Enunciado del ejercicio 2

En este ejercicio se propone simular 50000 observaciones de una variable aleatoria Z con distribución N(0,1), a partir de los valores simulados de una variable uniforme en (0,1), utilizando el método de Box-Muller.

1. Describir los pasos del procedimiento que le permitiría obtener los valores simulados de la N(0,1) e implementar el procedimiento en un algoritmo para la simulación.

2. Explicar la manera en que un histograma de frecuencias nos ayuda a representar la función de densidad de una variable continua. Representar el histograma de los valores simulados ¿Qué se puede decir de la forma del histograma obtenido?

3. Utilizar los resultados de la simulación para aproximar la probabilidad $P(Z > 1.645)$

Para realizar esta simulación utilizando el método de Box-Muller, deberemos recrear dos variables aleatorias de distribución uniforme, U_1 y U_2 para la generación de números aleatorios con la función runif(). Los valores de dichos números estarán comprendidos en el intervalo de 0 a 1.

A partir de estas dos variables aleatorias, generaremos X e Y como vectores de números aleatorios mediante las fórmulas de Box-Muller

$$X = \sqrt{-2\ lnU_2} \cos 2\pi U_1$$

$$Y = \sqrt{-2\ lnU_2} \sin 2\pi U_1$$

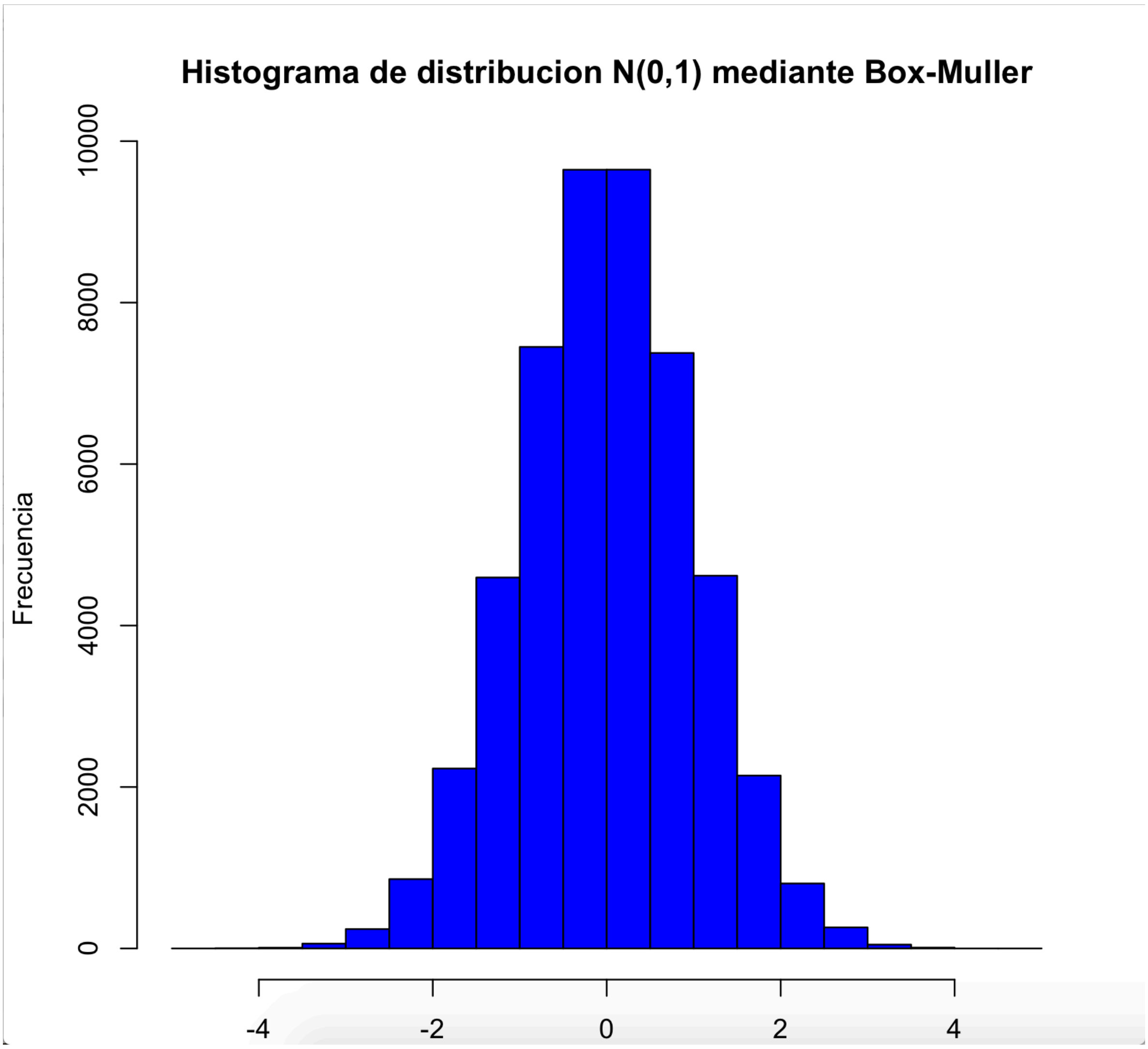
Para realizar la simulación he decidido usar el lenguaje de programación R, debido a su facilidad de lectura y al uso de sus funciones para la simulación.

En primer lugar se crea un parámetro para número de observaciones a realizar, que serán 25000 por variable aleatoria. Después se define la función que devolverá los valores de la simulación. Esta función repite un bucle por cada observación, almacenando en el array bidimensional el par de números aleatorios. Dicha variable simulará las variables aleatorias uniformes, que transformaremos con las fórmulas antes mencionadas de Box-Muller en la variable z, que es una matriz de 25000 filas y 2 columnas. La matriz z simulará las variables X e Y, almacenando un total de 50000 observaciones entre ambas. Una vez realizado esto se dispone el algoritmo a realizar al representación gráfica del histograma.

```r
#número de observaciones por variable (ambas suman 50000)
obs <- 25000;

#cuantil a aproximar
cuantil <- 1.645;

#funcion para la simulación
simul <- function(obs){
    z <-matrix(data=0, nr=obs, nc=2)

    for (i in 1:obs){
        #generamos un valor aleatorio
        u <- runif(2);

        #rellenamos de forma iterativa las posiciones de la matriz con datos devueltos por la fórmula de Box-Muller
        z[i,1] <- sqrt(-2*log(u[1]))*cos(2*pi*u[2]);
        z[i,2] <- sqrt(-2*log(u[1]))*sin(2*pi*u[2]);
    }

    return(z);
}

#la función devuelve la matriz de valores pasándole como parámetro el numero de observaciones por variable aleatoria
z <- simul(obs);

#llamamos a la función hist() para imprimir la función
hist(z, breaks=seq(-5,5,by=0.5), col="blue", freq=T,
    main="Histograma de distribucion N(0,1) mediante Box-Muller",
    xlab="variable aleatoria Z",
    ylab="Frecuencia");

#variable que almacena el valor de la probabilidad almacenada
prob = length(z[z>cuantil]) / length(z);

print(prob);
```

Si observamos la figura de abajo podemos ver el histograma generado por nuestro programa, donde la superficie de cada barra es proporcional a la frecuencia de los valores que ha tomado nuestra variable Z. El histograma es un estimador de la función de densidad de la variable. Debido a que hemos hecho un gran número de observaciones y los valores han sido elegidos aleatoriamente, podemos ir reduciendo los intervalos de clase (anchura de las barras) para ir formando cada vez una curva. La curva límite representará la curva de densidad de nuestra variable Z. Si reducimos el ancho de las barras (en nuestro ejemplo es 0.5) la curva se aproximará a la curva de la distribución normal con forma de campana de Gauss.

El histograma es un representación gráfica de la función de densidad de la variable aleatoria representada en ella. Para obtener el valor de la probabilidad sólo hay que obtener el valor de la superficie de las barras correspondientes al valor que se quiere obtener. En la siguiente imagen hemos escogido el ancho de bandas en 0.01 para que se aprecie mejor la aproximación a la forma de la campana de Gauss.



Para sacar una aproximación de P (Z > 1,645) a partir de los valores de Z, utilizaremos la Regla de Laplace:

P(A) = nº de casos favorables / nº de casos posibles

Por ejemplo, en una de nuestras simulaciones obtuvimos 2556 casos favorables, por lo que la formula se aplicaría de la forma siguiente:

P(Z < 1.645) = nº de valores de Z mayores que 1.645 / nº de observaciones = 2556 / 50000 = 0.05112