

Introducción a distribuciones de probabilidad

Ramon Ceballos

7/2/2021

Distribución Hipergeométrica

1. Conceptos teóricos y matemáticos

Consideremos el experimento “extraer a la vez (o una detrás de otra, sin retornarlos) n objetos donde hay N de tipo A y M de tipo B”. Si X es variable aleatoria que mide el “número de objetos del tipo A”, diremos que X se distribuye como una Hipergeométrica con parámetros N, M, n .

$$X \sim H(N, M, n)$$

Vamos a definir algunas de las propiedades que tendrá una distribución Hipergeométrica. En referencia a este tipo de distribución, citar que es un problema de combinatoria clásico.

- El **dominio** de X será $D_X = \{0, 1, 2, \dots, N\}$ (en general), siempre y cuando $n > N$.
- La **función de probabilidad** vendrá dada por:

$$f(x) = \frac{n! \text{comb}((N)(k)) * n! \text{comb}((M)(n-k))}{n! \text{comb}((N+M)(n))}$$

Así, $\binom{N}{k}$ es de los N objetos extraer k ; $\binom{M}{n-k}$ es de los M objetos extraer $n - k$; y $\binom{N+M}{n}$ es extraer de los $N + M$ es extraer los n .

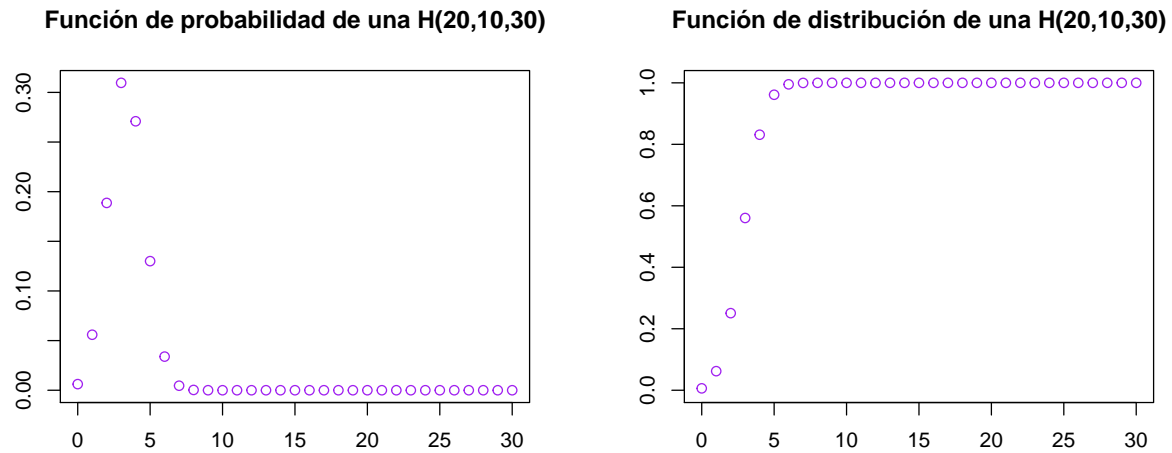
- La **función de distribución** vendrá dada por:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{k=0}^x f(k) & \text{si } 0 \leq x < n \\ 1 & \text{si } x \geq n \end{cases}$$

Va a ir sumando los valores de la función de probabilidad hasta alcanzar 1.

- **Esperanza** $E(X) = \frac{nN}{N+M}$.
- **Varianza** $Var(X) = \frac{nNM}{(N+M)^2} \cdot \frac{N+M-n}{N+M-1}$.

```
par(mfrow = c(1,2))
plot(0:30, dhyper(0:30,10,20,10),
     col = "purple", xlab = "", ylab = "",
     main = "Función de probabilidad de una H(20,10,30)")
plot(0:30, phyper(0:30,10,20,10),
     col = "purple", xlab = "", ylab = "",
     main = "Función de distribución de una H(20,10,30)",
     ylim = c(0,1))
```



```
par(mfrow= c(1,1))
```

Ejemplos: entre un grupo de animales (perros y gatos) la probabilidad de escoger uno u otro; al pescar en un lago hay carpas y tencas y sabiendo su proporción determinar la probabilidad con distribución hipergeométrica.

2. Distribución Hipergeométrica en R y Python

El código de la distribución Hipergeométrica:

- En **R** tenemos las funciones del paquete **Rlab**: **dhyper(x, m, n, k)**, **phyper(q, m, N, k)**, **qhyper(p, m, N, k)**, **rhyper(nn, m, N, k)** donde **m** es el número de objetos del primer tipo, **N** el número de objetos del segundo tipo y **k** el número de extracciones realizadas.
- En **Python** tenemos las funciones del paquete **scipy.stats.hypergeom**: **pmf(k,M, n, N)**, **cdf(k,M, n, N)**, **ppf(q,M, n, N)**, **rvs(M, n, N, size)** donde **M** es el número de objetos del primer tipo, **n** el número de objetos del segundo tipo y **x** el número de extracciones realizadas. El **nn** es el tamaño de la población aleatoria.

3. Ejemplos en código para la distribución Hipergeométrica (R y Python)

Supongamos que tenemos 20 animales, de los cuales 7 son perros. Queremos medir la probabilidad de encontrar un número determinado de perros si elegimos $k = 12$ animales al azar.

Ejemplo en R Empleamos las funciones anteriores.

```
library(Rlab)
```

```
## Rlab 2.15.1 attached.
```

```
##
## Attaching package: 'Rlab'

## The following objects are masked from 'package:stats':
##
##      dexp, dgamma, dweibull, pexp, pgamma, pweibull, qexp, qgamma,
##      qweibull, rexp, rgamma, rweibull

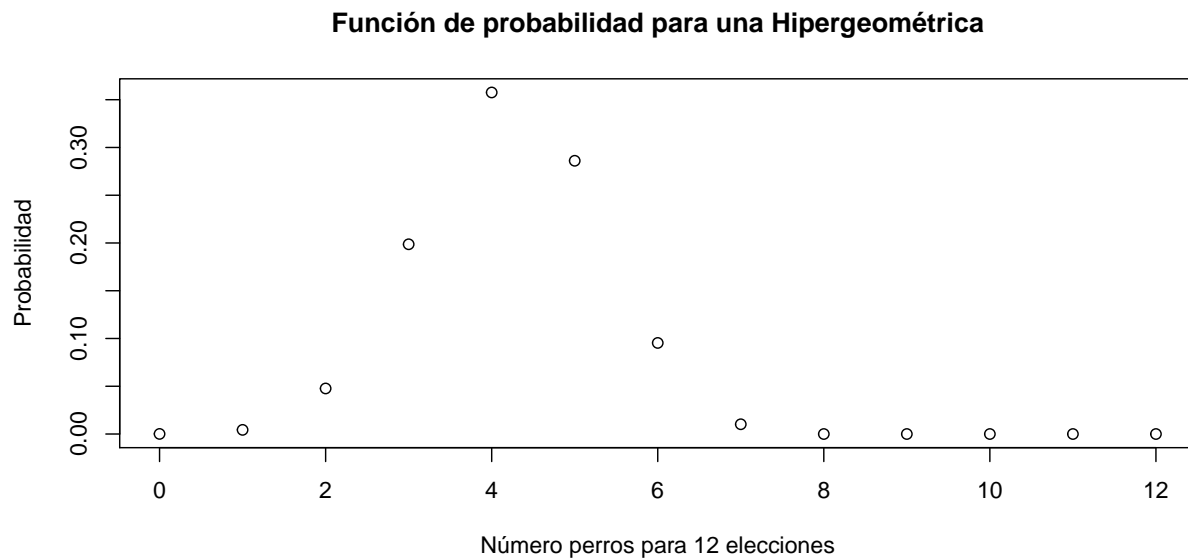
## The following object is masked from 'package:datasets':
##
##      precip
```

```
#Determinamos los perros (M , valor que queremos extraer)
M = 7

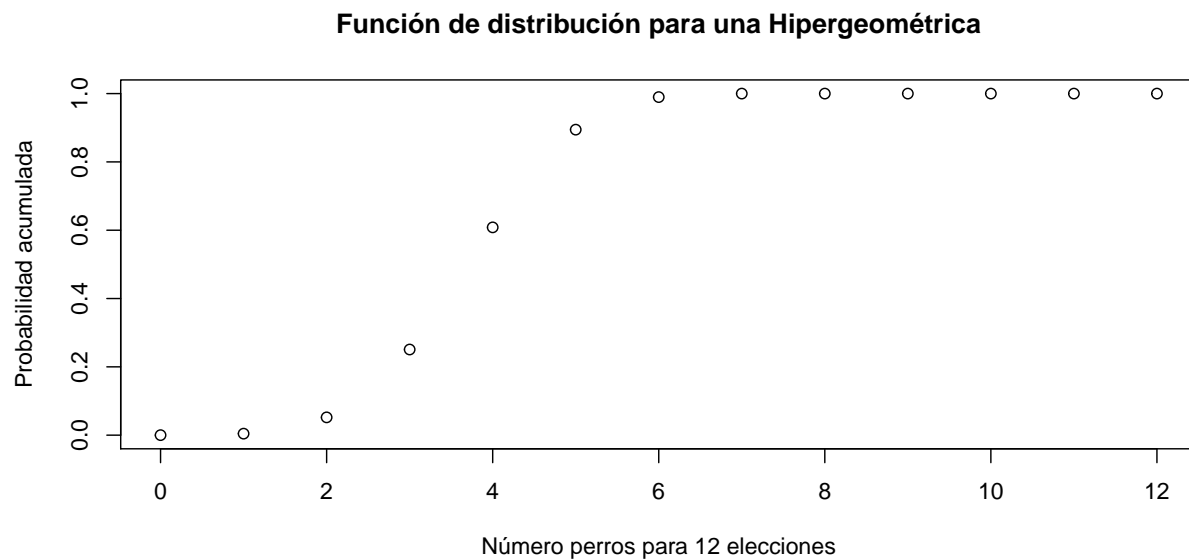
#Otro animal presente en el experimento
N = 13

#Nº de extracciones a realizar
k = 12

#Función de probabilidad para una Hipergeométrica
plot(0:12,dhyper(x = 0:12, m = M, n = N, k = k),
     title("Función de probabilidad para una Hipergeométrica"),
     xlab = "Número perros para 12 elecciones",
     ylab = "Probabilidad")
```



```
#Función de distribución para una Hipergeométrica (acumulada)
plot(0:12, phyper(q = 0:12, m = M, n = N, k = k),
     title("Función de distribución para una Hipergeométrica"),
     xlab = "Número perros para 12 elecciones",
     ylab = "Probabilidad acumulada")
```

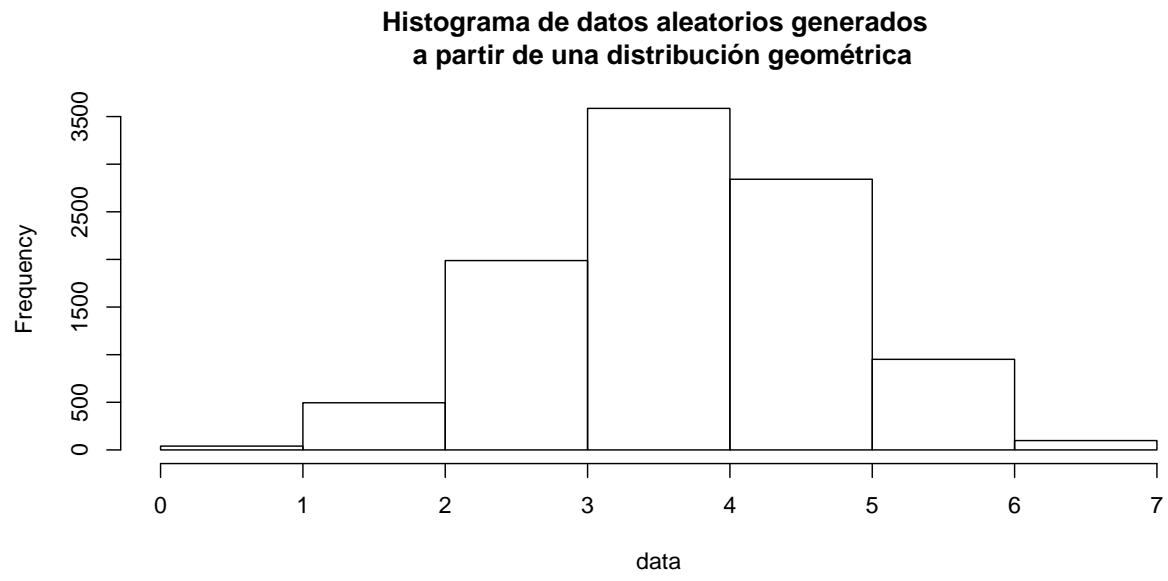


```
#La mediana
qhyper(p = 0.5, m = M, n = N, k = k)

## [1] 4

#Generamos datos aleatoria segun una hipergeométrica
rhyper(nn = 10000, m = M,
       n = N, k = k) -> data

#Representamos el histogrma de estos datos aleatorios
hist(data, breaks = 8,
     main = "Histograma de datos aleatorios generados \n a partir de una distribución geométrica")
```



Ejemplo en Python Vete al script de Python del presente tema 8.