Probability — Homework (4).

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 $a, b \in \mathbb{R}$.

Answer:

To begin with, we remember some things related with a random variable W which is uniformly distributed over the interval $[a,b]$:

(i) The probability density function $f: \mathbb{R} \to \mathbb{R}$ of the random variable W is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b], \\ 0 & \text{if } x \notin [a,b] \end{cases}$$

(r) The random variable W with density $f:\mathbb{R}\to\mathbb{R}$ has a mode in every point of the $[a,b]$ interval, because in this point f reaches its maximum value $\frac{1}{b-a}$.

(r) The cumulative distribution function $F:\mathbb{R}\to\mathbb{R}$ of the random variable W is:

$$F(x)=p(W\leq x)=\int_{-\infty}^{x} f(t)\,dt=\begin{cases} 0 & \text{if } x<a, \\ \frac{x-a}{b-a} & \text{if } a\leq x\leq b \\ 1 & \text{if } x>b \end{cases}$$

Therefore the median value m of the random variable W satisfaces that $F(m)=\frac{1}{2}$, where from:

$$F(m)=\frac{1}{2} \longleftrightarrow \frac{m-a}{b-a}=\frac{1}{2} \longleftrightarrow m=\frac{b-a}{2}+a$$

$$m=\frac{b+a}{2}$$

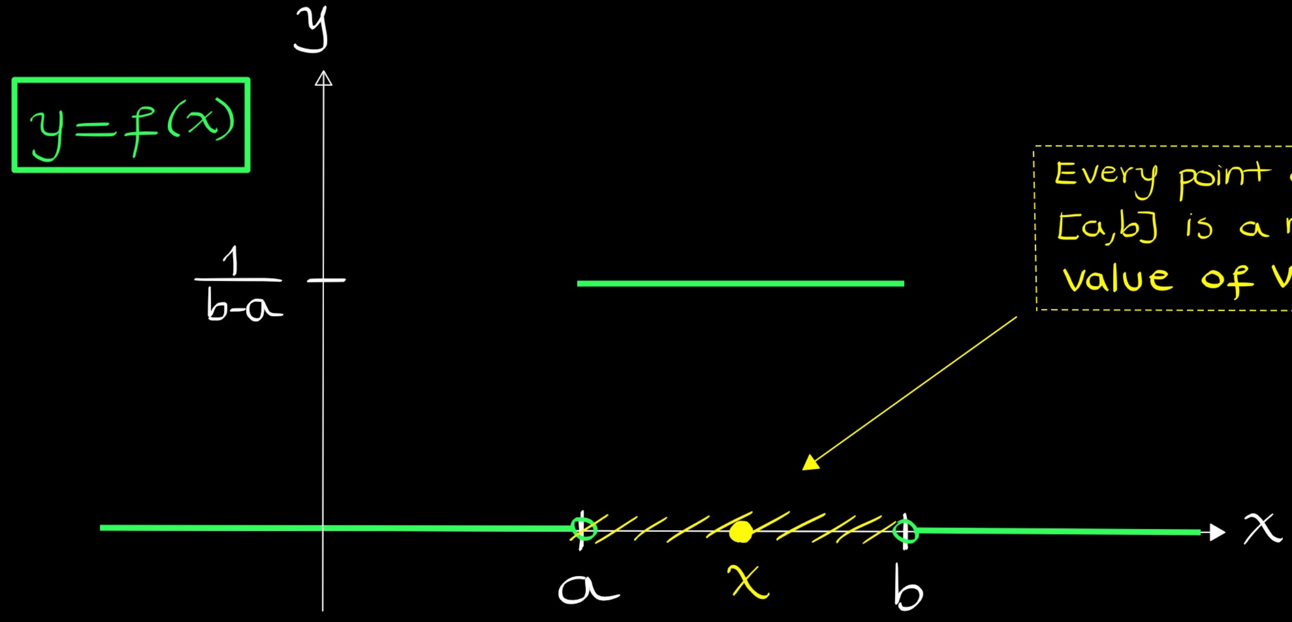
Thus, $m = \frac{b+a}{2}$ is the median value of W.

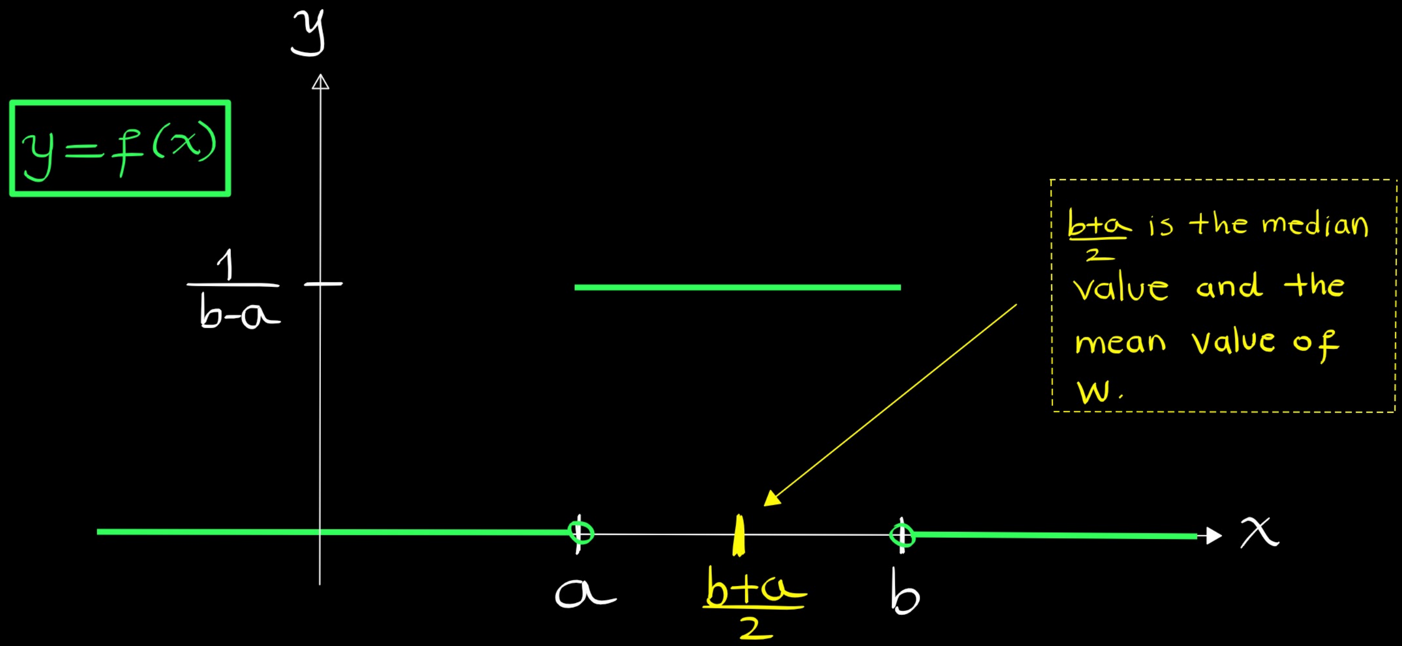
(v) The mean value $\mu = E(W)$ is

$$\mu = E(W) = \int_{-\infty}^{+\infty} t \cdot f(t)\,dt = \int_a^b \frac{t}{b-a}\,dt =$$

$$= \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \frac{b+a}{2}.$$

(v) The graph of the probability density function $f: \mathbb{R} \to \mathbb{R}$ is given below.





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

Answer:

If X is a random variable such that $X \sim \text{Nor}(\mu, \sigma^2)$, then:

(r) The probability density function $f: \mathbb{R} \to \mathbb{R}$ of X is given by:

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

for $x \in \mathbb{R}$.

(r) The mode value of X is $x = \mu$, because $f: \mathbb{R} \to \mathbb{R}$ reaches its maximum value in this point.

$$\boxed{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \leqslant 0} \longleftrightarrow \boxed{\frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}} \leqslant \boxed{\frac{1}{\sqrt{2\pi}\cdot\sigma} e^{0} = \frac{1}{\sqrt{2\pi}\cdot\sigma}}$$

$$\boxed{\frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \leqslant \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{0} = \frac{1}{\sqrt{2\pi}\cdot\sigma}} \longleftrightarrow \boxed{\begin{array}{l} f(x) \leqslant f(\mu) \\ \text{for all } x \in \mathbb{R} \end{array}}$$

(r) The cumulative distribution function $F: \mathbb{R} \to \mathbb{R}$ of the random variable X is:

$$\boxed{\begin{aligned} F(x) = P(X \leq x) &= \frac{1}{\sqrt{2\pi}\cdot\sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt. \end{aligned}}$$

(✓) $\int_{-\infty}^{0} e^{-\frac{1}{2}t^2}\,dt = \int_{0}^{+\infty} e^{-\frac{1}{2}t^2}\,dt = \frac{\sqrt{2\pi}}{2}.$

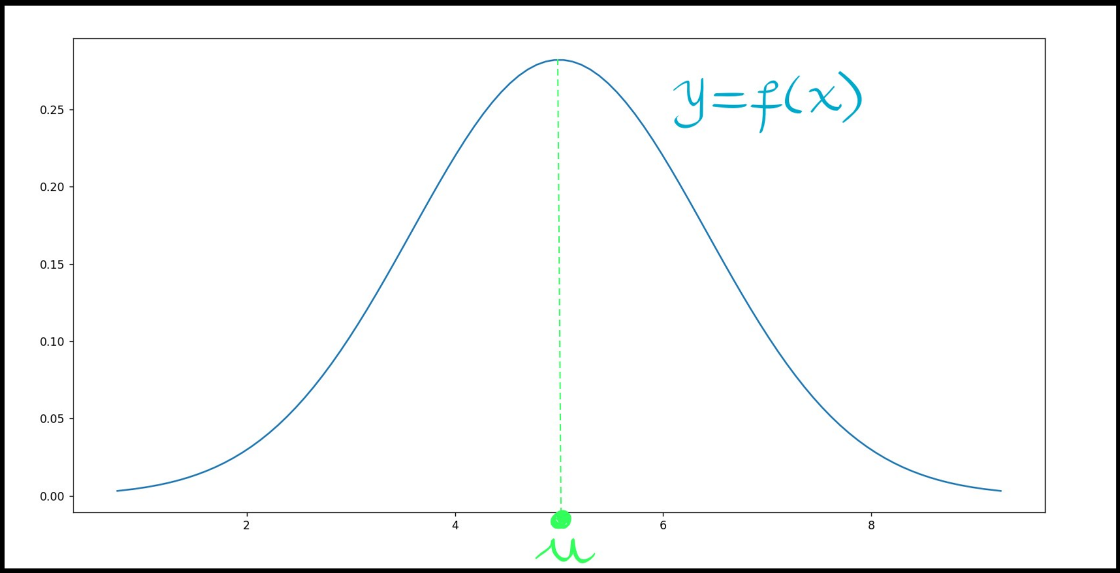
(✓) $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{1}{2}t^2}\,dt = \frac{1}{2}.$

(✓) The median value of x is the value $x=\mu$, because:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}t^2}\,dt$$

$$F(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\mu-\mu}{\sigma}} e^{-\frac{1}{2}t^2}\,dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{1}{2}t^2}\,dt = \frac{1}{2}$$

(✓) The mean value of x is $E(x)=\mu$.



μ is the median value the mode value and the mean value of X.

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

Answer:

If Y is a random variable such that

$Y \sim \exp(\lambda)$

(r) The probability density function $f : \mathbb{R} \to \mathbb{R}$ of Y is given by:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{if } x\geq 0,\\ 0 & \text{if } x<0.\end{cases}$$

(v) The mode value of $\underset{\sim}{y}$ is $x=0$, because the following reasons:

- $f(x)\geq 0$.

- If $g(x)=\lambda e^{-\lambda x}$, then $g'(x)=-(\lambda^2 e^{-\lambda x})<0$ and this implies that g is a decreasing function. In particular, we have that

$$\boxed{\begin{array}{l} g(0)\geq g(x) \\ \text{for } x\geq 0 \end{array}} \longleftrightarrow \boxed{\begin{array}{l} f(0)\geq f(x) \\ f(0)=\lambda e^{-\lambda(0)}=\lambda \\ f(x)=\lambda e^{-\lambda x} \quad x\geq 0 \end{array}}$$

Thus the maximum value of $\underline{f}$ is reached only in the point $x=0$, and this implies that $x=0$ is the mode value for the random variable $\underset{\sim}{y}$.

(r) The cumulative distribution function $F:\mathbb{R}\to\mathbb{R}$ of the random variable $\underset{\sim}{y}$ is:

$$F(x)=P(\underset{\sim}{y}\leq x)=\int_{-\infty}^{x} f(t)\,dt=$$

$$=\begin{cases} 0 & \text{if } x\leq 0, \\ \int_{0}^{x} f(t)\,dt & \text{if } x>0 \end{cases}$$

$$=\begin{cases} 0 & \text{if } x\leq 0, \\ \int_{0}^{x} \lambda e^{-\lambda t}\,dt & \text{if } x>0 \end{cases}$$

$$=\begin{cases} 0 & \text{if } x\leq 0, \\ 1-e^{-\lambda x} & \text{if } x>0 \end{cases}$$

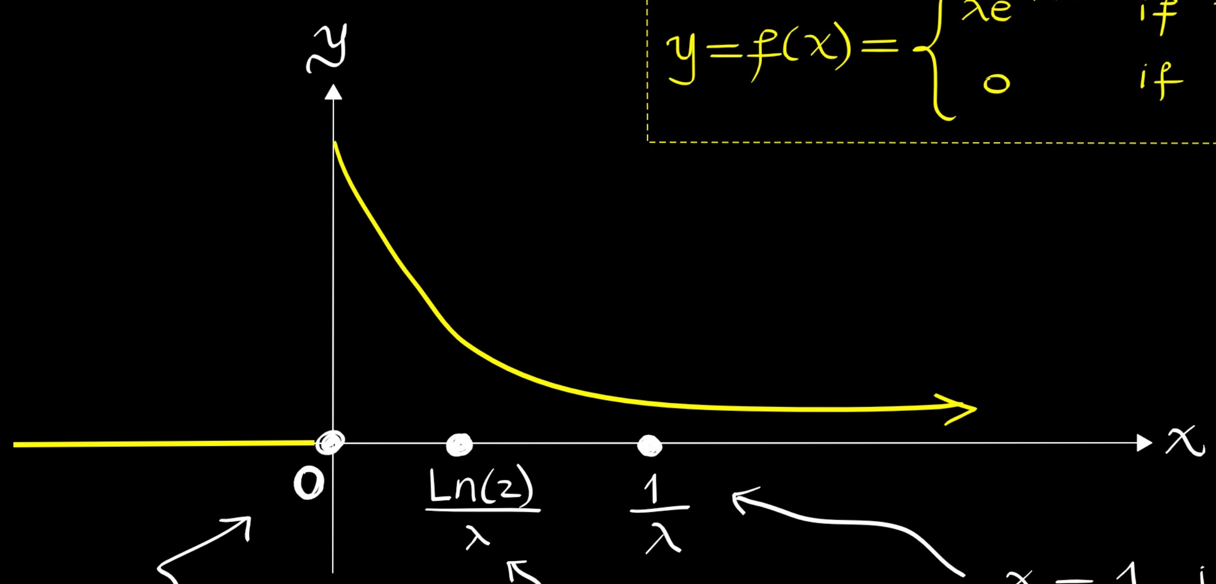
(v) $F(x) = \frac{1}{2} \Longleftrightarrow 1 - e^{-\lambda x} = \frac{1}{2} \Longleftrightarrow e^{-\lambda x} = \frac{1}{2}$

$$e^{-\lambda x} = \frac{1}{2} \Longleftrightarrow -\lambda x = \ln\left(\frac{1}{2}\right) \Longleftrightarrow x = \frac{\ln(2)}{\lambda}$$

Therefore the median value of the random variable $\underline{\underline{y}}$ is $x = \frac{\ln(2)}{\lambda}$.

(v) The mean value of $\underline{\underline{y}}$ is $E(y) = \frac{1}{\lambda}$.

$$y = f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$



$x = 0$ is the mode value

$x = \frac{\ln(2)}{\lambda}$ is the median value

$x = \frac{1}{\lambda}$ is the mean value

Problem 2

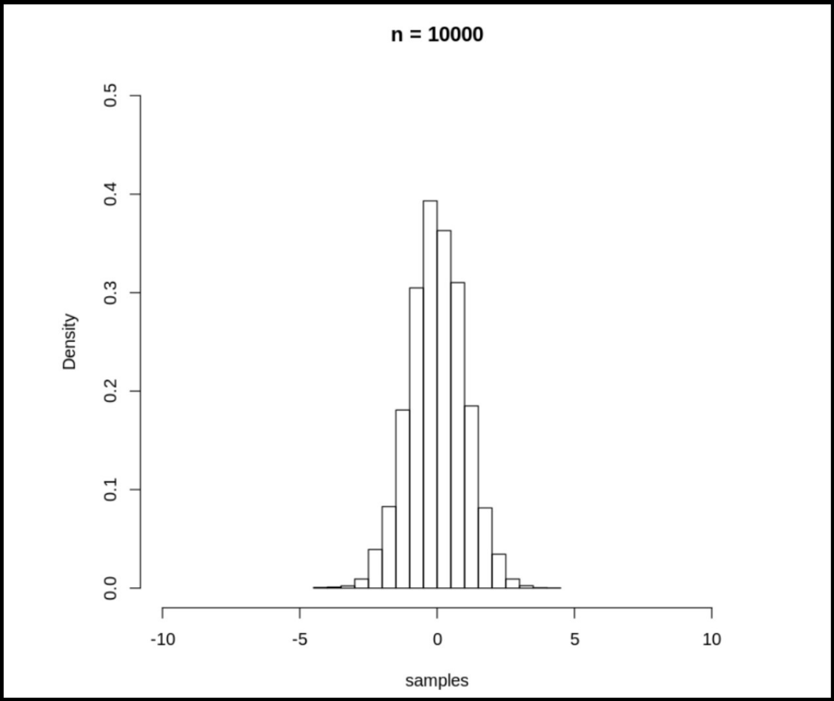
For this problem, we're going to visualize what's happening when we go between different normal distributions.

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**.

Answer:

```
samples=rnorm(10000, mean = 0, sd = 1)
hist(samples, main = "n = 10000",xlim = range(-10,10), ylim =range(0,0.5), prob = TRUE)
```

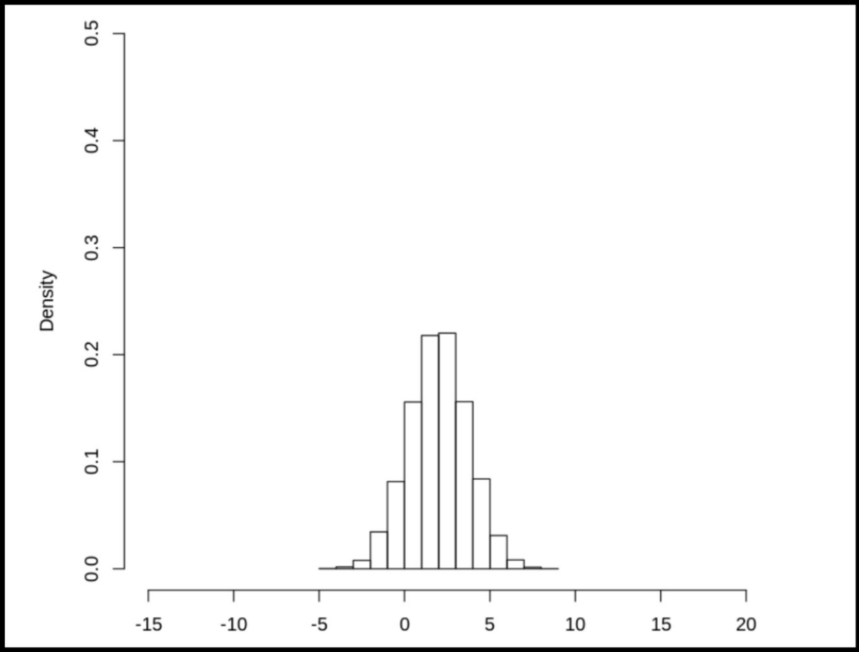


Part b) Now generate 10000 samples from a $N(2, 3)$ distribution and plot a histogram of the results, with the same x-limits and y-limits. Does the histogram make sense based on the changes to parameters?

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

Answer:

```
samples_2=rnorm(10000, mean = 2, sd = sqrt(3))
hist(samples_2, main = "n = 10000",xlim = range(-15,20), ylim =range(0,0.5), prob = TRUE)
```



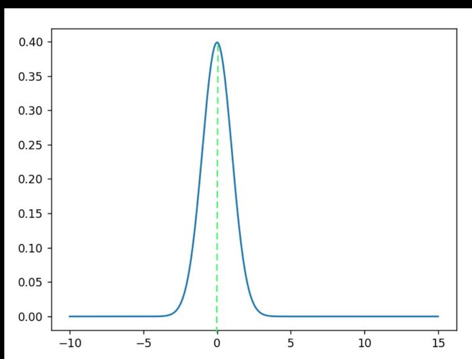
And this new histogram makes sense, because the probability density function $f: \mathbb{R} \to \mathbb{R}$ relative to the random variable $x \sim Nor(2,3)$ is described as:

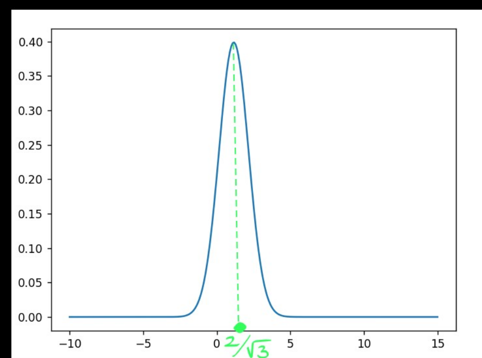
$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{3}} e^{-\frac{1}{2}\left(\frac{x-2}{\sqrt{3}}\right)^2} = \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{1}{\sqrt{3}}x - \frac{2}{\sqrt{3}}\right)^2}$$

and this function is obtained from the probality density función $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ relative to the random variable $y \sim N(0,1)$ as follows:

(✓) we shift horizontally to the right the function g by $\frac{2}{\sqrt{3}}$.

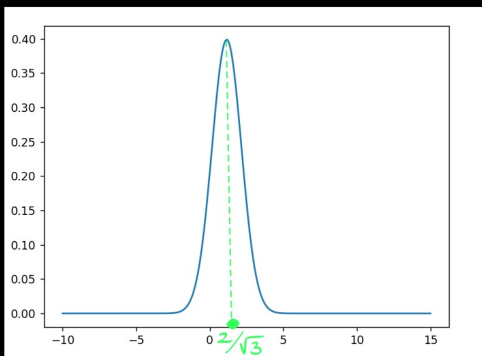
$$g(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \longrightarrow g_1(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(x-\frac{2}{\sqrt{3}}\right)^2}$$

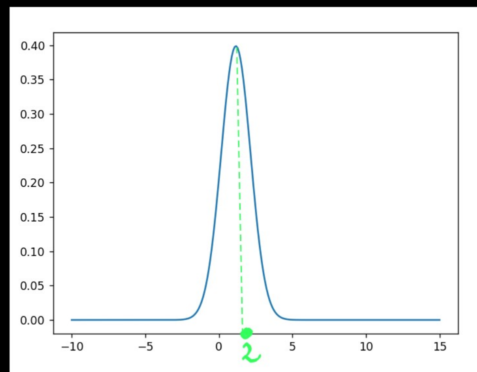




(✓) we dilate horizontally the function g_1 by $\sqrt{3}$.

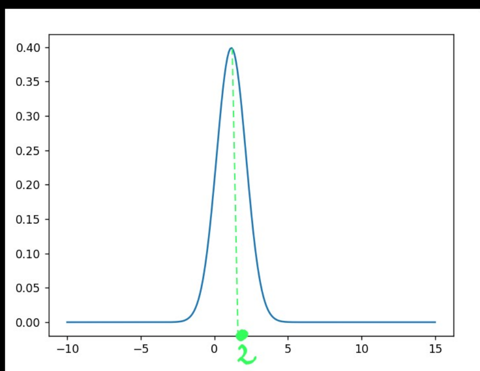
$$g_1(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(x-\frac{2}{\sqrt{3}}\right)^2} \longrightarrow g_2(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{1}{\sqrt{3}}x-\frac{2}{\sqrt{3}}\right)^2}$$

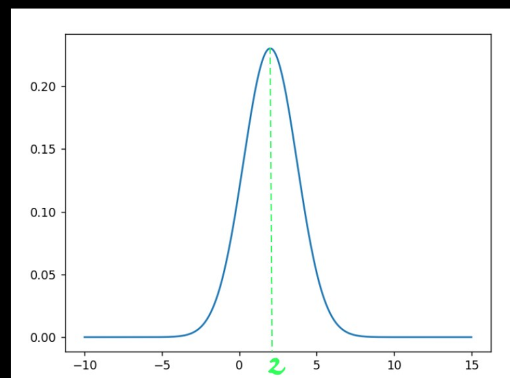




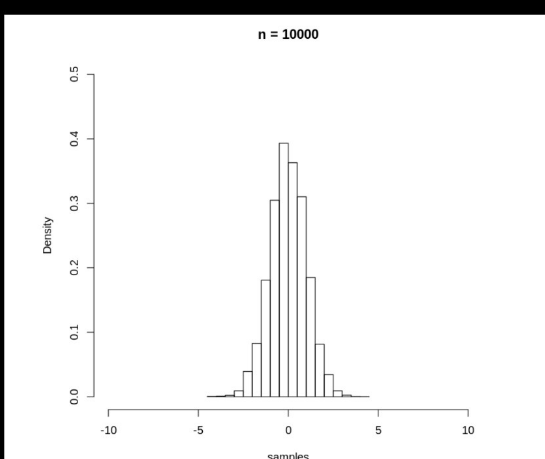
(v) And finally, we contract vertically the function g_2 by $\frac{1}{\sqrt{3}}$.

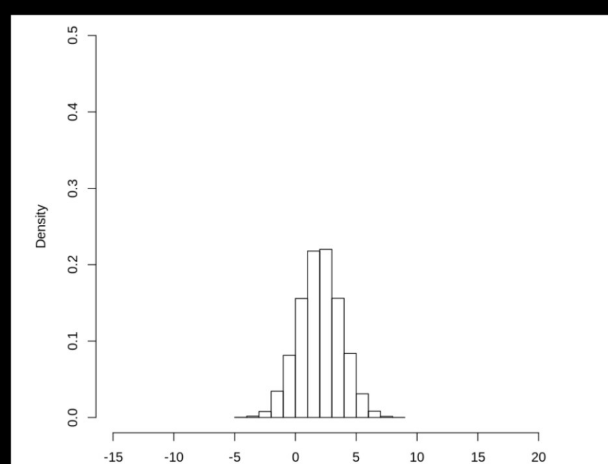
$$g_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{1}{\sqrt{3}}x - \frac{2}{\sqrt{3}}\right)^2} \longrightarrow f(x) = \frac{1}{\sqrt{2\pi}\cdot\sqrt{3}} e^{-\frac{1}{2}\left(\frac{x-2}{\sqrt{3}}\right)^2}$$





Therefore, the previous argument proves that our results make sense.





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**?

Answer:

To begin with, we note the following things:

(r)

```
samples=rnorm(10000, mean = 0, sd = 1)
samples
```

The variable samples generates 10000 observations from the random variable $x \sim N(0,1)$.

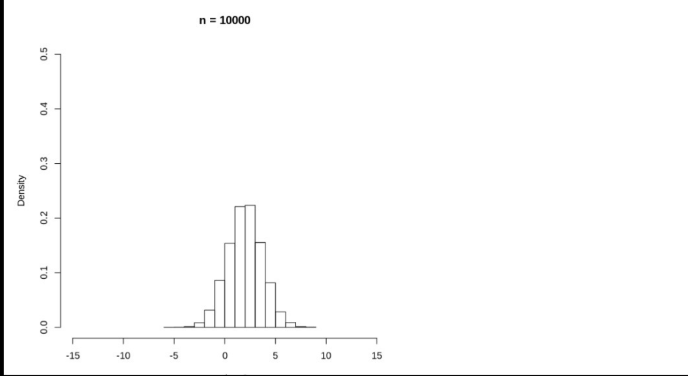
(r)

```
samples_2=sqrt(3)*samples+2
hist(samples_2, main = "n = 10000",xlim = range(-15,15), ylim =range(0,0.5), prob = TRUE)
```

The variable Samples_2 describes 10000 observations of the random variable $x \sim N(2,3)$.

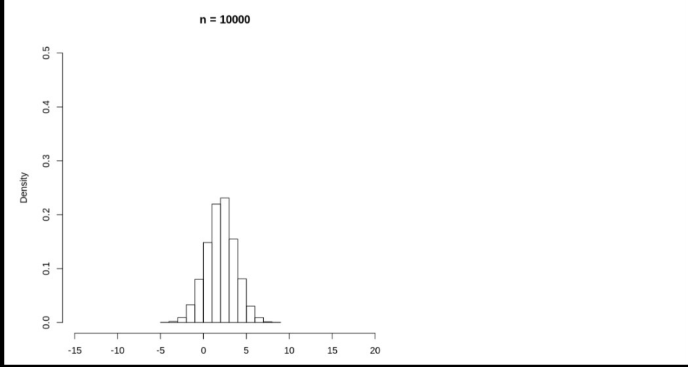
Histogram (c)

```r
samples_2=sqrt(3)*samples+2
hist(samples_2, main = "n = 10000",xlim = range(-15,15), ylim =range(0,0.5), prob = TRUE)
```



Histogram (b)

```r
samples_3=rnorm(10000, mean = 2, sd = sqrt(3))
hist(samples_3, main = "n = 10000",xlim = range(-15,20), ylim =range(0,0.5), prob = TRUE)
```



It is clear that the previous analysis shows that the histogram (b) and the histogram (c) are quite similar.

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**?

Answer:

we note to start the following things:

(r)

```r
samples=rnorm(10000, mean = 2, sd = sqrt(3))
samples
```

The variable samples generates 10000 observations from the random variable $x \sim N(2,3)$.

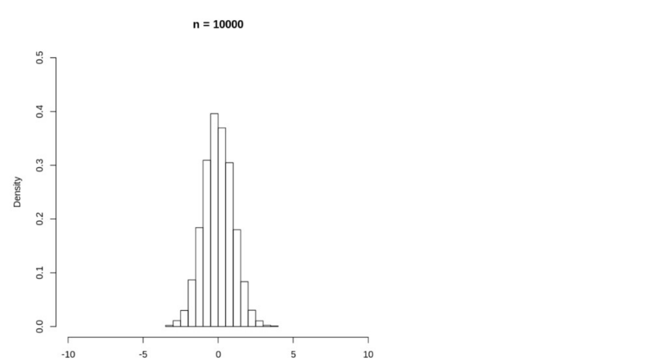
(r)

```
samples_2=(samples-2)/sqrt(3)
hist(samples_2, main = "n = 10000",xlim = range(-10,10), ylim =range(0,0.5), prob = TRUE)
```

The variable Samples_2 describes 10000 observations of the random variable $x \sim N(0,1)$.

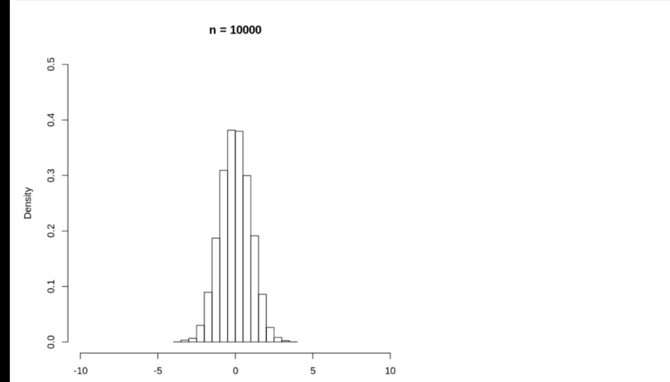
Histogram (d)

```
samples_2=(samples-2)/sqrt(3)
hist(samples_2, main = "n = 10000",xlim = range(-10,10), ylim =range(0,0.5), prob = TRUE)
```



Histogram (a)

```
samples=rnorm(10000, mean = 0, sd = 1)
hist(samples, main = "n = 10000",xlim = range(-10,10), ylim =range(0,0.5), prob = TRUE)
```



It is clear that the previous analysis shows that the histogram (a) and the histogram (d) are quite similar.