

PROBABILIDAD Y ESTADÍSTICA

Simulación de la definición frecuencial de Probabilidad

Resumen

A continuación se hace un resumen de algunas funciones disponibles en el software **R** que pueden utilizarse para simular computacionalmente los resultados teóricos de la definición frecuencial de Probabilidad.

1. Descripción de algunos comandos para el modelado de probabilidad

- **R** puede generar una secuencia de números enteros fácilmente. Digamos que queremos generar un arreglo con los posibles valores del experimento “arrojar un dado”. Simplemente tenemos que hacer:

```
> 1:6 ##dado de 6 caras
[1] 1 2 3 4 5 6
> a <- 1:6 ##o bien asignarle esos valores a una variable
> a
[1] 1 2 3 4 5 6
```

- Para generar el “muestreo”, es decir, el número que resultará al arrojar el dado, podemos usar la función **sample**:

```
> sample(a,1) ##aquí le decimos: tome el arreglo guardado en ‘a’
##y extraiga un elemento aleatoriamente.
[1] 6
> sample(a,6) ##podemos pedirle que extraiga al azar la cantidad
##de elementos que querramos.
[1] 2 3 5 1 6 4
```

como se puede observar en la último resultado, en el muestreo no hay repetición de valores. Es más, si pedimos:

```
> sample(a,10)
```

obtenemos un mensaje de error que básicamente nos dice que no podemos solicitar como resultado del muestreo más elementos que los que tenemos en el arreglo. Esto se debe a que por defecto la función **sample** realiza un muestreo “sin reposición”. Para convertirlo en un muestro con reposición, simplemente agregamos una tercera variable al comando:

```
> sample(a,10,rep=T)
[1] 3 2 4 3 1 6 1 3 2 4
```

cuya salida estaría simulando el experimento de arrojar un dado 10 veces (o 10 dados una sola vez).

- Una posibilidad interesante es armar una función basada en **sample** donde directamente podamos ingresar el número de veces que queremos arrojar el dado:

```
> arrojarDado = function(n) sample(1:6,n,rep=T)
```

donde declaramos a **arrojarDado** como una función que tiene a **sample** con todas sus variables internamente, con lo cual ya estaríamos modelando el lanzamiento del dado, necesitando como variable de entrada sólo la cantidad *n* de veces que necesitamos se arroje el dado (todo en una sola línea de instrucciones!).

Tendremos entonces por ejemplo:

```
> arrojarDado(100)
[1] 6 2 3 4 2 5 1 3 2 4 3 1 6 1 3 2 4 3 3 3 2 4 4 5 3 5 5
[29] 3 1 3 2 6 6 1 2 5 5 5 2 1 3 3 4 1 3 3 3 6 3 1 5 2 2 3 3
[57] 5 3 3 3 3 2 3 3 5 1 4 6 2 1 4 3 4 2 2 3 2 4 2 3 3 1 3 1
[85] 1 6 2 4 5 3 3 1 4 3 5 6 4 3 3 1
```

- Entonces si queremos saber las frecuencias de cada resultado, podemos usar la función `table`, la cual es útil para construir tablas de contingencia:

```
> table(A)
A
 1  2  3  4  5  6
15 17 35 13 12  8
```

A partir de aquí es sencillo transformar las frecuencias absolutas en relativas simplemente dividiendo la tercer fila de la tabla por la cantidad total de observaciones (en este caso 100):

```
> prop.table(table(A))
A
 1    2    3    4    5    6
0.15 0.17 0.35 0.13 0.12 0.08
```

donde mediante la función `prop.table` podemos expresar los valores en la tabla como fracciones de la totalidad de elementos. Para poder acceder a los valores de la tabla de manera sencilla, la transformamos en un “data frame”, que no es otra cosa que una lista de vectores de igual longitud con los datos de una tabla. Para ello hacemos:

```
> B = prop.table(table(A))
> t = as.data.frame(B)
## vemos que sale de t:
> t
  A Freq
1 1 0.15
2 2 0.17
3 3 0.35
4 4 0.13
5 5 0.12
6 6 0.08
```

donde vemos que tenemos una columna con el vector “A” con los posibles valores del dado (`t[,1]`) y otra con el vector “Freq” con los valores de las frecuencias relativas de cada resultado (`t[,2]`).

- Por último, podemos graficar el experimento mediante las funciones `plot` y `lines`:

```
##creamos el rango para abscisas y ordenadas del gráfico:
> xrango <- c(1,6)
> yrango <- c(0.05,0.3)
> plot(xrango, yrango, type="n")
> lines(t[,1], t[,2], type="l")
> lines(t[,1], t[,2], type="p")
##agregamos una línea horizontal indicando el valor de la
##probabilidad buscada:
> abline(h=1/6, col="red")
```

donde `type="n"`, `"l"` y `"p"` indican que “no se realice el gráfico”, “se dibujen líneas” y “se dibujen puntos”, respectivamente. En el código anterior se puede ver que se utilizó `c()` para crear el rango en x e y . En **R**, ese comando es una de las formas posibles de crear “arreglos”.

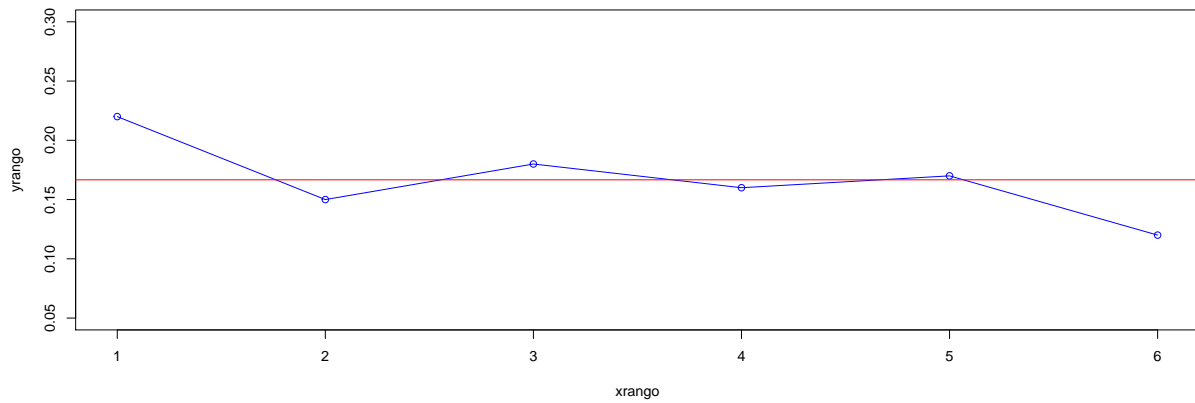


Figura 1: Probabilidad frecuencial con 100 observaciones.

Podemos ver que con 1000 veces que arrojemos el dado estamos bastante cerca de conseguir el valor real de la probabilidad:

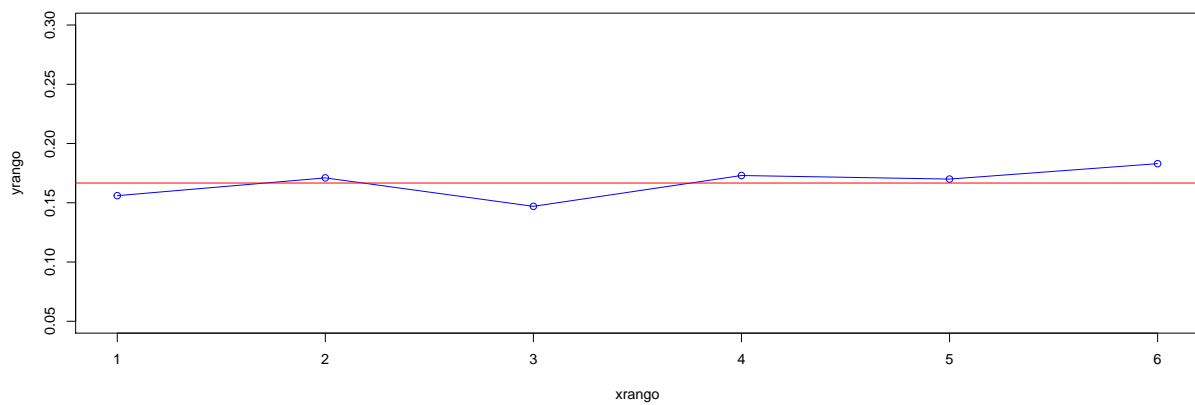


Figura 2: Probabilidad frecuencial con 1000 observaciones.

2. Modelado del experimento de la moneda

Realizar, basado en las funciones descritas en la sección anterior, el modelado de “arrojar una moneda al aire 10 veces”, donde el evento de interés es “número de caras obtenidas”. Para este caso tener en cuenta que en abscisas deben estar la cantidad de “caras” obtenidas y en ordenadas la frecuencia relativa (probabilidad en el límite) de “caras” observadas. Observar el comportamiento del experimento para una repetición de la experiencia de 100, 500, 1000 y 5000 veces. Discutir los resultados y todo procedimiento que crea conveniente.