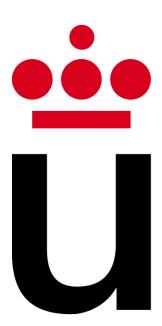
# CASO PRÁCTICO II Modelización y tratamiento de la incertidumbre

## Probabilidades y variables aleatorias

José Ignacio Escribano



Móstoles, 1 de noviembre de 2015

# Índice de figuras

1.	Funciones de masa de las inversiones	2
2.	Puntos de la función conjunta de masa	4
3.	Funciones de densidad y distribución de la distribución uniforme	6
4.	Funciones de densidad y distribución de la distribución exponencial	9
5.	Funciones de densidad e intervalo de probabilidad al 90 % de la distribu-	
	ción exponencial	11
6	Simulación del coste de 1000 llamadas	12

# Índice

1.	Introducción		
2.	Resolución del caso práctico	1	
	2.1. Tarea IV		
	2.2. Tarea V		
	2.3. Tarea VI	6	
	2.4. Tarea VII	8	
3.	Conclusiones	12	
4.	Código R	13	

### 1. Introducción

En este caso práctico resolveremos distintos problemas relativos a variables aleatorias, tanto discretas como continuas. Las dos primeras tareas se refieren a las primeras variables, y las dos últimas, a las segundas. En las cuatro tareas calcularemos probabilidades, realizaremos distintos tipos gráficos y responderemos a las preguntas planteadas.

## 2. Resolución del caso práctico

#### 2.1. Tarea IV

El beneficio de cada inversión viene dado por la diferencia entre los beneficios netos y la inversión (1 millón de euros). Así, por tanto, si se pierde todo lo invertido tenemos que el beneficio es -1 (los beneficios netos son cero y la inversión 1); cuando no se obtiene beneficio neto ni se pierde la inversión tenemos que los beneficios son 0 (se tiene un beneficio neto de un millón y un mi millón de inversión).

Si denotamos con  $X_i$  al beneficio neto obtenido en la inversión i con i = 1, 2, 3, tenemos que las funciones de masa vienen dadas de la siguiente forma:

$$f(X_1 = x) = \begin{cases} 0.1, & \text{si } x = 5 \\ 0.3, & \text{si } x = 1 \\ 0.6, & \text{si } x = -1 \end{cases}$$

$$f(X_2 = x) = \begin{cases} 0.2, & \text{si } x = 3 \\ 0.4, & \text{si } x = 1 \\ 0.4, & \text{si } x = -1 \end{cases}$$

$$f(X_3 = x) = \begin{cases} 0.1, & \text{si } x = 6 \\ 0.7, & \text{si } x = 0 \\ 0.2, & \text{si } x = -3 \end{cases}$$

En la Figura 1 se pueden ver las funciones de masa de las tres inversiones.

Calculamos la media y la varianza del beneficio de cada inversión.

$$E[X_1] = \sum_{i \in \Omega_1} i \cdot f(X_1 = i)$$
  
= 5 \cdot 0.1 + 1 \cdot 0.3 - 1 \cdot 0.6  
= 0.2

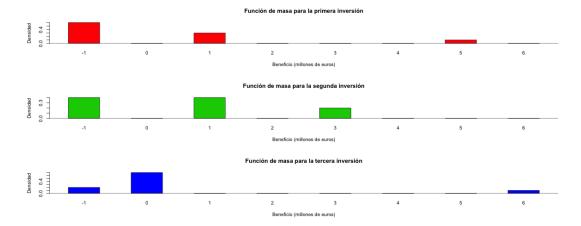


Figura 1: Funciones de masa de las inversiones

$$Var[X_1] = \sum_{i \in \Omega_1} f(X_1 = i)(x_i - E[X_1])^2$$

$$= 0.1(5 - 0.2)^2 + 0.3(1 - 0.2)^2 + 0.6(-1 - 0.2)^2$$

$$= 3.36$$

donde  $\Omega_1 = \{-1, 1, 5\}$  es el soporte de la variable  $X_1$ .

De igual forma, se obtiene que

$$E[X_2] = 0.6$$
$$Var[X_2] = 2.24$$

$$E[X_3] = 0.4$$
$$Var[X_3] = 3.64$$

con  $\Omega_2 = \{-1, 1, 3\}$  y  $\Omega_3 = \{-1, 0, 6\}$ , los soportes de las variables  $X_2$  y  $X_3$ , respectivamente.

Para ver la inversión más segura, calculamos la probabilidad de obtener beneficios, es decir,  $P(X_i > 0)$  con i = 1, 2, 3. Así, tenemos que

$$P(X_1 > 0) = 1 - P(X_1 \le 0)$$

$$= 1 - P(X_1 = -1)$$

$$= 1 - 0.6$$

$$= 0.4$$

$$P(X_2 > 0) = 1 - P(X_2 \le 0)$$

$$= 1 - P(X_2 = -1)$$

$$= 1 - 0.4$$

$$= 0.6$$

$$P(X_3 > 0) = 1 - P(X_3 \le 0)$$

$$= 1 - (P(X_3 = -1) + P(X_3 = 0))$$

$$= 1 - 0.7 - 0.2$$

$$= 0.1$$

Teniendo en cuenta los datos de la esperanza y la varianza, y las probabilidades calculadas anteriormente, vemos que la opción más segura para obtener algún beneficio es la segunda, ya que hay 60 % de probabilidades de ganar algo de dinero con esta inversión, junto con su mayor esperanza y su menor varianza.

Para ver la inversión más arriesgada, calculamos la probabilidad de no perder dinero, es decir,  $P(X_i \ge 0)$ ,

$$P(X_1 \ge 0) = 1 - P(X_0 < 0)$$

$$= 1 - P(X_1 = -1)$$

$$= 1 - 0.6$$

$$= 0.4$$

$$P(X_2 \ge 0) = 1 - P(X_2 < 0)$$

$$= 1 - P(X_2 = -1)$$

$$= 1 - 0.4$$

$$= 0.6$$

$$P(X_3 \ge 0) = 1 - P(X_3 < 0)$$

$$= 1 - P(X_3 = -1)$$

$$= 1 - 0.2$$

$$= 0.8$$

La opción más arriesgada es la tercera, donde hay más probabilidades de no perder dinero, pero hay un 70 % de no ganar nada, aunque en caso de ganar, ganaremos la mayor cantidad de dinero de todas las inversiones, 6 millones de euros.

Calculamos la probabilidad de que la inversión más arriesgada proporcione más beneficios que la más segura, es decir,  $P(X_3 > X_2)$ . Tenemos, una variable bidimensional. Calculamos la función de masa conjunta  $f(X_3 = x, X_2 = y) = f(X_3 = x) \cdot f(X_2 = y) \ \forall (x,y) \in \Omega_3 \times \Omega_2$ , suponiendo independencia entre las variables (en principio, se puede suponer que los beneficios de una empresa no dependen de la otra). La función de masa conjunta viene dada por la siguiente tabla:

$X_3 \setminus X_2$	-1	1	3
-1	$0.2 \cdot 0.4 = 0.08$	$0.2 \cdot 0.4 = 0.08$	$0.2 \cdot 0.2 = 0.04$
0	$0.7 \cdot 0.4 = 0.28$	$0.7 \cdot 0.4 = 0.28$	$0.7 \cdot 0.4 = 0.28$
6	$0.1 \cdot 0.4 = 0.04$	$0.1 \cdot 0.4 = 0.04$	$0.1 \cdot 0.2 = 0.02$

Tenemos que ver qué puntos de  $\mathbb{R}^2$  cumplen que  $X_3 > X_2$ . Si representamos cada punto en el plano (Figura 2), vemos que los únicos que cumplen la condición anterior son  $\Pi = \{(0, -1), (6, -1), (6, 1), (6, 3)\}.$ 

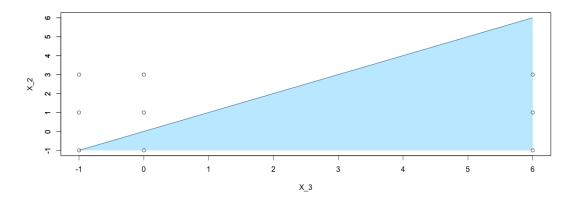


Figura 2: Puntos de la función conjunta de masa

Por tanto, la probabilidad pedida se calcula sumando las probabilidades de cada punto  $f(x,y) \ \forall (x,y) \in \Pi$ , calculadas en la tabla anterior. Esto es,

$$P(X_3 > X_2) = f(0, -1) + f(6, -1) + f(6, 1) + f(6, 3)$$
  
= 0.28 + 0.04 + 0.04 + 0.08  
= 0.44

Si 10 SICAV invierten cada una un millón de euros en la empresa con más riesgos y otro millón en las empresa más segura, calculamos cuántas de ellas es de esperar que obtengan más beneficios de la inversión más arriesgada que de la más segura. Tenemos una variable aleatoria binomial con n=10 y p=0.44, calculada en el apartado enterior. Así pues,

$$E[X] = n \cdot p$$
$$= 10 \cdot 0.44$$
$$= 4.4$$

Es decir, 4 de las 10 SICAVS obtendrán un beneficio mayor con la inversión más arriesgada que con la más segura.

#### 2.2. Tarea V

Si denotamos con X al número de trozos de cáscara de huevo que hay una tarta. Por el enunciado, deducimos que  $E[X]=\frac{3}{2}$ . Tenemos que  $X\sim \mathcal{P}ois(\lambda)$ . Por otro lado, la esperanza de una distribución Poisson de parámetro  $\lambda$  es igual a ese mismo parámetro. Por tanto, tenemos que  $\lambda=\frac{3}{2}$  y  $X\sim \mathcal{P}ois(\frac{3}{2})$ .

A tenor de lo anterior, el número esperado de trocitos en una tarta es  $E[X] = \frac{3}{2}$ .

La probabilidad de que no haya ningún trozo de cáscara de huevo en la tarta pedida es P(X=0),

$$P(X = 0) = \frac{e^{\frac{-3}{2}} \left(\frac{3}{2}\right)^0}{0!}$$
$$= 0.2231$$

Calculamos la probabilidad de que haya siete trozos de cáscara de huevo en una tarta, P(X = 7),

$$P(X = 7) = \frac{e^{\frac{-3}{2}} \left(\frac{3}{2}\right)^7}{7!}$$
$$= 0.0007$$

Por tanto, es posible que nos encontráramos una tarta con siete trozos de cáscara de huevo, aunque es muy poco probable: de cada 1000 tartas, 7 tienen siete trozos de cáscara de huevo.

Si pedimos 6 tartas, la probabilidad de que no haya ningún trocito en ninguna de las tartas, se calcula de la siguiente manera: tenemos una variable aleatoria  $Y \sim \mathcal{B}in(n=6,p=0.2231)$  donde p lo hemos calculado en los apartados anteriores. Así, la probabilidad pedida es P(Y=6),

$$P(Y = 6) = {6 \choose 6} 0.2231^{6} (1 - 0.2231)^{0}$$
$$= 0.00012$$

Es muy poco probable que en las 6 tartas pedidas, en ninguna de ellas no haya ningún trozo de cáscara de huevo.

#### 2.3. Tarea VI

Si denotamos con X el tiempo (en minutos) hasta que llegan las tartas encargadas a la sede de la SICAV, entonces  $X \sim \mathcal{U}(25,45)$ . La Figura 3 muestra las funciones de densidad y distribución de esta variable aleatoria.

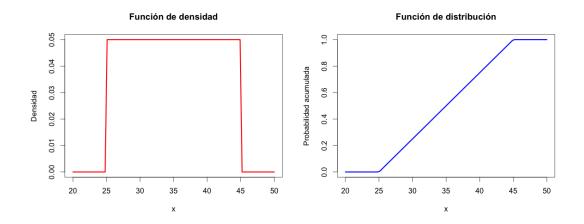


Figura 3: Funciones de densidad y distribución de la distribución uniforme

La esperanza de una de una distribución uniforme de parámetros (a, b) viene dada por

$$E[X] = \frac{a+b}{2}$$

$$Var[X] = \frac{(b-a)^2}{12}$$

Haciendo a = 25 y b = 45, se tiene que

$$E[X] = \frac{25 + 45}{2} = 35$$
$$Var[X] = \frac{(45 - 25)^2}{12} = \frac{100}{3}$$

La probabilidad de que X sea, como mucho, 30 minutos es  $P(X \le 30)$ .

$$P(X \le 30) = \int_{25}^{30} \frac{1}{45 - 25} dx$$
$$= \frac{1}{20} [x]_{25}^{30}$$
$$= \frac{1}{4}$$

La probabilidad de que X esté comprendido entre 30 y 40 minutos es  $P(30 \le X \le 40)$ .

$$P(30 \le X \le 40) = \int_{30}^{40} \frac{1}{45 - 25} dx$$
$$= \frac{1}{20} [x]_{40}^{30}$$
$$= \frac{1}{2}$$

Calculamos el percentil 90 de la distribución. Es decir tenemos que encontrar el número a tal que acumula el 90 % de la probabilidad. Es decir,

$$\int_{25}^{a} \frac{1}{45 - 25} \, dx = 0.9$$

$$\frac{1}{20} \int_{25}^{a} dx = \frac{1}{20} [x]_{25}^{a} = \frac{1}{20} (a - 25) = 0.9$$

Despejando a se tiene que  $a = 0.9 \cdot 20 + 25 = 43$ .

Es decir, el 90 % de las tartas llegará en un tiempo menor o igual a 43 minutos.

La probabilidad de que el tiempo X sea mayor de 40 minutos sabiendo que dicho tiempo es de al menos 30 minutos es P(X > 40|X > 30).

$$P(X > 40|X > 30) = \frac{P(\{X > 40\} \cap \{X > 30\})}{P(X > 30)}$$

$$= \frac{P(X > 40)}{P(X > 30)}$$

$$= \frac{1 - P(X \le 40)}{1 - P(X \le 30)}$$

$$= \frac{1 - \int_{25}^{40} \frac{1}{20} dx}{1 - \int_{25}^{30} \frac{1}{20} dx}$$

$$= \frac{1 - \frac{1}{20}(40 - 25)}{1 - \frac{1}{20}(30 - 25)}$$

$$= \frac{1 - \frac{3}{4}}{1 - \frac{1}{4}}$$

$$= \frac{1}{3}$$

Por tanto, sabiendo que el tiempo de llegada de las tartas va a ser mayor de 30 minutos, tenemos un 33.33 % de probabilidades de que tarden más de 40 minutos.

#### 2.4. Tarea VII

Si ahora denotamos con X al tiempo (en minutos) que dura una llamada cualquiera realizada por un trabajador de VACIS a un teléfono externo, tenemos que  $X \sim \mathcal{E}xp(\lambda)$ . Se sabe que la duración media de una llamada es de 8 minutos.

Así pues, sabemos que  $E[X] = \frac{1}{\lambda}$ . Por otro lado, tenemos que la duración media de las llamadas es de 8 minutos, es decir, E[X] = 8. Si igualamos y despejamos  $\lambda$  tenemos que  $\lambda = \frac{1}{8}$ . Por tanto,  $X \sim \mathcal{E}xp\left(\frac{1}{8}\right)$ . La Figura 4 muestra tanto la función de densidad como la de distribución de esta variable aleatoria.

La esperanza y la varianza de una distribución exponencial con parámetro  $\lambda$  viene dada por

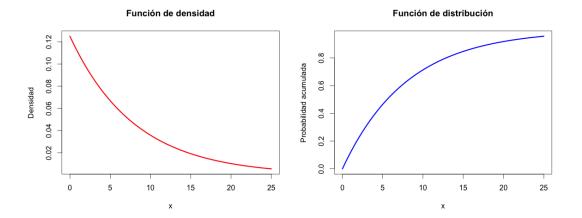


Figura 4: Funciones de densidad y distribución de la distribución exponencial

$$E[X] = \frac{1}{\lambda}$$

$$Var[X] = \frac{1}{\lambda^2}$$

Sustituyendo  $\lambda = \frac{1}{8}$ , tenemos que

$$E[X] = \frac{1}{\frac{1}{8}} = 8$$

$$Var[X] = \frac{1}{\left(\frac{1}{8}\right)^2} = 64$$

La probabilidad de que una llamada cualquiera a móvil dure menos de 9 minutos es P(X < 9),

$$P(X < 9) = \int_0^9 \frac{1}{8} e^{-\frac{x}{8}} dx$$
$$= -\left[e^{-\frac{x}{8}}\right]_0^9$$
$$= 1 - e^{-\frac{9}{8}}$$
$$= 0.675$$

La probabilidad de que una llamada cualquiera a móvil dure entre 7 y 9 miutos es  $P(7 \le X \le 9)$ ,

$$P(7 \le X \le 9) = \int_{7}^{9} \frac{1}{8} e^{-\frac{x}{8}} dx$$
$$= -\left[e^{-\frac{x}{8}}\right]_{7}^{9}$$
$$= -e^{-\frac{9}{8}} + e^{-\frac{7}{8}}$$
$$= 0.092$$

Calculamos el percentil 80 de esta distribución, es decir, el número a tal que la probabilidad acumulada de la distribución es del 80 %. Es decir,

$$\int_0^a \frac{1}{8} e^{-\frac{x}{8}} dx = 0.8$$

$$\int_0^a \frac{1}{8} e^{-\frac{x}{8}} dx = -\left[e^{-\frac{x}{8}}\right]_0^a$$
$$= -e^{\frac{a}{8}} + 1$$
$$= 0.8$$

Tomando logaritmos a ambos lados de la expresión anterior, se tiene que

$$a = -8 \cdot \ln 0.2 = 12.875$$

Por tanto, el 80 % de las llamadas tienen una duración inferior a 12.875 minutos.

Si tenemos un total de 1000 llamadas a móvil al día, el tiempo total de esas llamadas es de 8000 minutos diarios, ya que la duración media de una llamada cualquiera es de 8 minutos. Así 1000 llamadas  $\cdot$  8 minutos/llamada = 8000 minutos. Si queremos ser más precisos, con nuestra estimación podemos calcular un intervalo de probabilidad al 90%.

Normalmente, calculamos los intervalos de probabilidad bilaterales, pero en este caso, por tratarse de una distribución con alta asimetría hacia la derecha, calcularemos el intervalo con los percentiles 0 y 0.90. Si calculamos estas probabilidades como en el apartado anterior tenemos que el intervalo es (0,18.42) (Figura 5). Es decir, el 90 % de las llamadas estarán entre 0 y 18.42 minutos. Por tanto, el número de minutos a diario estará entre 0 y 18420 minutos diarios.

Si nuestra tarifa móvil actual (en minutos) es:

# 

Función de densidad e

Figura 5: Funciones de densidad e intervalo de probabilidad al 90 % de la distribución exponencial

$$Y = \begin{cases} 0.1 + 0.6X & \text{si } X < 1\\ 0.1 + 0.3X & \text{si } X \ge 1 \end{cases}$$

El coste medio de esta nueva variable aleatoria viene dado por su esperanza,

$$E[Y] = \begin{cases} 0.1 + 0.6E[X] & \text{si } X < 1 \\ 0.1 + 0.3E[X] & \text{si } X \ge 1 \end{cases} = \begin{cases} 4.9 & \text{si } X < 1 \\ 2.5 & \text{si } X \ge 1 \end{cases}$$

Por otro lado, la probabilidad de tener una llamada de menos de un minuto es de 11.7 %, y tener una llamada superior a un minuto es de 88.3 %. Usamos estas probabilidades para ponderar el coste medio de una llamada.

$$E[Y] = 0.117 \cdot 4.9 + 0.883 \cdot 2.5$$
$$= 2.7808$$

Por tanto, el coste medio para una llamada es de 2.78€. Si se realizan 1000 llamadas el coste es de 2780€.

Para calcular la varianza, procedemos como con la esperanza.

$$Var[Y] = \begin{cases} 0.6^2 \cdot Var[X] & \text{si } X < 1 \\ 0.3^2 \cdot Var[X] & \text{si } X \ge 1 \end{cases} = \begin{cases} 23.04 & \text{si } X < 1 \\ 5.76 & \text{si } X \ge 1 \end{cases}$$

Ponderando como con la media se obtiene que Var[Y]=7.781. Por tanto, la desviación típica es 2.789

Si repetimos lo anterior, pero haciendo una simulación, es decir, generamos 1000 muestras de una distribución exponencial con parámetro  $\lambda = \frac{1}{8}$ . Aplicamos nuestra función de coste a cada una de nuestras muestras, y calculamos la media y la varianza. En nuestro caso, el coste medio por llamada es de 2.601, el coste total es de 2601.67 y la desviación típica es de 2.530. La Figura 6 muestra la duración y el coste de las 1000 llamadas simuladas.

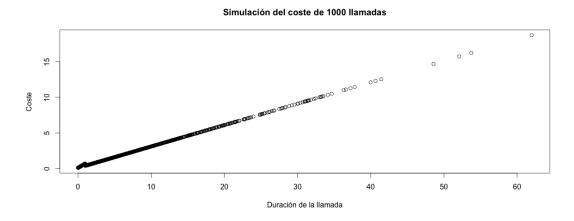


Figura 6: Simulación del coste de 1000 llamadas

#### 3. Conclusiones

En esta actividad hemos puesto en práctica lo aprendido en teoría sobre variables aleatorias –tanto continuas como discretas– y cálculo de probabilidades, y hemos aplicado estas variables aleatorias para modelizar distintos tipos de problemas. R nos ha permitido calcular probabilidades y representar gráficos de forma rápida, ahorrándonos una gran cantidad de tiempo.

## 4. Código R

```
#-----
# Tarea IV
# Definimos las varibles para las inversiones
# Primera inversión
x1 \leftarrow c(-1, 0, 1, 2, 3, 4, 5, 6)
m1 \leftarrow c(0.6, 0, 0.3, 0, 0, 0, 0.1, 0)
# Segunda inversión
x2 \leftarrow c(-1, 0, 1, 2, 3, 4, 5, 6)
m2 \leftarrow c(0.4, 0, 0.4, 0, 0.2, 0, 0, 0)
# Tercera inversión
# Segunda inversión
x3 \leftarrow c(-1, 0, 1, 2, 3, 4, 5,
m3 \leftarrow c(0.2, 0.7, 0, 0, 0, 0, 0.1)
# Representamos la función de masa de cada de las inversiones
par(mfrow=c(3,1))
barplot(names.arg = x1, m1, space = 1, xlab="Beneficio (millones de euros)",
  ylab="Densidad", main="Función de masa para la primera inversión", col = 2)
abline(h=0)
barplot(names.arg = x2, m2, space = 1, xlab="Beneficio (millones de euros)",
  ylab="Densidad", main="Función de masa para la segunda inversión", col = 3)
abline(h=0)
barplot(names.arg = x3, m3, space = 1, xlab="Beneficio (millones de euros)",
  ylab="Densidad", main="Función de masa para la tercera inversión", col = 4)
abline(h=0)
# Calculamos la media y la varianza para cada inversión
x1.mean <- x1 %*% m1
x1.var <- sum(m1*(x1-x1.mean)^2)
x2.mean <- x2 %*% m2
x2.var <- sum(m2*(x2-x2.mean)^2)
x3.mean <- x3 %*% m3
x3.var <- sum(m3*(x3-x3.mean)^2)
```

```
# Dibujamos todos los posibles puntos en el plano
x \leftarrow c(-1, -1, -1, 0, 0, 0, 6, 6, 6)
y \leftarrow c(-1, 1, 3, -1, 1, 3, -1, 1, 3)
x2 < -seq(-1,6)
y2 < -seq(-1,6)
plot(x,y, xlim = c(-1,6), ylim = c(-1,6), xlab = "X 3", ylab="X 2")
par(new = TRUE)
plot(x2,y2, xlim = c(-1,6), ylim = c(-1,6), type = "l", xaxt='n', ann=FALSE)
polygon(c(-1,6,6), c(-1,-1,6), col=rgb(0, 191/255, 255/255,0.3), border=NA)
par(new = FALSE)
#-----
# Tarea VI
#-----
# Definimos los parámetros de la distribución uniforme
a <- 25
b <- 45
# Representamos la función de densidad y la función de distribución
par(mfrow=c(1,2))
curve(dunif(x, a, b), from = a - 5, to = b + 5, main = "Función de densidad",
 ylab = "Densidad", col = 2, lwd=3)
curve(punif(x, a, b), from = a - 5, to = b + 5, main = "Función de distribución",
 ylab = "Probabilidad acumulada", col = 4, lwd=3)
#-----
# Tarea V
#-----
# Definimos el parámetro de una distribución de Poisson(3/2)
lambda <- 3/2
# Probabilidad de que no haya trozos de cáscara de huevo
ppois(0, lambda)
# Probabilidad de que haya siete trozos de huevo en una tarta
dpois(7, lambda)
# Probabilidad de que en 6 tartas no haya en ninguna ningún trozo de cáscaras de
# huevo
```

```
# Tarea VII
# Definimos el parámetro de la distribución exponencial
lambda <- 1/8
# Representamos la función de densidad y de distribución
par(mfrow=c(1,2))
curve(dexp(x, lambda), from = 0, to = 25, main = "Función de densidad",
  ylab = "Densidad", col = 2, lwd = 3)
curve(pexp(x, lambda), from = 0, to = 25, main = "Función de distribución",
  ylab = "Probabilidad acumulada", col = 4, lwd = 3)
# Calculamos el intervalo de probabilidad al 90%
lb <- qexp(0, lambda)</pre>
ub \leftarrow qexp(0.9, lambda)
# Representamos el intervalo de probabilidad sobre la función de densidad
par(mfrow=c(1,1))
curve(dexp(x, lambda), from = 0, to = 25, main = "Función de densidad e
  intervalo de probabilidad al 90%",
  ylab = "Densidad", col = 2, lwd = 3)
cords.x \leftarrow c(lb, seq(lb, ub, 0.01), ub)
cords.y \leftarrow c(0, dexp(seq(lb, ub, 0.01), lambda), 0)
polygon(cords.x, cords.y, col='skyblue', border=NA, add = TRUE)
# Simulamos el coste diario de 1000 llamadas
samples <- rexp(1000, lambda)</pre>
cost <- (samples <= 1)*(0.1 + 0.6*samples) + (samples > 1)*(0.1+0.3*samples)
plot(samples, cost, xlab="Duración de la llamada", ylab = "Coste",
  main = "Simulación del coste de 1000 llamadas")
mean(cost) # Media
sd(cost) # Desviación típica
sum(cost) # Coste total
```