

# Variables aleatorias

Selección de actividades  
resueltas

© Jose Fco. Martínez Boscá, Arnau Mir Torres, Lluís  
M. Pla Aragonés &  
Ángel A. Juan (Editor)

© FUOC 2009

## Introducción

Resolver problemas aleatorios usando la Teoría de las probabilidades puede resultar muy complicado debido a lo complejo que se puede convertir el espacio muestral.

Por ejemplo, imaginemos que estamos interesados en estudiar las peticiones a un servidor web de una empresa por parte de un conjunto de usuarios. Estudiar el espacio muestral de dicho estudio puede resultar muy complicado. Entonces, debemos preguntarnos, ¿en qué realmente estamos interesados? Supongamos que sea en la evolución de las peticiones diarias. Entonces, realmente, nuestro estudio se reduce a un cantidad numérica. Por tanto, debemos relacionar nuestro espacio muestral con la cantidad numérica objeto de nuestro estudio. A dicha relación se le denomina variable aleatoria.

O sea, una variable aleatoria es una función que nos relaciona el espacio muestral de nuestro experimento aleatorio en el conjunto de números de forma que hemos reducido el estudio de aplicar la teoría de las probabilidades al espacio muestral al estudio de aplicar las probabilidades a un conjunto ampliamente conocido, como es el conjunto de números.

Las aplicaciones de las variables aleatorias a las ciencias de la computación son múltiples y diversas. Destaquemos las siguientes: la simulación, los procesos estocásticos y la teoría de colas.

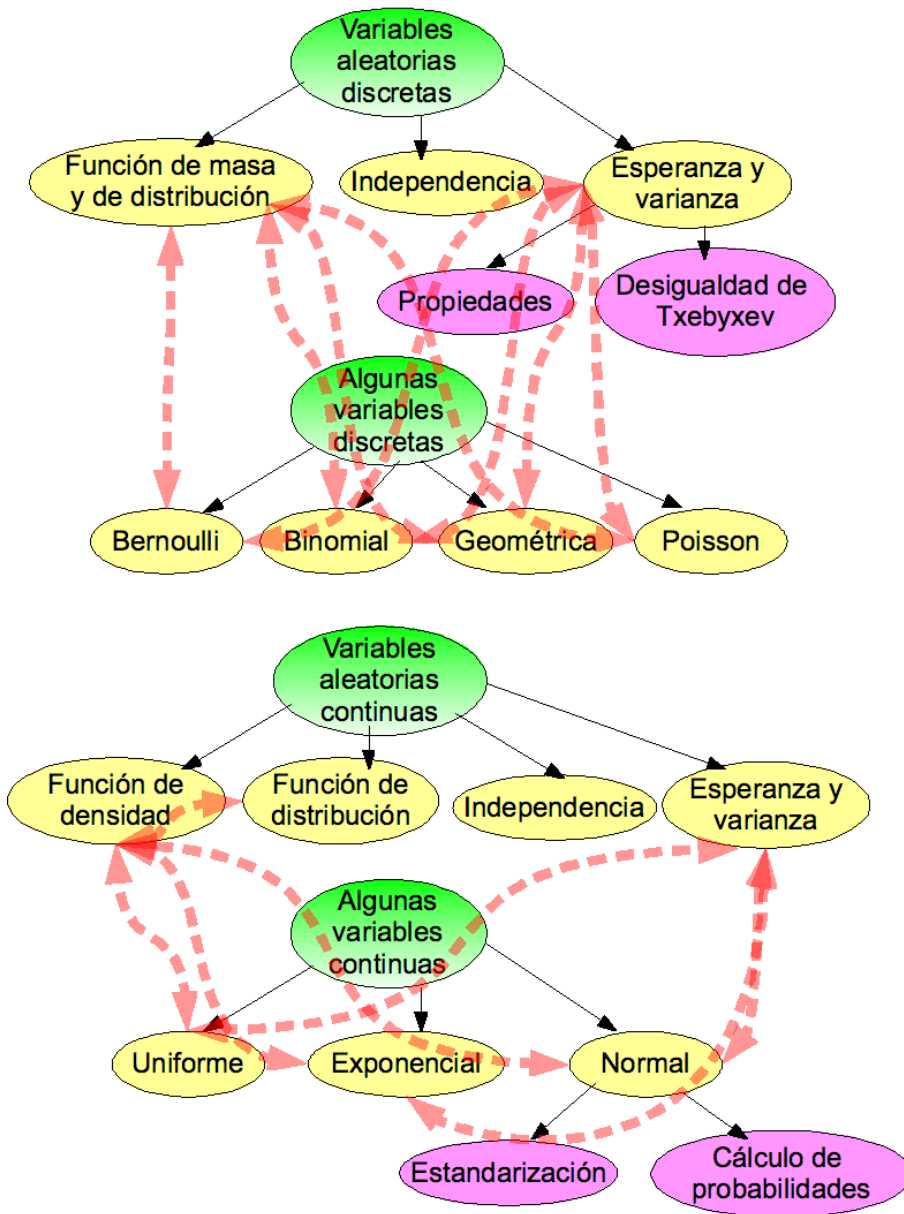
La simulación intenta hallar probabilidades de sucesos simulando el comportamiento de variables aleatorias. Veamos el ejemplo siguiente: supongamos que una empresa cuenta con 250 empleados. Cada minuto, la probabilidad de que un empleado envíe una petición al servidor de correo es 0.3. Conocemos la distribución de la variable aleatoria que nos da el número de peticiones enviadas por un usuario por minuto y la distribución de la variable aleatoria del tiempo que tarda el servidor en procesar una petición. Sabiendo que las peticiones se procesan consecutivamente, ¿cual es la probabilidad de que se procesen todas las peticiones enviadas en menos de una hora? Resolver el problema anterior directamente es muy complicado pero simulando las peticiones en un ordenador, puede obtenerse una respuesta aproximada de la probabilidad pedida.

Los procesos estocásticos estudian la evolución de un determinado proceso a lo largo del tiempo. Dicho proceso se modeliza mediante una variable aleatoria y se estudia la evolución de dicha variable aleatoria. Por ejemplo, el uso de la CPU de un ordenador puede considerarse un proceso estocástico.

La teoría de colas intenta resolver y asignar peticiones a servidores intentando optimizar el tiempo de espera de las peticiones. Por ejemplo, supongamos que tenemos una empresa con 4 servidores de correo saliente y tiene que gestionar las peticiones de envío de correo de todos los usuarios. Nuestro objetivo es asignar las peticiones entrantes de correo al servidor "más" adecuado de tal forma que el tiempo de espera "medio" de todo el sistema sea el menor posible.

# Mapa conceptual

## VARIABLES ALEATORIAS



## Actividades

### Actividad 1: Estudio de la subida en bolsa de las empresas de telecomunicaciones

Variables aleatorias discretas.

- Se sabe que el 60% de las empresas de telecomunicaciones han subido en la bolsa durante el último año. Si escogemos 50 empresas de este sector al azar, ¿cuál es la probabilidad de que menos de 20 hayan subido durante el último año?
- Supongamos que queremos realizar un estudio sobre las empresas del apartado anterior mediante llamadas telefónicas. ¿A cuántas empresas tendré que llamar hasta hallar una que no haya subido el año anterior?

### Solución

- Consideremos la variable aleatoria  $X$  que nos dice cuántas empresas han subido de entre las 50 escogidas el último año. Dicha variable aleatoria se distribuye según una Binomial de parámetros  $p=0,6$  y  $n=50$ . Nos piden hallar la probabilidad siguiente:  $P(X < 20) = P(X \leq 19)$ . Vamos a realizar el cálculo anterior con R. Podemos optar por usar la correspondiente instrucción o por usar RCommander.

- Si optamos por una instrucción, hemos de usar alguna de las relacionadas con la distribución binomial, que son:

```
dbinom(x, size, prob, log = FALSE)
pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
rbinom(n, size, prob)
```

Concretamente la primera (*dbinom*) es la función de densidad y calcula  $p(X=x)$ ; la segunda (*pbinom*) es la función de distribución y calcula  $P(X \leq x)$ ; la tercera (*qbinom*) es la función de distribución inversa, y dado  $q$  entre 0 y 1 calcula el valor de  $x$  tal que  $P(X \leq x) = q$ ; finalmente la última (*rbinom*) sirve para obtener muestras aleatorias de la distribución binomial.

Por lo tanto en nuestro caso usaremos *pbinom* con  $x=19$ ,  $size=50$  y  $p=0.6$  con lo que  $P(X \leq 19) = pbinom(19, 50, 0.6)$  y obtenemos

```
> pbinom(19,50,0.6)
[1] 0.001373762
```

- Si optamos por RCommander hemos de usar el menú Distribuciones->Distribuciones Discretas-> Distribución binomial -> Probabilidades Acumuladas Binomiales y debemos escribir la siguiente información en el cuadro de diálogo:

Con lo que obtenemos

```
> pbinom(c(19), size=50, prob=0.6, lower.tail=TRUE)
[1] 0.001373762
```

por lo que probabilidad buscada vale aproximadamente 0,00137.

- Consideremos la variable aleatoria  $Y$  que nos dice el número de empresas que hemos de llamar hasta que haya una que no haya subido en la bolsa el año anterior. La variable  $Y$  se distribuye según una variable geométrica de parámetro  $p=0,4$  ya que hay un 40% de probabilidad de que una empresa no haya subido en la bolsa durante el último año. En media, tendremos que llamar  $E(Y)$  veces. Usando que la esperanza

de una distribución geométrica vale  $\frac{1}{p}$ , tendremos que en nuestro caso tendremos que llamar a un total de  $\frac{1}{0,4} = 2,5$  empresas. O sea, tendremos que llamar a tres empresas hasta encontrar a una que no haya subido en la bolsa el año anterior.

## Actividad 2: Número de mensajes con virus que llegan a un servidor.

### Variable de Poisson.

Supongamos que un servidor recibe, en promedio, 48 mensajes con virus cada día. Si el nombre de virus que llegan al servidor es una variable de Poisson, hallar la probabilidad de que en una hora lleguen 3 virus, la probabilidad de que no lleguen virus en 3 horas y la probabilidad de que lleguen más de 500 mensajes con virus en una semana.

### Solución

Consideremos la variable  $X_t$  dependiente de un parámetro  $t$  que nos da el número de mensajes con virus que llegan en  $t$  horas. La distribución de la variable  $X_t$  es de Poisson de parámetro  $2t$  ya que si llegan, en promedio, 48 mensajes con virus en 24 horas, llegan, en promedio  $\frac{48}{24}t = 2t$  mensajes con virus en  $t$  horas. Nos piden  $P(X_1 = 3)$  (probabilidad de

que lleguen 3 virus en 1 hora,  $P(X_3 = 0)$  (probabilidad de que no lleguen virus en 3 horas) y  $P(X_{168} > 500)$  (probabilidad de que lleguen más de 500 mensajes en una semana que tiene  $24 \cdot 7 = 168$  horas). Vamos a calcular las probabilidades anteriores con R.

Si queremos usar instrucciones de R; tendremos que usar alguna de las relacionadas con la distribución de Poisson, que son

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

donde como en el caso anterior, la que comienza por “d” corresponde a la función de densidad, la que comienza por “p” corresponde a la función de distribución, la que comienza por “q” a la inversa de la distribución, y la que comienza por “r” genera muestras aleatorias de la distribución. Veamos pues los cálculos que hay que hacer

- $P(X_1 = 3) = \text{dpois}(3, \text{lambda}=2)$  ya que sigue una distribución de Poisson de parámetro  $2t = 2 \cdot 1 = 2$ , con lo que obtenemos

```
> dpois(3, lambda=2)
[1] 0.180447
```

- $P(X_3 = 0) = \text{dpois}(0, \text{lambda}=6)$  ya que sigue una distribución de Poisson de parámetro  $2t = 2 \cdot 3 = 6$ , con lo que obtenemos

```
> dpois(0, lambda=6)
[1] 0.002478752
```

- $P(X_{168} > 500) = \text{ppois}(500, \text{lambda}=336, \text{lower.tail}=\text{FALSE})$  ya que queremos una probabilidad acumulada de una distribución de Poisson de parámetro  $2 \cdot 168 = 336$  con lo que se obtiene

```
> ppois(500, lambda=336, lower.tail=FALSE)
[1] 2.910482e-17
```

El mismo resultado se obtiene con

```
> 1-ppois(499, lambda=336)
[1] 0
```

ya que la probabilidad buscada es prácticamente 0, Por tanto, es muy improbable que lleguen más de 500 mensajes con virus en una semana.

**Actividad 3: Tiempo en crear un antivirus para un virus determinado.**

Variable normal.

En una empresa dedicada a crear antivirus se sabe que el tiempo que se tarda en hallar un antivirus adecuado para cada virus nuevo sigue una distribución normal de media 10 días y desviación típica o estándar de 4 días. Se pide:

- Calcular la probabilidad de que se tarde entre 8 y 12 días en hallar un antivirus para un cierto virus.
- Imaginemos que la empresa quiere preparar un eslogan del tipo: "El 95% de las veces hallamos un antivirus en menos de x días". ¿Qué tiene que valer x?

**Solución**

- a) Consideremos la variable  $T$  que nos da el tiempo que tarda la empresa en hallar el antivirus para cada virus nuevo. Nos dicen que la distribución de  $T$  es normal de media  $\mu = 10$  días y desviación típica  $\sigma = 4$  días. Nos piden hallar la probabilidad siguiente:  $P(8 < T < 12)$ . Antes de hallar la probabilidad anterior con R, escribámosla en términos de la función de distribución:  $P(8 < T < 12) = P(T \leq 12) - P(T \leq 8)$ . Por tanto, tenemos que hallar la función de distribución de una normal de media  $\mu = 10$  días y desviación típica  $\sigma = 4$  días en los puntos 12 y 8. Para hacerlo, podemos usar las funciones relacionadas con la normal, que son, en la línea de las introducidas para la binomial y la Poisson, las siguientes
- ```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

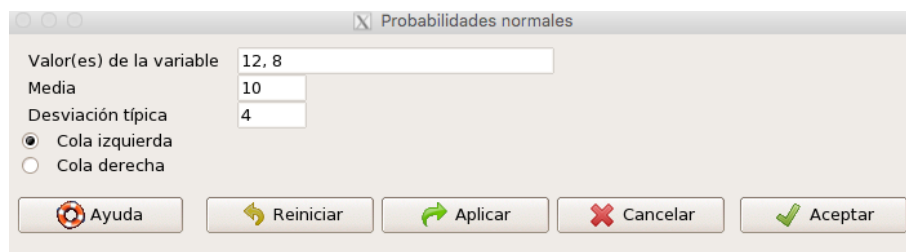
Así pues hay que calcular

$$P(8 < T < 12) = P(T \leq 12) - P(T \leq 8) =$$

= `pnorm(12, mean=10, sd=4) - pnorm(8, mean=10, sd=4)` con lo que obtenemos

```
pnorm(12, mean=10, sd=4) - pnorm(8, mean=10, sd=4)
[1] 0.3829249
```

Si queremos usar RComander iremos al menú vamos al menú Distribuciones->Distribuciones Continuas-> Distribución Normal -> Probabilidades Normales Acumuladas, rellenamos las casillas según se muestra a continuación



y obtendremos

```
> pnorm(c(12,8), mean=10, sd=4, lower.tail=TRUE)
[1] 0.6914625 0.3085375
```

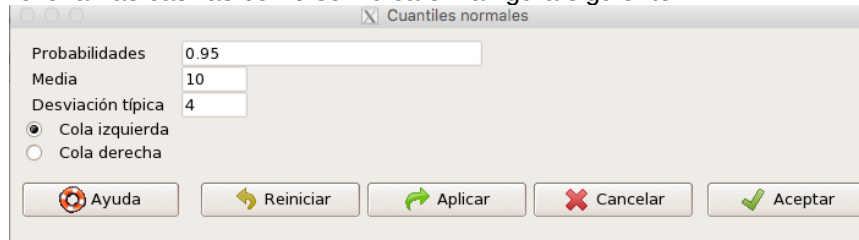
con lo que la probabilidad buscada es  $0,6914625 - 0,3085375 = 0,382925$ , tal y como hemos obtenido a partir de las instrucciones.

- b) En términos de probabilidades, nos dicen que  $P(T < x) = 0,95$ . Por tanto, tenemos que hallar x tal que la función de distribución de la variable normal de media  $\mu = 10$

d as y desviaci n t pica  $\sigma = 4$  d as en  $x$  vale 0,95. Para hallar este  $x$ , podemos usar directamente la funci n *qnorm* (que es la inversa de la funci n de distribuci n) de la siguiente forma

```
> qnorm(0.95,mean=10,sd=4)
[1] 16.57941
```

En RCommander hemos de ir al men  Calc->Probability distributions->Normal..., rellenar las casillas como se indica en la figura siguiente:



con lo que se obtiene

```
> qnorm(c(0.95), mean=10, sd=4, lower.tail=TRUE)
[1] 16.57941
```

Con lo podemos afirmar que el 95% de las veces hallamos un antivirus en menos de 16,58 d as.

#### Actividad 4: N mero de mensajes "SPAM" que llegan a un servidor. Variables aleatorias discretas.

De cada 1000 mensajes que llegan a un servidor de correo, 700 son "SPAM". Si hemos recibido 50 mensajes en una ma ana,

- calcular la probabilidad de que se hayan recibido al menos 20 mensajes que no sean "SPAM",
- calcular la probabilidad de que todos los mensajes recibidos sean "SPAM",
- hallar el n mero esperado de mensajes "SPAM" as  como su varianza.

#### Soluci n

- Consideremos la variable  $X$  que nos da el n mero de mensajes SPAM de entre los 50 mensajes que hemos recibido. La distribuci n de la variable  $X$  ser  binomial de par metros  $n = 50$  y  $p = \frac{700}{1000} = 0,7$ . Nos piden la probabilidad de que al menos 20 mensajes no sean SPAM, que es lo mismo que decir la probabilidad de recibir como m ximo 30 mensajes SPAM. Escrito en t rminos de probabilidades, nos piden  $P(X \leq 30)$ . Para calcularlo en R podemos hacerlo con la instrucci n *pbinom* de la siguiente forma:  

```
> pbinom(30,50,0.7)
[1] 0.0848026
```

 La probabilidad pedida vale, aproximadamente, 0,085.

- Nos piden la probabilidad siguiente  $P(X = 50)$ . Para calcularlo en R podemos usar la funci n *dbinom* de la siguiente forma:

```
> dbinom(50,50,0.7)
[1] 1.798465e-08
```

La probabilidad, por tanto, es pr cticamente nula.

- En general, la media de una variable Binomial de par metros  $n$  y  $p$  vale  $E(X) = np$ ,  $\text{Var}(X) = np(1 - p)$ . Sustituyendo los par metros en nuestro caso, tenemos que el n mero esperado de mensajes SPAM es  $E(X) = np = 50 \cdot 0,7 = 35$  mensajes, mientras que su varianza vale  $\text{Var}(X) = np(1 - p) = 50 \cdot 0,7 \cdot 0,3 = 10,5$ .

**Actividad 5: Cálculo del coste por la conexión a Internet.****VARIABLES ALEATORIAS DISCRETAS.**

Un proveedor de Internet (ISP) cobra a sus clientes por el tiempo en que éstos están conectados redondeando a la última hora. En la tabla siguiente se muestra la probabilidad de que un cliente de dicho proveedor escogido al azar esté conectado X horas y se cobre Y euros por las X horas:

|         |   | X horas |      |      |      |
|---------|---|---------|------|------|------|
|         |   | 1       | 2    | 3    | 4    |
| Y euros | 1 | 0       | 0.06 | 0.06 | 0.10 |
|         | 2 | 0.10    | 0.10 | 0.04 | 0.04 |
|         | 3 | 0.40    | 0.10 | 0    | 0    |

Por ejemplo, la probabilidad de que un cliente esté conectado 3 horas (en el redondeo) y se le cobre 2 horas por hora vale 0.04. Y así sucesivamente. El cobro total al cliente será  $Z=X*Y$ . Hallar la función de probabilidad y de distribución de la variable Z así como su esperanza y su varianza.

**Solución**

Calculemos primero los valores que puede alcanzar la variable aleatoria Z. Éstos serán:  $Z(\Omega) = \{1, 2, 3, 4, 6, 8, 9, 12\}$ . Dichos valores se han obtenido multiplicando los valores de la variable X por los valores de la variable Y:

|         |   | X horas |   |   |    |
|---------|---|---------|---|---|----|
|         |   | 1       | 2 | 3 | 4  |
| Y euros | 1 | 1       | 2 | 3 | 4  |
|         | 2 | 2       | 4 | 6 | 8  |
|         | 3 | 3       | 6 | 9 | 12 |

En la tabla anterior vemos de qué forma se pueden obtener los valores de Z. Hallemos a continuación la función de probabilidad de la variable Z:

$$p(Z=1) = p(X=1, Y=1) = 0,$$

$$p(Z=2) = p(X=1, Y=2) + p(X=2, Y=1) = 0,10 + 0,06 = 0,16,$$

$$p(Z=3) = p(X=1, Y=3) + p(X=3, Y=1) = 0,40 + 0,06 = 0,46,$$

$$p(Z=4) = p(X=2, Y=2) + p(X=4, Y=1) = 0,10 + 0,10 = 0,20,$$

$$p(Z=6) = p(X=2, Y=3) + p(X=3, Y=2) = 0,10 + 0,04 = 0,14,$$

$$p(Z=8) = p(X=4, Y=2) = 0,04$$

$$p(Z=9) = p(X=3, Y=3) = 0,$$

$$p(Z=12) = p(X=4, Y=3) = 0.$$

A posteriori, vemos que Z sólo alcanza los valores 2,3,4,6 y 8 y la función de probabilidad queda indicada en la tabla siguiente:

| X        | 2    | 3    | 4    | 6    | 8    |
|----------|------|------|------|------|------|
| $p(X=x)$ | 0,16 | 0,46 | 0,20 | 0,14 | 0,04 |

La función de distribución será:



$$F_Z(x) = \begin{cases} 0, & \text{si } x < 2, \\ 0,16, & \text{si } 2 \leq x < 3, \\ 0,16 + 0,46 = 0,62, & \text{si } 3 \leq x < 4, \\ 0,62 + 0,20 = 0,82, & \text{si } 4 \leq x < 6, \\ 0,82 + 0,14 = 0,96, & \text{si } 6 \leq x < 8, \\ 0,98 + 0,04 = 1, & \text{si } x \geq 8. \end{cases}$$

La esperanza de la variable Z será:

$$E(Z) = 2 \cdot 0,16 + 3 \cdot 0,46 + 4 \cdot 0,20 + 6 \cdot 0,14 + 8 \cdot 0,04 = 3,66.$$

El cliente paga de media 3,66 euros.

Su varianza será:

$$\text{Var}(Z) = 2^2 \cdot 0,16 + 3^2 \cdot 0,46 + 4^2 \cdot 0,20 + 6^2 \cdot 0,14 + 8^2 \cdot 0,04 - 3,66^2 = 2,1844.$$

### Actividad 6: Transmisión de mensajes.

#### Variables aleatorias discretas.

Un sistema interactivo consiste en 50 terminales todas conectadas a un ordenador central. En cualquier momento, cada terminal está preparada para transmitir mensajes con una probabilidad de 0.7, independientemente de las demás terminales. Hallar la probabilidad de que exactamente 16 terminales estén preparadas para transmitir a las 8.00 horas de la mañana, la probabilidad de que como mínimo 25 terminales estén preparadas para transmitir a las 12.00 horas de la mañana y de que haya entre 20 y 40 terminales ambas inclusive preparadas para transmitir a las 14.00 horas.

#### Solución

Consideremos la variable X que nos dice el número de terminales preparadas para transmitir mensajes en un momento cualquiera del día de las 50 terminales. La distribución de X será Binomial de parámetros  $n=50$  y  $p=0,7$ .

En primer lugar nos piden la probabilidad de que exactamente 16 terminales estén preparadas para transmitir a las 8.00 horas de la mañana, o, escrito en función de X,  $P(X=16)$ . Para calcularlo en R usamos

```
> dbinom(16,50,0.7)
[1] 2.728865e-08
```

Como podemos observar, la probabilidad es prácticamente nula.

A continuación nos piden la probabilidad de que como mínimo 25 terminales estén preparadas para transmitir a las 12.00 horas de la mañana. Escrito en términos de la variable X, nos piden  $P(X \geq 25)$ . Si escribimos la probabilidad anterior en términos de la función de distribución de X, tenemos  $P(X \geq 25) = 1 - P(X \leq 24)$ . En R podemos usar directamente la función *pbinom* de la siguiente forma

```
> 1-pbinom(24,50,0.7)
[1] 0.9990668
```

Concluimos que es muy probable que haya como mínimo 25 terminales preparadas para transmitir a las 12.00 horas.

Por último, nos piden la probabilidad de que haya entre 20 y 40 terminales ambas inclusive preparadas para transmitir a las 14.00 horas. Si escribimos dicha probabilidad en términos de X, tenemos  $P(20 \leq X \leq 40)$ . Escribiendo dicha probabilidad en términos de la función de

distribución de  $X$ , tenemos:  $P(20 \leq X \leq 40) = P(19 < X \leq 40) = P(X \leq 40) - P(X \leq 19)$ . Hemos de calcular la función de distribución de  $X$  para los valores 40 y 19. Procediendo de la misma forma como en el caso del cálculo de la probabilidad anterior, tenemos los siguientes resultados que nos da R:

```
>pbinom(40,50,0.7)-pbinom(19,50,0.7)
[1] 0.9597655
```

### Actividad 7: Tiempo de proceso de un trabajo por una impresora.

Variable exponencial.

El tiempo que tarda una impresora en procesar un trabajo sigue una variable exponencial de media 12 segundos. Se envía un trabajo a la impresora a las 10.00 horas. ¿Cuál es la probabilidad de que esté finalizado a las 10.01 horas? ¿Cuál es la probabilidad de que a las 10.02 h. aún no esté acabado el trabajo?

### Solución

En primer lugar, debemos calcular la probabilidad de que un trabajo esté finalizado a las 10.01 h., o  $P(X \leq 1)$ , donde  $X$  recordemos que nos da el tiempo que tarda la impresora en procesar un trabajo. En R se calcula de la siguiente forma:

```
> pexp(1,1/12)
[1] 0.07995559
> |
```

Para calcular la probabilidad de que a las 10.02 h. aún no esté finalizado un trabajo, debemos calcular  $P(X > 2)$ :

```
> 1-pexp(2,1/12)
[1] 0.8464817
>
```

### Actividad 8: Tiempo de proceso de un trabajo por una impresora.

Variable exponencial.

Dos expertos en ordenadores trabajan para una cierta empresa de informático. El primer experto se encarga del 60% de las reparaciones y el segundo del 40%. El tiempo que tarda el primer experto en reparar un ordenador sigue una distribución exponencial de parámetro  $\lambda_1 = 3 \text{ horas}^{-1}$  y el tiempo que tarda el segundo experto en reparar un ordenador sigue una distribución exponencial de parámetro  $\lambda_2 = 2 \text{ horas}^{-1}$ . Se ha mandado a reparar un ordenador hace 30 minutos y aún sigue averiado. ¿Cuál es la probabilidad de que el primer experto trabaje en la reparación de dicho ordenador? ¿Cuál es el tiempo medio de reparación de un ordenador para la empresa?

### Solución

Consideremos la variable  $T_1$  que nos dice el tiempo que tarda el primer experto y  $T_2$ , el tiempo que tarda el segundo experto.

Definimos los sucesos  $E_1$ : "reparación realizada por el primer experto" y  $E_2$ : "reparación realizada por el segundo experto".

Nos dan las probabilidades siguientes:  $P(E_1) = 0,6$ ,  $P(E_2) = 0,4$ .

Nos dicen que hace 30 minutos que un ordenador está en reparación y aún no se ha reparado. Sea  $T$  la variable que nos da el tiempo de reparación de un ordenador por la empresa. Nos piden hallar la probabilidad siguiente:  $P(E_1 | T > 0,5)$ . O sea, nos piden la probabilidad de que el ordenador sea reparado por el primer experto suponiendo que se tarda más de 30 minutos (0,5 horas) en repararlo. Para hallar la probabilidad anterior, usaremos la regla de Bayes:

$$p(E_1 | T > 0,5) = \frac{p(E_1 \cap T > 0,5)}{p(T > 0,5)} = \frac{p(E_1) \cdot p(T > 0,5 | E_1)}{p(E_1) \cdot p(T > 0,5 | E_1) + p(E_2) \cdot p(T > 0,5 | E_2)}$$

La probabilidades  $p(T > 0,5 | E_1)$  y  $p(T > 0,5 | E_2)$  son fáciles de calcular ya que sabemos la distribución de los tiempos que tarda cada uno de los expertos:

$$p(T > 0,5 | E_1) = p(T_1 > 0,5) = e^{-\lambda_1 \cdot 0,5} = e^{-3 \cdot 0,5} = e^{-1,5} \approx 0,2231,$$

$$p(T > 0,5 | E_2) = p(T_2 > 0,5) = e^{-\lambda_2 \cdot 0,5} = e^{-2 \cdot 0,5} = e^{-1} \approx 0,3679.$$

La probabilidad pedida será:

$$p(E_1 | T > 0,5) = \frac{0,6 \cdot 0,2231}{0,6 \cdot 0,2231 + 0,4 \cdot 0,3679} \approx 0,4764.$$

Para hallar el tiempo medio, usaremos la expresión siguiente:

$$E(T) = p(E_1) \cdot p(T | E_1) + p(E_2) \cdot p(T | E_2) =$$

$$0,6 \cdot E(T_1) + 0,4 \cdot E(T_2) = 0,6 \cdot \frac{1}{3} + 0,4 \cdot \frac{1}{2} = 0,4 \text{ horas. O sea, 24 minutos.}$$

La expresión anterior usa que el 60% de las veces hace la reparación el primer experto y el 40% restante, el segundo experto.

### Actividad 9: Estudio de la falla de un sistema informático.

#### Variable exponencial.

Un sistema informático está basado en dos módulos independientes A y B. Una falla de cualquiera de los dos módulos causa una falla del sistema. La duración de cada uno de los módulos sigue una distribución exponencial de parámetros  $\lambda_1 = 1 \text{ año}^{-1}$  y  $\lambda_2 = 2 \text{ años}^{-1}$ , respectivamente. Se pide:

- ¿cuál es la probabilidad de que el sistema funcione al menos 2 años sin tener una falla?
- Si el sistema falló en los dos primeros años, ¿cuál es la probabilidad de que la falla haya sido debida al módulo B pero no al módulo A?

### Solución

- a) Consideremos las variable  $S$ ,  $T_A$  y  $T_B$  que nos da el tiempo de duración del sistema hasta que hay una falla, el tiempo de duración del módulo A y el tiempo de duración del módulo B, respectivamente. Intentemos averiguar si la distribución de la variable  $S$  es conocida. Para ello, calculamos primeramente:  $p(S > t) = p(T_A > t \cap T_B > t)$ , ya que si el sistema ha durado más de  $t$  años, los módulos A y B también. Usando que los módulos A y B son independientes, tenemos que la probabilidad anterior se puede calcular como:

$$p(S > t) = p(T_A > t \cap T_B > t) = p(T_A > t) \cdot p(T_B > t) = e^{-t} \cdot e^{-2t} = e^{-3t}. \text{ Concluimos que la variable } S \text{ sigue la distribución exponencial de parámetro } \lambda = 3 \text{ años}^{-1}.$$

Nos piden la probabilidad siguiente:  $p(S > 2) = e^{-3 \cdot 2} = e^{-6} \approx 0,00248$ .

- b) Nos piden hallar la probabilidad siguiente:  $p(T_B \leq 2 \cap T_A > 2 | S \leq 2)$ , ya que sabemos que el sistema falló en los dos primeros años ( $S \leq 2$ ) y nos piden la probabilidad de que la falla sea debida el módulo B ( $T_B \leq 2$ ) y no al módulo A ( $T_A > 2$ ). Usando la fórmula de la probabilidad condicionada tenemos que la probabilidad anterior puede calcularse como:

$$p(T_B \leq 2 \cap T_A > 2 \mid S \leq 2) = \frac{p(T_B \leq 2 \cap T_A > 2 \cap S \leq 2)}{p(S \leq 2)}.$$

$T_B \leq 2 \cap T_A > 2$  está incluido en el suceso  $S \leq 2$ , la probabilidad anterior vale:

$$p(T_B \leq 2 \cap T_A > 2 \mid S \leq 2) = \frac{p(T_B \leq 2 \cap T_A > 2 \cap S \leq 2)}{p(S \leq 2)} = \frac{p(T_B \leq 2 \cap T_A > 2)}{p(S \leq 2)}.$$

Por último, usando la independencia de A y B y la distribución de las variables  $S$ ,  $T_A$  y  $T_B$ , tenemos que la probabilidad anterior vale:

$$p(T_B \leq 2 \cap T_A > 2 \mid S \leq 2) = \frac{(1 - e^{-2 \cdot 2}) \cdot e^{-2}}{1 - e^{-3 \cdot 2}} = \frac{e^{-2} - e^{-6}}{1 - e^{-6}} \approx 0.133.$$

### Actividad 10: Duración de un procesador.

Variable normal.

La duración de un procesador sigue una distribución normal de media 50000 horas y desviación típica 100 horas. Si queremos que sólo fallen el 3%, que vale la  $x$  en la siguiente frase de la campaña promocional del procesador: "¡Procesadores que duran más de  $x$  horas!". ¿Cuál es la probabilidad de que un procesador de este tipo dure más de 7 años en un servidor que funciona las 24 horas?

### Solución

Si queremos que sólo fallen el 3%, la  $x$  en la campaña promocional debe cumplir  $p(T > x) = 0,97$ , donde recordemos que  $T$  es la variable que nos da la duración del procesador. En R debemos realizar la operación siguiente:

```
> qnorm(0.03, 50000, 100)
[1] 49811.92
>
```

A continuación, nos piden la probabilidad de que un procesador de este tipo dure más de 7 años en un servidor que funciona las 24 horas, o  $p(T > 61320) = 1 - p(T \leq 61320)$ :

```
> 1-pnorm(7*365*24, 50000, 100)
[1] 0
>
```

### Actividad 11: Tiempo de procesamiento de trabajos de impresión en un laboratorio.

Variable exponencial. Variable uniforme.

Un laboratorio tiene dos impresoras. La impresora I procesa el 40% de los trabajos y su tiempo de impresión sigue una distribución exponencial de media 2 minutos. La impresora II procesa el resto y su tiempo de impresión sigue una distribución uniforme entre 0 minutos y 5 minutos.

- Un trabajo es impreso en menos de 5 minutos, ¿cuál es la probabilidad de que haya sido impreso por la impresora II?
- Si un trabajo tarda menos de 2 minutos, ¿cuál es la probabilidad de que haya sido impreso por la impresora I?
- Hallar la probabilidad de que un trabajo cualquiera tarde menos de 3 minutos en ser impreso en dicho laboratorio.

### Solución

a) Sea  $T$  el tiempo de impresión de un trabajo cualquiera por el laboratorio. Nos dicen que  $T < 5$  minutos y nos piden la probabilidad de que haya sido impreso por la impresora II. Escrito matemáticamente:  $p(II \mid T < 5)$ . Podemos hallar la probabilidad anterior usando la fórmula

de

Bayes:

$$p(II | T < 5) = \frac{p(II \cap T < 5)}{p(T < 5)} = \frac{p(II) \cdot p(T < 5 | II)}{p(I) \cdot p(T < 5 | I) + p(II) \cdot p(T < 5 | II)}. \quad \text{Las}$$

probabilidades  $p(T < 5 | I)$  y  $p(T < 5 | II)$  se pueden calcular al conocerse la distribución de los tiempos de las impresoras I y II:

$$p(T < 5 | I) = 1 - e^{-1/2 \cdot 5} = 1 - e^{-2,5} \approx 0,918, \quad p(T < 5 | II) = \frac{5-0}{5-0} = 1. \quad \text{La probabilidad}$$

$$\text{anterior vale: } p(II | T < 5) = \frac{0,6 \cdot 1}{0,4 \cdot 0,918 + 0,6 \cdot 1} \approx 0,620.$$

b) Usamos el mismo método que hemos usado en el apartado anterior para calcular la probabilidad pedida:

$$p(I | T < 2) = \frac{p(I \cap T < 2)}{p(T < 2)} = \frac{p(I) \cdot p(T < 2 | I)}{p(I) \cdot p(T < 2 | I) + p(II) \cdot p(T < 2 | II)} =$$

$$\frac{0,4 \cdot (1 - e^{-1/2 \cdot 2})}{0,4 \cdot (1 - e^{-1/2 \cdot 2}) + 0,6 \cdot \frac{(2-0)}{(5-0)}} = \frac{0,4 \cdot (1 - e^{-1})}{0,4 \cdot (1 - e^{-1}) + 0,6 \cdot 0,4} \approx 0,513.$$

c) La probabilidad de que la impresora del laboratorio tarde menos de 3 minutos en procesar un trabajo vale:

$$p(T < 3) = p(I) \cdot p(T < 3 | I) + p(II) \cdot p(T < 3 | II) =$$

$$0,4 \cdot (1 - e^{-1/2 \cdot 3}) + 0,6 \cdot \frac{(3-0)}{(5-0)} = 0,4 \cdot (1 - e^{-1,5}) + 0,6 \cdot 0,6 \approx 0,671.$$

### Actividad 12: Tiempo de conexión a Internet.

Variable exponencial. Variable uniforme.

Un proveedor de Internet tiene dos líneas de conexión para sus clientes. El 80% de sus clientes se conectan a través de la línea I, y el resto, a través de la línea II. El tiempo de conexión por la línea I sigue una distribución exponencial de parámetro  $\lambda = 2 \text{ minutos}^{-1}$ . El tiempo de conexión por la línea II sigue la distribución uniforme con parámetros  $a = 20$  segundos y  $b = 40$  segundos.

- Hallar la probabilidad de que un cliente cualquiera tarde más de 30 segundos en conectarse a Internet.
- Si un cliente ha tardado menos de 20 segundos en conectarse a Internet, ¿cuál es la probabilidad de que se haya conectado por la línea I?

### Solución

- Sea  $T$  el tiempo de conexión por un cliente cualquiera. Nos piden la probabilidad siguiente:  $p(T > 0,5)$ , ya que 30 segundos son 0,5 minutos. La probabilidad anterior se calcula de la forma siguiente:

$$p(T > 0,5) = p(I) \cdot p(T > 0,5 | I) + p(II) \cdot p(T > 0,5 | II) =$$

$$0,8 \cdot e^{-2 \cdot 0,5} + 0,2 \cdot \frac{(0,5 - 20 / 60)}{(40 / 60 - 20 / 60)} \approx 0,394.$$

En la expresión anterior hemos usado que la distribución del tiempo de conexión por la línea II es una variable uniforme de parámetros  $a = 20 / 60$  minutos y  $b = 40 / 60$  minutos.

- Nos dicen que  $T < 20 / 60 \approx 0,33$  minutos y nos piden  $p(I | T < 0,33)$ . Hallaremos la probabilidad anterior usando la fórmula de Bayes:

$$\begin{aligned}
 P(I | T < 0,33) &= \frac{P(I \cap T < 0,33)}{P(T < 0,33)} = \frac{P(I) \cdot P(T < 0,33 | I)}{P(I) \cdot P(T < 0,33 | I) + P(II) \cdot P(T < 0,33 | II)} = \\
 &= \frac{0,8 \cdot (1 - e^{-2 \cdot 0,33})}{0,8 \cdot (1 - e^{-2 \cdot 0,33}) + 0,2 \cdot \frac{(0,33 - 0,33)}{(0,66 - 0,33)}} = \frac{0,8 \cdot (1 - e^{-0,66})}{0,8 \cdot (1 - e^{-0,66}) + 0} = 1.
 \end{aligned}$$

El resultado anterior es el esperado ya que si un cliente tarda menos de 20 segundos en conectarse, es imposible que se conecte a través de la línea II ya que los clientes que se conectan a través de esta línea tardan entre 20 y 40 segundos en conectarse.

## Direcciones de interés

[http://en.wikipedia.org/wiki/Random\\_variable](http://en.wikipedia.org/wiki/Random_variable)

Definición y tratamiento matemático de variables aleatorias en la Wikipedia.

[http://en.wikipedia.org/wiki/Normal\\_distribution](http://en.wikipedia.org/wiki/Normal_distribution)

Definición y propiedades de la distribución normal

[http://wiki.stat.ucla.edu/socr/index.php/AP\\_Statistics\\_Curriculum\\_2007\\_Distrib\\_RV](http://wiki.stat.ucla.edu/socr/index.php/AP_Statistics_Curriculum_2007_Distrib_RV)

Libro on line sobre probabilidades y estadística. Muy recomendable. El enlace anterior nos envía al capítulo sobre variables aleatorias.

[http://www.math.csusb.edu/faculty/stanton/m262/intro\\_prob\\_models/intro\\_prob\\_models.html](http://www.math.csusb.edu/faculty/stanton/m262/intro_prob_models/intro_prob_models.html)

Simulación de la variable aleatoria que nos da la suma de los resultados del lanzamiento de varios dados.

<http://math.dartmouth.edu/~prob/prob/prob.pdf>

Libro on line sobre probabilidades y variables aleatorias con un enfoque muy práctico y con muchos ejemplos.