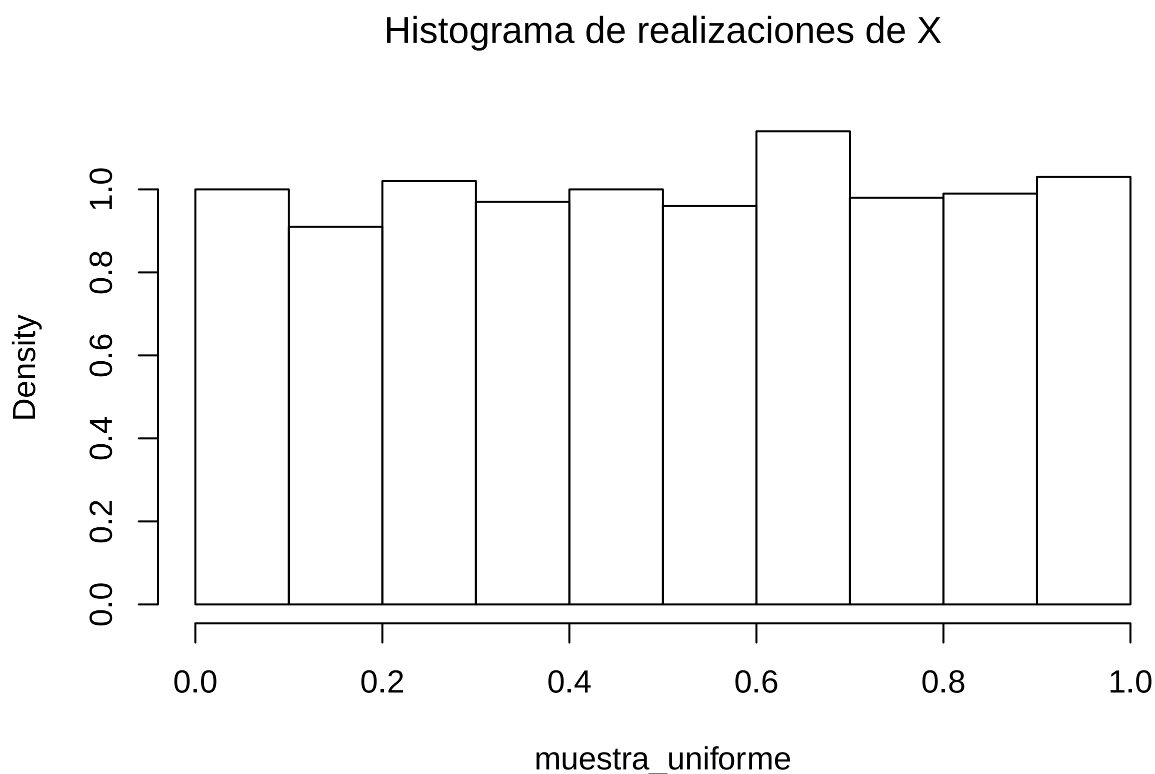
Trabajo Practico Ley de los Grandes Números y Teorema Central del Límite

Leo Mansini (318/19), Andres Mauro (39/17)

Ejercicio A.

$$X \sim U(0, 1)$$

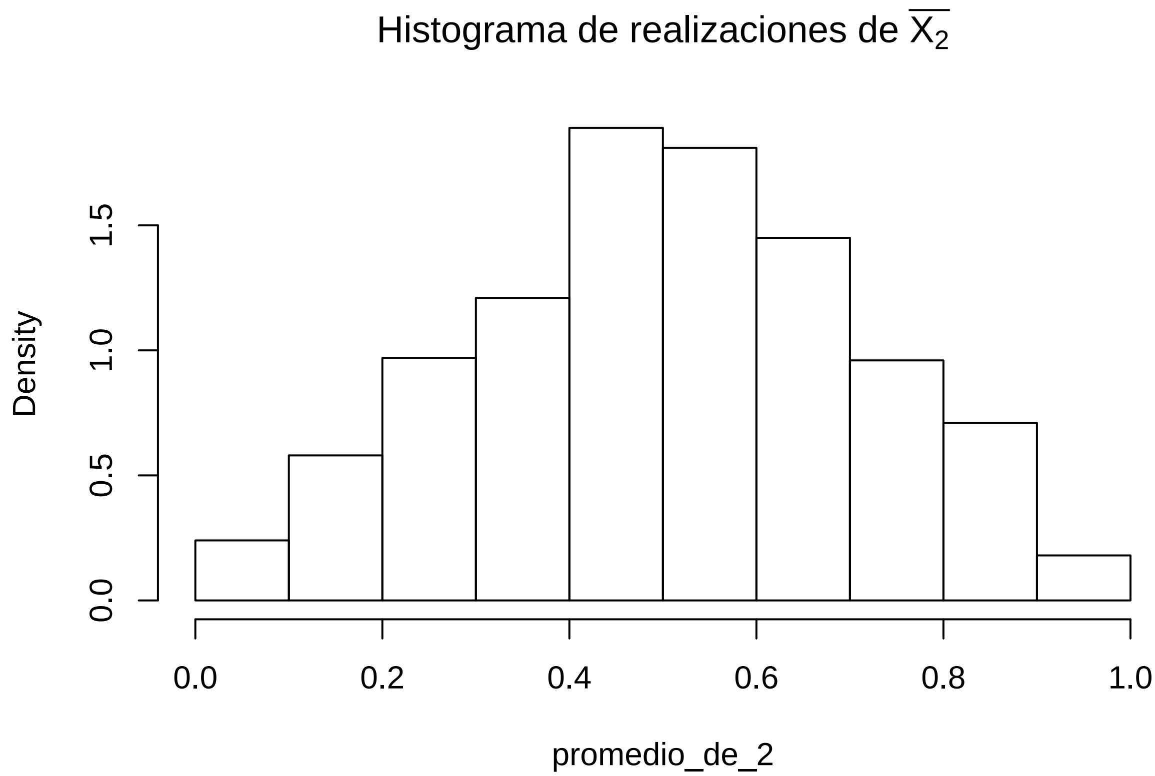


El histograma se asmeja al gráfico de una densidad uniforme entre 0 y 1.

Ejercicio B.

Considero la variable aleatoria

$$\overline{X}_2 = \frac{X_1 + X_2}{2}, \quad X_1, X_2 \sim U(0,1)$$

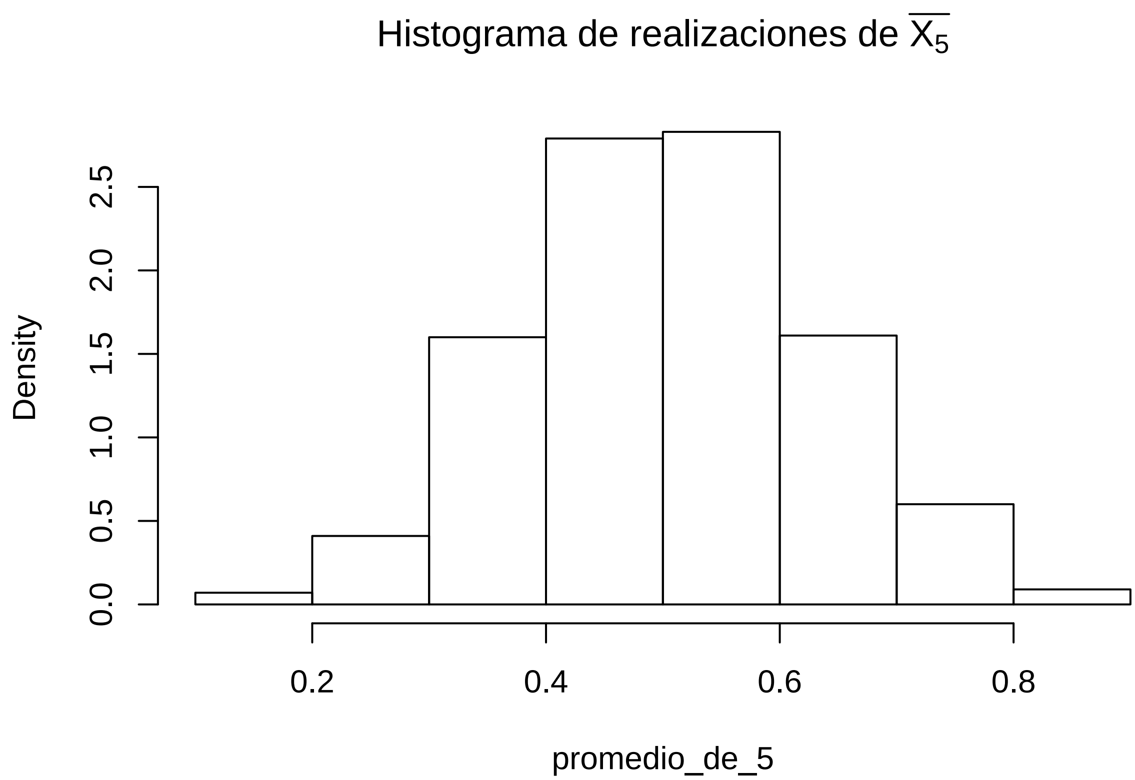


El histograma se parece a una densidad normal centrada en el 0.5.

Ejercicio C.

Defino

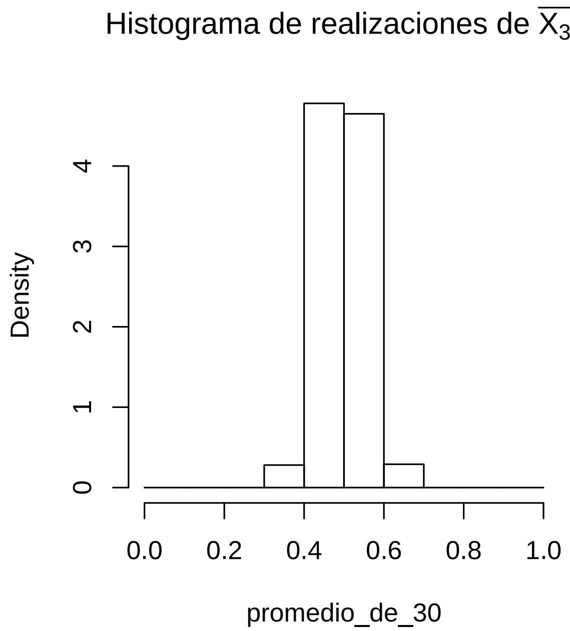
$$\overline{X}_5 = \frac{1}{5}\sum_{i=1}^{5} X_i, \quad X_i \sim U(0,1)$$

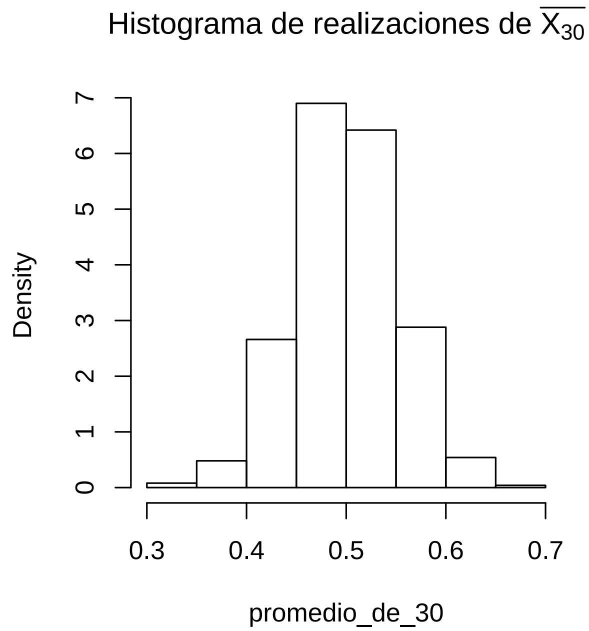


Este histograma es similar al anterior pero muestra una dispersión menor en los datos. También parece una normal centrada en 0.5, pero con una varianza menor que en el ejercicio b.

Ejercicio D.

$$\overline{X}_{30} = \frac{1}{30}\sum_{i=1}^{30} X_i, \quad X_i \sim U(0,1)$$

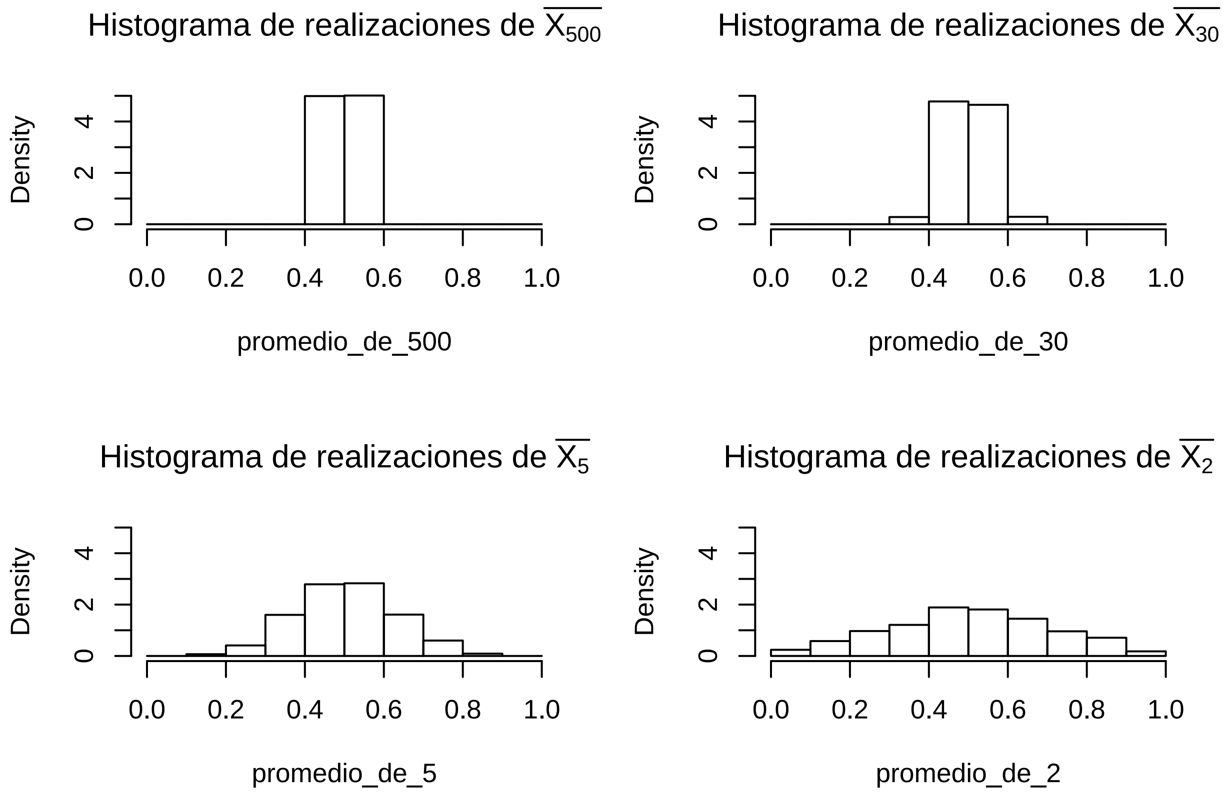




En el primer histograma se graficaron las realizaciones de la variable uniforme en la misma escala que los histogramas anteriores, lo cual muestra que la concentración es mucho mayor. En el segundo gráfico se grafica de 0.3 a 0.7, donde están la gran mayoria de los datos.

Ejercicio E.

$$\overline{X}_{500} = \frac{1}{500}\sum_{i=1}^{500} X_i, \quad X_i \sim U(0,1)$$

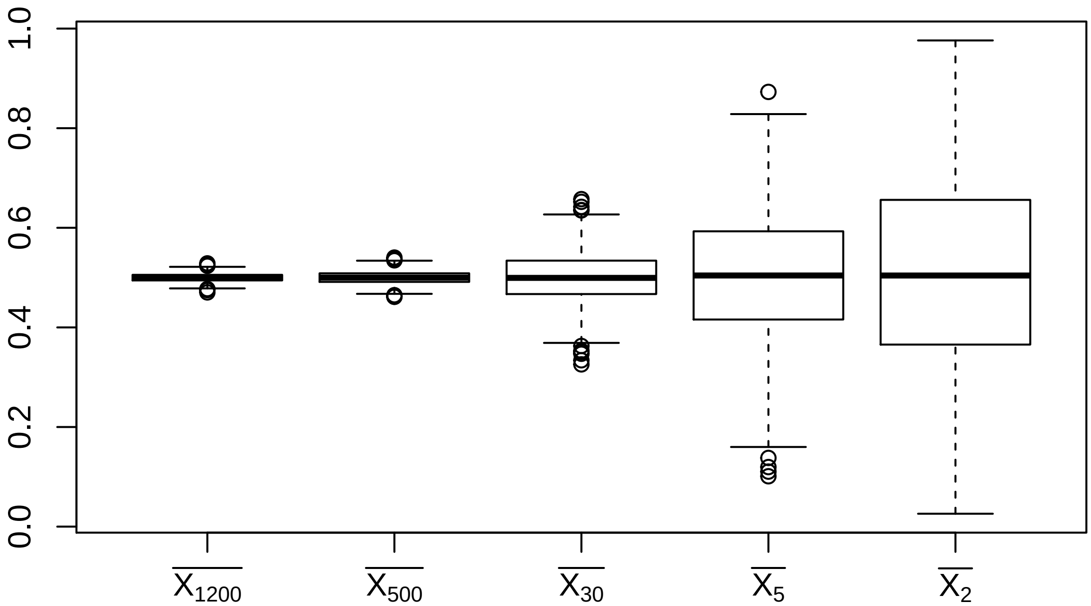


Los histogramas están todos en la misma escala.

Si se aumenta el tamaño de la muestra los valores estan aún más concentrados.

Ejercicio F.

$$\overline{X}_{1200} = \frac{1}{1200}\sum_{i=1}^{1200} X_i, \quad X_i \sim U(0,1)$$



En orden, estan los boxplot del promedio de 1200 muestras, 500 muestras, 30 muestras, 5 muestras y 2 muestras. Se ve claramente como a medida que aumenta la cantidad de muestras, la dispersión de datos decrece. Los datos se encuentran cada vez más acumulados en 0.5.

Las medias muestrales de los promedios de 2, 5, 30, 500, y 1200 datos, son:

```
## [1] 0.504712
```

```
## [1] 0.5059515
```

```
## [1] 0.500213
```

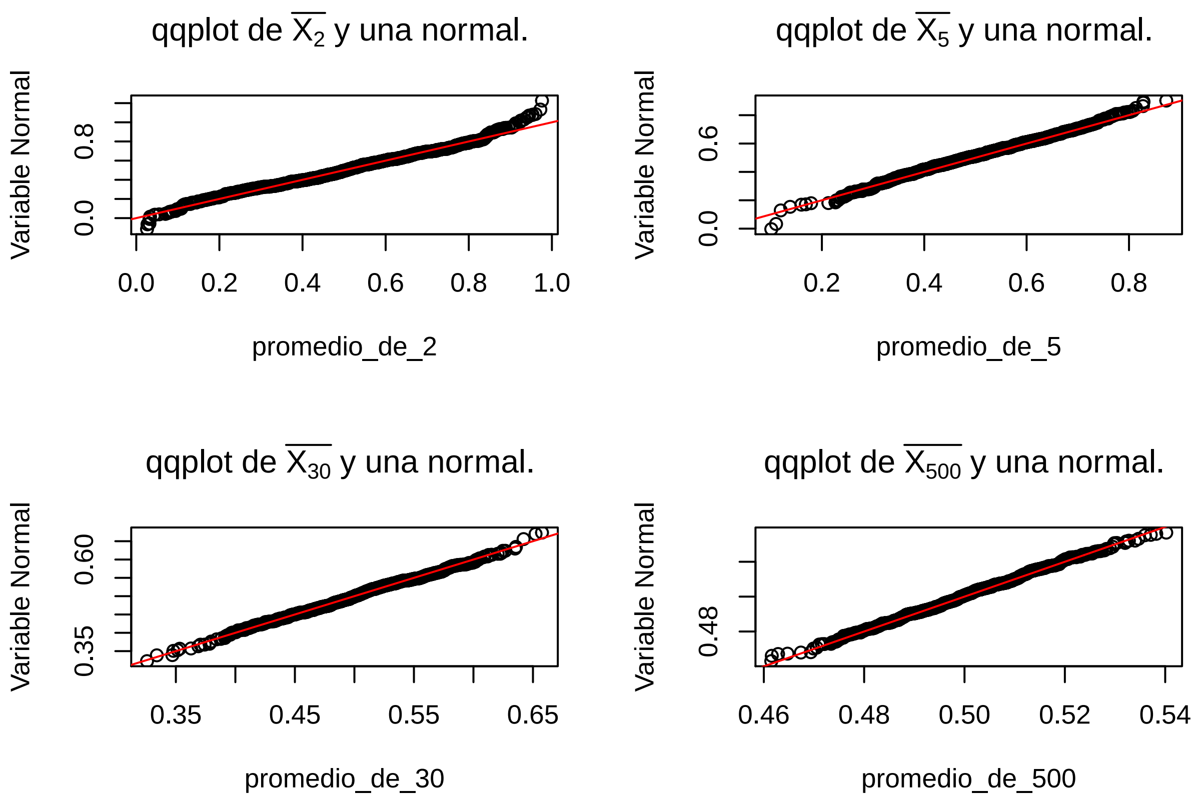
```
## [1] 0.4999383
```

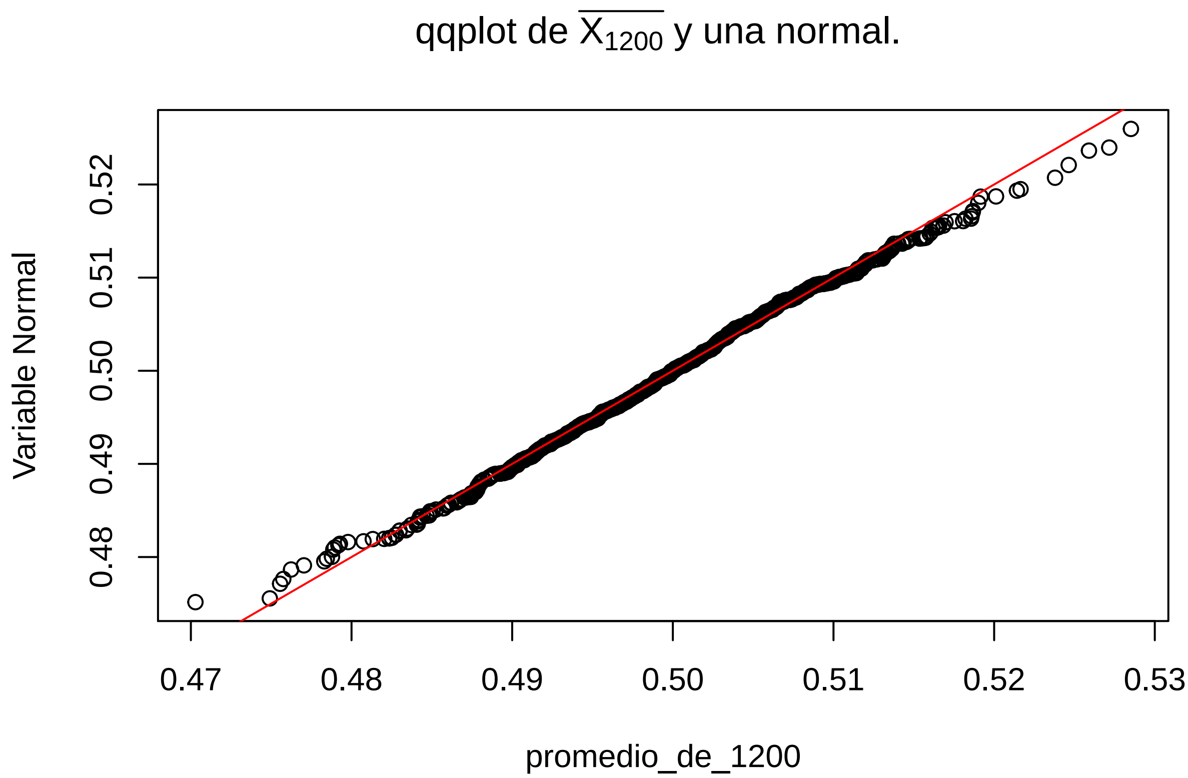
```
## [1] 0.4999171
```

Las varianzas muestrales de los promedios de 2, 5, 30, 500, y 1200 datos, son:

```
## [1] 0.04261539

## [1] 0.01635311

## [1] 0.002659915

## [1] 0.0001582653

## [1] 6.987465e-05
```

Los valores teóricos a los que la media muestral debe tender a medida que aumenta n son la esperanza de una variable aleatoria con distibucion uniforme entre 0 y 1, es decir, 0.5. La varianza muestral debe parecerse a la varianza de la variable aleatoria dividido n, $\frac{1}{12n}$. Por eso es que a medida que aumenta n, la varianza muestral tiende a 0.





Aquí se realizaron qqplot de las variables de promedio de uniformes, y de variables normales con media y varianza igual a la media y varianza muestral de los promedios.

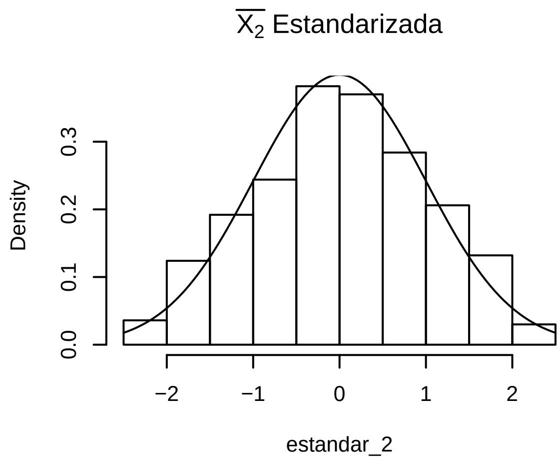
También se superpuso a los qqplot una recta identidad en rojo.

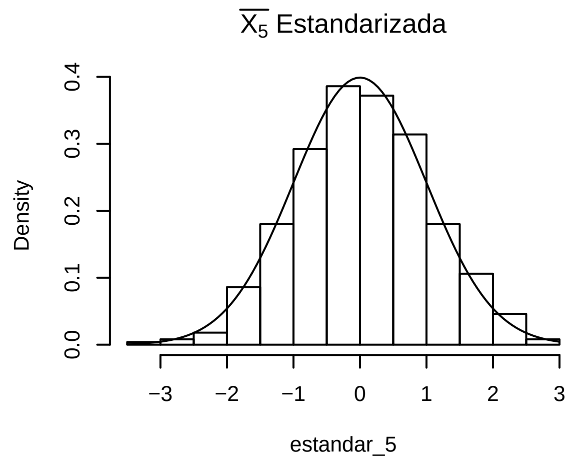
Cuando el qqplot de un conjunto de datos y otro se asemeja a una recta identidad, esto significa que sus distribuciones son similares.

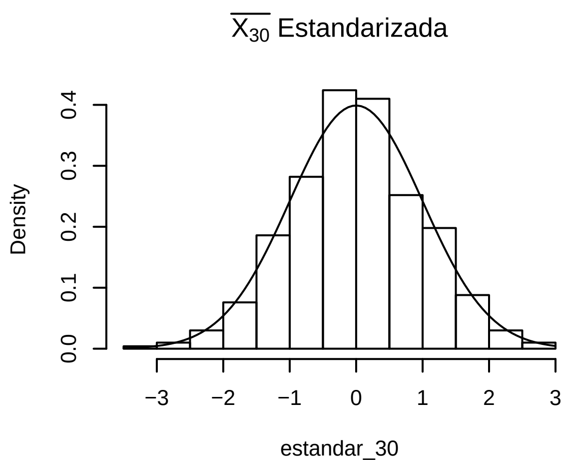
El hecho de que se parezcan a una distribución normal se condice con el Teorema Central de Límite ya que las variables graficadas son promedios, y estas tienden en probabilidad a una variable aleatoria con distribución normal de media $E(X)$ y varianza $\frac{V(X)}{n}$, a medida que aumenta n.

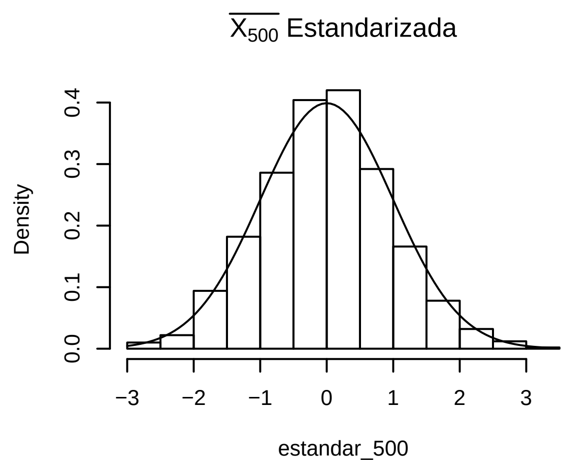
Este Teorema es aplicable porque tanto la esperanza como la varianza de una variable aleatoria uniforme son finitas.

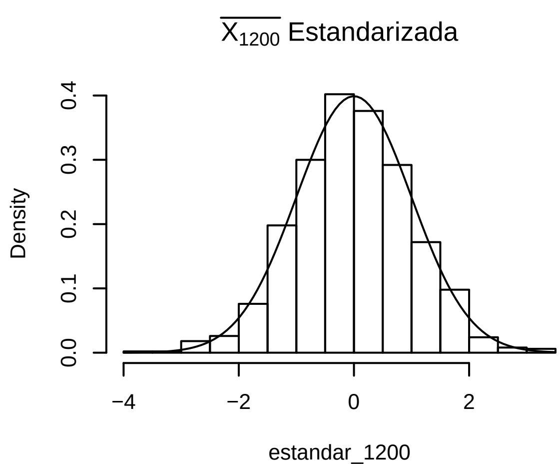
Ejercicio G.

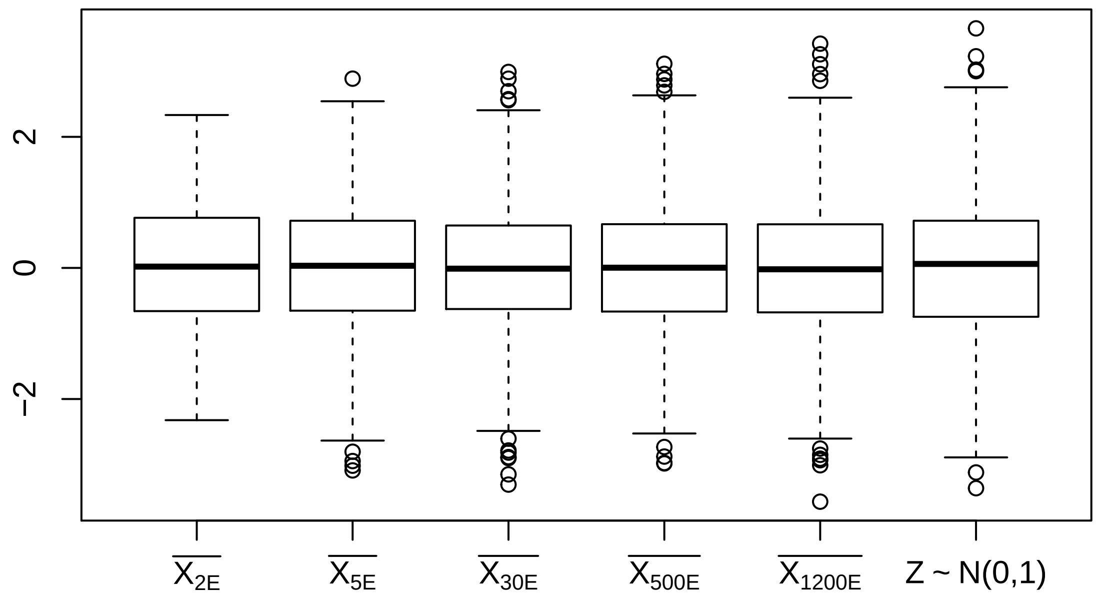

$\overline{X_2}$ Estandarizada
Density
estandar_2


$\overline{X_5}$ Estandarizada
Density
estandar_5


$\overline{X_{30}}$ Estandarizada
Density
estandar_30


$\overline{X_{500}}$ Estandarizada
Density
estandar_500

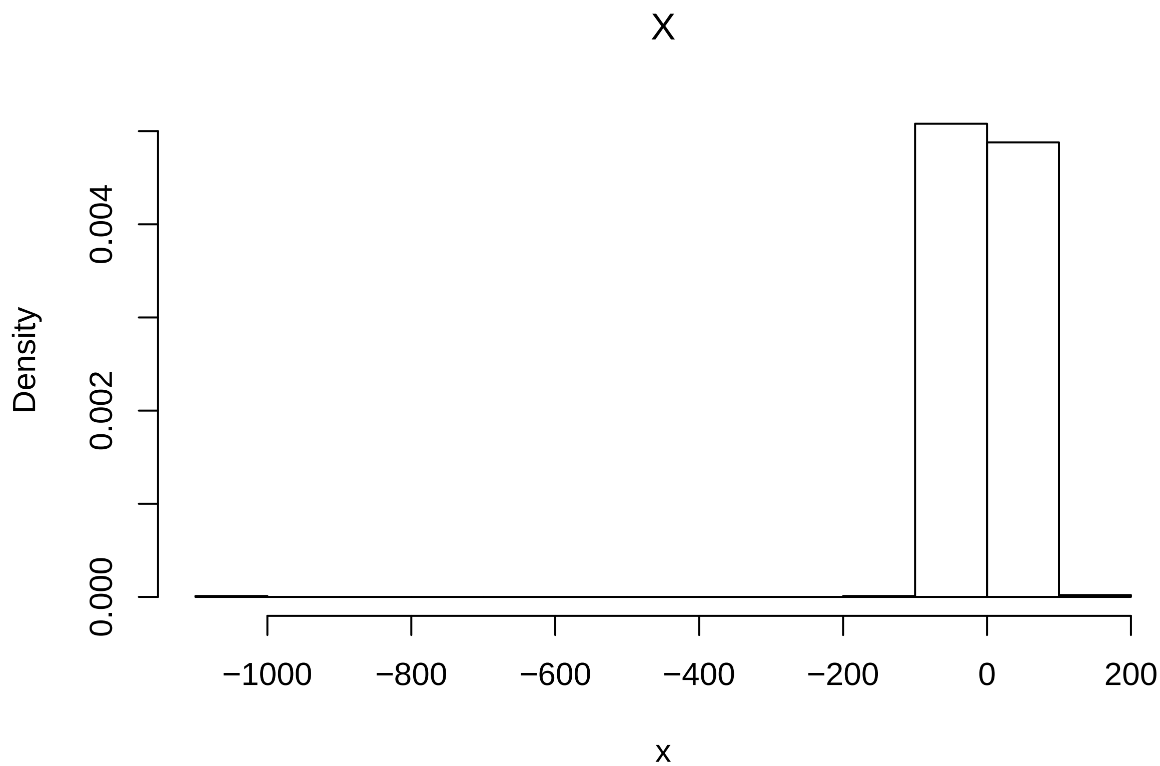

$\overline{X_{1200}}$ Estandarizada
Density
estandar_1200



Podemos decir que todos los conjuntos de datos se pueden estandarizar y asi parecen una normal estándar. En los histogramas se puede ver con la curva de densidad normal sobre las barras, y en los boxplot esta graficado un conjunto de 1000 variables con distribución normal para comparar.

Ejercicio H.

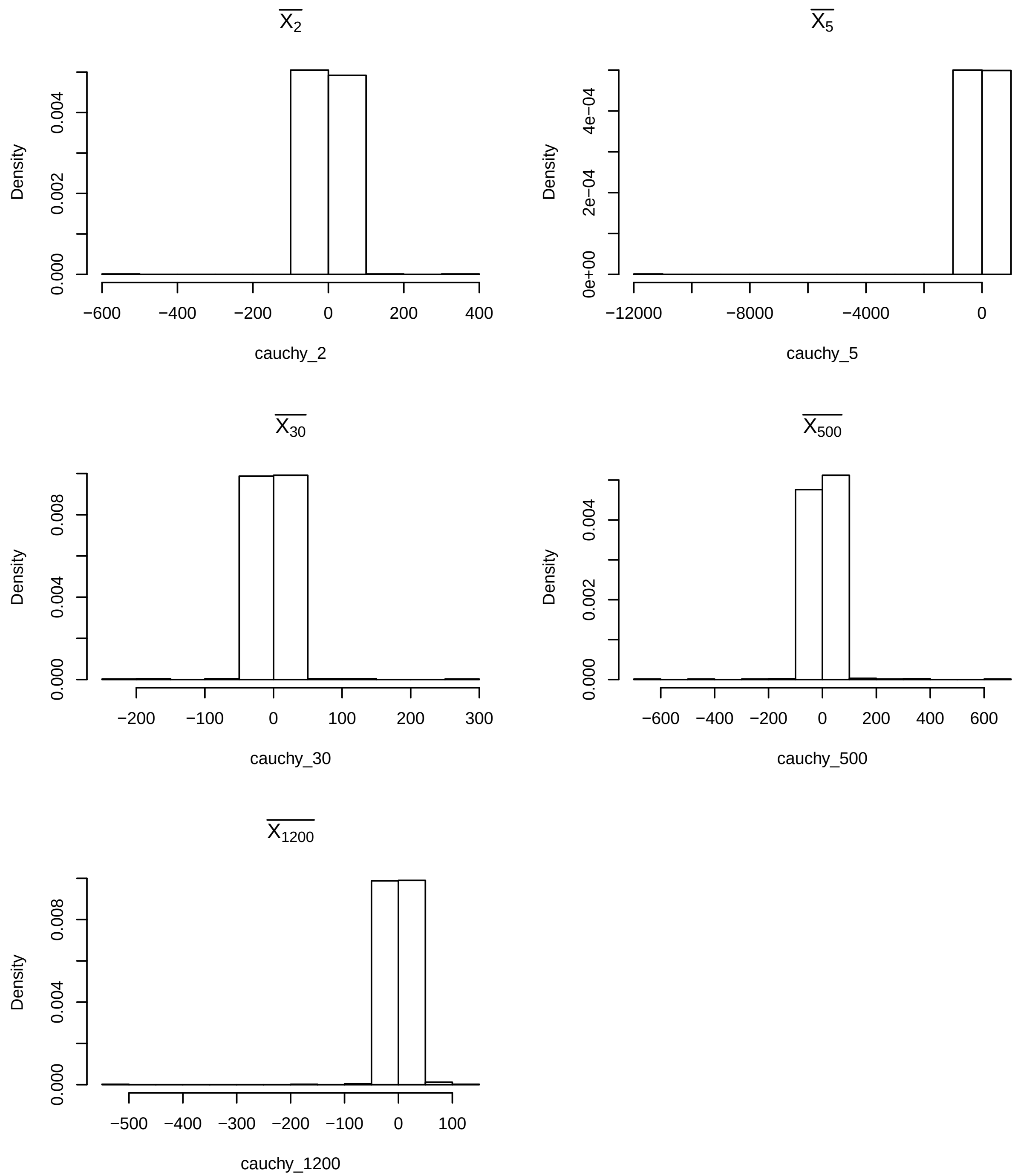
$$X \sim C(0,1), \text{ siendo } f_X(x) = \frac{1}{\pi(1+x^2)}$$



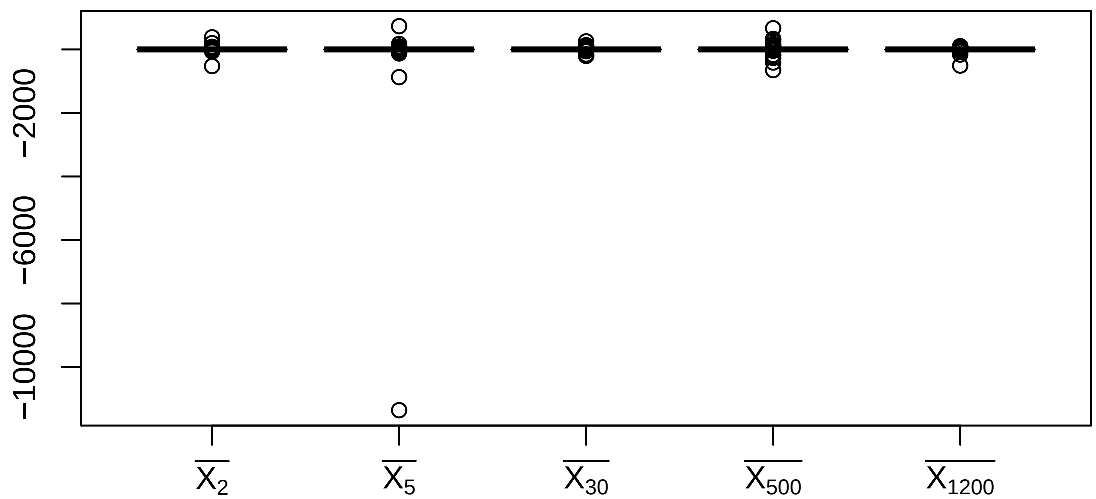
No se parece a ninguna otra densidad que hayamos estudiado.

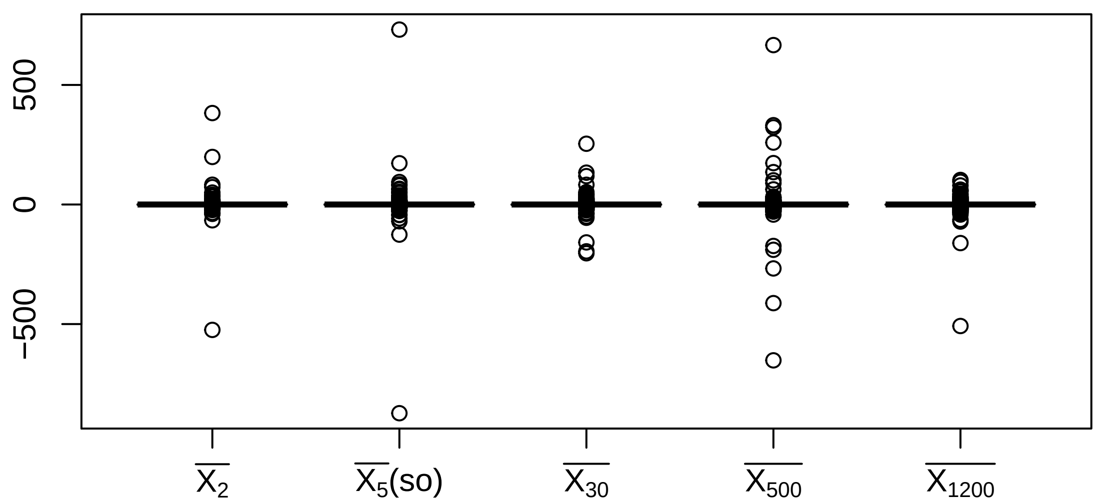
Defino:

$$\overline{X_i} = \frac{1}{i}\sum_{j=1}^{i} X_j, \quad X_j \sim C(0,1)\forall j$$



Aquí estan graficados los histogramas de los conjuntos de datos de promedios de 2, 5, 30, 500, y 1200 variables con distribución de cauchy. Ninguno de estos histogramas se parece a una normal. Aumentando el tamaño de la muestra no parece haber cambios en la dispersión.





Los boxplots están graficados 2 veces por la presencia de un outlier en el conjunto de datos cauchy5. En el segundo gráfico se removió ese outlier, cambiando $\overline{X_5}$ por $\overline{X_5}(so)$ aunque aún así no se pudo comparar los conjuntos de datos, y ninguno se parece al boxplot de una normal.

Las medias muestrales de los promedios de 2, 5, 30, 500, y 1200 realizaciones de una variable con distribución Cauchy, son:

```
## [1] 0.21678
```

```
## [1] -11.1939
```

```
## [1] -0.04510932
```

```
## [1] 0.6625824
```

```
## [1] -0.2734701
```

Las varianzas muestrales de los promedios de 2, 5, 30, 500, y 1200 realizaciones de una variable con distribución Cauchy, son:

```
## [1] 506.1722
```

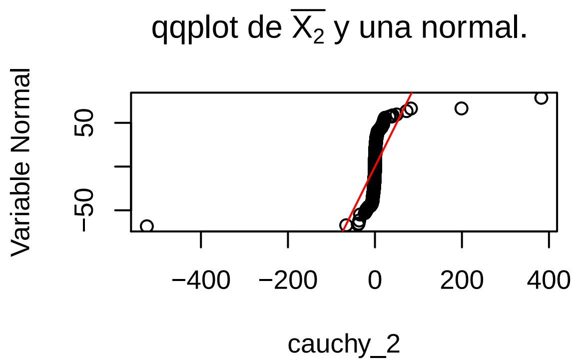
```
## [1] 130457.6
```

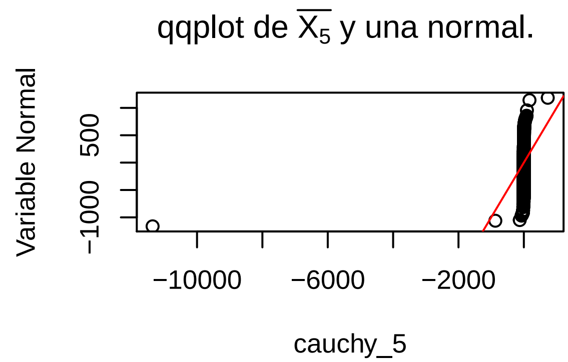
```
## [1] 244.1498
```

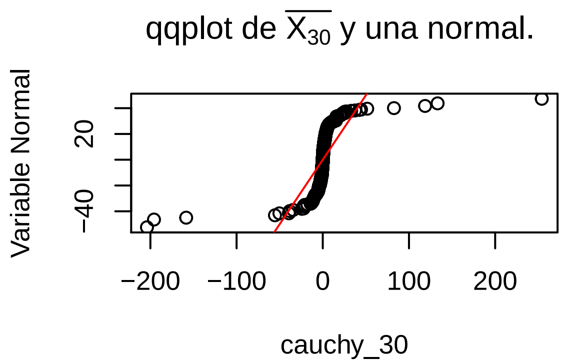
```
## [1] 1553.12
```

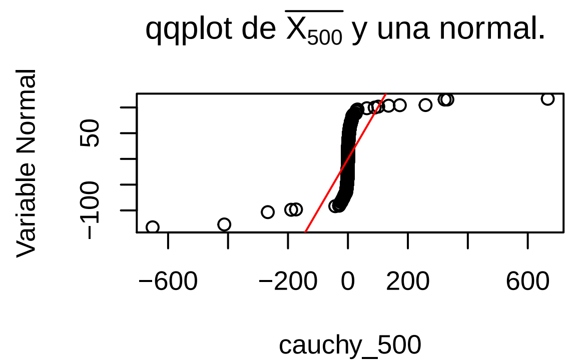
```
## [1] 368.3424
```

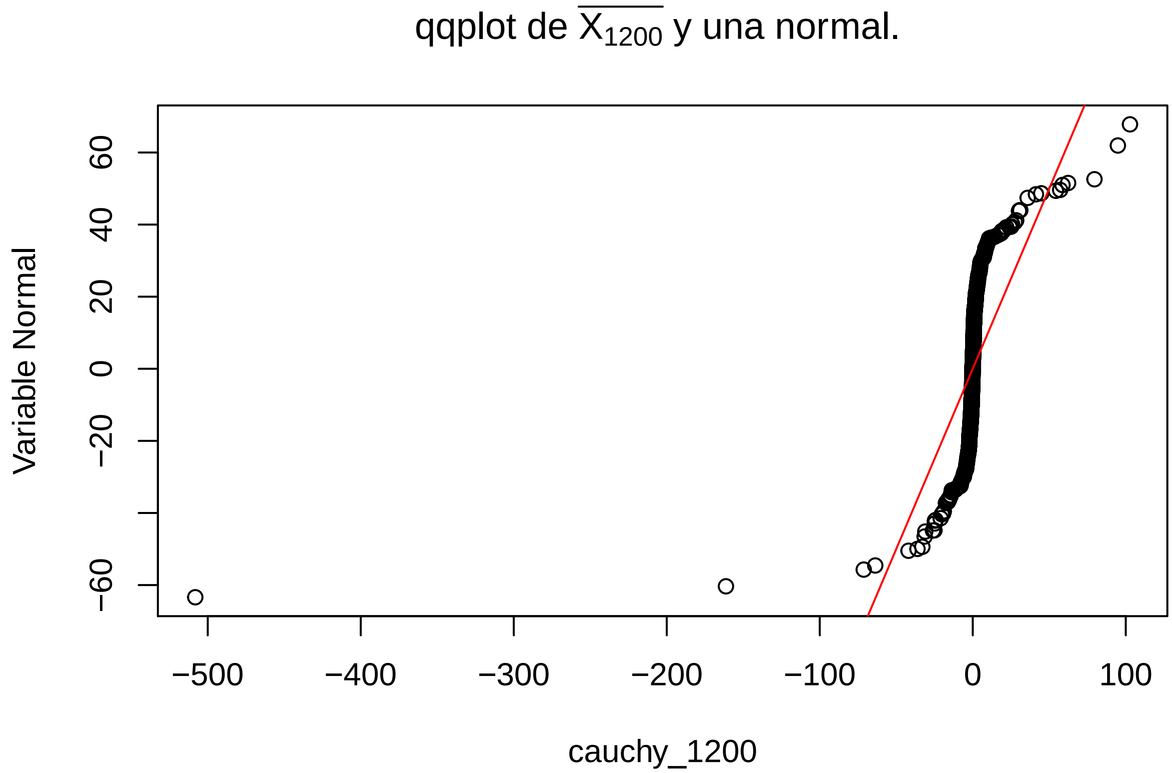
No es posible dar valores teóricos a los que deberían tender las medias o varianzas muestrales dado que la distribución Cauchy no tiene una varianza o una esperanza finita.


qqplot de $\overline{X_2}$ y una normal.
Variable Normal
50
−50
−400
−200
0
200
400
cauchy_2


qqplot de $\overline{X_5}$ y una normal.
Variable Normal
500
−1000
−10000
−6000
−2000
cauchy_5


qqplot de $\overline{X_{30}}$ y una normal.
Variable Normal
20
−40
−200
−100
0
100
200
cauchy_30


qqplot de $\overline{X_{500}}$ y una normal.
Variable Normal
50
−100
−600
−200
0
200
600
cauchy_500



Se realizaron qqplots comparando los conjuntos de datos de promedios de variables de Cauchy y variables normales. Para las variables normales se utilizaron como parametros la media y varianza muestral de los conjuntos de datos.

Aquí se ve que los gráficos no son rectas identidad, por lo que los promedios de variables de Cauchy no siguen una distribución normal.

El Teorema Central del Límite no nos puede decir nada sobre si este resultado es esperado o no, dado que la distribución Cauchy no tiene una esperanza o una varianza teórica finita, una de las condiciones que tiene el Teorema.

Estandarizar el promedio de Cauchy no es posible, dado que necesitamos saber la esperanza y la varianza de una variable Cauchy, y estas no son finitas.