Ley de los Grandes Números

Gómez Jiménez Aaron Mauricio

23/1/2022

1.El objetivo de este ejercicio es utilizar la ley de los grandes números para aproximar integrales de funciones continuas en intervalos finitos. Esta vez procuraremos aproximar la probabilidad de que una variable Z ~ N (0, 1) tome valores en un intervalo [a, b]. Es decir, queremos aproximar numéricamente la siguiente integral

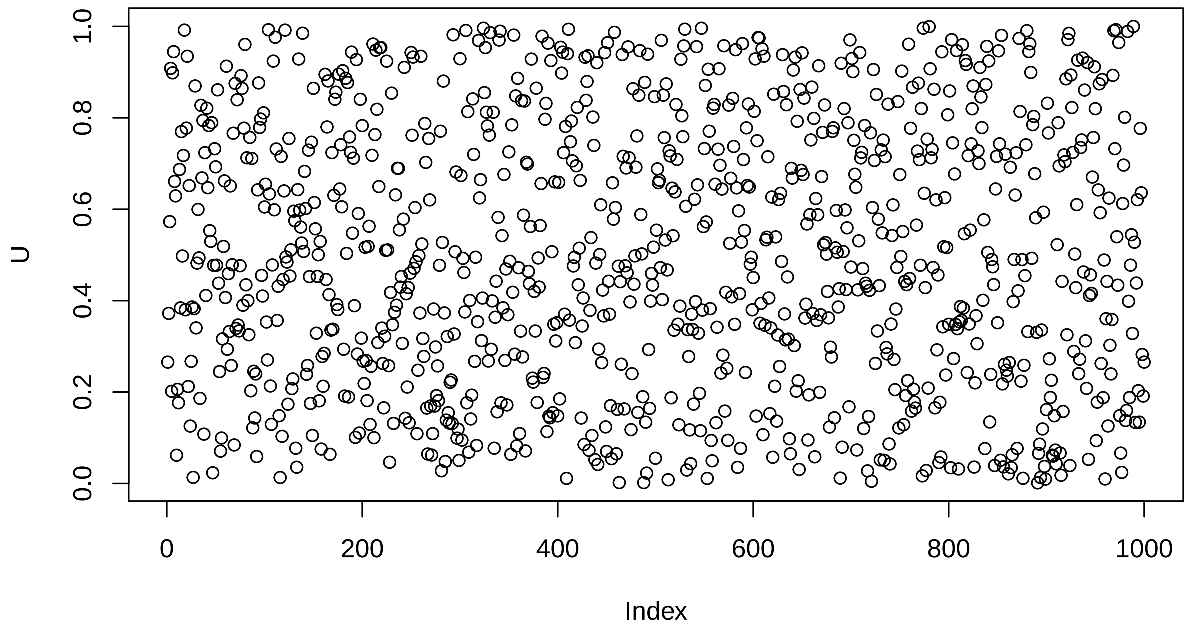
$$\int_a^b \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx$$

Para ello consideremos la siguiente propuesta:

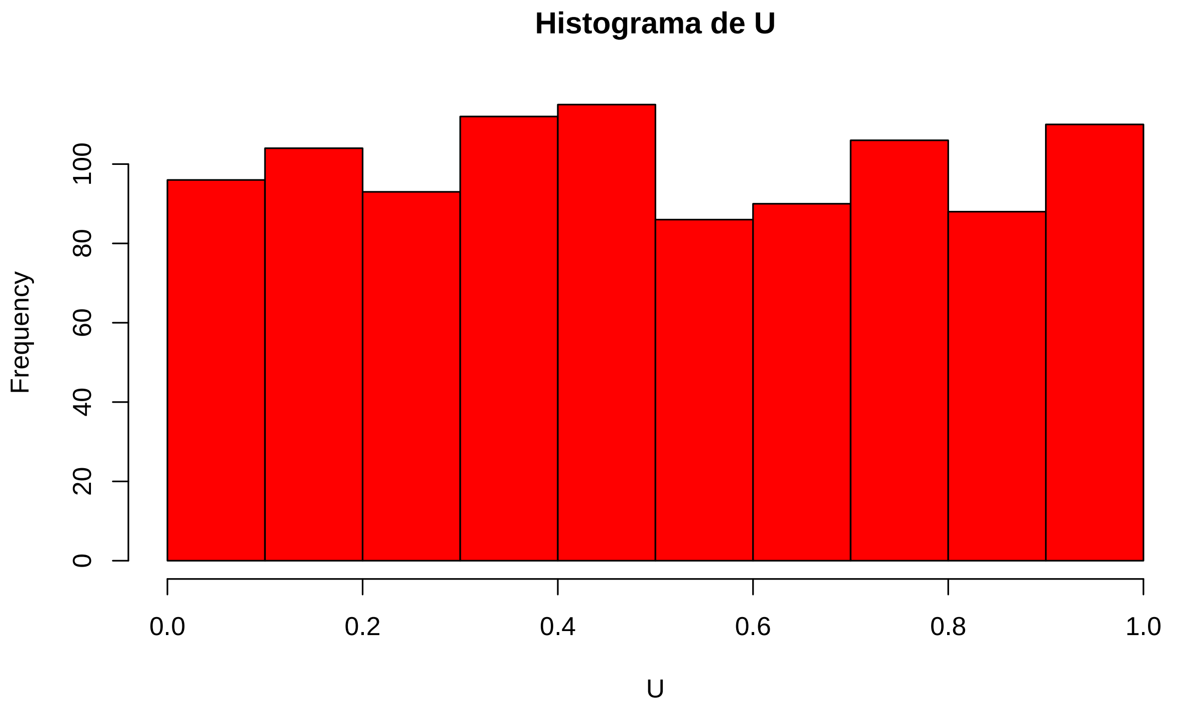
a)Sea U ~ U[0,1]. Dados a, b R, encontrar α ***y*** β ***de forma tal que*** $Y = \alpha U + \beta$ ***tenga distribución uniforme en el intervalo*** $[a, b]$***.***

Notemos que al hacer una nueva variable $Y = \alpha U + \beta$ es como si estuvieramos haciendo una transformación lineal donde la gráfica de la variable solo cambia su dominio, es decir la gráfica se traslada β y el intervalo aumenta en α, veamos un ejemplo, veamos una variable con distribución uniforme en $[0, 1]$

```
set.seed(1)
U<-runif(1000,0,1)
plot(U, type='p')
```

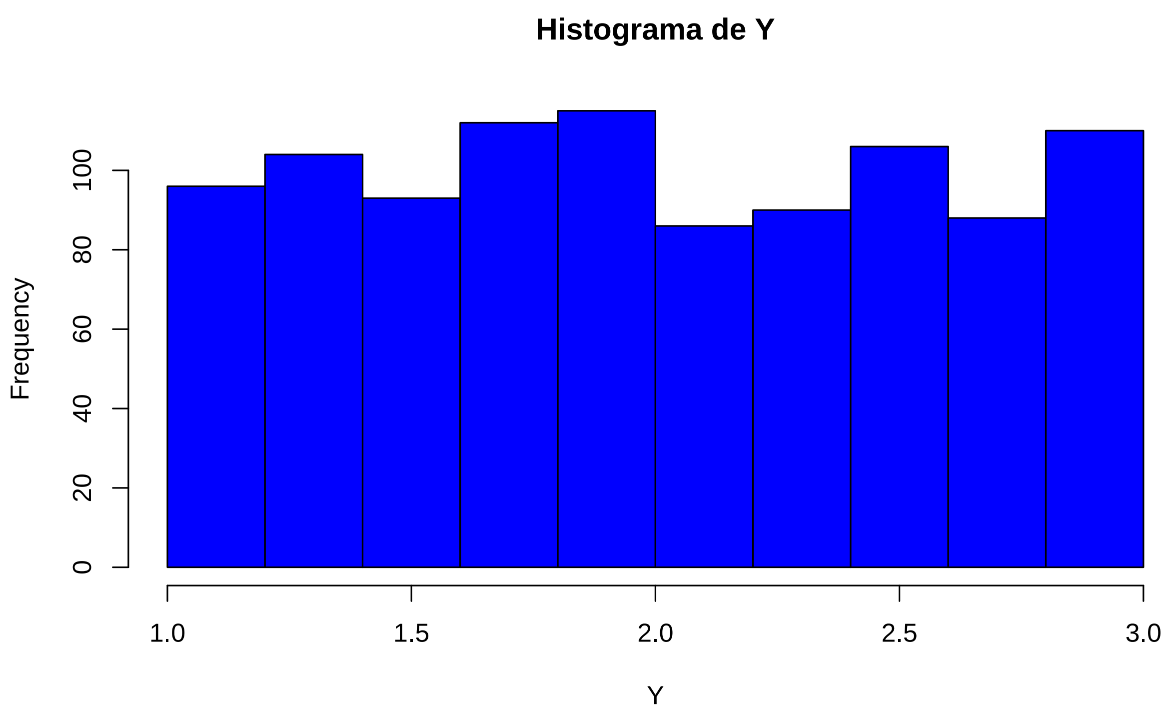



```r
hist(U, main='Histograma de U', c='red')
```



Ahora veamos a la variable $Y = \alpha U + \beta$ con $\alpha = 2$ y $\beta = 1$

```r
alpha <- 2
beta <- 1
Y <- alpha*U + beta
hist(Y, main='Histograma de Y', c='blue')
```

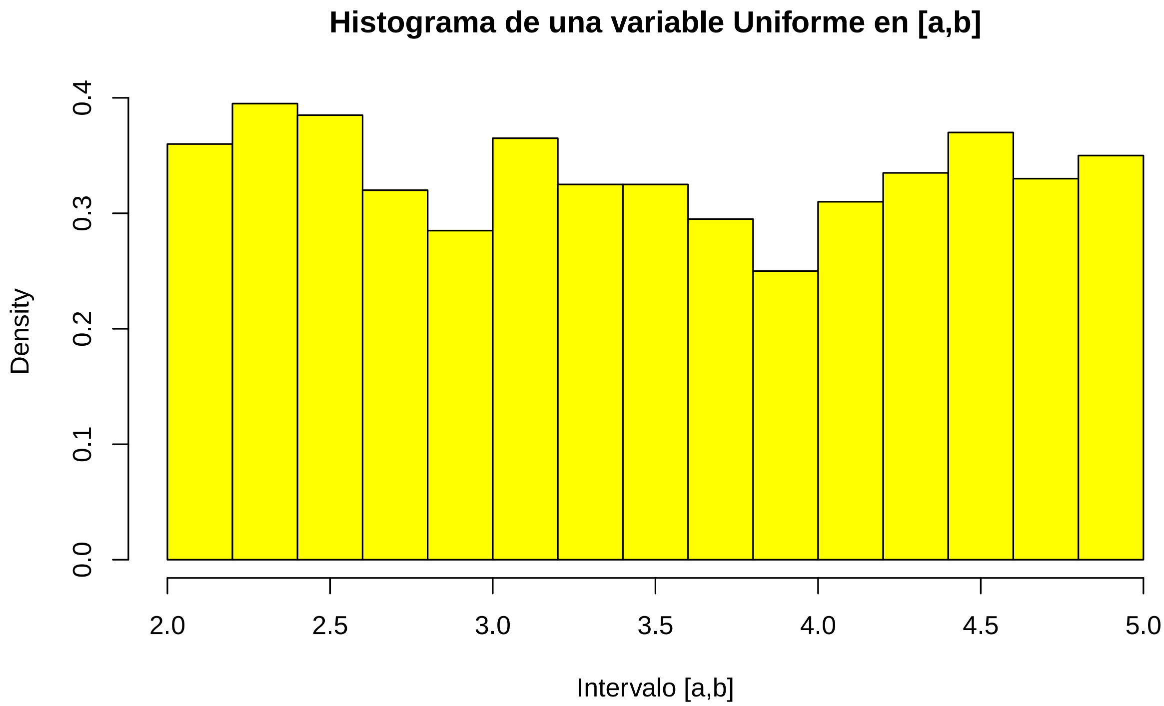


Efectivamente se cumple lo anteriormente explicado, se traslada β y el intervalo aumenta en α.

Observemos que si $0 < x < 1$ entonces $Y = 2(x)+1$ tomando la frontera del intervalo obtenemos $2(0)+1 = 1$ $\longrightarrow$ a=
analogamente con $b = 1$ obtenemos $2(1) + 1 = 3 \longrightarrow b = \alpha + \beta$ entonces despejando $\alpha = b - a$, por lo tanto $\beta = a$ y $\alpha = b - a$.

Veamos otra distribución que nos muestra lo antes dicho.

```r
N <- 1000
Ys <- rep(0,N)

a=2
b=5
#alpha=3
#beta=5
for (i in 1:N){
  U <- runif(1,0,1)
  Ys[i] <- (b-a)*U + a
}
hist(Ys, freq= F, xlab= 'Intervalo [a,b]',main='Histograma de una variable Uniforme en [a,b]', col='yel
```



b) A partir de $U1, U2, \ldots$ variables i.i.d., $Ui \sim U[0,1]$ construir variables aleatorias $Y1, Y2, \ldots$ i.i.d. tales que $Yi \sim U[a,b]$. ¿A que tiende en probabilidad la siguiente expresión?

$$\frac{\sum_{i=1}^{n} \frac{e^{\frac{-Y_i^2}{2}}}{\sqrt{2\pi}}}{n(b-a)}$$

Construyendo las Y_i analogamente como se hizo en el inciso anterior y calculando la integral como una suma de n variables uniformes, evaluando cada Y_i obtenemos lo siguiente

```
mi_unif <- function(a,b){
  U <- runif(1,0,1)
  return((b-a)*U+a)
}

mi_unif(2,5)
```

```
## [1] 4.615415
```

```
integral <- function(a,b,n){
  suma <- 0
  for (i in 1:n){
   Y <- mi_unif(a,b)
   suma <- suma + exp(-((Y^2)/2)) /sqrt(2*pi)
  }
  return( suma/n *(b-a) )
}
```

Ley de los grandes ńumeros

Sea $X_1, X_2, ...$ una sucesíon infinita de variables aleatorias independientes e idénticamente distribuidas con media finita μ. Entonces, cuando $n \mapsto \infty$

$$\frac{1}{n}\sum_{i=0}^{n} X_i \mapsto \mu$$

en donde la convergencia se verifica en el sentido casi seguro (ley fuerte) y tambíen en probabidad (ley débi).

Por la Ley de los Grande Números

$$\frac{1}{n}\sum_{i=0}^{n} Y_i \mapsto \mu$$

Así podemos concluir que la expresión tiende a la integral, es decir

$$\frac{\sum_{i=1}^{n} \frac{e^{\frac{-Y_i^2}{2}}}{\sqrt{2\pi}}}{n(b-a)} \mapsto \int_a^b \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx$$

c)Aproximar la integral por simulación usando el punto anterior para los siguientes valores de a y b, y para $n = 100$, 1000 y 50000.

***i)** $a = -1,96$ **y** $b = 1,96$. **ii)** $a = -2$ **y** $b = 1$. **iii)** $a = 0$ **y** $b = 2,34$.*

```
valores <- matrix(NA, ncol = 3, nrow = 3)

rownames(valores) <- c("n=100", "n=1000", "n=50000" )
colnames(valores) <- c("Intervalo 1", "Intervalo 2", "Intervalo 3")

a= -1.96
b= 1.96
valores[1,1] <- integral(a,b,100)
valores[2,1] <- integral(a,b,1000)
valores[3,1] <- integral(a,b,50000)


a=-2
b=1
valores[1,2] <- integral(a,b,100)
valores[2,2] <- integral(a,b,100)
valores[3,2] <- integral(a,b,50000)


a=0
b=2.34
valores[1,3] <- integral(a,b,100)
valores[2,3] <- integral(a,b,1000)
valores[3,3] <- integral(a,b,50000)

valores
```

```
##         Intervalo 1 Intervalo 2 Intervalo 3
## n=100     0.9262836   0.8117681   0.4596456
```

```
## n=1000     0.9427851   0.7682256   0.4829000
## n=50000    0.9502434   0.8162241   0.4908601
```

d) Comparar estos resultados con aquellos que obtendría a partir de la tabla de la función Φ o utilizando un comando provisto por el software.

```r
valores <- matrix(NA, ncol = 3, nrow = 4)

rownames(valores) <- c("n=100", "n=1000", "n=50000", "Software R")
colnames(valores) <- c("Intervalo 1", "Intervalo 2", "Intervalo 3")

a= -1.96
b= 1.96
valores[1,1] <- integral(a,b,100)
valores[2,1] <- integral(a,b,1000)
valores[3,1] <- integral(a,b,50000)
valores[4,1] <- pnorm(b) - pnorm(a)


a=-2
b=1
valores[1,2] <- integral(a,b,100)
valores[2,2] <- integral(a,b,100)
valores[3,2] <- integral(a,b,50000)
valores[4,2] <- pnorm(b) - pnorm(a)


a=0
b=2.34
valores[1,3] <- integral(a,b,100)
valores[2,3] <- integral(a,b,1000)
valores[3,3] <- integral(a,b,50000)
valores[4,3] <- pnorm(b) - pnorm(a)

valores
```

```
##            Intervalo 1 Intervalo 2 Intervalo 3
## n=100        0.9407710   0.8331806   0.4710875
## n=1000       0.9490939   0.8421380   0.4861337
## n=50000      0.9489760   0.8202476   0.4903811
## Software R   0.9500042   0.8185946   0.4903581
```

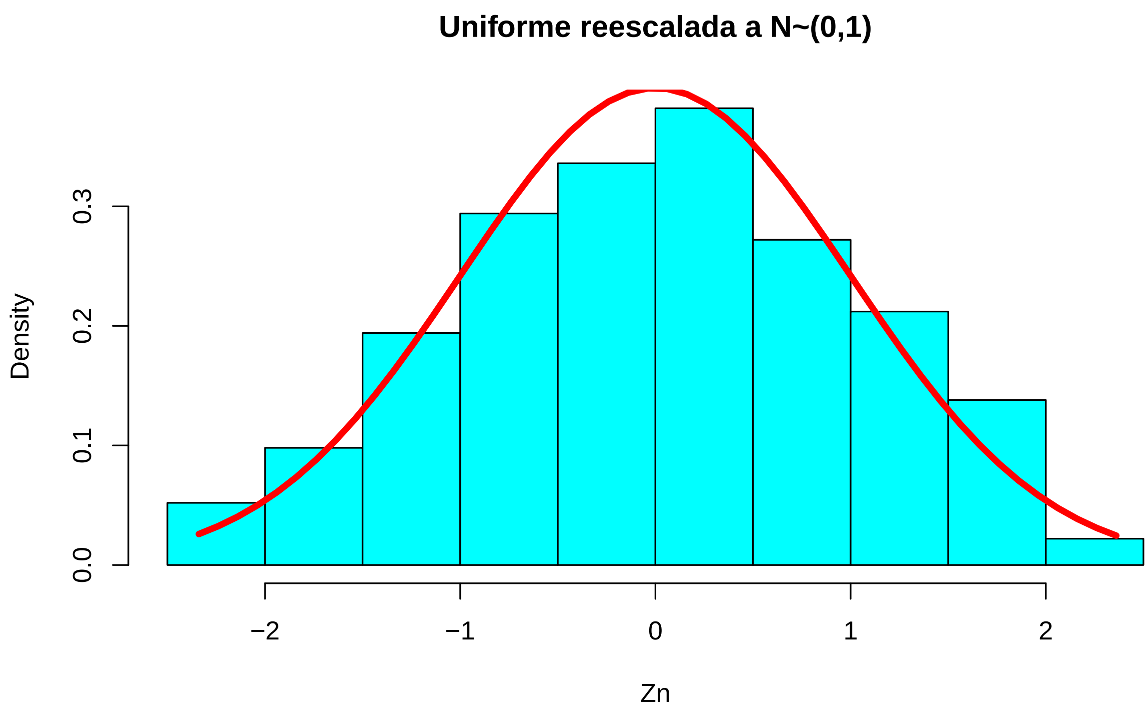
Como podemos observar entre mayor sea n mayor es la exactitud con la que se aproxima a la integral.

2. En este ejercicio estudiaremos la distribución del promedio de variables independientes e idénticamente distribuidas reescaladas según la desviación estandar, y a través de los histogramas correspondientes analizaremos el comportamiento de estas distribuciones a medida que promediamos un número creciente de variables aleatorias. Para ello generaremos una muestra de variables aleatorias con una distribución dada y luego calcularemos el promedio de cada muestra. Replicaremos ésto mil veces, es decir, generaremos una muestra aleatoria de la variable X de tamaño 1000. Observe que, en principio, desconocemos la distribución de X. A partir de todas las replicaciones realizaremos un histograma para los promedios obtenidos para obtener una aproximación de la densidad o la función de probabilidad de X.

a) Considerar dos variables aleatorias X_1 y X_2 independientes con distribución $U(-1,1)$ y el promedio reescalado de ambas, es decir

$$Z_2 = \frac{\sqrt{3}(x_1 + x_2)}{\sqrt{2}}$$

```
Uniforme_reescalada <- function(n,m,a,b){
  media <- (a+b)/2
  var <- (b-a)^2/12
  Sn <- vector('numeric',length = m)

  for (i in 1:m){
    Sn[i] <- sum(runif(n,a,b))
  }
  Zn <- (Sn -n*media) / (sqrt(var)*sqrt(n) )

  hist(Zn, c = 'cyan',main= 'Uniforme reescalada a N~(0,1)' ,freq = F)
  x <- seq(min(Zn),max(Zn),by=0.1)
  lines(x, dnorm(x, mean=0,sd=1),lwd=4, col = 'red' )
}
Uniforme_reescalada(2,1000,-1,1)
```

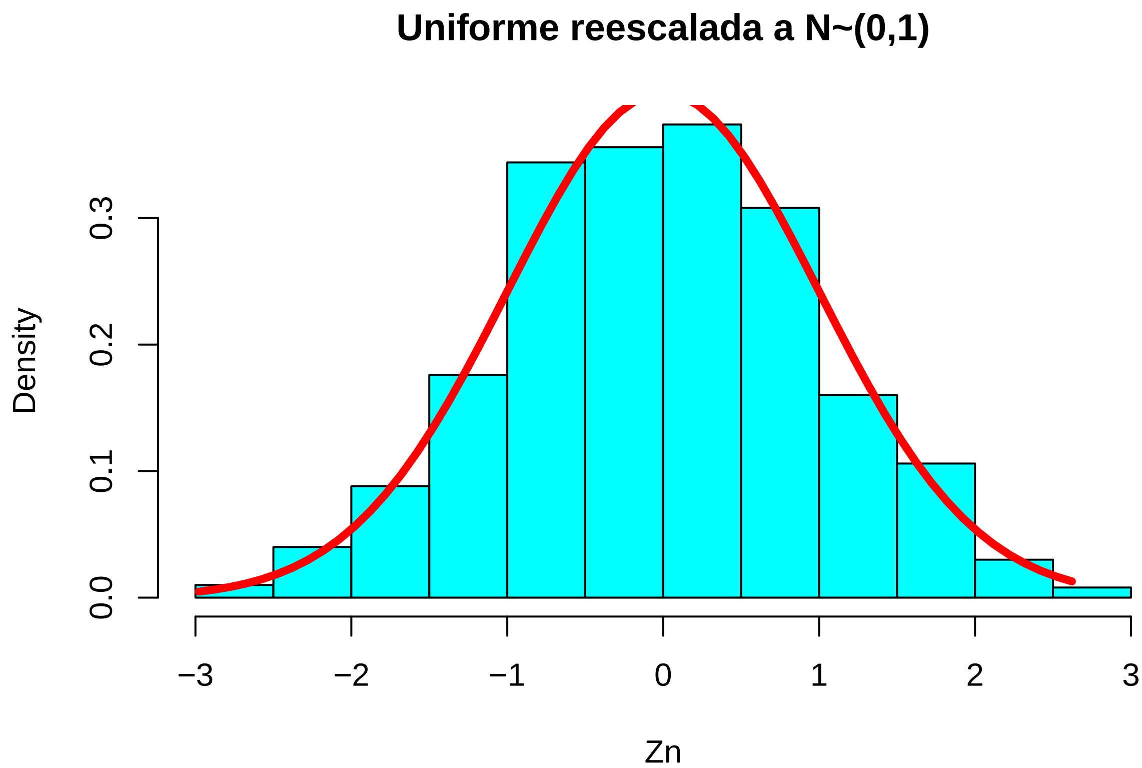


b)Aumentemos a cinco las variables promediadas. Considerar ahora 5 ***variables aleatorias uniformes independientes, es decir*** X_1***,***X_2***,...,***X_5 ***i.i.d. con*** $X_i \sim U(-1,1)$ ***y definir***

$$Z_5 = \frac{\sqrt{3}\sum_{i=1}^{5}}{\sqrt{5}}$$

Generando muestras de cinco variables aleatorias con distribución $U(-1,1)$ ***computar*** Z_5***. Repetir*** 1000 ***veces y realizar un histograma para los valores obtenidos. Comparar con el histograma anterior. ¿Qué se observa?***

```
Uniforme_reescalada(5,1000,-1,1)
```



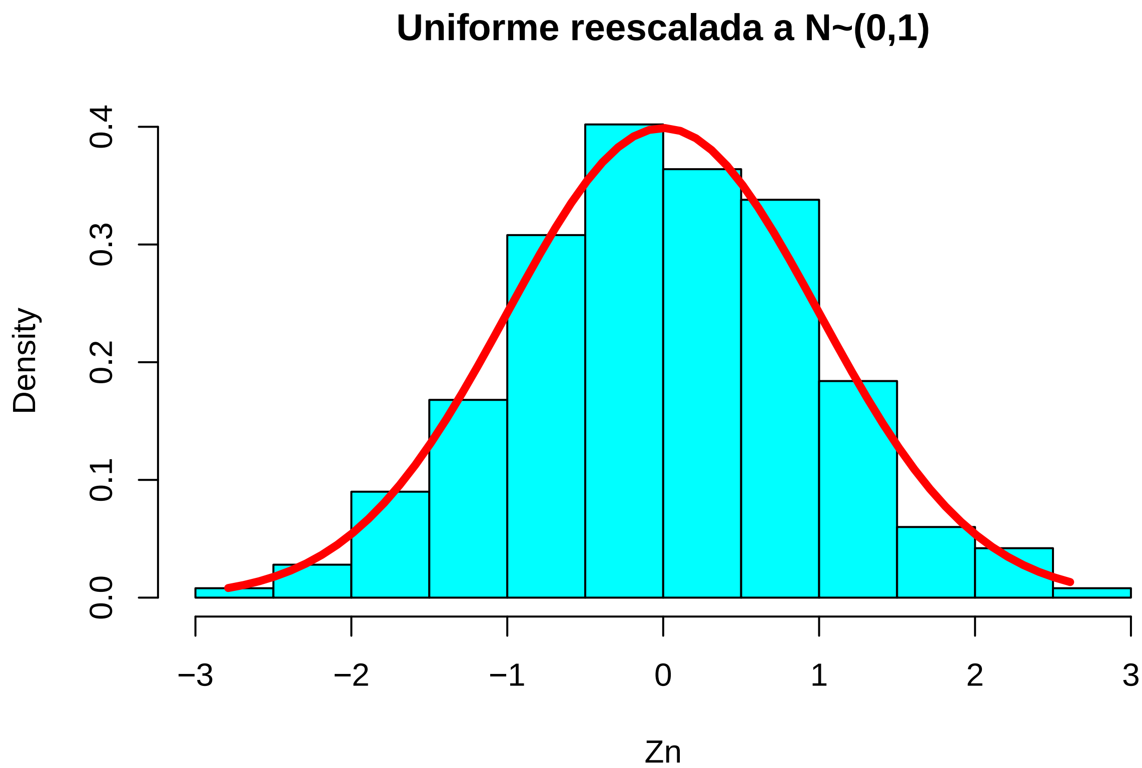
Observamos una mejor aproximación respecto a la simulación anterior, esto se debe a que aumentaron las variables aleatorias.

c)Aumentemos aún más la cantidad de variables promediadas. Generando muestras de* 30 *variables aleatorias con distribución* $U(-1,1)$ *y computar

$$Z_{30} = \frac{\sqrt{3}\sum_{i=1}^{30}}{\sqrt{30}}$$

¿Qué se observa?

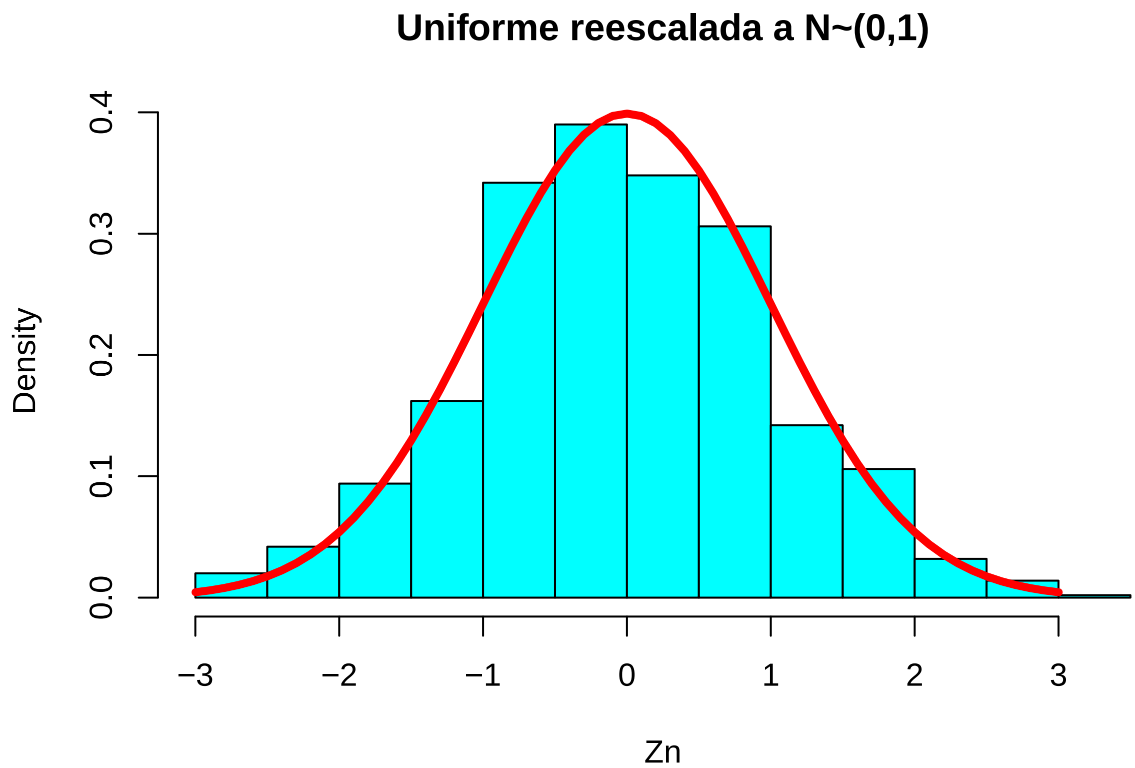
```
Uniforme_reescalada(30,1000,-1,1)
```



Podemos notar que entre mayor es n, mejor es la aproximación a la normal stándar, que es lo que nos dice el Teorema del Límite Central.

d)Idem anterior generando muestras de 200 ***variables aleatorias. ¿Qué pasa si se aumenta el tamaño de la muestra?***

```
Uniforme_reescalada(200,1000,-1,1)
```



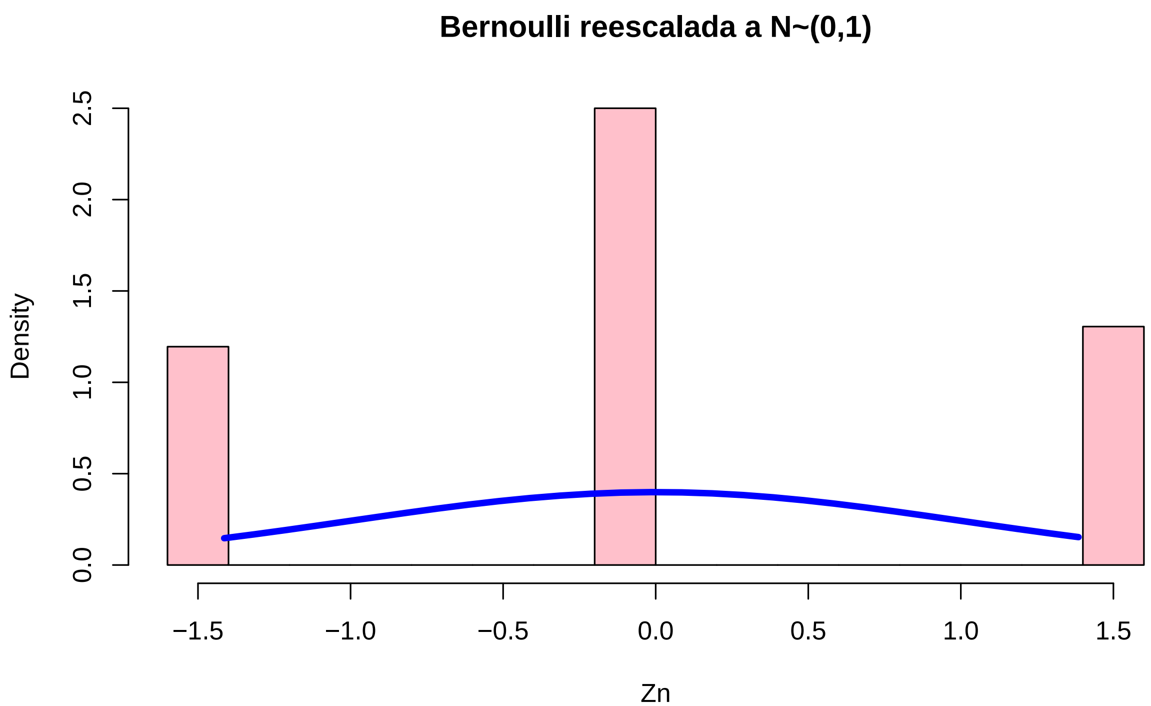
Observamos que entre mayor sea la muestra de variables aleatorias, mejor es la aproximación a la normal stándar, y aunque 200 no es un número muy grande, ya nos da una buena aproximación.

e)Repetir los ítems a) a d) generando ahora variables con distribución $Be(p)$ ***para*** $p = 0,5$***;*** $0,01$ ***y*** $0,0001$***; realizando el histograma para***

$$Z_n = \frac{\sum_{i=1}^{n}(X_i - p)}{\sqrt{p(1-p)n}}$$

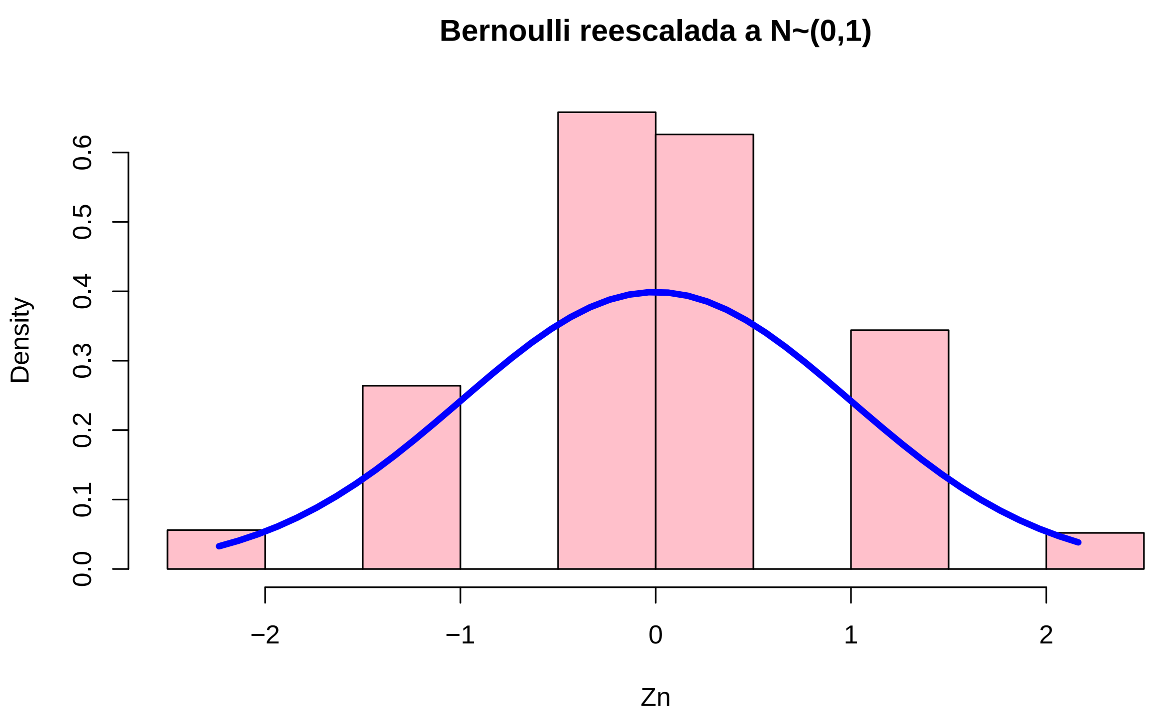
Para $n = 2$ ***y*** $p = 0.5$ ***obtenemos***

```
Bernoulli_reescalada <- function(n,m,p){
  media <- p
  var <- p - p^2
  Sn <- vector('numeric',length = m)

  for (i in 1:m){
    Sn[i] <- sum(rbinom(n,1,0.5))
  }
  Zn <- (Sn -n*media) / (sqrt(var)*sqrt(n) )

  hist(Zn, c = 'pink',main= 'Bernoulli reescalada a N~(0,1)' ,freq = F)
  x <- seq(min(Zn),max(Zn),by=0.1)
  lines(x, dnorm(x, mean=0,sd=1),lwd=4, col = 'blue' )
}
Bernoulli_reescalada(2,1000,0.5)
```



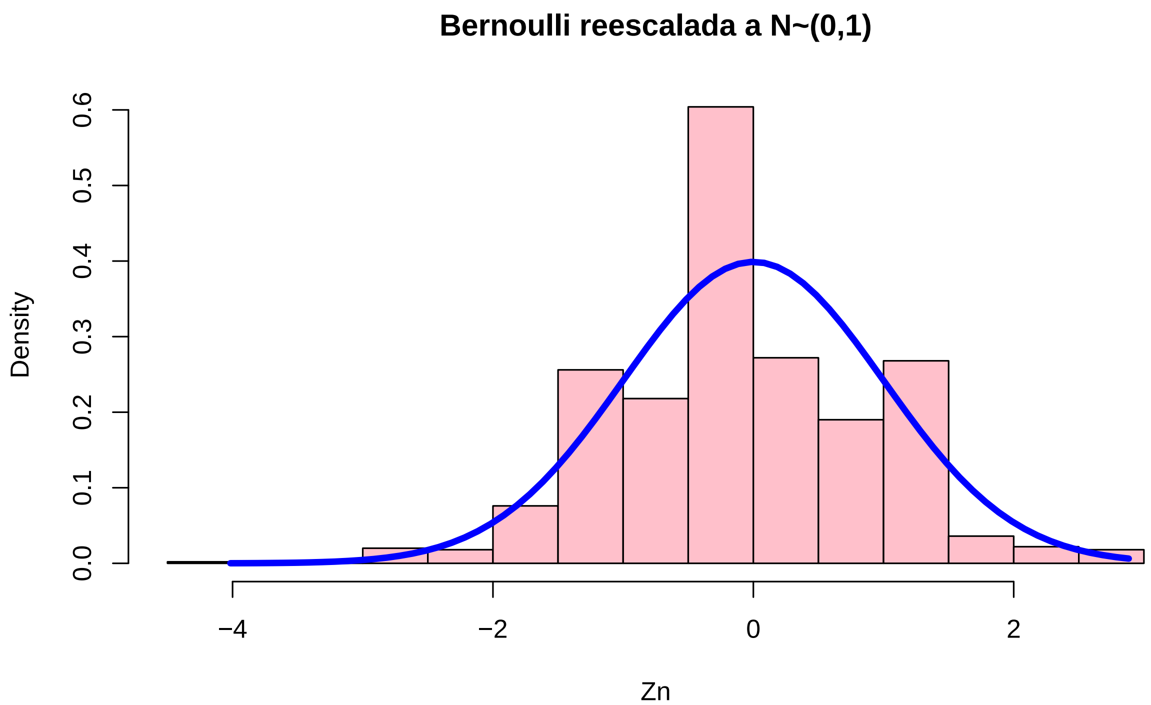
Para $n = 5$ ***y*** $p = 0.5$ ***obtenemos***

```
Bernoulli_reescalada(5,1000,0.5)
```



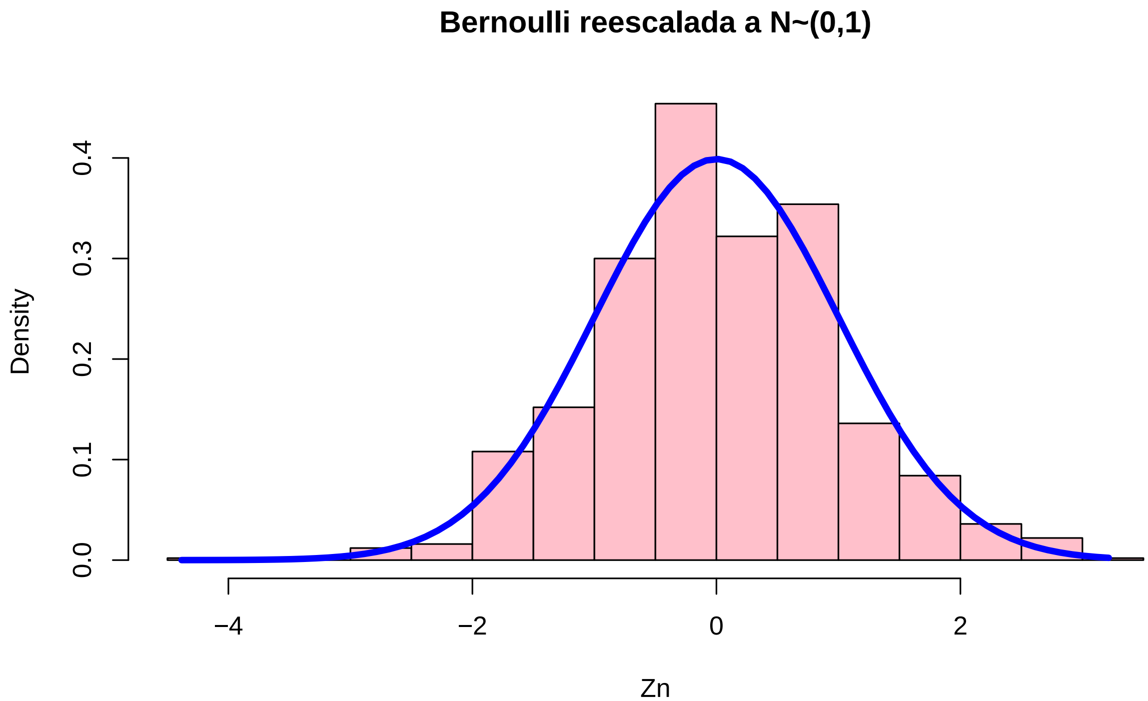
Para $n = 30$ ***y*** $p = 0.5$ ***obtenemos***

```
Bernoulli_reescalada(30,1000,0.5)
```



Para $n = 200$ ***y*** $p = 0.5$ ***obtenemos***

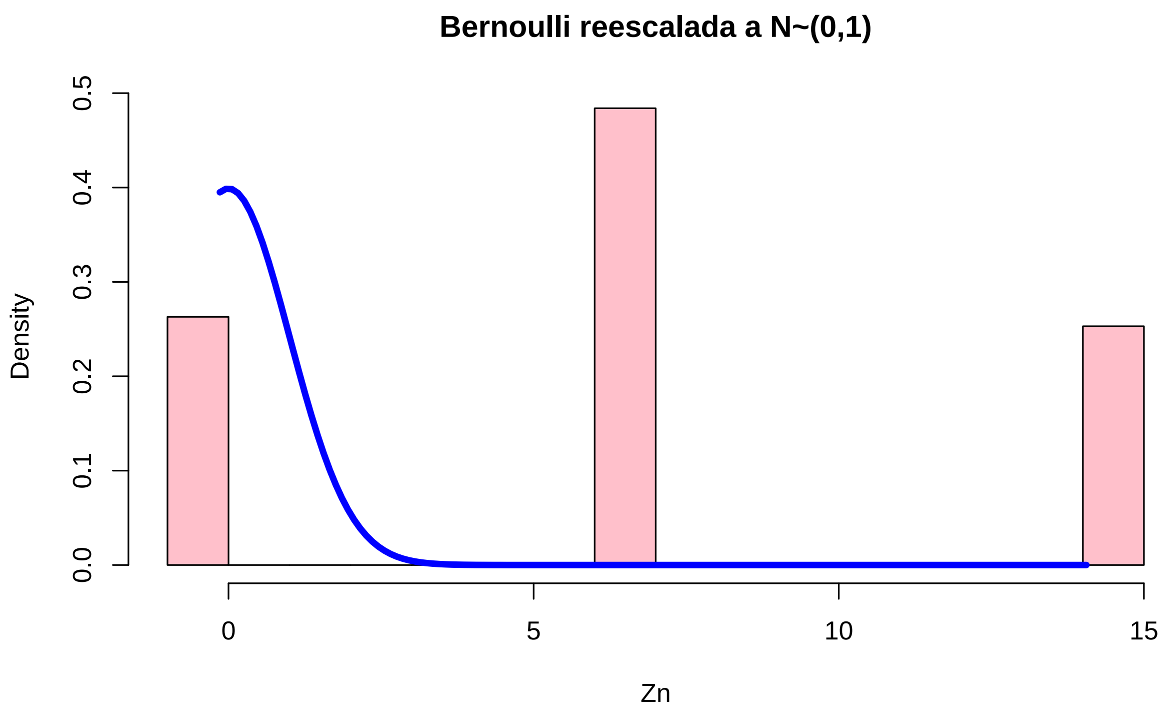
```
Bernoulli_reescalada(200,1000,0.5)
```



Ahora haremos la simulación de todos los casos anteriores con $p = 0.01$

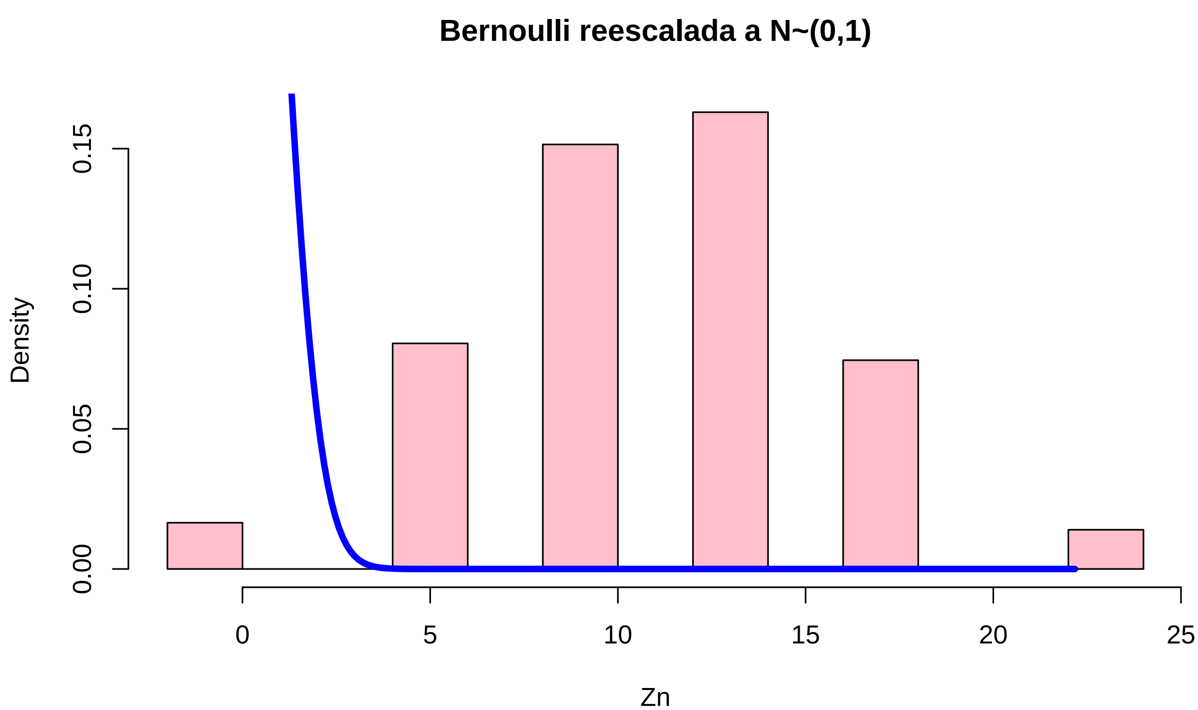
Para $n = 2$

```
Bernoulli_reescalada(2,1000,0.01)
```



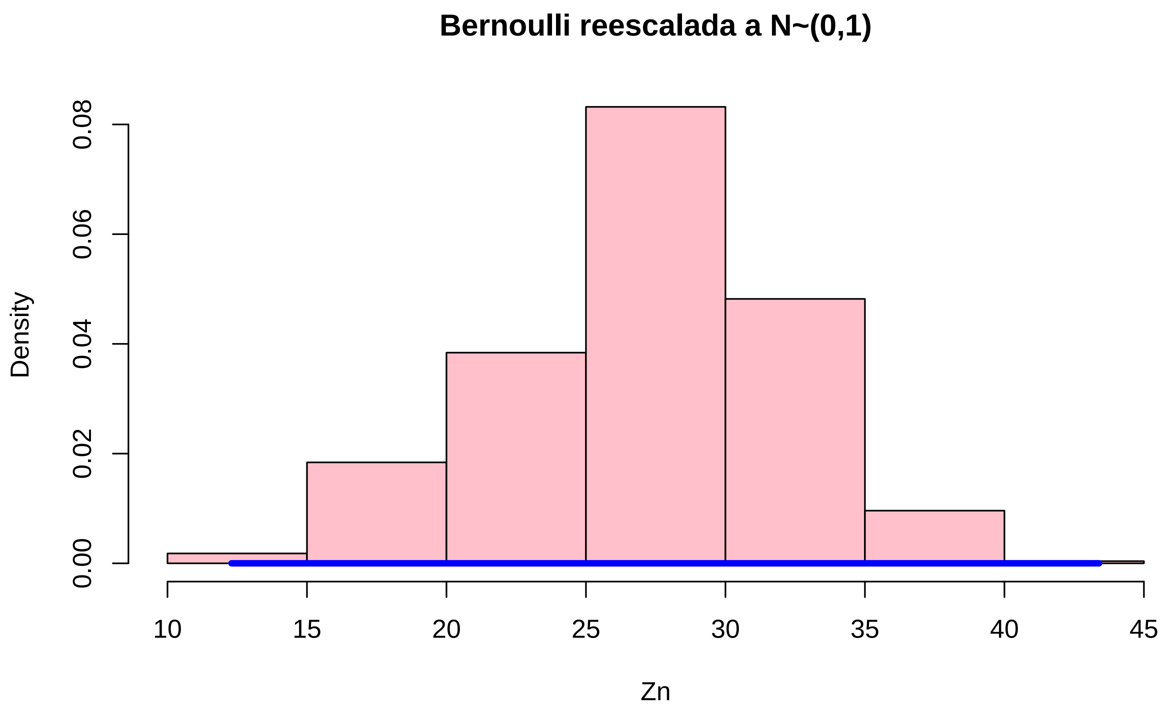
Para $n = 5$

```
Bernoulli_reescalada(5,1000,0.01)
```



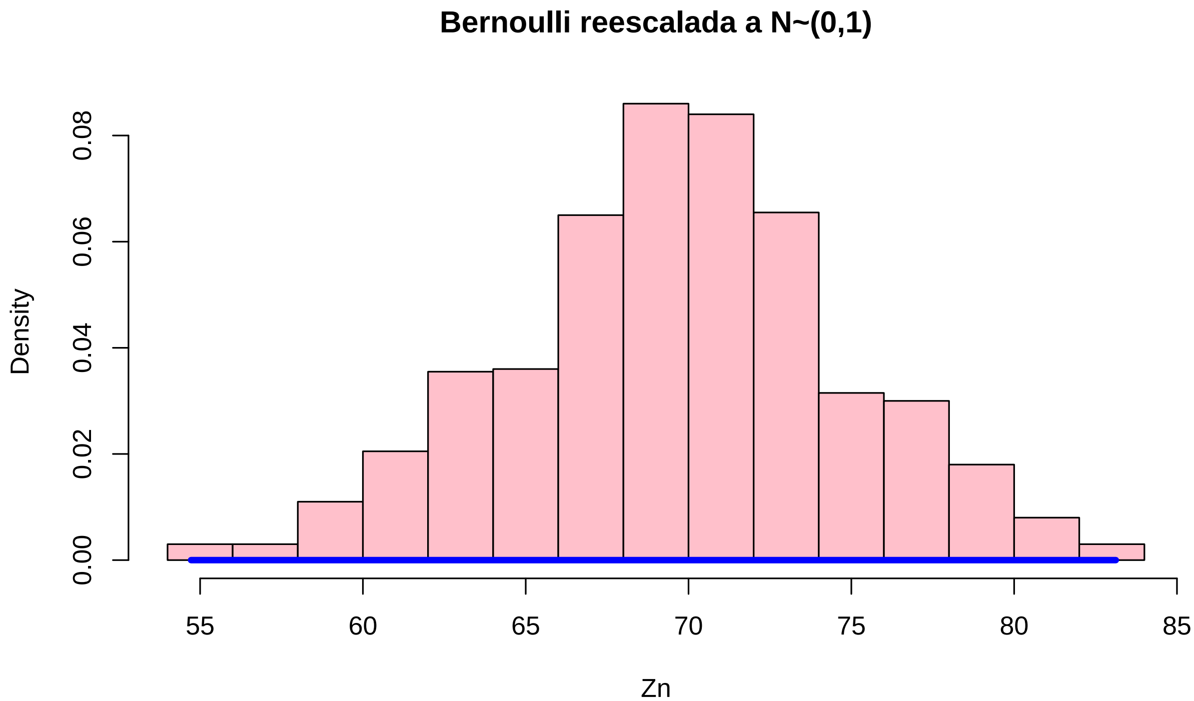
Para $n = 30$

```
Bernoulli_reescalada(30,1000,0.01)
```



Para $n = 200$

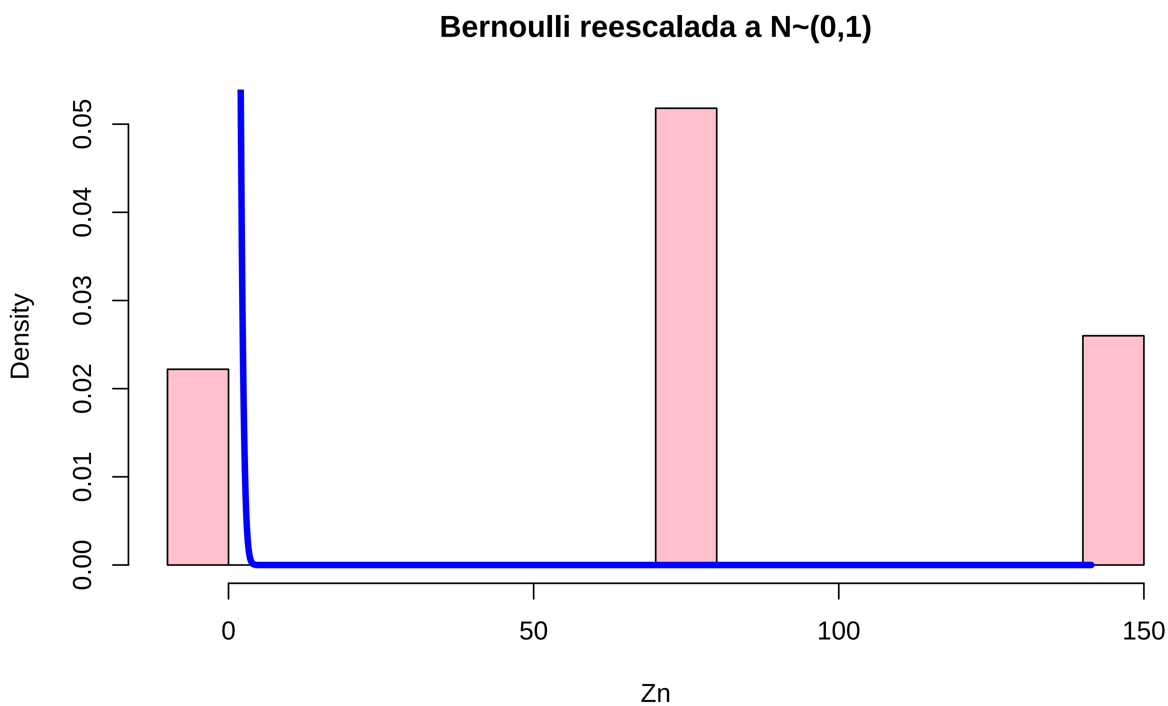
```
Bernoulli_reescalada(200,1000,0.01)
```



Ahora haremos la simulación de todos los casos anteriores con $p = 0.0001$

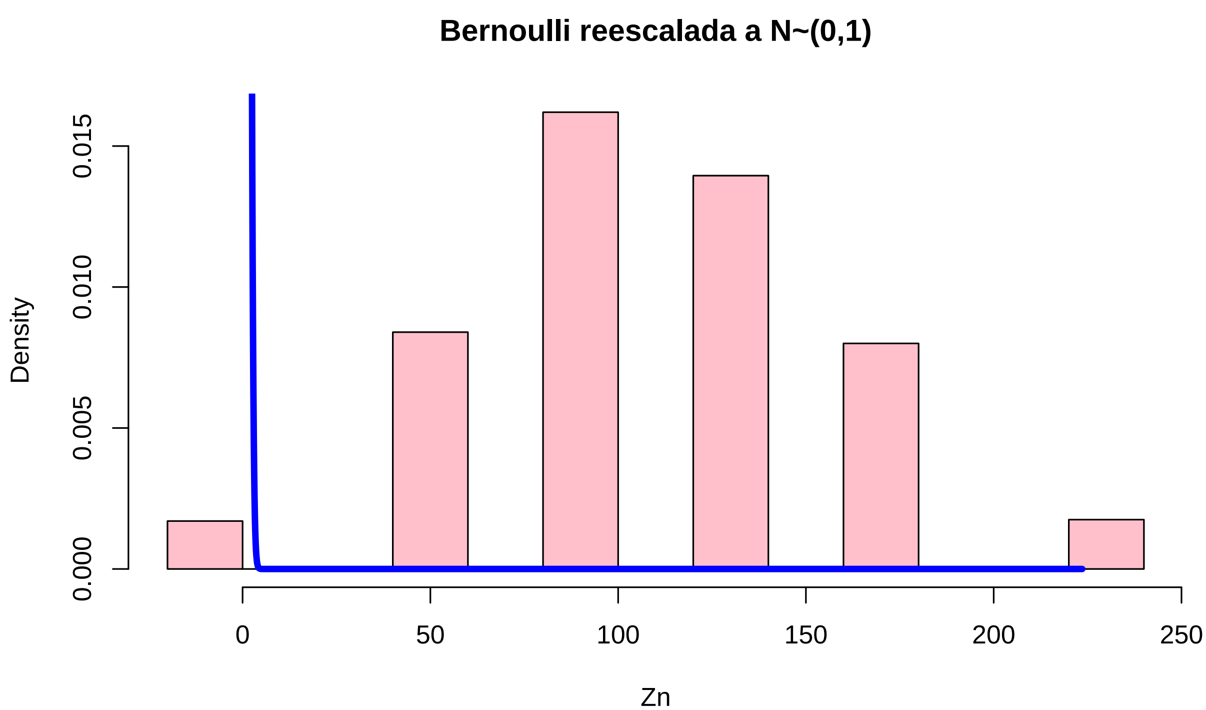
Para $n = 2$

```
Bernoulli_reescalada(2,1000,0.0001)
```



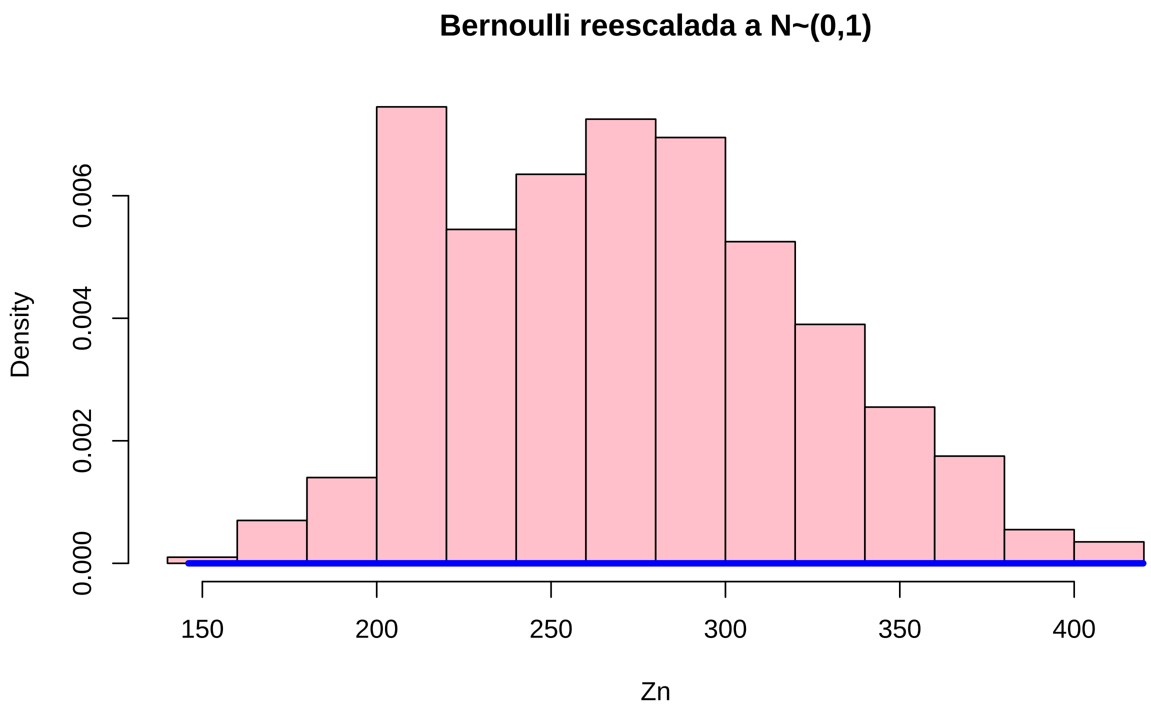
Para $n = 5$

```
Bernoulli_reescalada(5,1000,0.0001)
```



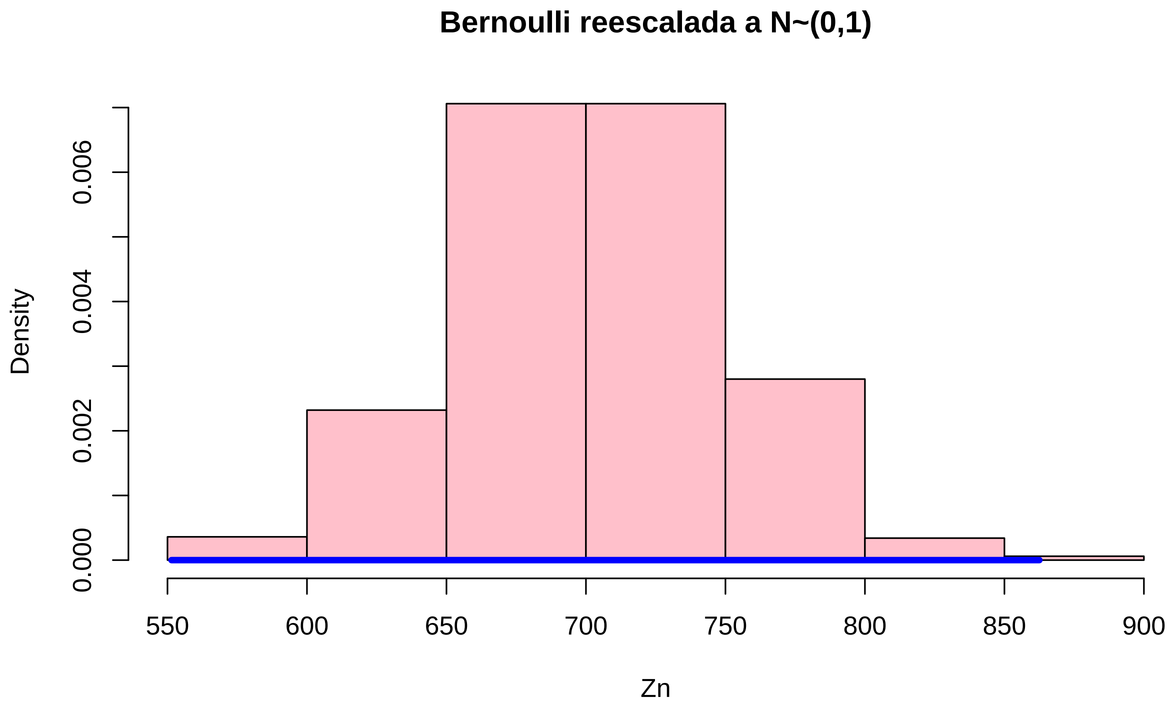
Para $n = 30$

```
Bernoulli_reescalada(30,1000,0.0001)
```



Para $n = 200$

```
Bernoulli_reescalada(200,1000,0.0001)
```



Podemos observar que para $p = 0.5$ entre mayor es n mejor es la aproximación a la normal stándar, pero para $p = 0.01$ y $p = 0.0001$ las gráficas se asemejan a una campana de Gauss pero mas pequeña, esto debido al valor del párametro p, debido a como está definida la variable aleatoria Bernoulli, a su media, y a su varianza.

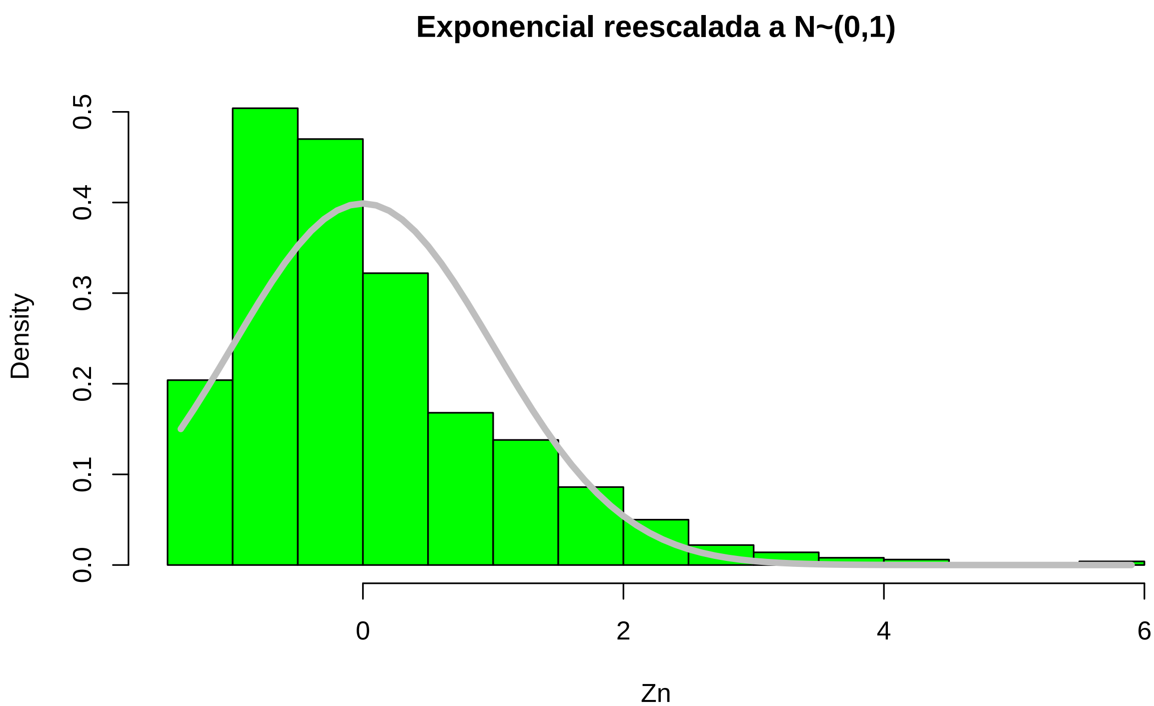
f)Repetir los ítems a) a d) generando ahora variables con distribución $Exp(\lambda)$ para $\lambda = 1$; $0,5$ y $0,1$; realizando el histograma para

$$Z_n = \frac{\sum_{i=1}^{n}(X_i - \frac{1}{\lambda})}{\frac{\sqrt{n}}{\lambda}}$$

Haremos la simulación de todos los casos anteriores con $\lambda = 1$

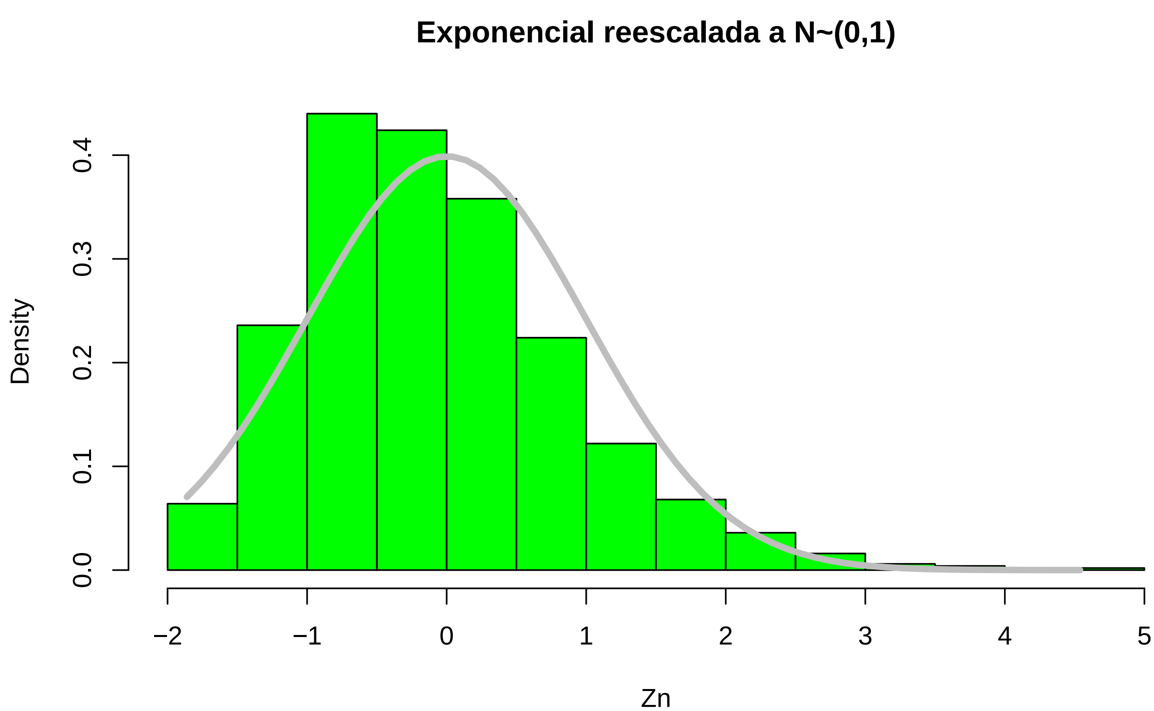
Para $n = 2$

```
Exponencial_reescalada <- function(n,m,lambda){
  media <- 1 / lambda
  var <- 1 / lambda
  Sn <- vector('numeric',length = m)

  for (i in 1:m){
    Sn[i] <- sum(rexp(n, lambda))
  }
  Zn <- (Sn -n*media) / (sqrt(var)*sqrt(n) )

  hist(Zn, c = 'green',main= 'Exponencial reescalada a N~(0,1)' ,freq = F)
  x <- seq(min(Zn),max(Zn),by=0.1)
  lines(x, dnorm(x, mean=0,sd=1),lwd=4, col = 'gray' )
}
Exponencial_reescalada(2,1000,1)
```



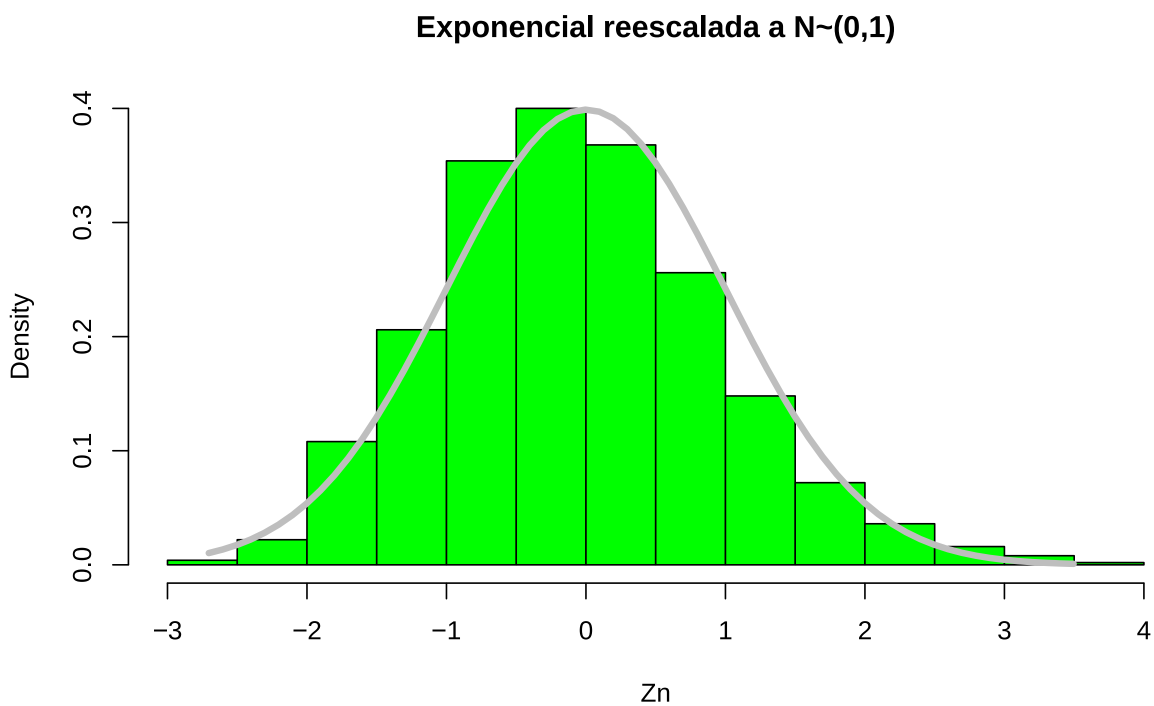
Para $n = 5$

```
Exponencial_reescalada(5,1000,1)
```



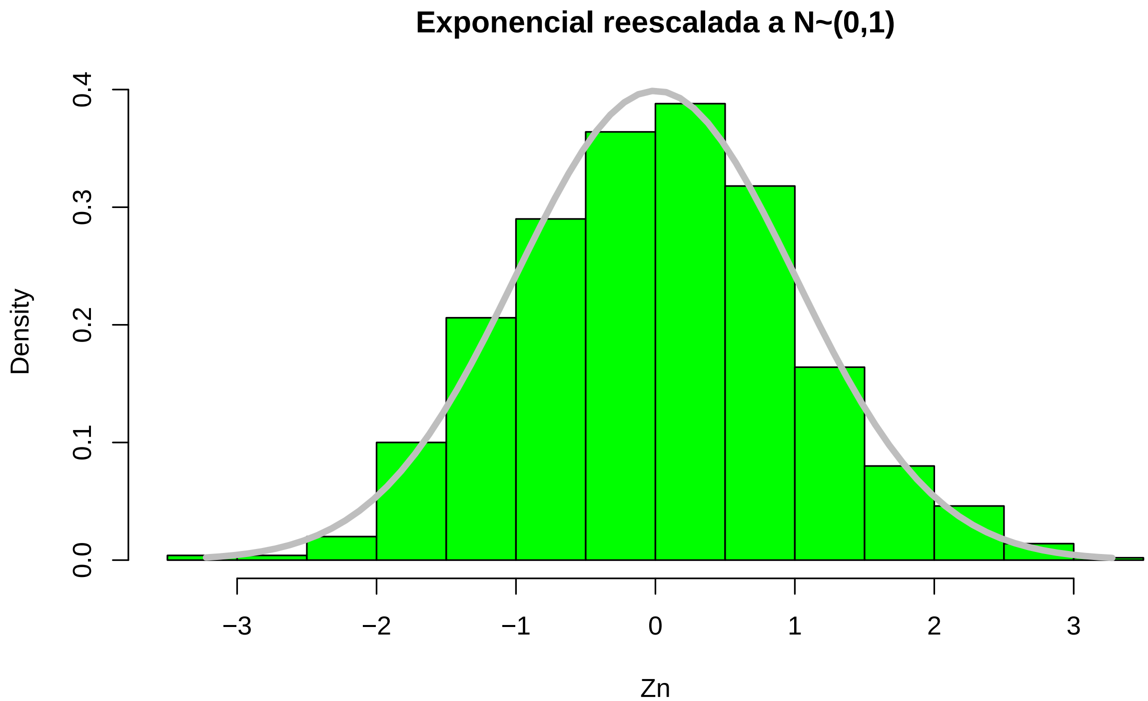
Para $n = 30$

```
Exponencial_reescalada(30,1000,1)
```



Para $n = 200$

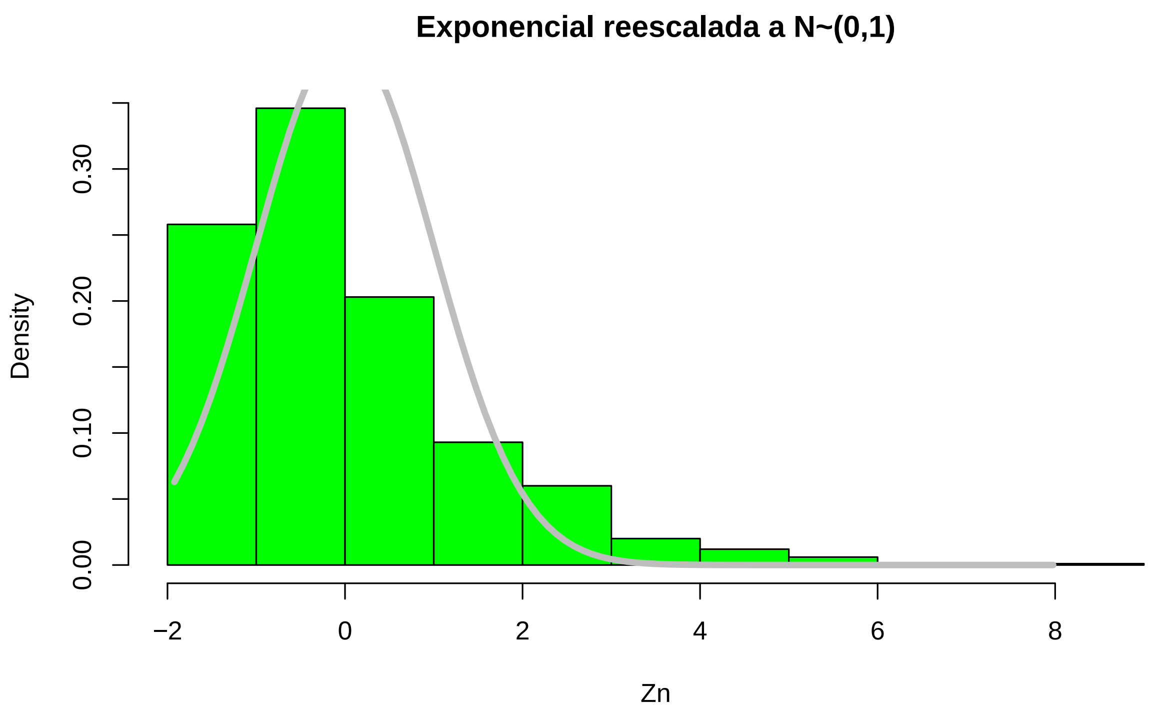
```
Exponencial_reescalada(200,1000,1)
```



Haremos la simulación de todos los casos anteriores con $\lambda = 0.5$

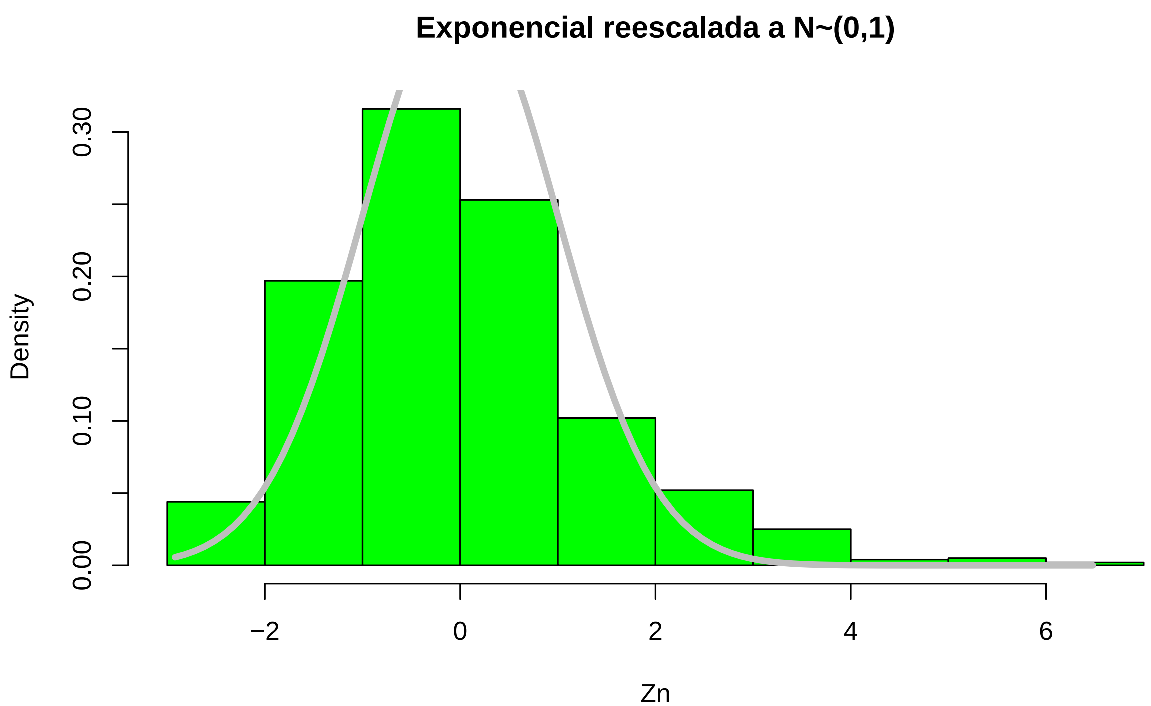
Para $n = 2$

```
Exponencial_reescalada(2,1000,0.5)
```



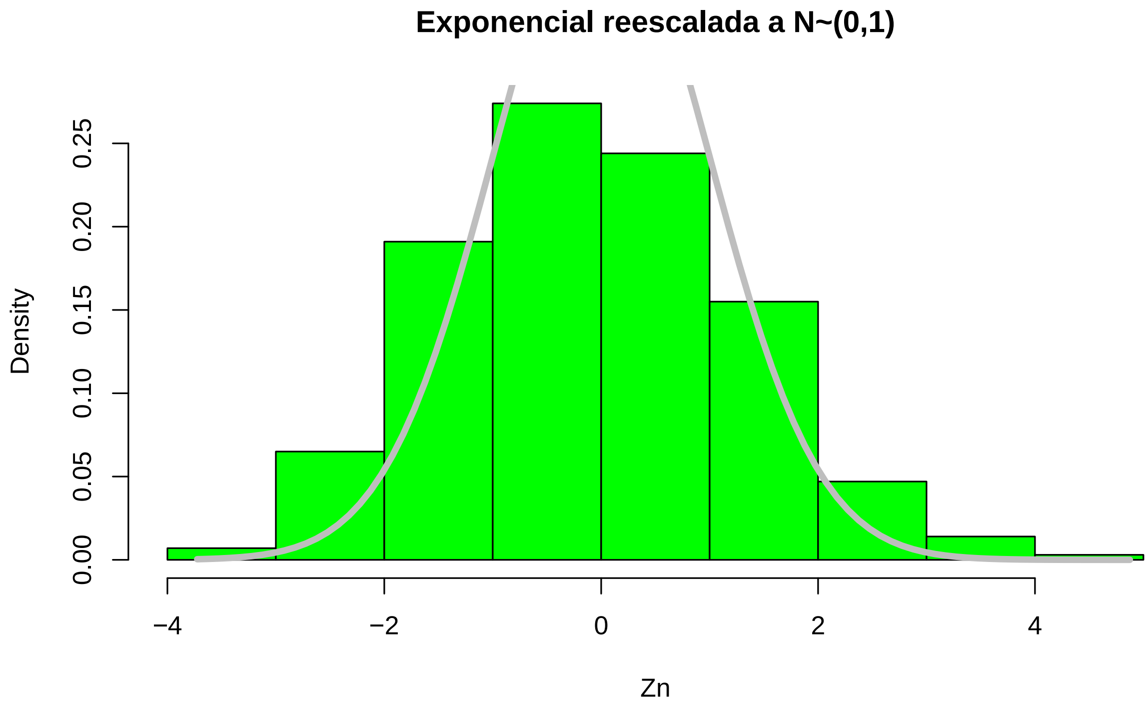
Para $n = 5$

```
Exponencial_reescalada(5,1000,0.5)
```



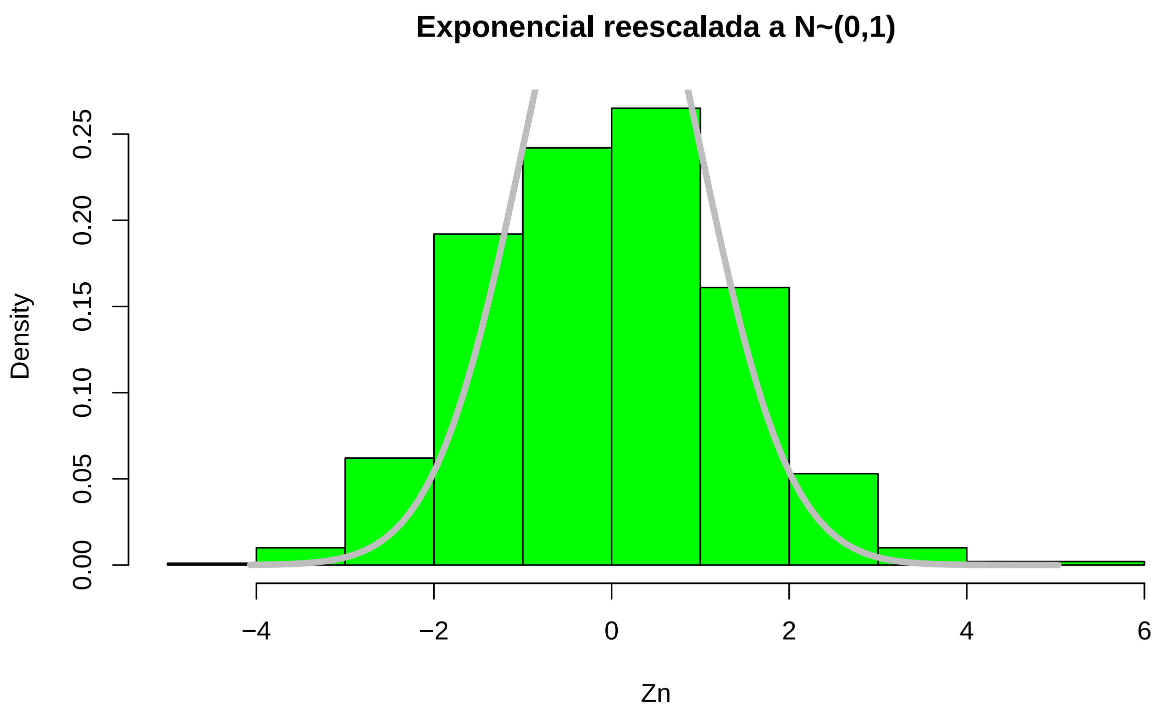
Para $n = 30$

```
Exponencial_reescalada(30,1000,0.5)
```



Para $n = 200$

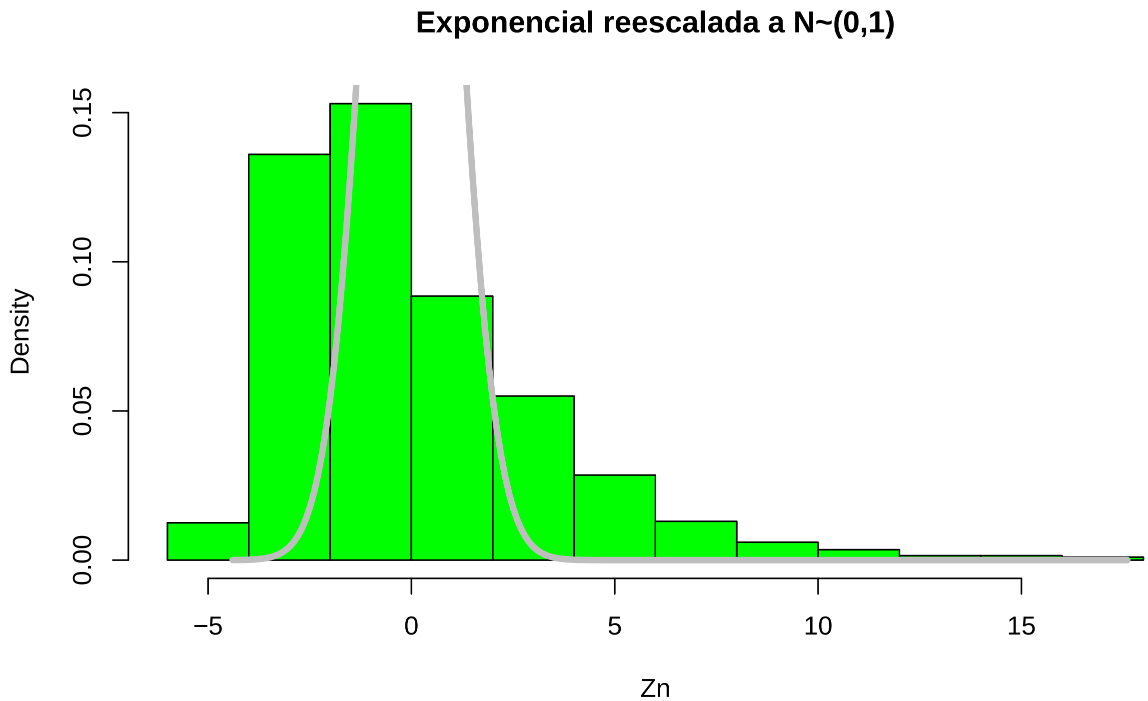
```
Exponencial_reescalada(200,1000,0.5)
```



Para la simulación de todos los casos anteriores con $\lambda = 0.1$

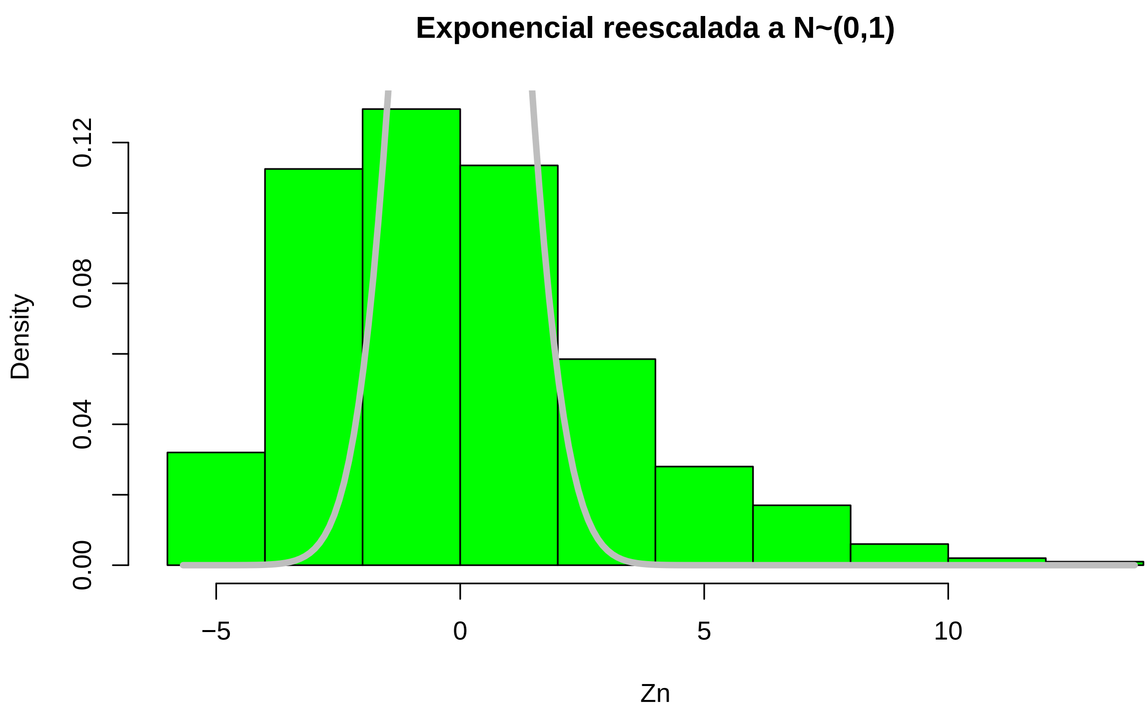
Para $n = 2$

```
Exponencial_reescalada(2,1000,0.1)
```



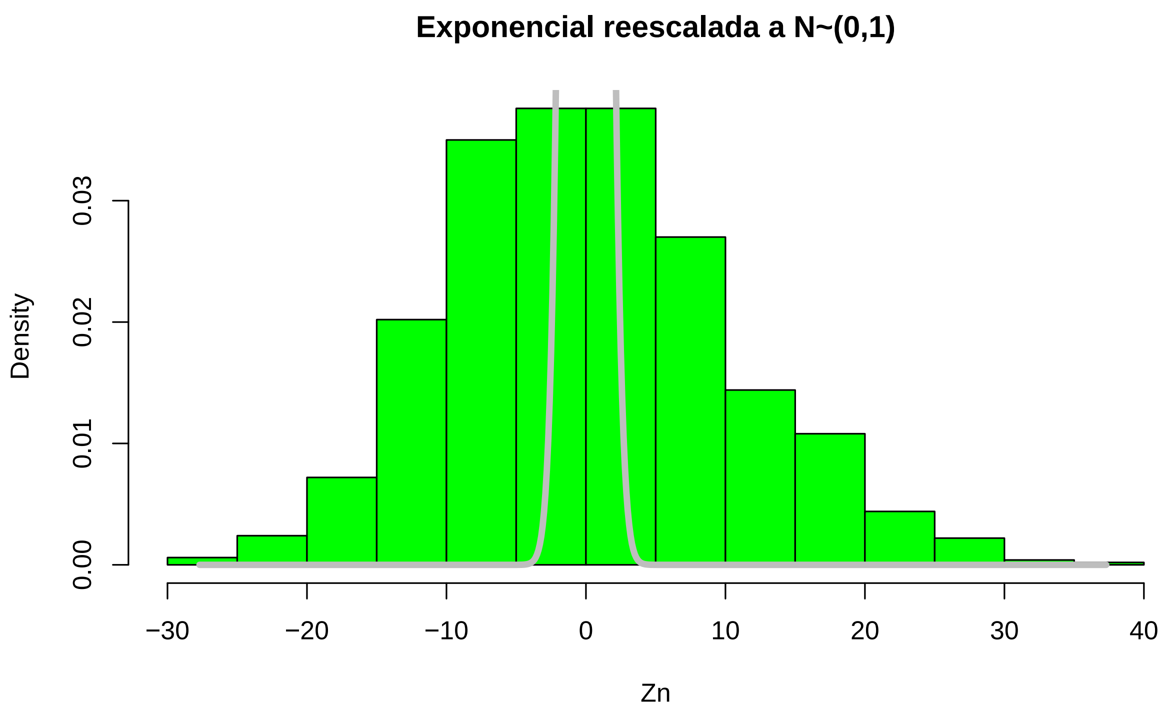
Para $n = 5$

```
Exponencial_reescalada(5,1000,0.1)
```



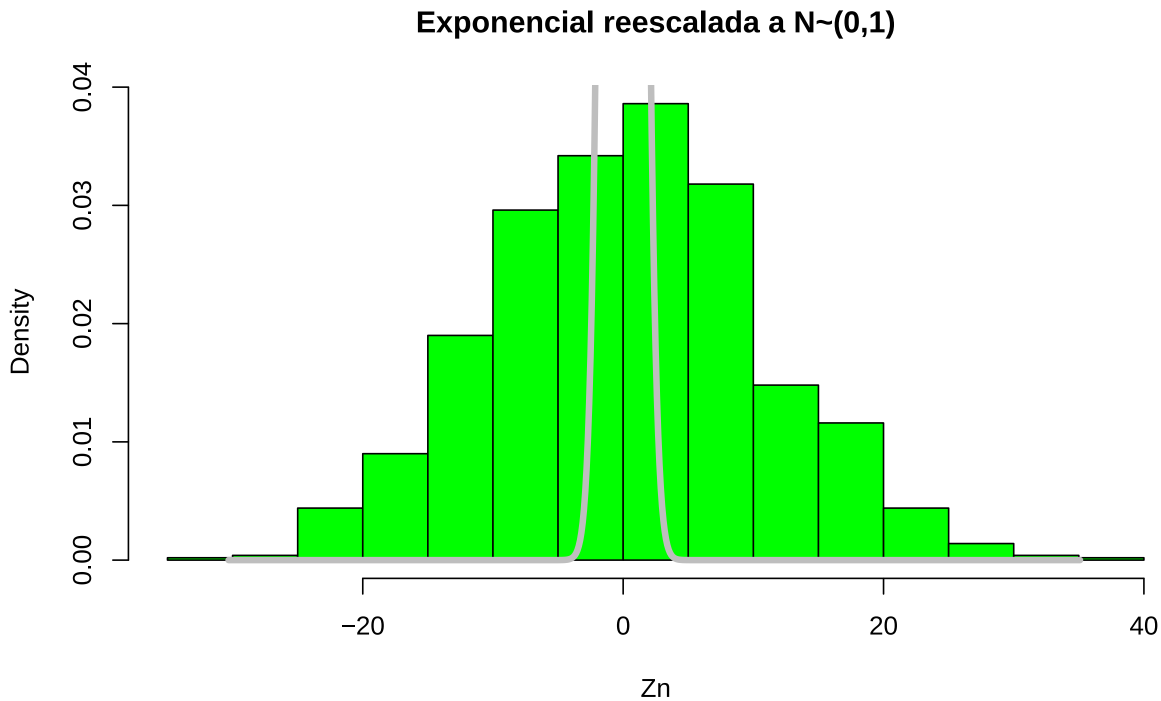
Para $n = 30$

```r
Exponencial_reescalada(30,1000,0.01)
```



Para $n = 200$

```r
Exponencial_reescalada(200,1000,0.01)
```



¿Cuáles son sus conclusiones?

Se puede observar que para $\lambda = 1$ cuando n crece mejor es la aproximación a la normal stándar, pero para $\lambda = 0.5$ y $\lambda = 0.1$ cuando n crece la gráfica de la sucesión de variables sobrepasa la gráfica de la distribucion normal stándar, debido a como está definida su media y varianza.

Podemos ver como se verifican mas allá de la teoría El Teorema Central Del Límite y La Ley de los Grandes Números, con los ejercicios que hemos realizado.