Teorema del Límite Central

Linden y yo

26/10/2015

Teorema del límite central.

Vamos a sumar N variables aleatorias *uniformes*, cuya función de distribución no se parece en nada a una gaussiana, para comprobar el teorema del límite central. Es decir, si definimos:

$$S = U_1 + U_2 + \cdots + U_N$$

donde $U_i \sim Uniforme(0, 1)$, es decir:

$$p(u_i) = 1, \quad i \ = 1,\ 2, \ldots, N$$

vamos a comprobar que $p(S)$ es una campana de *Gauss*.

Además añadiremos a la simulación numérica la solución teórica de $p(S)$ para cualquier valor de N. (Linden-BayesianProbabilityTheory, epígrafe 8.1.2, página 141).

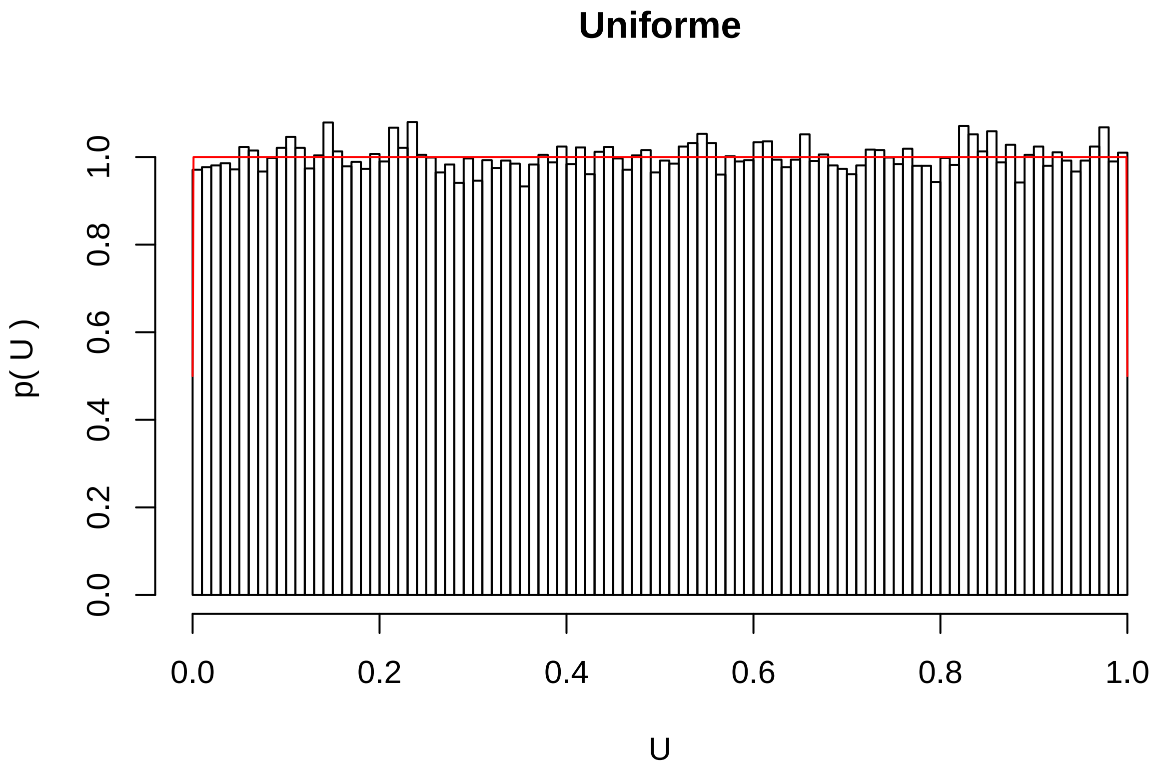
Función de distribución uniforme.

Genero un número grande M de números aleatorios con la función *runif* de **R**. Los valores generados estarán comprendidos entre 0 y 1:

```r
library("latex2exp", lib.loc="~/R/i686-pc-linux-gnu-library/3.2")

# Solución teórica S = U1 + U2 + ... + UN
pS <- function( S ){
  N <- max( S )
  fS <- 0 * S
  K <- 0 : N
  for ( i in seq( along = S ) ){

    aux <- (1 / 2) * ( - 1 )^N
    fS[ i ] <- aux * sum( choose( N, K) * ( -1 )^K * ( K - S[ i ] )^(N - 1) *
               sign( K - S[ i ] ) / factorial( N - 1) )
  }

  return(fS)

}
```

```r
M <- 100000

U <- runif( M )

hist( U, 100,
      ylab = "p( U )",
      probability = T,
      main = "Uniforme")
```

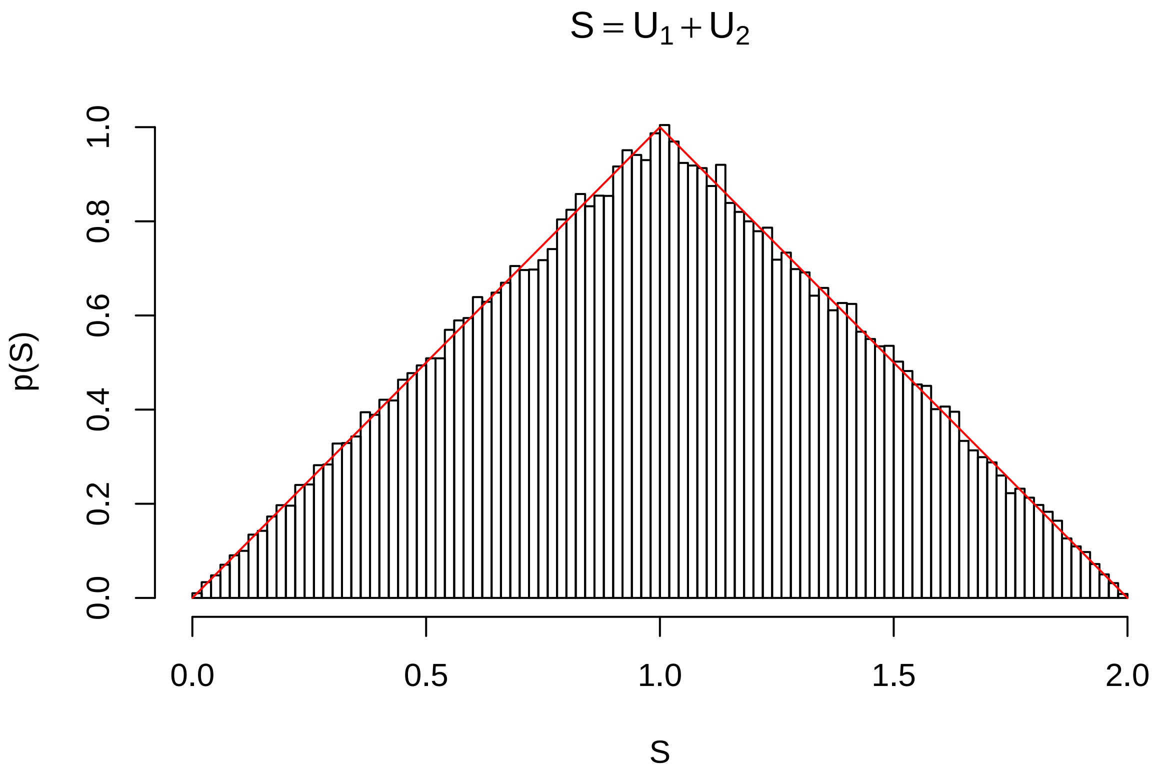
```r
S <- seq( 0, 1, len = 1000)
lines( S, pS( S ), col = 2 )
```



Como puede verse, los números generados tienen una distribución uniforme en el intervalo $0, 1$. La media es $\langle U \rangle = 1/2$ y la varianza $\sigma_U^2 \equiv var(U) = \langle U^2 \rangle - \langle U \rangle^2 = 1/3 - (1/2)^2 = 1/12$.

Repito con la suma de dos variables aleatorias uniforme:

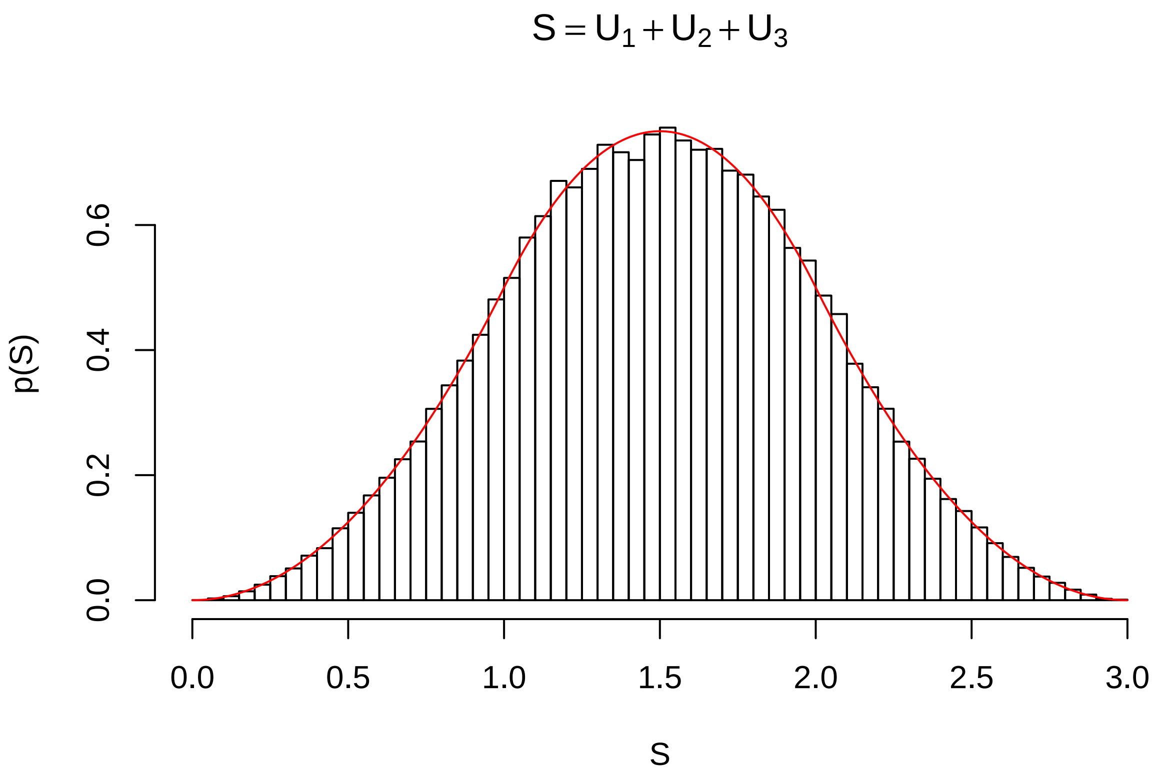
```r
U1 <- runif( M )
U2 <- runif( M )

S <- U1 + U2

titulo <- latex2exp("$\ S = U_1 + U_2$")

hist( S, 100,
      ylab = "p(S)",
      probability = T,
      main = titulo )

N <- 2
S <- seq( 0, N, len = 1000)
lines( S, pS( S ), col = 2 )
```



Vemos que no todos los valores posibles de S tienen la misma probabilidad. Sigamos.

Ahora con tres variables:

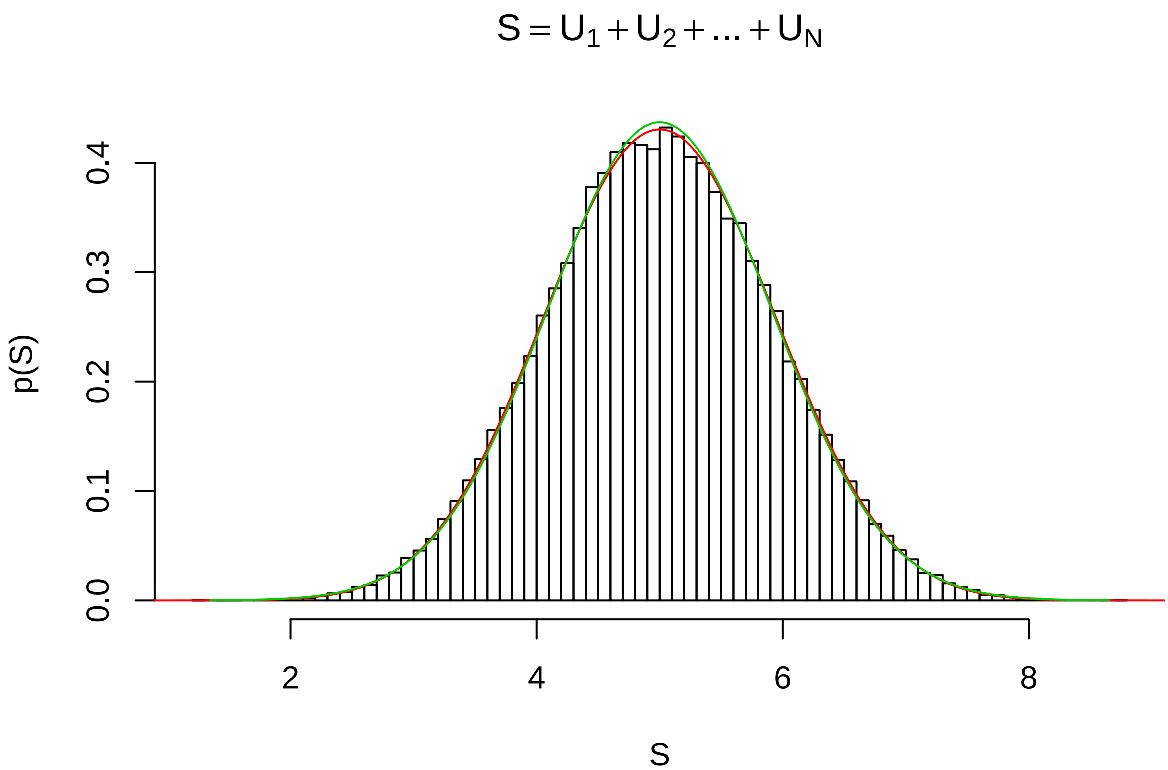
```
U1 <- runif( M )
U2 <- runif( M )
U3 <- runif( M )

U <- U1 + U2 + U3

titulo3 <- latex2exp("$\\S = U_1 + U_2 + U_3$")

hist( U, 100,
      xlab = "S",
      ylab = "p(S)",
      probability = T,
      main = titulo3 )

N <- 3
S <- seq( 0, N, len = 1000)
lines( S, pS( S ), col = 2 )
```



Repetimos con N tendiendo a infinito:

```
N <- 10
U <- rep(0,M)

for (i in 1 : N ){
  Ui <- runif( M )
  U <- U + Ui
}

tituloN <- latex2exp("$\\S = U_1 + U_2 + ... + U_N$")

hist( U, 100,
      xlab = "S",
      ylab = "p(S)",
      probability = T,
      main = tituloN )

S <- seq( 0, N, len = 1000)
lines( S, pS( S ), col = 2 )

# La gaussiana como límite
#Media
muS <- N / 2
#Varianza
varS <- N / 12
#desviación estándar
deS <- sqrt( varS )
```

```r
x1 <- muS - 4 * deS
x2 <- muS + 4 * deS
x <- seq( x1, x2, length.out = 1000 )

lines( x, dnorm( x, muS, deS ), col = 3 )
```



Teorema.

De acuerdo con el teorema del límite central, la variable aleatoria suma S tendrá una distribución normal de media $\langle S\rangle = N\langle U\rangle = \frac{N}{2}$ y varianza $\sigma_S^2 \equiv var(S) = N * var(U) = \frac{N}{12}$ que hemos dibujado en verde.