# Simulación de variables aleatorias y la practica de LGN y TLC (con R)

## 1 Familiarización con R y simulación de algunas variables aleatorias

El lenguaje R permite calcular en líneas de comando.

```
> 10+2
[1] 12
```

Podemos entrar algunas operaciones de una vez y añadir un comentario.

```
> 10+2; 5+3; # blabla
[1] 12
[1] 8
```

Lo cual es equivalente a

```
10+2
5+3
[1] 12
[1] 8
```

Para la definición de variables.

```
> n <- 10
> A=n;B=7;ls() # ls() lista las variables en memoria
[1] "A" "B" "n"
> rm(A);ls() # rm(X) borra la variable X
[1] "B" "n"
> n
[1] 10
```

Podemos simular algunas variables aleatorias con algunas funciones de R. En general se escriben `r(leydeprobabilidad)` y toman algunos parametros. Por ejemplo

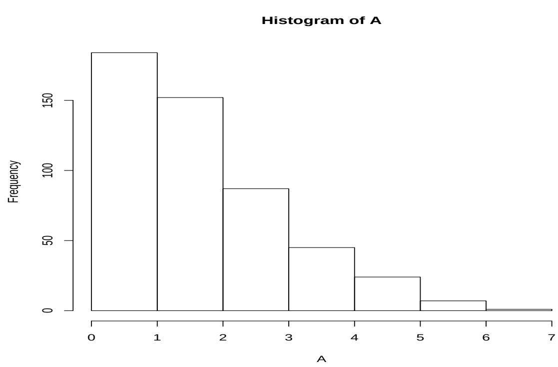
```
> A=rpois(10,2);B=rbinom(8,50,0.3);C=rgeom(12,0.1)
> print(A);print(B);print(C);
[1] 3 1 0 1 1 3 0 3 1 0
[1] 19 12 18 16 10 18 17 14
[1] 12 19 17  0  8  0 14  0 19 11  1  1
```

## 2 Manejo de histogramas y visualización de la función de distribución

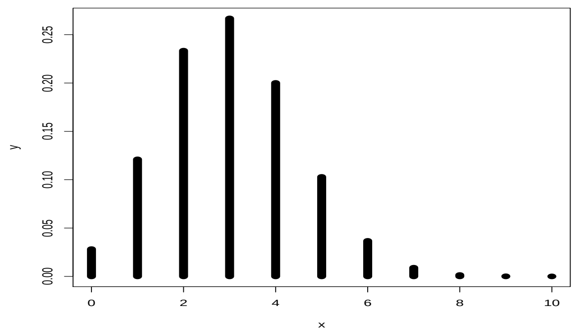
Vamos a usar la función `hist`. Si uno quiere decidir el número de barras en el histograma, podemos usar un segundo argumento con el número a especificar.

```
> pdf("histo_2020_poisson.pdf") # Eso es opcional
> A=rpois(500,2);hist(A,7)
> dev.off() # Eso tambi\'en
```

que nos proporciona una grafica con el histograma de los valores de 500 repeticiones de una variable de $\mathcal{P}(2)$. El resultado es un documento pdf de nombre `histo_2020_poisson.pdf` en la carpeta corriente.

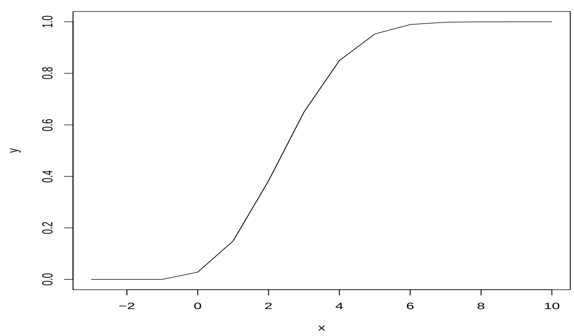


Podemos también dibujar las probabilidades de los eventos elementales. En otras palabras, vamos a dibujas la ley $P_X$ de una variable aleatoria. En el ejemplo de una variable de Binomiale $\mathcal{B}(10, 0.3)$,




Para la función de distribución, podemos usar

```
> x=-3:10
> y=pbinom(x,10,0.3)
> plot(x,y,type="l")
```




## 3 Media empírica y convergencia hacia la esperanza

Vamos a usar la funcionalidades de R para visualizar un efecto importante de la media empirica. Si uno tiene una colección $X_1, \ldots, X_n$ de variables aleatorias independientes y de misma ley. La media empírica es dada por al formula

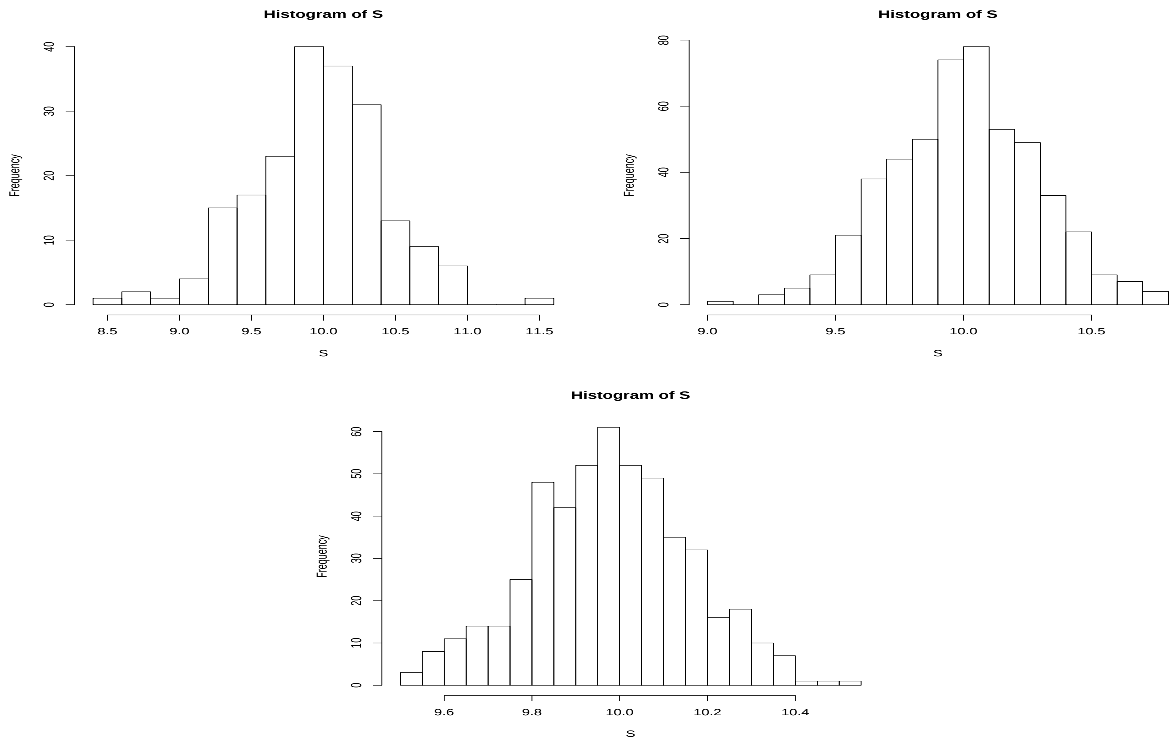
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Por la linealidad del operador de esperanza, podemos calcular la esperanza de $\bar{X}_n$ si la esperanza de cada $X_i$ existe. Como tienen todas la misma ley, $\mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n] =: m$

$$\mathbb{E}\left[\bar{X}_n\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} m = m.$$

Podemos simular $K$ veces la variable $\mathbb{E}\left[\bar{X}_n\right]$ y observar su histograma.

```
> K=200;n=30;
> S=0
> x=1:n
> for (i in 1:K){
+   y=rbinom(x,25,0.4)
+   s=sum(y)/n
+   S[i]=s
+ }
> hist(S,20)
```




De la izquierda a la derecha $(K = 200, n = 30)$, $(K = 500, n = 80)$ y $(K = 500, n = 200)$.

Observamos que cuando $n$ crece, la variable $\bar{X}_n$ tiene la tendencia de concentrarse alrededor de $\mathbb{E}\left[\bar{X}_n\right] = 10$.

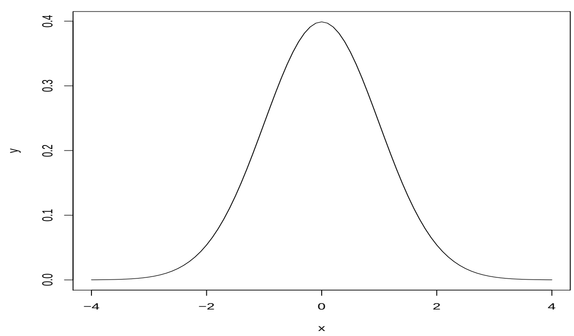
# 4 Variable gaussiana y visualización del TLC

La variable gaussiana $N$ tiene dos parámetros $\mu$ y $\sigma^2$ y se nota $\mathcal{N}(\mu, \sigma^2)$. No es una variable que entrar en la definición del Día_2. Es una variable de soporte igual a $\mathbb{R}$ entero. Cuando el soporte de una variable es un intervalo de $\mathbb{R}$ digamos que la variable en cuestión es una **variable continua**. La curva análoga de la curva de la ley discreta se llama **densidad** de la variable. Por la variable $N \sim \mathcal{N}(\mu, \sigma^2)$, su función de densidad está dada por

$$f_N : x \mapsto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Podemos dibujar esa función con

```
> mu=0;sigma2=1
> x=seq(-4,4,0.1)
> y=dnorm(x,mu,sigma2)
> plot(x,y,type="l")
```




Esa curva tiene el mismo perfil que los histogramas que dibujamos en la parte anterior. Este hecho es el resultado famoso de un teorema central en la teoría de la probabilidad : el teorema del limite central

(TCL). Este teorema importante nos dice que si renormalizamos $\bar{X}_n$ de la siguiente manera

$$Y_n = \sqrt{n}\frac{\bar{X}_n - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}}$$

entonces la variable $Y_n$ se acerca de la ley de una variable gaussiana $N$ cuando $n \to \infty$.

```
> lambda=2;K=500;n=200;
> S=0
> for (i in 1:K){
+   y=rpois(n,lambda)
+   s=(sum(y)/n-lambda)*sqrt(n)/sqrt(lambda)
+   S[i]=s
+ }
> hist(S,breaks=20,probability=TRUE)
> x=seq(-4,4,0.1)
> y=dnorm(x,0,1)
> lines(x,y,type="l")
```