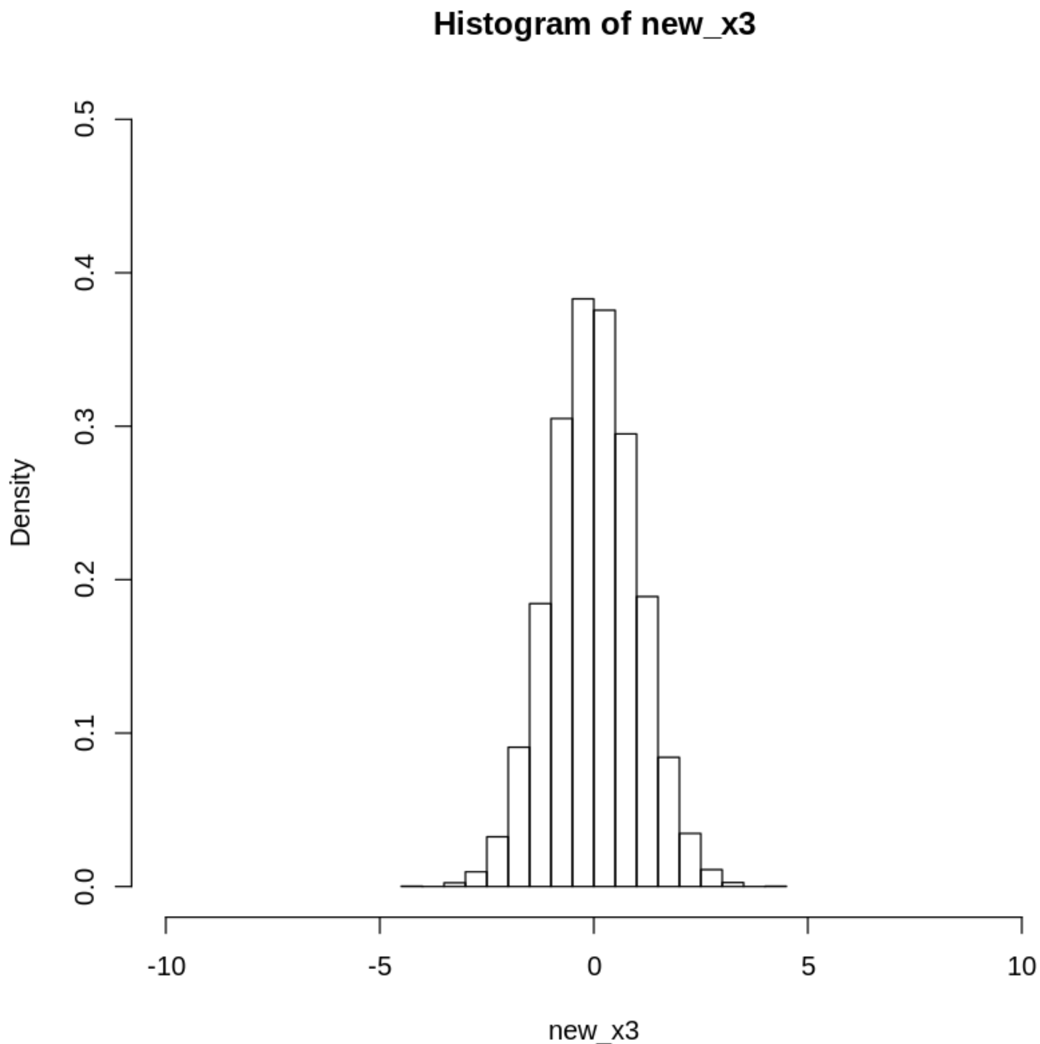
$$\frac{x-2}{\sqrt{3}}$$

Where $X \sim N(2,3)$. As we can see, the desnity plot in part d is similar to part b. Not sure if asking for the $X \sim N(3, 4)$ distribution is an error. Will plot below since the problem statement asked for it.

[71]:
```
set.seed(23567)
x3_vals = rnorm(10000, 3, sqrt(4))
hist(x3_vals, prob=TRUE, xlim=c(-10,10), ylim=c(0,0.5))
```



[72]:
```
new_x3 = (x3_vals - 3)/ sqrt(4)
hist(new_x3, prob=TRUE, xlim=c(-10,10), ylim=c(0,0.5))
```



Again, this distribution looks just like the normal distribution density plot in part b.

[ ]:

```

```