Máster en Data Science. URJC
Técnicas y Métodos de Ciencia de Datos

Práctica Variable Aleatoria Unidimensional

Leda Duelo, Javier Llorente, Candela Vidal, Jaime Zamorano

18/11/2017

Índice

1 Problema **1**

1.1 Ejercicio 1 2

1.2 Ejercicio 2 3

1.3 Ejercicio 3 4

1.4 Ejercicio 4 5

1.5 Ejercicio 5 6

1.6 Ejercicio 6 7

1.7 Ejercicio 7 8

1 Problema

Una empresa de servicios de Internet quiere hacer una campaña para aplicar entre un 5% y un 25% de descuento a sus clientes de forma aleatoria y lineal, y entonces la probabilidad de que un cliente reciba un determinado descuento se puede modelizar mediante la siguiente función de densidad:

$$f(x) = \begin{cases} k(25 - x) & si\, 5 \leq x \leq 25 \\ 0 & resto \end{cases}$$

Responde a las siguientes cuestiones:

1.1 Ejercicio 1

Calcula k para que $f(x)$ sea realmente una funcion de densidad

La integral de la función de densidad en todo el dominio en el que este definida ha de valer 1.

$$1 = \int_{-\infty}^{\infty} f(x)dx$$

$$1 = \int_{5}^{25} k(25 - x)dx$$

Despejando se obtiene el valor de k:

$$k = \frac{1}{\int_5^{25}(25 - x)dx}$$

```
1 / integrate(function(x) (25 - x), 5, 25)$value
```

```
## [1] 0.005
```

Comprobación de que es una función de densidad:

```
func = function(x) 0.005 * (25 - x)

integrate(func, 5, 25)
```

```
## 1 with absolute error < 1.1e-14
```

Nota: hemos definido la función dentro del intervalo [5, 25], ya que fuera de este intervalo su valor es cero.

Para poder integrar la función entre [$-\infty$, ∞], habría que definir la función de está manera.

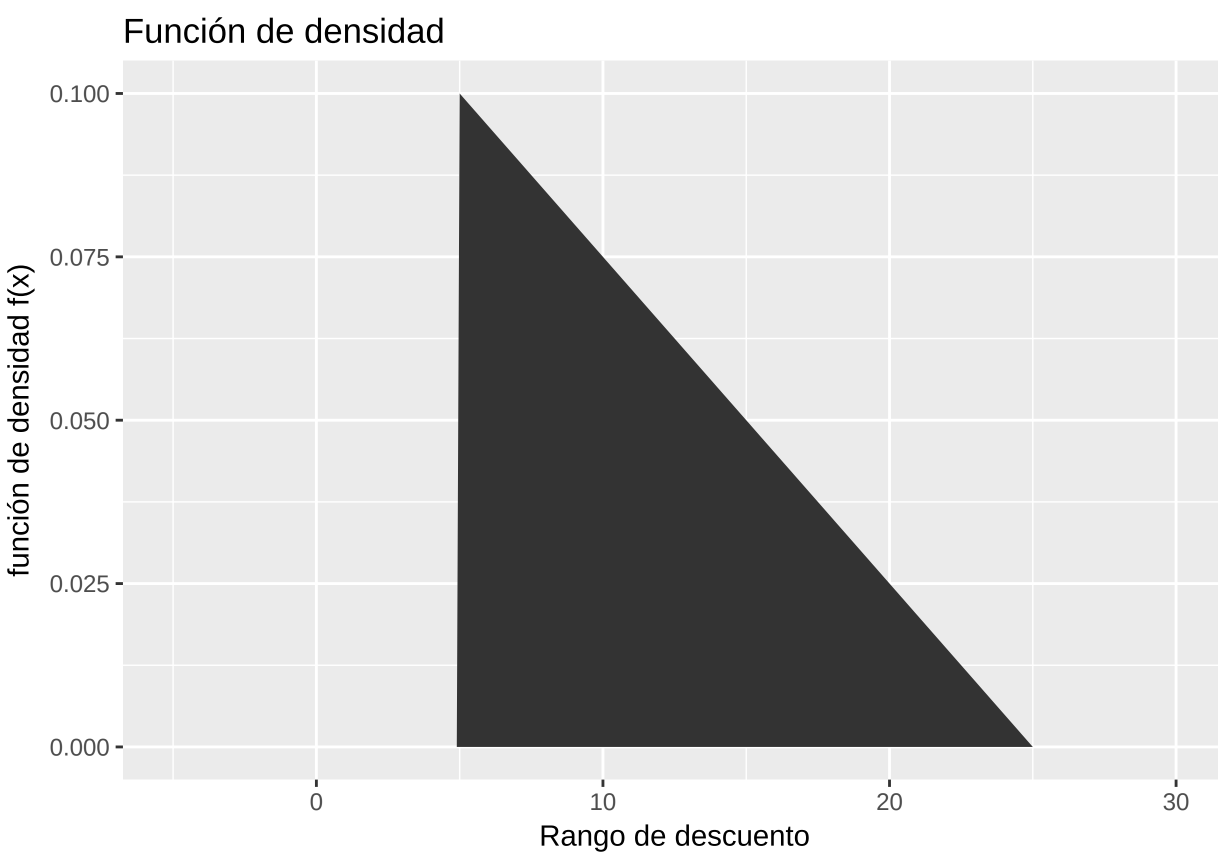
```
fun_inf <- function(x){
  k <- 1/200
  ifelse(x>=5&x<=25 ,(25-x)*k, 0)
}
```

Visualización de la función de densidad:

```
dto <- seq(-5,30,0.1)
densidad <- fun_inf(dto)

p <- ggplot(data.frame(dto,densidad),aes(x=dto,y=densidad)) +
        geom_area() +
        ggtitle("Función de densidad") +
        xlab("Rango de descuento") + ylab("función de densidad f(x)")

plot(p)
```



1.2 Ejercicio 2

Calcula la probabilidad de que un cliente obtenga mas de un 20% de descuento

$$P[X > a] = 1 - F(a) = \int_a^{\infty} f(x)dx$$

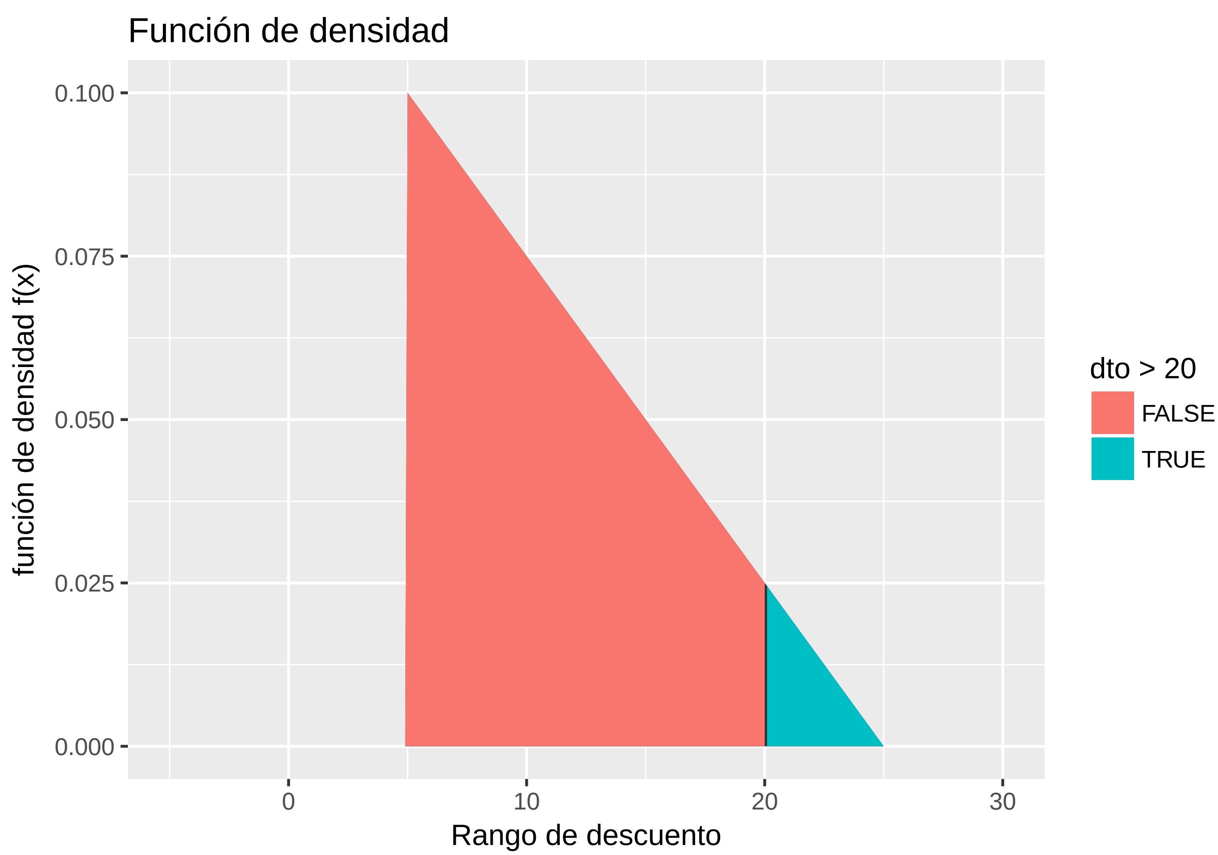
$$P[X > a] = 1 - F(a) = \int_{20}^{25} 0.005(25 - x)dx$$

```
integrate(func, 20, 25)
```

```
## 0.0625 with absolute error < 6.9e-16
```

Está probabilidad se muestra en la gráfica, color azul (es el valor de la integral).

```
p + geom_area(aes(fill=dto>20))
```



1.3 Ejercicio 3

¿Cuál es el descuento medio que se espera aplicar?

La esperanza matemática es el valor central entorno al cual se concentra la distribución de probabilidad. Se simboliza por μ = E[X], y se corresponde con su valor medio.

$$\mu = E[X] = \int_{-\infty}^{\infty} x.f(x)dx$$

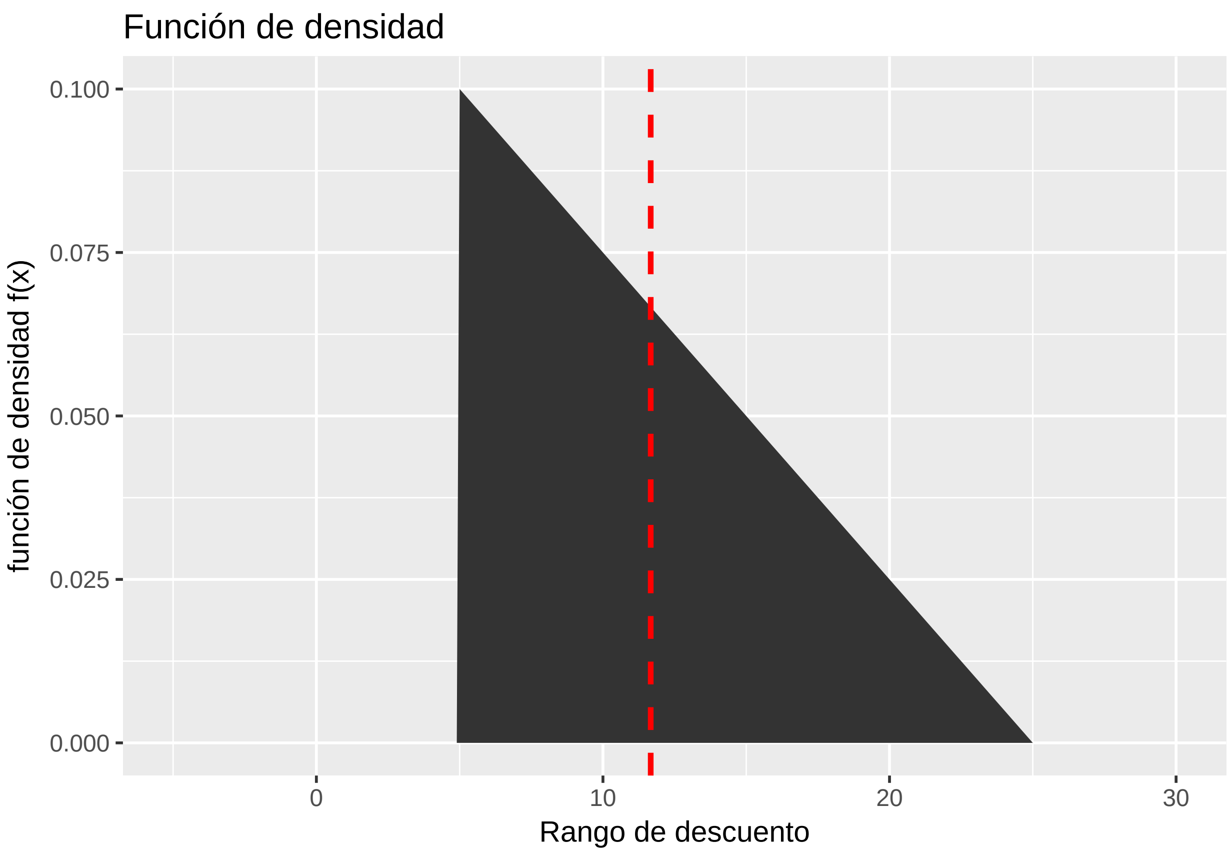
$$\mu = E[X] = \int_{5}^{25} x.0.005(25 - x)dx$$

```
esperanza_matematica = integrate(function(x) x*func(x), 5, 25)
esperanza_matematica
```

```
## 11.66667 with absolute error < 1.3e-13
```

El siguiente gráfico situamos el valor de la media sobre la función de densidad

```
p + geom_vline(aes(xintercept=esperanza_matematica$value), color="red", linetype="dashed
```



1.4 Ejercicio 4

Calcula la varianza de la variable aleatoria

La varianza es el parámetro de dispersión de la variable aleatoria. Se define como el momento de orden 2 respecto de la media.

$$Var[X] = \sigma^2 = \mu_2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

$$Var[X] = \sigma^2 = \mu_2 = \int_5^{25} (x-\mu)^2 0.005(25-x)dx$$

```
varianza = integrate(function(x) (x-esperanza_matematica$value)^2 * func(x), 5, 25)
varianza
```

```
## 22.22222 with absolute error < 2.5e-13
```

1.5 Ejercicio 5

¿Entre qué valores estarán probablemente la mitad de los descuentos realizados (en la zona central de la distribución)?

Calculo de primer y tercer cuartil:

"Necesitamos obtener la función de distribución y después su inversa. Así, podremos obtener cualquier cuantil, incluida la mediana."

```
midistribucion = function(q) integrate(func, 5, q)[[1]]
micuantil = function(p) uniroot(function(x) midistribucion(x) - p, c(5, 25))[[1]]
primer_quartil = micuantil(0.25)
primer_quartil
```

```
## [1] 7.679501
```

```
tercer_quartil = micuantil(0.75)
tercer_quartil
```

```
## [1] 15
```

Comprobamos

```
integrate(function(x) 0.005*(25 - x), 5, primer_quartil)
```

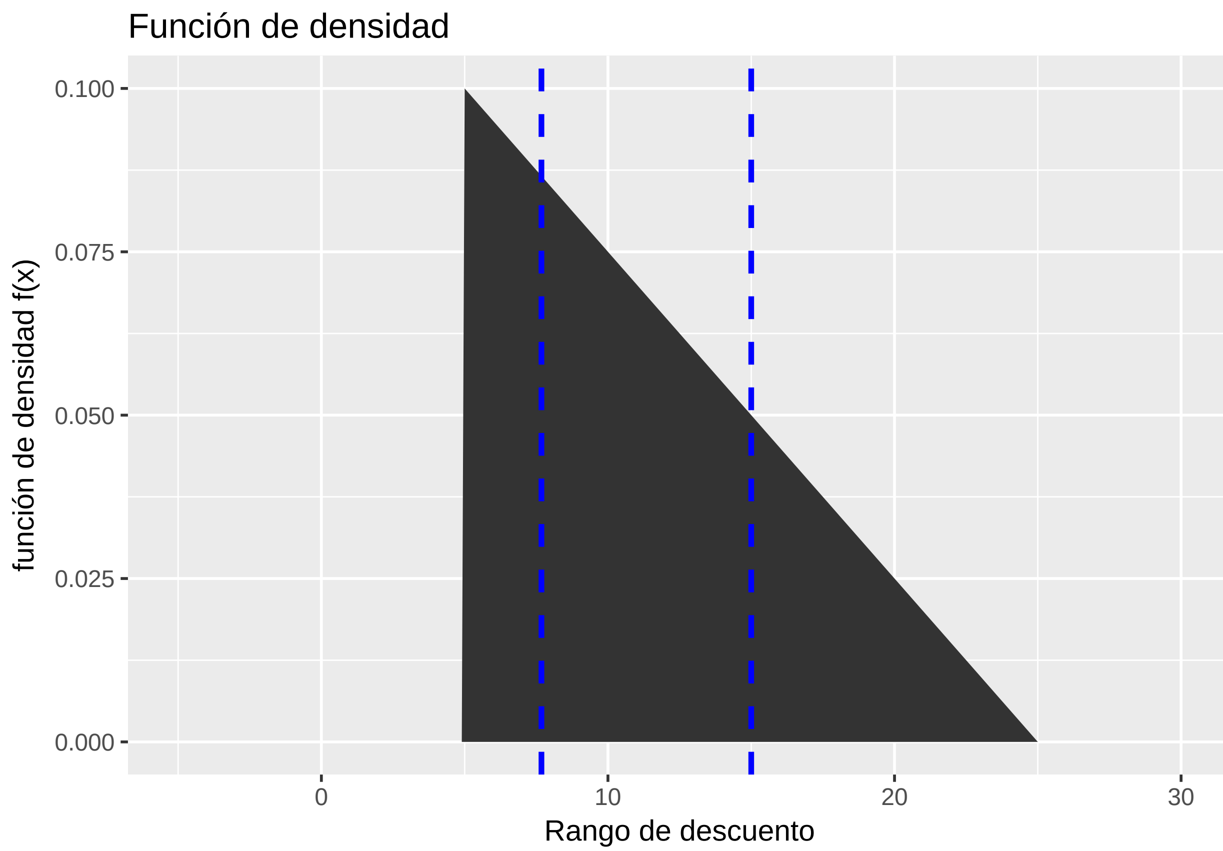
```
## 0.2500008 with absolute error < 2.8e-15
```

```
integrate(function(x) 0.005*(25 - x), 5, tercer_quartil)
```

```
## 0.7500001 with absolute error < 8.3e-15
```

Gráficamente:

```
p + geom_vline(aes(xintercept=primer_quartil), color="blue", linetype="dashed", size=1)
```



1.6 Ejercicio 6

¿Cuál es la moda de la variable aleatoria?

$$Mo(X) = \max_x f(x)$$

$$f'[x] = 0$$

```
expr = expression(0.005 * (25 - x))
D(expr,"x")
```

```
## -0.005
```

Como la funcion de densidad es lineal, el maximo estará en los extremos ya que la derivada primera valdrá -0.005 a lo largo de toda la curva.

Calculamos el maximo utilizando la funcion optimize:

```r
optimize(func, c(5,25), maximum = TRUE)
```

```
## $maximum
## [1] 5.000074
##
## $objective
## [1] 0.09999963
```

El máximo de la función se encuentra en 5, por lo que éste será el valor de la moda.

1.7 Ejercicio 7

Calcula la mediana de la variable aleatoria

$$Mo(X) = x/F(x) = 0,5$$

```r
mediana = micuantil(0.5)
mediana
```

```
## [1] 10.85787
```

Comprobacion:

```r
integrate(function(x) 0.005*(25 - x), 5, mediana)
```

```
## 0.5000002 with absolute error < 5.6e-15
```