Urna bayesiana

J. Abellán

7/11/2019

1. Distribución hipergeométrica

Conocido el contenido de la urna (N bolas, R de las cuales son rojas), si extraemos n bolas sin volver a meterlas

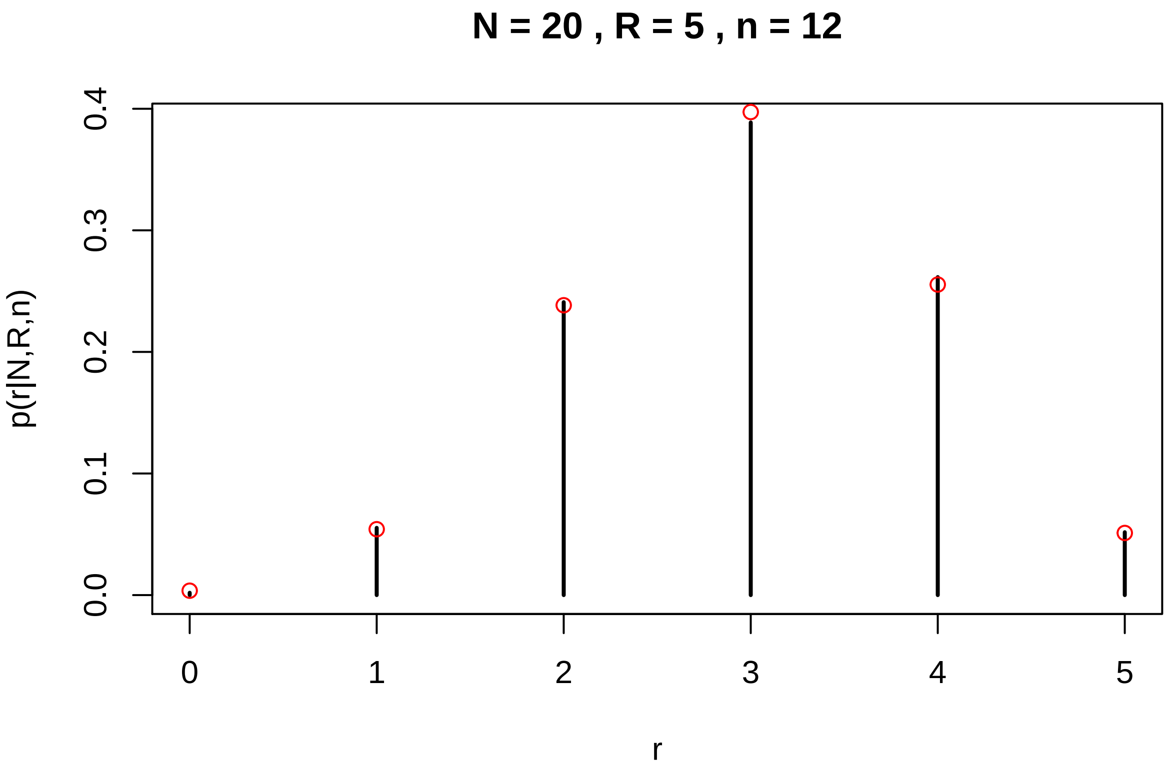
- ¿Cuál será la probabilidad de obtener r bolas rojas?
- La respuesta es bien conocida: la distribución hipergeométrica:

$$p(r|N,R,n) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

Comprobemos experimentalmente la teoría:

```
N <- 20  #bolas en la urna

R <- 5  #número de bolas rojas

# por tanto, probabilidad de rojo
pR <- R / N

n <- 12  #bolas extraidas

# la urna contiene R bolas rojas y N-R bolas negras
urna <- c( rep( 1, R ), rep( 0, N - R ) )

# Al sumar estamos contando el número de bola rojas
Nexp <- 10000

X <- replicate( Nexp, sum( sample( urna, n ) ) )

plot( table( X ) / Nexp,

      type = "h",

      #xlim = c( 0, n ),

      xlab = "r",

      ylab = "p(r|N,R,n)",

      main = paste("N =",N,", R =",R,", n =",n)

    )

r <- 0:n  #posibles valores para r (número de bolas rojas extraídas)
```

```r
# la solución teórica: la hipergeométrica
pr <- choose( R, r ) * choose( N - R, n - r ) / choose( N, n )

points( r, pr, col = 2 )
```



2. Inversión del problema.

Conocidos $n,\ r$ ¿qué podemos decir de los parámetros $N,\ R$?

- Supongamos que inicialmente no sabemos nada sobre R salvo $0 \le R \le N$, es decir, todos los valores de R son igualmente probables:

$$p(R|N\ I_o) = \frac{1}{N+1}$$

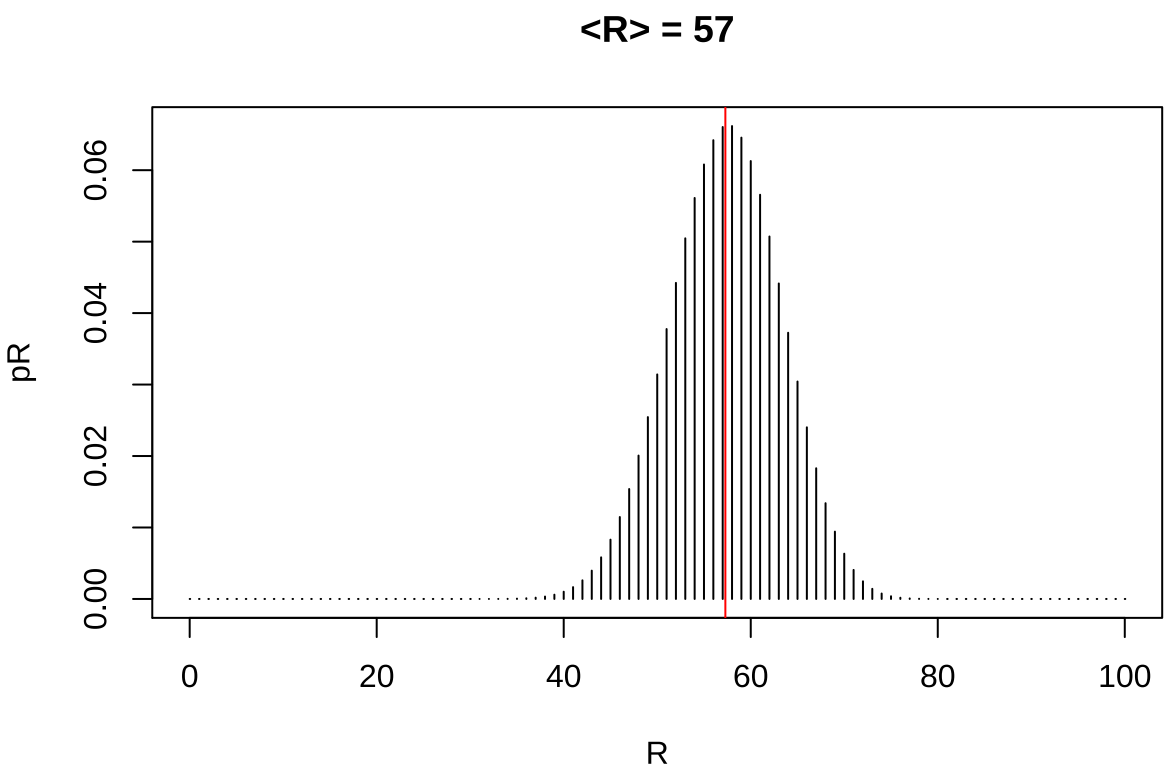
La respuesta es:

$$p(R|n\ r\ N\ I_o) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N+1}{n+1}}$$

Pongamos un ejemplo:

```r
N <- 100    #Bolas en la urna

n <- 40     #bolas extraidas

r <- 23     #de las extraidas r son rojas
```

```r
R <- 0:N   #posibles valores para R, el número de bolas rojas

pR <- choose( R, r) * choose( N - R, n - r ) / choose( N + 1, n + 1 )

Rm <- sum( R * pR )  #número promedio de bolas rojas

plot( R, pR,

      type = "h",

      main = paste( "<R> =", round( Rm ) )

    )

abline( v = Rm, col = 2 )
```



- Supongamos que a priori sabemos algo sobre R de la forma:

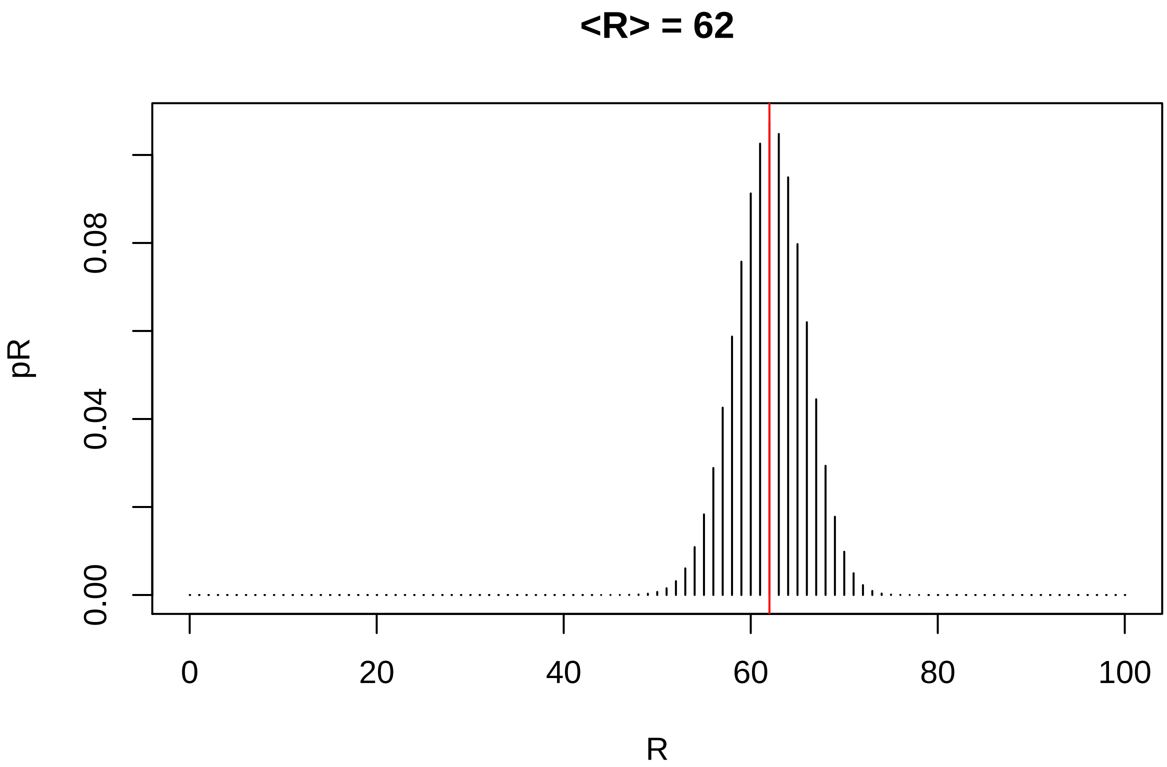
$$p(R|N\ \theta\ I_o) = \binom{N}{R}\theta^R(1-\theta)^{N-R}$$

donde θ es conocida. Dicho de otro modo, la urna ha sido preparada introduciendo las bolas rojas lanzando una moneda de probabilidad θ.

- La solución es de la forma:

$$p(R|n\ r\ N\ \theta\ I_o) = \binom{N-n}{R-r}\theta^{R-r}(1-\theta)^{(N-n)-(R-r)}$$

```r
N <- 100   #Número de bolas en la urna

p <- 0.65   #probabilidad con la que se rellena de bolas rojas

n <- 40    #número de bolas extraidas

r <- 23    #r de las extraidas son rojas

R <- 0:N   #los valores posibles de bolas rojas

pR <- choose( N - n, R - r ) * p^( R - r ) * ( 1 - p )^( N - n - R + r )

Rm <- sum( R * pR )  #Valor promedio de bolas rojas

plot( R, pR,

      type = "h",

      main = paste( "<R> =", round( Rm ) )

    )

abline( v = Rm, col = 2 )
```



3. Experimento binomial.

Lanzamos n veces una moneda y obtenemos r éxitos (caras)

- ¿Cuál es la probabilidad de éxito θ en cada lanzamiento?
- ¿Está trucada la moneda?

La respuesta, si no tenemos ninguna información previa sobre θ, será:

$$p(\theta|n, r) = \frac{(n+1)!}{r!(n-r)!}\theta^r(1-\theta)^{n-r}$$

```r
#valores posibles de tita: 0 <= tita <= 1
ntita <- 100       # valores 'continuos' de tita

tita <- seq( 0, 1, length.out = ntita )

dtita <- 1 / ntita  # separación entre valores de tita

n <- 20             #lanzamientos de moneda

r <- 8             #número de caras o éxitos obtenidos

#probabilidad teórica para p
ptita <- ( n + 1 ) * choose( n, r ) * tita^r * ( 1 - tita )^( n - r )

#normalización (caso continuo)
ptita <- ptita / sum( ptita * dtita )

# valor medio o más probable de tita
tita.mp <- sum( tita * ptita * dtita )

plot( tita, ptita,

      type = "l",

      xlab = expression( theta ),

      ylab = expression( p( theta / n, r ) ),

      main = paste( "<tita> =", round( tita.mp, 3 ), ", n =", n, ", r =", r )

    )

abline( v = tita.mp, col = 2 )
```

<tita> = 0.409 , n = 20 , r = 8