

**PRÁCTICAS DE
ESTADÍSTICA
CON *R***

PRÁCTICA 3: DISTRIBUCIONES DE PROBABILIDAD DISCRETAS Y CONTINUAS

3.1 Distribuciones discretas

Las principales ideas que vamos a ir desarrollando a lo largo de la primera parte de esta práctica hacen referencia al ejercicio siguiente:

■ Ejemplo 3-1: Simular 1000 veces el experimento consistente en lanzar 5 dados. Considérese la variable aleatoria *número de unos y doses obtenidos entre los 5 dados*. Hacer un gráfico de frecuencias relativas y compararlo con la función de probabilidad de una variable aleatoria $B(5, 1/3)$.

En R es posible calcular probabilidades para las principales variables aleatorias discretas. Los nombres reservados a algunas de esas distribuciones son:

- Binomial: `binom`
- Hipergeométrica: `hyper`
- Poisson: `pois`
- Binomial negativa: `nbinom`
- Geométrica: `geom.`

Los nombres anteriores, sin embargo, no son sentencias de R que produzcan una salida válida. Es necesario anteponerles los prefijos “**d**” para la función de masa o función de probabilidad, “**p**” para la función de distribución acumulada, “**r**” para generar valores aleatorios y “**q**” para la función cuantil. A continuación vemos algunos ejemplos.

Calcular la probabilidad de que una variable aleatoria binomial de parámetros $n=10$, $p=0.3$ tome el valor 4:

```
> dbinom(4,size=10,prob=0.3)
[1] 0.2001209
```

Probabilidad acumulada en el valor 5 de una variable aleatoria de Poisson de parámetro $\lambda=2$:

```
> ppois(5,2)
[1] 0.9834364
```

Generar 10 valores aleatorios de una distribución de Poisson de parámetro 3,52:

```
> rpois(10,3.52)
[1] 4 3 3 3 4 5 4 1 2 1
```

3.2 Cómo simular en R el lanzamiento de un dado

En primer lugar vamos a simular el lanzamiento de un dado una vez. Para ello utilizamos la función `sample`. Mediante esta función se escogen al azar un número de elementos de tamaño especificado entre todos los elementos de un cierto vector. Por ejemplo, podemos usarla para escoger un número al azar entre los naturales del 1 al 6 (lanzamiento de un dado una vez). Previamente definiremos el vector `dado`, que representa todos los posibles resultados que pueden producirse al lanzar un dado:

```
> dado<-c(1:6)
> sample(dado,1)
[1] 5
```

Para simular el lanzamiento más de una vez, por ejemplo 10, evidentemente debemos indicar la opción con reemplazamiento (por defecto extrae sin reemplazamiento):

```
> sample(dado,10,replace=T)
[1] 1 5 3 6 6 4 2 3 2 2
```

3.3 Función de masa

En el ejemplo 4-1 se necesita simular el lanzamiento de cinco dados 1000 veces, lo que hacemos generando una matriz de 1000 filas por 5 columnas del siguiente modo:

```
>lanzamientos<-
matrix(sample(dado,5000,replace=T),nrow=1000)
> # Por ejemplo, los resultados correspondientes a las
filas 250 a 260 son los que vienen a continuación
> lanzamientos[250:260,]
      [,1] [,2] [,3] [,4] [,5]
[250,]    2    1    2    5    4
[251,]    1    1    2    3    5
[252,]    4    5    4    1    1
[253,]    2    3    4    5    6
[254,]    5    1    6    5    4
[255,]    3    3    6    4    6
[256,]    5    6    2    1    2
[257,]    5    6    1    5    3
[258,]    4    5    4    2    1
[259,]    5    6    2    6    2
[260,]    6    1    2    1    5
```

El resultado anterior no es del todo práctico en nuestro caso, ya que más que en los resultados concretos de los 1000 lanzamientos de 5 dados estamos interesados en el número de unos y doses que se obtienen en cada lanzamiento. Para ello podemos usar la distribución $B(1,2/6)$ que toma el valor 1 si ha salido uno o dos y 0 en otro caso, y generar 5000 valores al azar de esta distribución.

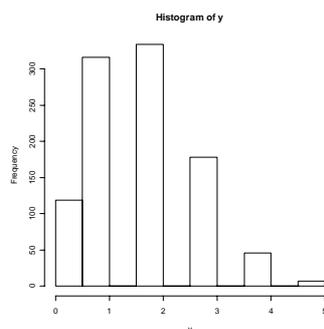
```
> x<-matrix(rbinom(5000,1,2/6),nrow=1000)
> # Por ejemplo, los resultados correspondientes a
las diez primeras filas son los siguientes
> x[1:10,]
      [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    1    0    1
[2,]    0    0    1    0    0
[3,]    0    1    1    1    1
[4,]    0    0    1    1    1
[5,]    0    1    1    0    0
[6,]    0    0    1    1    1
[7,]    1    1    0    1    1
[8,]    1    0    0    1    0
[9,]    1    0    0    1    0
[10,]   0    0    1    0    0
```

Construimos a continuación el vector formado por la suma de los valores de las filas de x:

```
> y<-c(margin.table(x,1))
> y
 [1] 2 1 4 3 2 3 4 2 2 1 2 0 4 0 3 1 3 3 2 0 3 3
 0 4 3 3 2 2 3 1 4 2 2 1 3 3 0
...
[963] 4 0 1 3 1 2 2 2 2 3 1 2 0 2 2 1 2 1 4 2 1 1
 3 2 1 2 2 2 3 1 2 2 2 3 3 0 3
[1000] 2
```

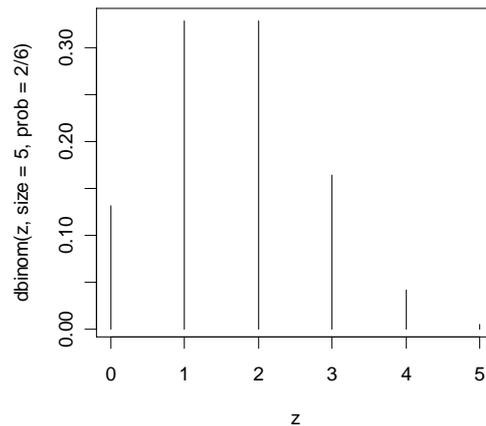
Construimos un histograma con los valores anteriores:

```
> hist(y)
```



Vamos a dibujar ahora la función de masa de una distribución $B(5,2/6)$ y a compararla con el histograma obtenido. Si queremos conservar el gráfico anterior para que esté disponible más adelante podemos abrir una nueva ventana gráfica con la instrucción `windows()`:

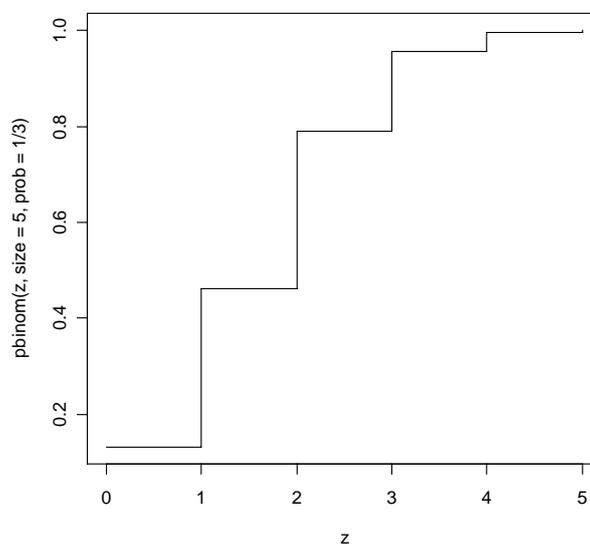
```
> windows()  
> z<-0:5  
> plot(z,dbinom(z,size=5,prob=1/3),type="h")
```



3.4 Función de distribución de una variable aleatoria discreta

Utilizando la función `pbinom` podemos dibujar la función de distribución acumulada de la variable aleatoria que representa el número de unos y doses que aparecen al lanzar cinco dados. Utilizamos la opción `type="s"` para construir la función en forma de escalera.

```
> plot(z,pbinom(z,size=5,prob=1/3),type="s")
```



3.5 Distribuciones continuas

En *R* es posible calcular probabilidades para las principales variables aleatorias continuas. Los nombres reservados a las distribuciones continuas más importantes son:

- Uniforme: **unif**
- Exponencial: **exp**
- Normal: **norm**
- χ^2 : **chisq**
- t de Student: **t**
- F de Snedecor: **f**

Los nombres anteriores, igual que para las distribuciones discretas, no son sentencias de *R* que produzcan una salida válida. Es necesario anteponerles los prefijos "d" para la función de densidad, "p" para la función de distribución acumulada, "r" para generar valores aleatorios y "q" para la función cuantil (inversa de la función de distribución).

Por ejemplo, si queremos conocer la ordenada de la función de densidad de una variable aleatoria $N(-2,4)$, valor que por otro lado no tiene ninguna utilidad práctica, haremos:

```
> dnorm(3, -2, 4)
[1] 0.04566227
```

Este resultado lo podemos obtener así mismo mediante la función de densidad de la variable aleatoria normal:

```
> exp(-(1/2)*((3+2)/4)^2)/(4*sqrt(2*pi))
[1] 0.04566227
```

Generemos 10 valores aleatorios de una distribución $U(-5,3)$:

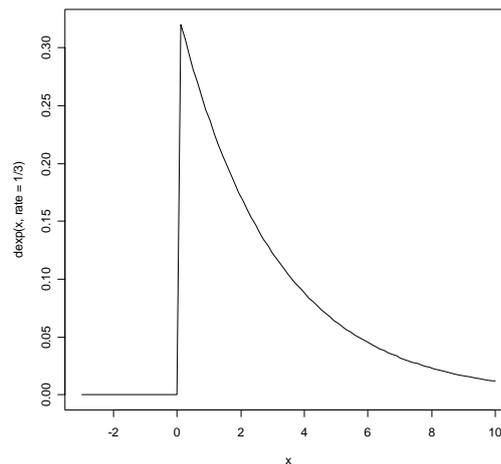
```
> runif(10, -5, 3)
[1] -1.8616074 -0.2718027 -3.4057091  1.3315307
-1.6892947  2.6328703
[7] -2.8891604 -1.1120779 -0.1570343 -0.1074194
```

3.6 Función de densidad

En el primer ejemplo vamos a dibujar la función de densidad de una variable aleatoria exponencial de parámetro 3. Posteriormente, veremos cómo se aproxima una distribución binomial por una normal.

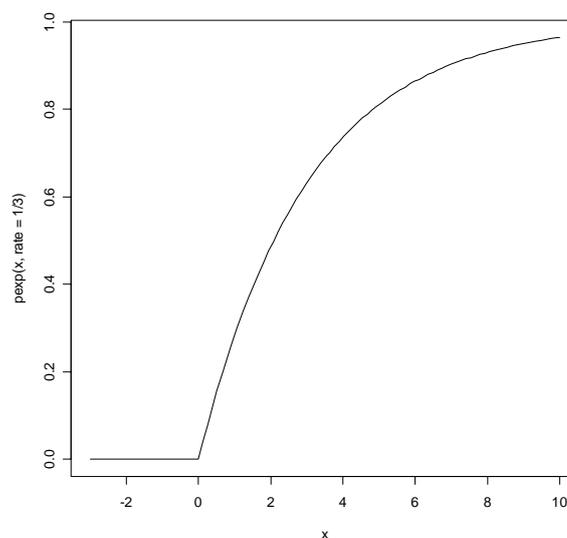
■ **Ejemplo 3-6-1:** Dibujar la función de densidad y la función de distribución de una variable aleatoria exponencial de media 3.

```
> # Una forma muy simple de dibujar la función de densidad es usar la función "curve"  
> curve(dexp(x,rate=1/3),from=-3,to=10)
```



```
> # Una alternativa es utilizar la función "plot",  
pero previamente debemos definir un vector z en el que  
evaluaremos la función de densidad  
> z<-seq(-3,10,0.1)  
> plot(z,dexp(z,rate=1/3),type="l")  
> # Con la opción type="l" conseguimos dibujar el  
gráfico como una línea
```

```
> # Dibujamos ahora la función de distribución  
> curve(pexp(x,rate=1/3),from=-3,to=10)
```



■Ejemplo 3-6-2: La probabilidad de que un cierto componente electrónico falle a lo largo de un mes es 0,1. Si hay 500 componentes en funcionamiento, calcular la probabilidad de que en un mes: 1º) fallen como mucho 50 componentes, 2º) fallen entre 30 y 60, 3º) fallen más de 65. Resolver, en primer lugar, el problema utilizando la distribución binomial y despues mediante la aproximación de la binomial a la normal. Comprobar gráficamente tal aproximación.

```
> # Calculamos las probabilidades exactas pedidas
mediante la distribución binomial
```

```
> pbinom(50,500,0.1)
[1] 0.5375688
```

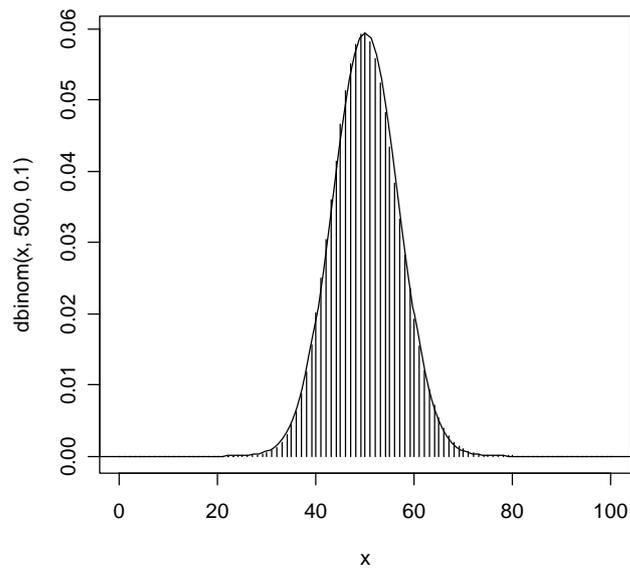
```
> pbinom(60,500,0.1)-pbinom(29,500,0.1)
[1] 0.9376227
```

```
> 1-pbinom(65,500,0.1)
[1] 0.01269429
```

Como es sabido, la distribución $B(n, p)$ se aproxima a la $N(np, \sqrt{np(1-p)})$, y la aproximación es buena si $np > 5$ y $n(1-p) > 5$. Aplicando este resultado obtendremos, de manera aproximada, las probabilidades anteriores.

```
> n<-500;p<-0.1
> n*p;n*(1-p)
[1] 50
[1] 450
> media<-n*p;destip<-sqrt(n*p*(1-p))
> pnorm(50.5,media,destip)
[1] 0.5297079
> pnorm(60.5,media,destip)-
+ pnorm(29.5,media,destip)
[1] 0.9401159
> 1-pnorm(65.5,media,destip)
[1] 0.01042738
> # En todos los casos hemos tenido en cuenta la
corrección por continuidad

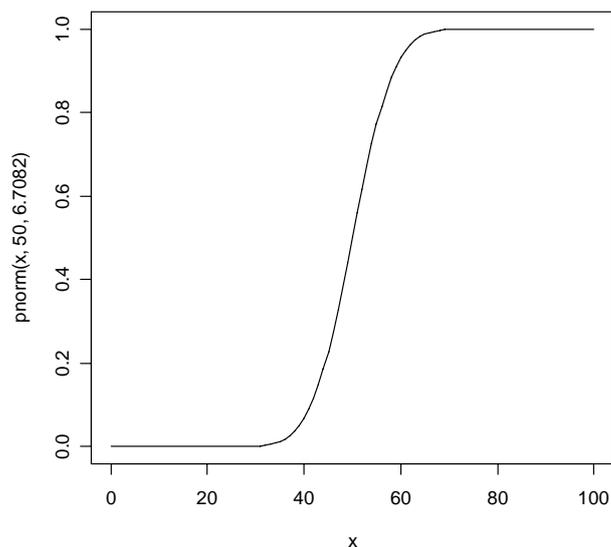
> # Dibujamos la función de masa de la
distribución B(500,0.1) y le superponemos la
función de densidad de la N(media,destip)
> z<-0:100
> plot(z,dbinom(z,500,0.1),type="h")
> curve(dnorm(x,media,destip),add=T)
> # Notemos que en la sentencia anterior hemos
utilizado la opción add=T (TRUE) para indicarle a
R que superponga este gráfico al anterior
```



3.7 Función de distribución de una variable aleatoria continua

Vamos a dibujar ahora la función de distribución de la variable aleatoria normal del ejemplo 3-6-2. Como se trata de una variable aleatoria continua utilizaremos la opción **type="l"** para dibujarla como una línea, obteniendo la clásica curva en forma de S:

```
> plot(x, pnorm(x, 50, 6.7082), type="l")
```



Alternativamente, podríamos construir este gráfico mediante la sentencia siguiente:

```
> curve(pnorm(x, media, destip), from=0, to=100)
```

PRÁCTICA 3: EJERCICIOS

■ **Ejercicio 3-1:** Representar gráficamente la función de probabilidad y la función de distribución de una variable aleatoria que sigue una distribución de Poisson de parámetro 4,32.

■ **Ejercicio 3-2:** De un lote de 100 piezas, se sabe que 5 de ellas no cumplen con los requisitos de aceptación. Si se toman al azar 10, ¿cuál es la probabilidad de encontrar uno o menos elementos defectuosos en la muestra? Resolverlo para dos casos:

- a) El muestreo se hace sin reemplazamiento, y
- b) el muestreo se hace con reemplazamiento.

■ **Ejercicio 3-3:** Dibujar la función de densidad, y la función de distribución de una variable aleatoria uniforme continua de parámetros (-3,4).

■ **Ejercicio 3-4:** Dibujar, superpuestas en un mismo gráfico, las funciones de densidad de las siguientes variables aleatorias normales: $N(3,0.75)$, $N(0,1)$, $N(-2,1)$, y $N(-2,2)$.

■ **Ejercicio 3-5:** Dibujar, superpuestas en un mismo gráfico, las funciones de densidad de las siguientes variables aleatorias: t_2 y $N(0,1)$.