Cálculo de probabilidades con el Lenguaje R

© José Luis Romero Béjar

2023-09-03

En este guión breve se pretende que el lector se familiarice con el **cálculo de probabilidades, cuantiles y generación de datos aleatorios** para distintas distribuciones de probabilidad unidimensionales discretas y continuas, mediante el uso del **Lenguaje R**.

1. Distribución Binomial

1.1. *Función masa de probabilidad*

Para evaluar la función masa de probabilidad de una distribución binomial de parámetros n y p,

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}; x = 0, 1, 2, \dots, n,$$

se hace uso de la siguiente función disponible en el paquete base del lenguaje R:

$$dbinom(x, size = n, prob = p)$$

Está función devuelve la probabiliad $P[X = x]$ de una variable X distribuida mediante una ley binomial, $X \sim B(size = n, prob = p)$.

Ejemplo 1 Para poner en práctica esta función se propone el siguiente ejercicio.

En un estudio sociológico se ha tomado una muestra de 50 individuos. Por medio de estudios inferenciales se ha estimado que la probabilidad de que el individuo sea votante de un determinado partido político es 0.35. Es evidente que la variable aleatoria que cuenta el número de individuos que votan a este partido político sigue una distribución binomial de parámetros $n = 50$ y $p = 0.35$. Calcular la probabilidad de que exactamente 20 individuos sean votantes de este partido.

Aquí se pide calcular $P[X = 20]$ siendo X una variable aleatoria que representa el número de votantes de este partido político, $X \sim B(size = 50, prob = 0.35)$.

```r
n=50
p=0.35
valor=20
dbinom(valor,size=n,prob=p)
```

```
## [1] 0.08750881
```

En vista del resultado obtenido se deduce que la probabilidad de que 20 individuos voten a este partido político es de 0.08750881.

1.2. *Función de distribución*

Para evaluar la función de distribución de probabilidad de una distribución binomial de parámetros n y p se hace uso de la función disponible en el paquete base del lenguaje R siguiente:

$$pbinom(x, size = n, prob = p, lower.tail = TRUE)$$

Está función devuelve la probabiliad $P[X \leq x]$ de una variable X distribuida mediante una ley $B(size = n, prob = p)$. Si el parámetro $lower.tail = FALSE$ calcula la probabilidad complementaria a la función de distribución, esto es: $P[X > x]$.

Ejemplo 2 Continuando con el ejemplo anterior, se pide obtener la probabilidad de que al menos 15 individuos sean votantes de este partido político.

En este caso se pide obtener la probabilidad $P[X \geq 15] = 1 - P[X < 15] = 1 - P[X \leq 14]$

```
valor=14
1-pbinom(valor,size=n,prob=p, lower.tail=TRUE)
```

```
## [1] 0.812223
```

También se podría haber obtenido esta probabilidad considerando que $P[X \geq 15] = P[X > 14]$ utilizando así el parámetro $lower.tail = FALSE$.

```
valor=14
pbinom(valor,size=n,prob=p, lower.tail=FALSE)
```

```
## [1] 0.812223
```

En cualquiera de los dos casos, la probabilidad de que al menos 15 individuos voten a este partido político es de 0.812223.

En este ejemplo, también nos podríamos plantear obtener la probabilidad de que entre 10 y 15 individuos (ambos inclusive) sean votantes de este partido. En este caso se pide $P[10 \leq X \leq 15] = P[X \leq 15] - P[X \leq 9]$.

```
valor1=15
valor2=9
pbinom(valor1,size=n,prob=p, lower.tail=TRUE)-pbinom(valor2,size=n,prob=p, lower.tail=TRUE)
```

```
## [1] 0.2734065
```

A la vista del resultado se tiene que la probabilidad de que a este partido político lo voten entre 10 y 15 individuos es 0.2734065.

1.3. Obtención de *cuantiles* y *generación* de valores aleatorios binomiales

También suele ser de interés obtener distintos cuantiles de una distribución de probabilidad. Un **cuantil** es aquel valor a que deja a su izquierda una proporción q de valores, es decir, un valor a tal que $P[X \leq a] = q$.

La función que determina el(los) cuantil(es) de una distribución binomial de parámetros n y p, disponible en el paquete base del lenguaje R, es la siguiente:

$$qbinom(a, size = n, prob = p, lower.tail = TRUE)$$

De la misma forma que antes, si el parámetro $lower.tail = FALSE$ se obtendría el valor a que deja por encima un porcentaje q de valores, esto es: $P[X > a] = q$.

Ejemplo 3 Continuando con el ejemplo anterior, se pide calcular el valor de la variable tal que deja a su derecha un 70% de las observaciones.

Es evidente que el valor de la variable que deja a su derecha un 70% de las observaciones es el mismo que deja a su izquierda el 30% restante, por lo que se pide calcular el valor a tal que $P[X \leq a] = 0.30$.

```
q=0.30
qbinom(q,size=n,prob=p, lower.tail=TRUE)
```

```
## [1] 16
```

Finalmente, también puede ser de interés generar una muestra aleatoria de un tamaño concreto de una distribución binomial. La función de R, disponible en su paquete base, que permite extraer un conjunto de r números aleatorios de una distribución binomial de parámetros n y p, es la siguiente:

$$rbinom(r, size = n, prob = p)$$

Ejemplo 4 Continuando con el ejemplo anterior, se pide generar 10 valores aleatorios de esta distribución binomial.

```
r=10
rbinom(r,size=n,prob=p)
```

```
##  [1] 20 12 20 15 13 16 16 10 21 18
```

2. Distribución de Poisson

2.1. *Función masa de probabilidad*

Para evaluar la función masa de probabilidad de una distribución de Poisson de parámetro λ,

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}; x = 0, 1, 2, \ldots,$$

se hace uso de la función disponible en el paquete base del lenguaje R siguiente:

$$dpois(x, lambda)$$

Está función devuelve la probabiliad $P[X = x]$ de una variable X distribuida mediante una ley de Poisson, $X \sim P(\lambda)$.

Ejemplo 5 Para poner en práctica esta función se propone el siguiente ejercicio.

En un determinado servidor el número medio de conexiones entrantes rechazadas cada hora debidas a uno concreto de sus puertos es 0.90. Es evidente que la variable que cuenta el número de conexiones rechazadas por hora debidas a este puerto sigue una ley de Poisson de parámetro $\lambda = 0.90$.

Se pide calcular la probabilidad de que en una hora, este puerto rechace exactamente 4 conexiones.

Se pide calcular $P[X = 4]$.

```
lambda=0.90
valor=4
dpois(valor,lambda = lambda)
```

```
## [1] 0.0111146
```

En vista del resultado obtenido se deduce que la probabilidad de que este servidor rechace 4 conexiones entrantes en una hora es de 0.0111146.

2.2. *Función de distribución*

Para evaluar la función de distribución de probabilidad de una distribución de Poisson de parámetro λ se hace uso de la función disponible en el paquete base del lenguaje R siguiente:

$$ppois(x, lambda, lower.tail = TRUE)$$

Está función devuelve la probabiliad $P[X \leq x]$ de una variable X distribuida mediante una ley $P(\lambda)$. Si el parámetro *lower.tail = FALSE* calcula la probabilidad complementaria a la función de distribución, esto es: *P[X > x]*.

Ejemplo 6 Continuando con el ejemplo anterior, se pide obtener la probabilidad de que al menos se rechacen 15 conexiones en una hora debidas a éste puerto del servidor.

En este caso se pide obtener la probabilidad $P[X >= 15] = 1 - P[X < 15] = 1 - P[X <= 14]$

```
valor=14
1-ppois(valor,lambda = lambda, lower.tail=TRUE)
```

```
## [1] 6.783463e-14
```

También se podría haber obtenido esta probabilidad considerando que $P[X >= 15] = P[X > 14]$ utilizando así el parámetro *lower.tail = FALSE*.

```
valor=14
ppois(valor,lambda = lambda, lower.tail=FALSE)
```

```
## [1] 6.781511e-14
```

En cualquiera de los dos casos, la probabilidad de que al menos se rechacen 15 conexiones en una hora debidas a éste puerto del servidor es de $6.781511e - 14$ (muy pequeña).

En este ejemplo, también nos podríamos plantear obtener la probabilidad de que se rechacen entre 10 y 15 conexiones (ambos inclusive) por éste puerto del servidor. En este caso se pide $P[10 <= X <= 15] = P[X <= 15] - P[X <= 9]$.

```
valor1=15
valor2=9
ppois(valor1,lambda = lambda, lower.tail=TRUE)-ppois(valor2,lambda = lambda, lower.tail=TRUE)
```

```
## [1] 4.251957e-08
```

A la vista del resultado se tiene que la probabilidad de que se rechacen entre 10 y 15 conexiones (ambos inclusive) por éste puerto del servidor es $4.251957e - 08$ (también muy pequeña).

2.3. Obtención de *cuantiles* y *generación* de valores aleatorios de Poisson

Como se definió anteriormente, un **cuantil** es aquel valor a que deja a su izquierda una proporción q de valores, es decir, un valor a tal que $P[X \leq a] = q$.

La función que determina el(los) cuantil(es) de una distribución de Poisson de parámetro λ, disponible en el paquete base del lenguaje R, es la siguiente:

$$qpois(a, lambda, lower.tail = TRUE)$$

De la misma forma que antes, si el parámetro *lower.tail = FALSE* se obtendría el valor a que deja por encima un porcentaje q de valores, esto es: $P[X > a] = q$.

Ejemplo 7 Continuando con el ejemplo anterior, se pide calcular el valor de la variable tal que deja a su derecha un 10% de las observaciones.

Es evidente que el valor de la variable que deja a su derecha un 10% de las observaciones es el mismo que deja a su izquierda el 90% restante, por lo que se pide calcular el valor a tal que $P[X \leq a] = 0.90$.

```
q=0.90
qpois(q,lambda = lambda, lower.tail=TRUE)
```

```
## [1] 2
```

Finalmente, puede ser también de interés generar una muestra aleatoria de tamaño concreto de una distribución de Poisson. La función de R, disponible en su paquete base, que permite extraer un conjunto de r números aleatorios de una distribución de poisson de parámetro λ, es la siguiente:

$$rpois(r, lambda)$$

Ejemplo 8 Continuando con el ejemplo anterior, se pide generar 10 valores aleatorios de esta distribución de poisson.

```
r=10
rpois(r,lambda = lambda)
```

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

3. Distribución Normal

3.1. *Función de densidad*

Para evaluar la función de densidad de probabilidad de una variable aleatoria X que sigue una distribución Normal de parámetros μ y σ, $X \sim N(\mu, \sigma)$,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R},$$

se hace uso de la función disponible en el paquete base del lenguaje R siguiente

$$dnorm(x, mean = media, sd = desv.tip)$$

Es importante advertir que si no se pasan a estas función los parámetros de media y desviación típica, ésta adopta $\mu = 0$ y $\sigma = 1$, es decir que trabajaría con la distribución *Normal Estándar o Tipificada*.

Ejemplo 9 Para practicar con esta función y las que se introducen a continuación consideramos el siguiente ejemplo.

Se sabe que la distribución de probabilidad que sigue la cantidad de agua(l) que es recogida en una determinada estación meteorológica es una Normal del media $\mu = 3$ y desviación típica $\sigma = 0.3$.

Se pide evaluar la función de densidad de esta distribución en el punto 2.9.

```
media=3
desv.tip=0.30
valor=2.9
dnorm(valor, mean=media, sd=desv.tip)
```

```
## [1] 1.257944
```

Es directo comprobar que la función de densidad para la medición indicada es 1.257944.

3.2. *Función de distribución*

Para evaluar la función de distribución de probabilidad de una distribución Normal de parámetros μ y σ se hace uso de la función disponible en el paquete base del lenguaje R siguiente:

$$pnorm(x, mean = media, sd = desv.tip, lower.tail = TRUE)$$

Esta función devuelve la probabilidad $P[X \leq x]$ de una variable X distribuida mediante una ley $N(\mu, \sigma)$. Si el parámetro *lower.tail = FALSE* calcula la probabilidad complementaria a la función de distribución, esto es: *P[X > x]*.

Ejemplo 10 Continuando con el ejemplo anterior se pide hallar la probabilidad de que la cantidad de precipitación recogida en la estación meteorológica esté entre 2.8 y 3 litros.

En este caso se pide calcular $P[2.8 \leq X \leq 3]$.

```
valor1=3
valor2=2.8
pnorm(valor1,mean=media,sd=desv.tip)-pnorm(valor2,mean=media,sd=desv.tip)
```

```
## [1] 0.2475075
```

A la vista del resultado se observa que la probabilidad de que se recojan entre 2.8 y 3 litros es 0.2475075.

3.3. Obtención de *cuantiles* y *generación* de valores aleatorios de una Normal

Como en las distribuciones discretas anteriores (binomial y poisson) también se pueden obtener distintos cuantiles de la distribución normal. Un **cuantil** es aquel valor a que deja a su izquierda una proporción q de valores, es decir, un valor a tal que $P[X \leq a] = q$.

La función que determina el(los) cuantil(es) de una distribución normal de parámetros μ y σ, disponible en el paquete base del lenguaje R, es la siguiente:

$$qnorm(a, mean = media, sd = desv.tip, lower.tail = TRUE)$$

Como antes, si el parámetro *lower.tail = FALSE* se obtendría el valor a que deja por encima un porcentaje q de valores, esto es: $P[X > a] = q$.

Ejemplo 11 Continuando con el ejemplo anterior se quiere calcular el 15% de las precipitaciones más grandes caídas en la estación meteorológica. Éste es aquel valor que deja a su derecha el 15% de las observaciones y, por tanto, a su izquierda el 85% de las observaciones restantes.

En este caso se pide calcular el cuantil $q = 0.85$.

```
q=0.85
qnorm(q,mean=media,sd=desv.tip)
```

```
## [1] 3.31093
```

A la vista del resultado este cuantil corresponde a una precipitación total de 3.31093 litros.

Finalmente, la función del paquete base del lenguaje R que permite la generación de un conjunto de valores aleatorios provenientes de una distribución Normal de parámetros μ y σ es la siguiente:

$$rqnorm(r, mean = media, sd = desv.tip)$$

Ejemplo 12 Continuando con el ejemplo anterior se pide generar 15 cantidades aleatorias de la distribución normal con la que se está trabajando.

```
r=15
rnorm(r,mean=media,sd=desv.tip)
```

```
##  [1] 2.825225 2.727484 3.160000 2.947644 3.232215 2.906802 2.845398 2.622243
##  [9] 2.616394 2.998108 2.771471 2.699943 3.129482 3.649031 2.551571
```

Referencias

1.- The R Project for Statistical Computing (https://www.r-project.org/) (Accedido el 1 de octubre de 2022).

2.- Guía práctica de Bioestadística con R (https://digibug.ugr.es/bitstream/handle/10481/70802/Practica 4_bioestad%c3%adsticaR.pdf?sequence=1&isAllowed=y) (Accedido el 1 de octubre de 2022).