

Probabilidades I

jmm

4 de Septiembre del 2015

Objetivos:

- Familiarizarse con distintas distribuciones de probabilidades y con la manera de usarlas en R.
- Simular procesos estocásticos.
- ver https://sites.google.com/site/modelosydatos/Bestiario_sp.pdf

R tiene funciones para todas las distribuciones de probabilidad estándares, y para cada una de estas distribuciones tenemos funciones para:

- generar valores,
- calcular probabilidades,
- probabilidades acumuladas y
- cuantiles

Estas funciones comienzan con las letras **r**, **d**, **p** y **q** respectivamente. Por ejemplo, para la distribución de Poisson tenemos: `rpois`, `dpois`, `ppois`, y `qpois`.

En muchos casos es útil poder generar muestras de una distribución en particular. Asumimos que estas muestras generadas en la computadora son una “muestra aleatoria” pero en realidad provienen de un generador de números aleatorios por lo que es más correcto decir que son “pseudo-aleatorios”. Un aspecto importante, sobre todo pensando en la reproducibilidad de nuestro trabajo, es que si iniciamos al generador de números aleatorios siempre en el mismo valor, la secuencia de números pseudo-aleatorios se va a repetir y por lo tanto podemos reproducir exactamente una simulación estocástica.

En R usamos `set.seed(12345)` para inicializar el generador de números aleatorios en 12345. El número que le ponemos a `set.seed` es arbitrario.

ejemplo: simulamos una muestra de 10 valores de una distribución de Poisson

```
set.seed(12345)
rpois(n=10,lambda=1)
```

```
## [1] 1 2 2 2 1 0 0 1 1 4
```

Nuestros scripts pueden ser más fáciles de leer y modificar si definimos variables fuera de las funciones. Por ejemplo:

```
n <- 100      # tamaño de la muestra
lambda <- 1.2
```

Ahora simulamos datos y vemos la frecuencia en un histograma

```
y <- rpois(n = n, lambda = lambda)
hist(y, xlab="valores", main="")
```

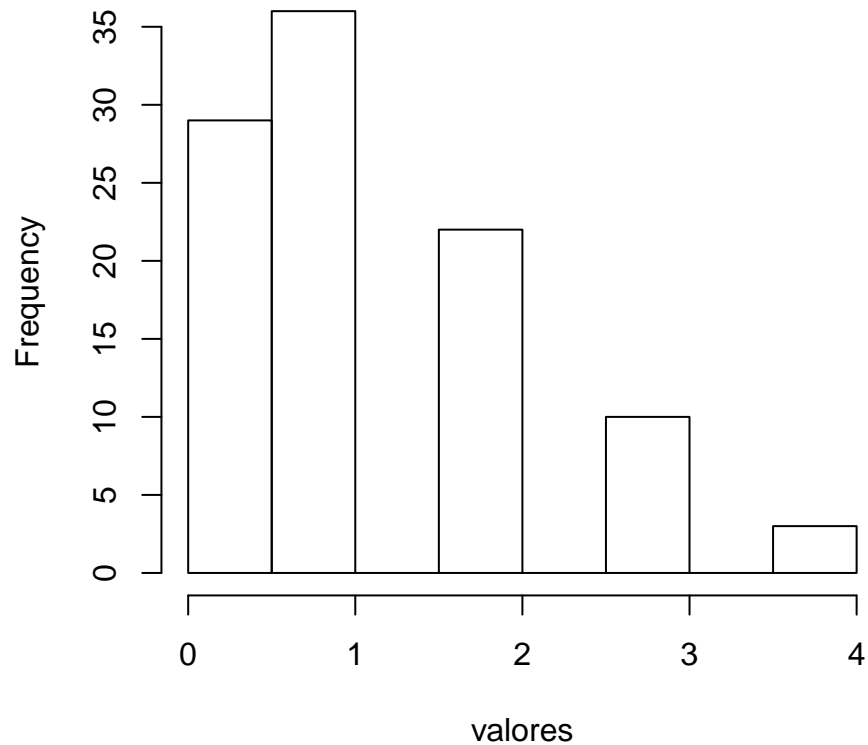


Figure 1: Histograma de datos simulados de una Poisson con $\lambda = 1.2$.

Una vez que tenemos “datos” como estos, podemos ver las frecuencias relativas y compararlas con la distribución teórica usando la función `dpois`.

```
f <- factor(y, levels=0:max(y))
obsprobs <- table(f)/n
plot(obsprobs, xlab="valores", ylab="proporción")
tprobs <- dpois(x = 0:max(y), lambda = lambda)
points(0:max(y), tprobs, pch=16, col=2)
```

Otra función útil es la [Función de Distribución](#), que nos da la probabilidad de encontrar un valor menor o igual a un valor dado. Por ejemplo, la probabilidad de obtener $y \leq 3$ con $\lambda = 1$ es:

```
ppois(3, lambda=lambda)
```

```
## [1] 0.966231
```

Si queremos saber la probabilidad de obtener un valor *mayor* a 3 hacemos

```
1 - ppois(3, lambda=lambda)
```

```
## [1] 0.03376897
```

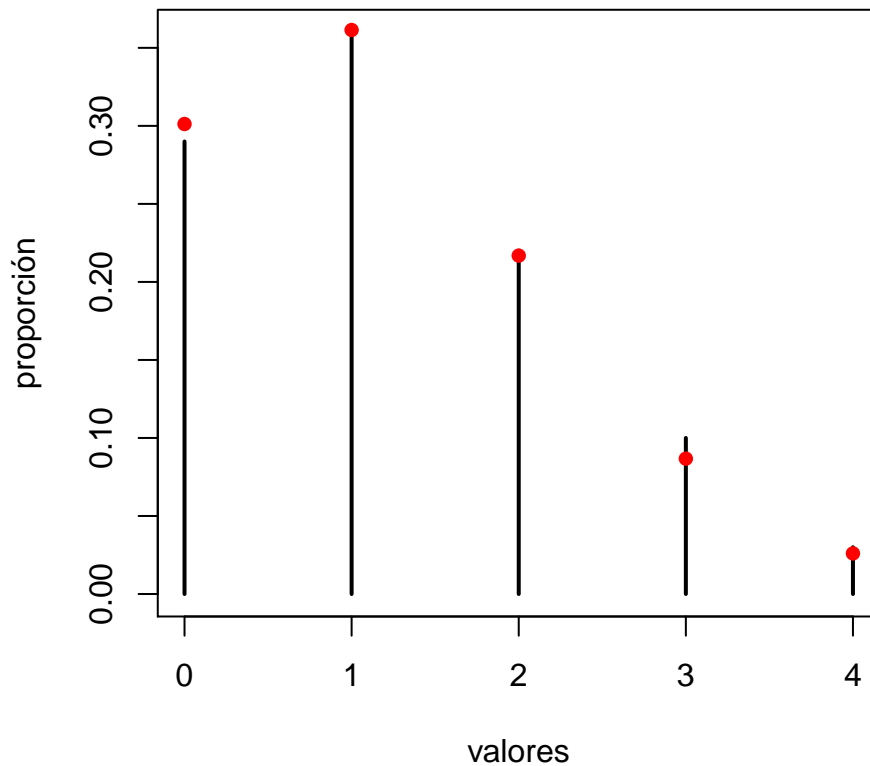


Figure 2: distribución empírica y teórica para Poisson con $\lambda = 1.2$

Para ver la probabilidad de un valor en particular, por ej $y = 3$:

```
ppois(3,lambda=lambda) - ppois(2, lambda=lambda)
```

```
## [1] 0.08674393
```

¿Por qué tiene sentido hacer lo de arriba? Podemos corroborar este resultado usando `dpois`

```
dpois(3,lambda=lambda)
```

```
## [1] 0.08674393
```

El equivalente empírico de la función acumulada es la función `ecdf`:

```
eF <- ecdf(y)
plot(eF, xlab="valores", ylab="Función de Probabilidad Empírica", main="")
lines( 0:6, ppois(0:6, lambda=1), type="s", col=2)
```

Por último, la función cuantil `qpois` es la inversa de la función de distribución y para un valor de probabilidad acumulada nos devuelve el valor de la variable

```
qpois(0.95, lambda=lambda)
```

```
## [1] 3
```

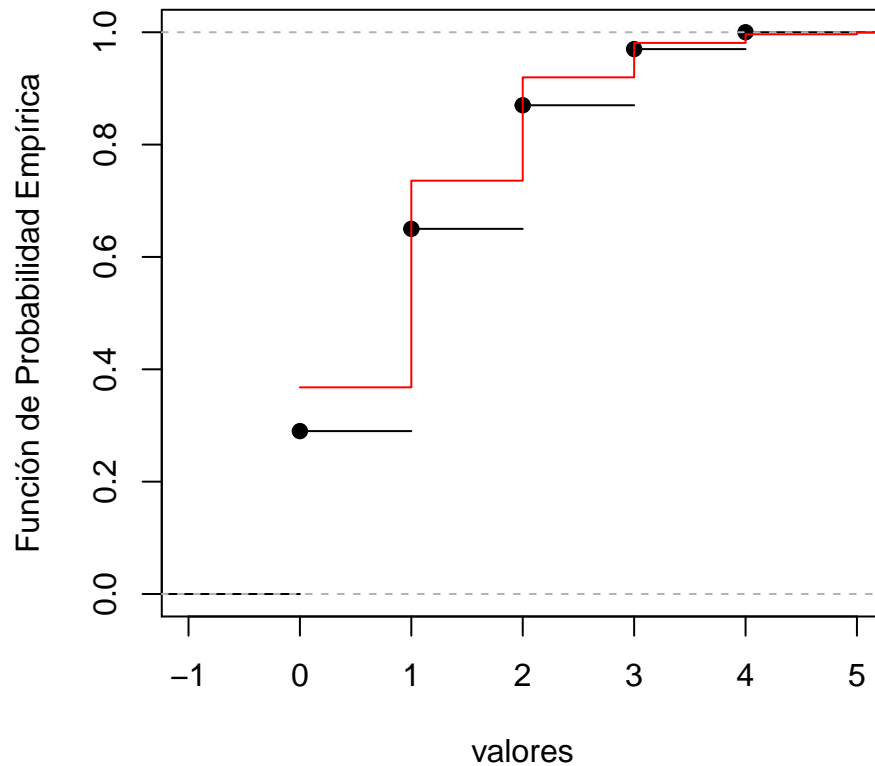


Figure 3: Acumulada empírica y teórica para Poisson con $\lambda = 1.2$

La función `qpois` sirve también para calcular intervalos que contienen un porcentaje de los valores de la distribución. Por ejemplo, el 95% de los valores están entre `qpois(c(0.025,0.975), lambda)`. El equivalente empírico es la función `quantile`

```
qpois(c(0.025,0.975), lambda)
```

```
## [1] 0 4
```

```
quantile(y, probs = c(0.025,0.975))
```

```
## 2.5% 97.5%
```

```
## 0.000 3.525
```

Ejercicios:

1. Asumiendo una distribución de Poisson:

- ¿cuál es la probabilidad teórica de obtener $y = 0$ para $\lambda = 1$?
- ¿cuál es la probabilidad teórica de $y > 2$?
- comparar la probabilidad de $y=0$ para $\lambda = 1$ de la distribución teórica con la obtenida empíricamente con muestras de 10, 100, 1000 y 10000 observaciones
- comparar las diferencias entre el valor esperado de y para la distribución teórica con los valores empíricos de muestras de 10, 100, 1000 y 10000 observaciones
- lo mismo para los intervalos de 95%

2. ¿cómo cambia la forma de la distribución de Poisson a medida que cambia λ ? (pueden ver esto simulando datos y haciendo histogramas o graficando las probabilidades teóricas)