UNIVERSIDAD DE EL SALVADOR

FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE

DEPARTAMENTO DE MATEMÁTICA

Licenciatura en Estadística

Control Estadístico del Paquete R

”UNIDAD TRES”

Alumna:
Erika Beatríz Guillén Pineda

Fecha de elaboración
Santa Ana - 27 de noviembre de 2015

1. INTRODUCCIÓN A LAS DISTRIBUCIONES DE PROBABILIDAD.

La teoría de la probabilidad y de variable aleatoria van a permitir establecer un amplio catálogo de modelos te?ricos, tanto discretos como continuos, con los cuales se van a poder asimilar muchas de las situaciones de la vida real. El estudio de los modelos teóricos, incluyendo la caracterización a través de sus parámetros, el cálculo de probabilidades en sus distintos formatos y la generación de números aleatorios, van a facilitar enormemente el an?lisis de estas situaciones reales, algunos ejemplos de estos fenómenos son:

- Si se contesta al azar un examen tipo test de 10 preguntas, donde cada una de ellas tiene 4 posibilidades siendo sólo una de ellas la correcta, Qué número de aciertos es más probable?
- Se sabe que las bombillas de bajo consumo de 14 w tienen una vida media útil de 10,000 horas, mientras que las bombillas clásicas por incandescencia de 60 w tienen una vida media útil de 1,000 horas. Si cada d?a se encienden unas 4 horas Cuál es la probabilidad de que después de un aõ están funcionando las dos?, ?ninguna de las dos?, ?al menos una de las dos?

El primer problema a resolver será la elección del modelo teórico apropiado para cada caso en estudio. Para tener un buen manejo matemático de las distintas situaciones que se puedan plantear dada la distinta naturaleza y la diversidad de los resultados que proporcionan los experimentos, se necesita realizar una abstracción cuantificada del experimento. Esto lleva a una primera gran clasificación entre modelos de probabilidad discretos y continuos.

Las probabilidades asociadas a cada uno de los valores de la variable aleatoria pueden ser organizadas como una distribución de probabilidad, expresándose mediante una tabla, una gráfica o una fórmula, denominándose en este último caso, a la regla de correspondencia-valores probabilidades, función de probabilidad.

Como sabemos, los números aleatorios son descritos por una distribución. Esto es, alguna función la cual especifica la probabilidad que un número aleatorio este en algún rango, por ejemplo P(a<X<b). Frecuentemente es dada por una densidad de probabilidad (en el caso continuo) o por una función masa de probabilidad P(X=x)=p(x) en el caso discreto. Con R podemos obtener n?meros seleccionados aleatoriamente de diferentes distribuciones, para ello sólo tenemos que familiarizarnos con los parámetros que hay que dar a las funciones tal como la media, o una proporción, etc (dependiendo de la distribucipn que se esté considerando y de lo que se esté analizando).

2. DISTRIBUCIONES DISCRETAS

1. **Distribución Binomial.**

   - **Parámetros**

     x = n?mero de ?xitos

     size = n?mero de ensayos

     p = proporci?n de ?xitos

     lower.tail = TRUE P(X<=x)

     lower.tail = FALSE P(X>=x)

     n = tamaño de la muestra

   - **Sintasis en R.**

     dbinom(x, size, prob, log = FALSE)

```
dbinom(2, 4, 0.5, log = FALSE)

## [1] 0.375
```

     pbinom(x, size, prob, lower.tail = TRUE, log.p = FALSE)

```
pbinom(7, 10, 0.3, lower.tail = TRUE, log.p = FALSE)

## [1] 0.9984096
```

     qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)

```
qbinom(0.3, 4, 0.5, lower.tail = TRUE, log.p = FALSE)

## [1] 1
```

     rbinom(n, size, prob)

```
rbinom(10, 4, 0.3)

##  [1] 0 2 0 1 0 3 2 1 3 1
```

2. **Distribución Geométrica**

   - **Parámetros**

     x = ensayos necesarios para obtener el primer éxito
     p = proporci?n de ?xitos
     lower.tail = TRUE P(X<=x)
     lower.tail = FALSE P(X>=x)
     n = tamaño de la muestra

- **Sintasis en R.**

  dgeom(x, prob, log = FALSE)

```
dgeom(4, 0.25, log = FALSE)

## [1] 0.07910156
```

  pgeom(x, prob, lower.tail = TRUE, log.p = FALSE)

```
pgeom(4, 0.25, lower.tail = TRUE, log.p = FALSE)

## [1] 0.7626953
```

  qgeom(p, prob, lower.tail = TRUE, log.p = FALSE)

```
qgeom(0.75, 0.25, lower.tail = TRUE, log.p = FALSE)

## [1] 4
```

  rgeom(n, prob)

```
rgeom(15, 0.25)

##  [1] 4 0 5 0 0 7 7 1 1 6 3 2 3 5 0
```

3. **Distribución Hipergeométrica**

- **Parámetros**

$$\text{x} = \text{objetos seleccionados tipo m (primer tipo)}$$
$$\text{m} = \text{total de objetos (primer tipo)}$$
$$\text{n} = \text{total de objetos (segundo tipo)}$$
$$\text{y} = \text{el número total de objetos seleccionados primer tipo y segundo tipo}$$
$$\text{size} = \text{tamaño de la muestra}$$

- **Sintasis en R**
  dhyper(x, m, n, y, log = FALSE)

```
dhyper(8, 8, 5, 13, log = FALSE)

## [1] 1
```

  phyper(x, m, n, y, lower.tail = TRUE, log.p = FALSE)

```
phyper(5, 5, 14, 19, lower.tail = TRUE, log.p = FALSE)

## [1] 1
```

  qhyper(p, m, n, y, lower.tail = TRUE, log.p = FALSE)

```
qhyper(0.25, 6, 5, 11, lower.tail = TRUE, log.p = FALSE)

## [1] 6
```

rhyper(size, n, m, n,y)

```
rhyper(20, 4, 6, 10)

##  [1] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
```

4. **Distribución Poisson**

- **Parámetros**

$$\text{x} = \text{valor cualqueira}$$
$$\text{p} = \text{probabilidad}$$
$$\text{lambda} = \text{media de la distribución}$$
$$\text{size} = \text{tamaño de la muestra}$$

- **Sintasis en R.**

dpois(x, lambda, log = FALSE)

```
dpois(3, 0.8, log = FALSE)

## [1] 0.03834274
```

ppois(x, lambda, lower.tail = TRUE, log.p = FALSE)

```
ppois(3, 7.5, lower.tail = TRUE, log.p = FALSE)

## [1] 0.05914546
```

qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)

```
qpois(0.5162, 3.75, lower.tail = TRUE, log.p = FALSE)

## [1] 4
```

Para cada una de las distribuciones discretas o continuas est?n disponibles las siguientes opciones:

- Gráfica de la distribución: Genera la gráfica de la función de probabilidad.
- Probabilidades: Determina la probabilidad deque la variable tome un valor dado.
- Probabilidades Acumuladas: Calcula bien el valor de P(X<=x)(cola de la izquierda), o bien, P(X>x)(cola de la derecha) para cada cuantil(q) de X.

- Cuantiles: Permite calcular el valor de la variable que deja a derecha o a izquierda (según se seleccione) una determinada probabilidad.
- Muestra de la distribución: Genera muestras aleatorias extraídas de la distribución.

El Paquete R proporciona 4 funciones para cada distribución (ya sea continua o discreta) que pueden usarse escribiendo el nombre de la distribución, anteponi?ndole una d, si se quiere la función de densidad o la probabilidad de que la variable tome el valor especificado en x, es decir P(X=x); una p para la función de distribuci?n acumulada, es decir P(X<=x); una q para los cuantiles, es decir, el valor x de la distribución acumulada que deja un área igual p P(X<=x)=p, y una r para generar una muestra aleatoria de la distribución; únicamente hay que tener en cuenta la sintaxis presentada en el cuadro anterior.

3. CÁLCULO DE PROBABILIDADES.

- **Ejemplo 1:**

  Si un estudiante responde al azar a un examen de 8 preguntas de verdadero o falso.

  a) Cuál es la probabilidad de que acierte 4? P[X=4]

  La variable X=”número de aciertos”sigue una distribución Binomial de parámetros n = 8 y p=1/2(p probabilidad de ?xito).

```
# Usando la funciones propias de R

dbinom(4,8,0.5)

## [1] 0.2734375

# dbinom calcula la probabilidad en un valor concreto
```

  b) Cuál es la probabilidad de que acierte a lo sumo 2? P[X<=2]

```
x <- 2; n=8; p=1/2

pbinom(x, size = n, prob = p, lower.tail=TRUE)

## [1] 0.1445313

# pbinom es la funcion de distribucion acumulada
```

  b) ?Cuál es la probabilidad de que acierte 5 o m?s? P[X>=5]

```
x <- 4; n=8; p=1/2

#primera forma

F <- 1 - pbinom(x, n, p, lower.tail=TRUE); F

## [1] 0.3632813
```

```
#segunda forma

pbinom(4, size=8, prob=0.5, lower.tail=FALSE)

## [1] 0.3632813
```

- **Ejemplo 2:**

  Una cierta área de Estados Unidos es afectada, en promedio, por 6 huracanes al a?o. Encuentre la probabilidad de que en un determinado a?o esta área sea afectada por:

  a) Menos de 4 huracanes.P[X<4]=F(3)

  Se define la variable X = "número de huracanes por aõz asumiendo que dicha variable tiene una distribución de Poisson de par?metro lambda igual a 6, porque describe el número de éxitos por unidad de tiempo y porque son independientes del tiempo desde el último evento. Se calcularán ahora las probabilidades:

```
x <- 3; mu <- 6

ppois(x, lambda = mu, lower.tail=TRUE)

## [1] 0.1512039
```

  b) Entre 6 y 8 huracanes.P[6<=X<=8]=P[X<=8]-P[X<=5]=F(8)-F(5)

  Para calcular la probabilidad de que ocurran entre 6 y 8 huracanes, se pueden sumar las probabilidades P(X=6)+P(X=7)+P(X=8)

  primera forma

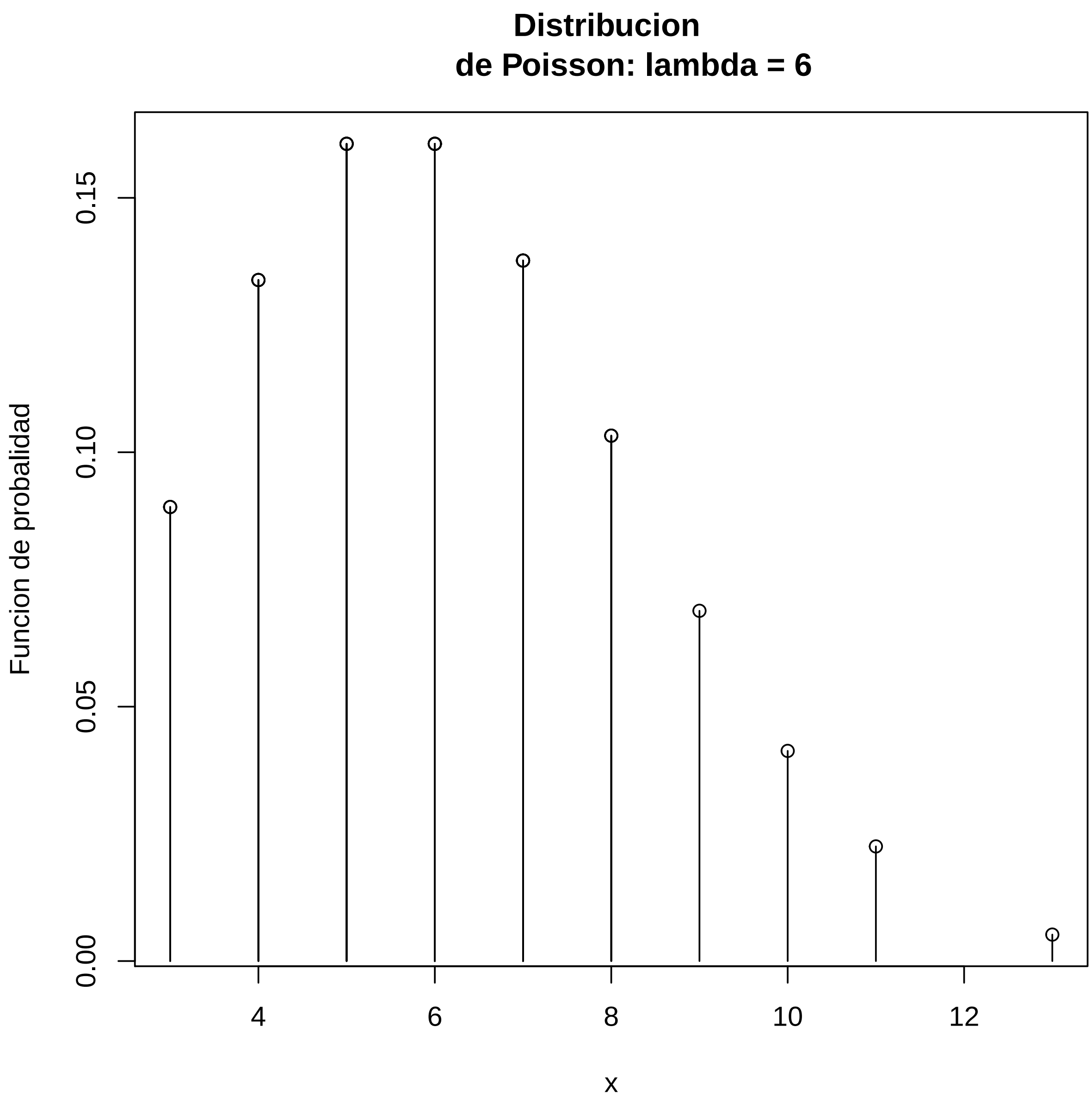
```
sum(dpois(c(6,7,8),lambda = 6))

## [1] 0.4015579
```

  segunda forma

```
# O restar las probabilidades acumuladas, con la opcion Cola izquierda,
# P(X<=8)-P(X<=5).Como antes se realizan en primer lugar las probabilidades
# acumuladas y se restan los resultados obtenidos:

F8 <- ppois(8, lambda = 6, lower.tail=TRUE)
F5 <- ppois(5,lambda = 6, lower.tail=TRUE)
F8 - F5

## [1] 0.4015579
```

c) Represente gráficamente la función de probabilidad de la variable aleatoria que mide el número de huracanes por año.

```
n <- 30

# genera 30 valores de una distribucion de Poisson con lambda igual a 6

x <- rpois(n, lambda=mu)
x <- rpois(30, lambda=6)

#calcula las probabilidades para cada valor generado

y <- dpois(x, lambda=mu)
y <- dpois(x, lambda=6)


#genera el grafico de distribuci?n

plot(x, y, xlab="x", ylab="Funcion de probalidad", main="Distribucion
     de Poisson: lambda = 6", type="h")

#une los puntos a las lineas

points(x, y, pch=21)
```



- **Ejemplo 3:**

  En un juego se disponen 15 globos llenos de agua, de los que 4 tienen premio. Los participantes en el juego, con los ojos vendados, golpean los globos con un bate por orden hasta que cada uno consigue romper 2.

  a) Cuál es la probabilidad de que el primer participante consiga un premio?

  mero de premios conseguidos entre 2 posibles"sigue una distribución hipergeométrica de parámetros m=11, n=4, K=2.

```
x <- 0:2; m = 11; n <- 4; k=2

# x define el n\'umero de globos con premio
# se construye la distribucion de frecuencias del numero de premios

Tabla <- data.frame(Probabilidad=dhyper(x, m, n, k))
rownames(Tabla) <- c("Ningun premio","Solamente uno", "Dos premios")
Tabla

##               Probabilidad
## Ningun premio   0.05714286
## Solamente uno   0.41904762
## Dos premios     0.52380952
```

b) Si el primer participante ha conseguido sólo un premio, cuál es la probabilidad de que el segundo participante consiga otro?

Para el segundo participante la variable seguirá una hipergeométrica de parámetros m= 10, n= 3 y k= 2,

```
x = 1; m= 10; n= 3; k= 2;
dhyper(x, m, n, k)

## [1] 0.3846154
```

- **Ejemplo 4:**
  Un vendedor de alarmas de hogar tiene éxito enuna casa de cada diez que visita. Calcula:

  a) La probabilidad de que en un da determinado consiga vender la primera alarma en la sexta casa que visita.

Se define la variable X="número de casas que visita antes de conseguir vender la primera alarma", que sigue una distribución Geométrica con Probabilidad de éxito=0.1

Habría que calcular la probabilidadde que tenga 5 fracasos antes del primer éxito, obteniendo de la tabla la probabilidad P(X=5)=0.059049

```
# x define el n\'umero de intentos fallidos

x <- 0:5; p=0.1

# creando la tabla de distribucion de frecuencias del n\'umero de intentos
# fallidos antes de obtener la primera venta.
```

```
Tabla <- data.frame(Probabilidad=dgeom(x, prob=p))

# nombrando las filas de la distribuci\'on de frecuencias

rownames(Tabla) <- c("Venta en el primer intento",
                     "Venta en el segundo intento",
                     "Venta en el tercer intento",
                     "Venta en el cuarto intento",
                     "Venta en el quinto intento",
                     "Venta en el sexto intento")

Tabla

##                              Probabilidad
## Venta en el primer intento       0.100000
## Venta en el segundo intento      0.090000
## Venta en el tercer intento       0.081000
## Venta en el cuarto intento       0.072900
## Venta en el quinto intento       0.065610
## Venta en el sexto intento        0.059049
```

b) La probabilidad de que no venda ninguna después de siete viviendas visitadas.

La variable X="número de alarmas vendidas en 7 viviendas"sigue una distribución Binomial con n=7 Ensayos binomiales y Probabilidad de éxito p=0.1, luego en nuestro caso se tiene P(X=0)=0.4782969

```
x=0; n=7; p=0.1
dbinom(x, n, p, log = FALSE)

## [1] 0.4782969
```

c) Si se plantea vender tres alarmas, cuál es la probabilidad deque consiga su objetivo en la octava vivienda que visita?

Para abordar esta cuestión, se define la variable Y= "número de casas que visita antes de conseguir vender la tercera alarma. Esta variable sigue una distribución Binomial Negativa de parámetros Número de éxitos= 3, Probabilidad de éxito p=0.1, de donde: P(Y=5)=0.01240029

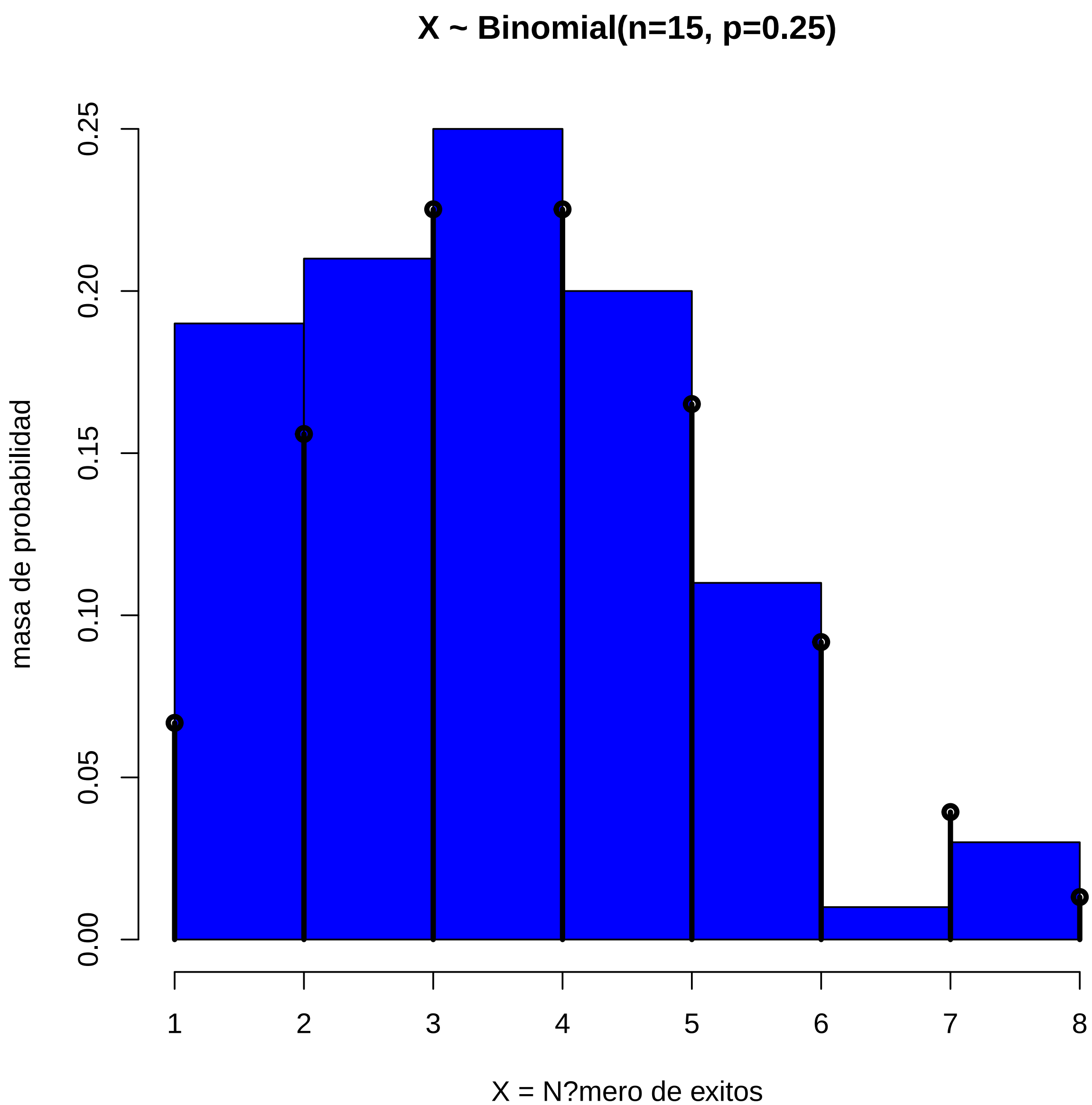
```
y <- 0:5; r=3; p <- 0.1

Tabla <- data.frame(Probabilidad=dnbinom(y, size=r, prob=p))
```

```
rownames(Tabla) <- 0:5

Tabla

##   Probabilidad
## 0   0.00100000
## 1   0.00270000
## 2   0.00486000
## 3   0.00729000
## 4   0.00984150
## 5   0.01240029
```

4. GENERACIÓN DE MUESTRAS ALEATORIAS DE LAS DISTRIBUCIONES.

- **Ejemplo 1:** Generar 100 números aleatorios de una distribuci?n Binomial de parámetros n=15 ensayos o pruebas y una probabilidad de éxito de 0.25.

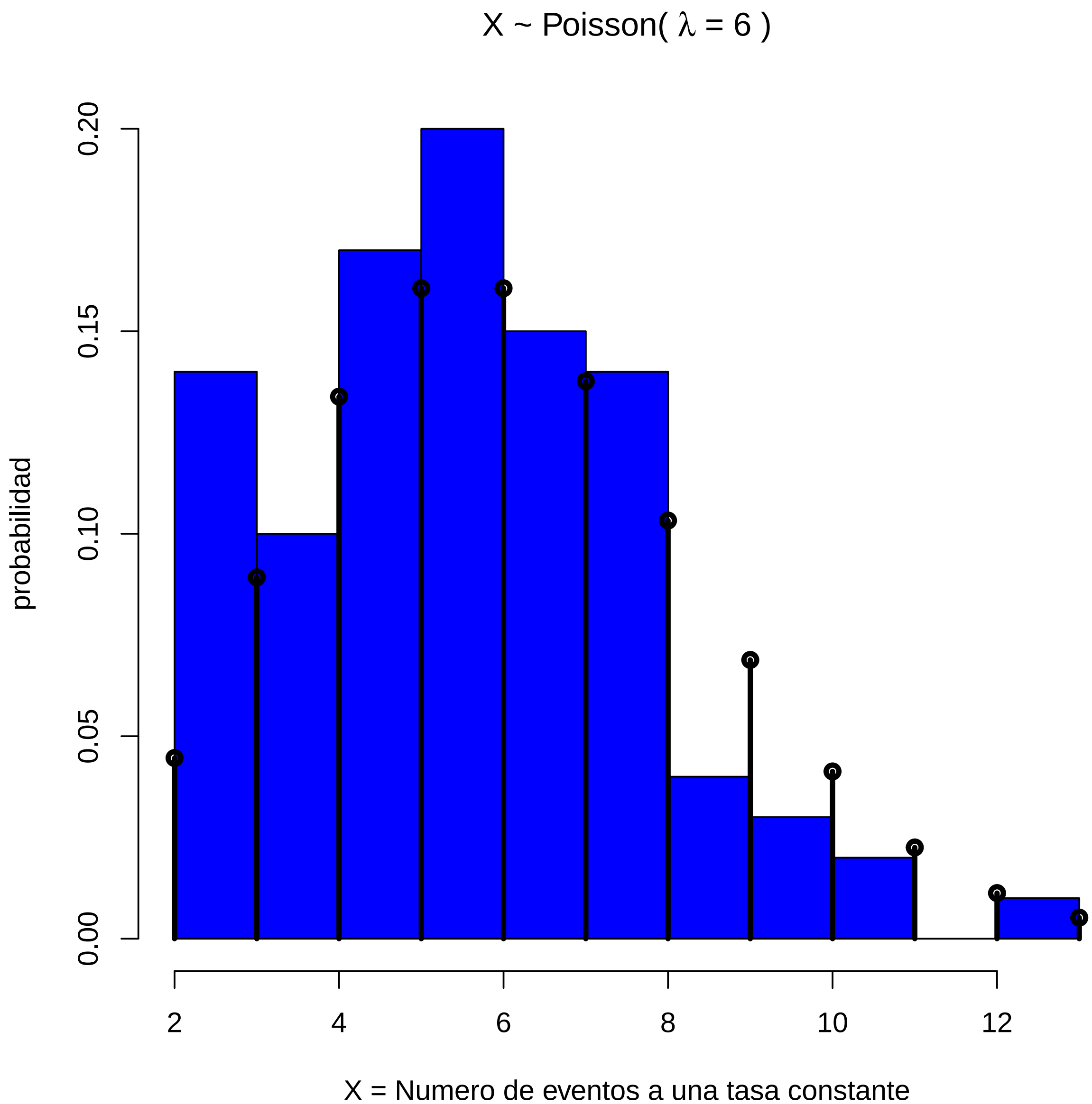
```
# Definir los parametros apropiados

n <- 15; p <- 0.25

# generar 100 numeros aleatorios binomiales

x = rbinom(100, n, p); x

##   [1] 6 8 2 4 3 5 4 1 4 4 3 3 5 2 5 3 1 6 1 5 4 3 4 3 2 5 4 5 3 6 2 1 4 3 2
##  [36] 6 4 1 4 2 3 3 4 5 5 3 7 5 2 5 2 4 3 5 3 3 4 4 8 4 5 1 2 4 3 5 4 3 3 3
##  [71] 5 5 4 6 8 4 6 4 6 5 5 5 3 3 2 4 4 2 6 4 5 6 6 5 4 6 3 2 4 2

# Histograma para la muestra aleatoria de tamaÃ±o 100

hist(x, main="X ~ Binomial(n=15, p=0.25)", xlab="X = N?mero de exitos",
ylab="masa de probabilidad", probability=TRUE, col="blue")

# Graficar la funci?n de probabilidad te?rica, use la funci?n points(), no
# debe cerrar el gr?fico obtenido con la instrucci?n anterior

xvals=0:n; points(xvals, dbinom(xvals, n, p), type="h", lwd=3)
points(xvals, dbinom(xvals, n, p), type="p", lwd=3)
```



- **Ejemplo 2:**
  Generar 100 números aleatorios de una distribución Poisson con 200000 ensayos o pruebas y una probabilidad de éxito de 3/100000

```
# Definir los par?metros apropiados

n <- 200000; p <- 3/100000; lambda=n*p

# generar 100 n\'umeros aleatorios de la distribuci\'on

x = rpois(100, lambda); x

##   [1]  7  8  4  5  9  4  6 10  7  7  3  8  5  5  3  4  4  3  6  5  2  2  4
```

```
## [24]  3  5  3  5  5  6  8  5  6  5  6  2  8  9 11  6  7  6  9  8  5  3  3
## [47]  4  9  4  8  7 10  8  7  6  8  6  5  6  7  8  7  6  3 13  6  6  6  7
## [70]  5  6  4  6  6  5  5  8  5  7 10  7  8  7  4  7  4  3  7  6  6 11  8
## [93]  5  3  3  6  8  8  5  7

# Histograma para la muestra aleatoria de tama\~no 100

hist(x, main=expression(paste("X ~ Poisson( ", lambda, " = 6 )")),
xlab="X = Numero de eventos a una tasa constante", ylab="masa de
probabilidad", probability=TRUE, col="blue")

# Graficar la funci\'on de probabilidad teorica, use la funci\'on points()

xvals=0:n; points(xvals, dpois(xvals, lambda), type="h", lwd=3)
points(xvals, dpois(xvals, lambda), type="p", lwd=3)
```


X ~ Poisson(λ = 6)
probabilidad
0.00
0.05
0.10
0.15
0.20
2
4
6
8
10
12
X = Numero de eventos a una tasa constante