Capítulo 2

Simulación de distribuciones de probabilidad utilizando la función cuantil

En la práctica anterior hemos visto cómo generar valores aleatorios que provienen de algunas de las distribuciones más habituales, además de ver cómo calcular probabilidades o los valores de las funciones de distribución e inversa. En esta práctica vamos a ver cómo manejar otro tipo de distribuciones y, para ello, vamos a hacer uso de la función cuantil.

2.1. Función cuantil

En la primera práctica introducimos el concepto de función cuantil, que viene dada por:

$$F^{(-1)}(p) = \text{ínf}\{x \in \mathbb{R} \mid F(x) \geq p\}.$$

$F^{(-1)}(p)$ indica el menor valor t que deja por debajo suyo una probabilidad de al menos p. A continuación, mostramos dos ejemplos detallando cómo calcular la función cuantil.

Ejemplo 2.1. El primer ejemplo es muy sencillo: consideramos una variable aleatoria X cuya función de distribución viene dada por:

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0. \\ x^2 & \text{si } t \in [0,1). \\ 1 & \text{si } t \geq 1. \end{cases}$$

Esta función de distribución es continua, y en particular es estrictamente creciente en el intervalo $(0,1) = \{t \in \mathbb{R} \mid F(t) \in (0,1)\}$. Eso significa que la función cuantil $F_X^{(-1)}$ no es más que la inversa de F_X en dicho intervalo:

$$F_X^{(-1)}(p) = \sqrt{p}, \qquad \forall p \in (0,1).$$

Ejemplo 2.2. Consideramos a continuación una variable aleatoria discreta X cuya distribución de probabilidad viene dada por:

X	1	2	3
p_X	0.3	0.3	0.4

La función de distribución asociada a esta variable es:

$$F_X(t) = \begin{cases} 0 & \text{si } t < 1. \\ 0.3 & \text{si } t \in [1,2). \\ 0.6 & \text{si } t \in [2,3). \\ 1 & \text{si } t \geq 3. \end{cases}$$

Vamos a calcular su función cuantil. Comencemos tomando un valor $p \in (0, 0.3)$. Para este valor, se cumple que:

$$\{t \in \mathbb{R} \mid F_X(t) \geq p\} = [1, \infty).$$

Por lo tanto, el ínfimo de dicho conjunto será el 1: $F_X^{(-1)}(p) = 1$. Lo mismo ocurre si tomamos el valor $p = 0.3$. Por tanto, $F_X^{(-1)}(p) = 1$ para $p \in (0, 0.3]$.

Tomamos a continuación un valor $p \in (0.3, 0.6]$. Para este valor, tenemos que:

$$\{t \in \mathbb{R} \mid F_X(t) \geq p\} = [2, \infty),$$

y por lo tanto $F_X^{(-1)}(p) = 2$. Por último, si tomamos $p \in (0.5, 1)$, se cumple que:

$$\{t \in \mathbb{R} \mid F_X(t) \geq p\} = [3, \infty),$$

Por lo que $F_X^{(-1)}(p) = 3$. Concluimos por tanto que la función cuantil viene dada por:

$$F_X^{(-1)}(p) = \begin{cases} 1 & \text{si } p \in (0, 0.3]. \\ 2 & \text{si } p \in (0.3, 0.6]. \\ 3 & \text{si } p \in (0.6, 1). \end{cases}$$

Ejemplo 2.3. Por último, consideremos una variable aleatoria X que no es ni continua ni discreta. Su función de distribución viene dada por:

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0. \\ \frac{1}{4}t & \text{si } t \in [0,1). \\ \frac{1}{4}t + \frac{1}{2} & \text{si } t \in [1,2). \\ 1 & \text{si } t \geq 2. \end{cases}$$

Esta función de distribución representa que la probabilidad se distribuye de la siguiente manera: el 50 % de la probabilidad se asigna al valor 1, mientras que el otro 50 % se distribuye de manera uniforme en el intervalo $(0, 2)$.

Para calcular la función cuantil, hemos de notar que en el intervalo $(0, 1) = \{t \in \mathbb{R} \mid F_X(t) \in (0, 0.25)\}$, F_X es estrictamente creciente. Por lo tanto, para $p \in (0, 0.25)$ se cumple que $F_X^{(-1)}(p) = 4p$.

Por otra parte, para $p \in [0.25, 0.75]$, se cumple que:

$$\{t \in \mathbb{R} \mid F_X(t) \geq p\} = [1, \infty),$$

así que $F_X^{(-1)}(p) = 1$ para $p \in [0.25, 0.75]$. Por último, en el intervalo $(1, 2) = \{t \in \mathbb{R} \mid F_X(t) \in (0.75, 1)\}$, F_X es estrictamente continua, por lo tanto para $p \in (0.75, 1)$ se tiene que $F_X^{(-1)}(p) = 4p - 2$, es decir, la inversa de F_X. Concluimos que:

$$F_X^{(-1)}(p) = \begin{cases} 4p & \text{si } p \in (0, 0.25). \\ 1 & \text{si } p \in [0.25, 0.75]. \\ 4p - 2 & \text{si } p \in (0.75, 1). \end{cases}$$

La función cuantil tiene una propiedad muy importante y que será esencial a lo largo de esta práctica. La enunciamos a continuación:

Teorema. *Sea X una variable aleatoria con función de distribución F_X. Si denotamos por $F_X^{(-1)}$ a su función cuantil, y por U a una variable aleatoria con distribución $U = \mathcal{U}(0,1)$, se cumple que $F_X^{(-1)}(U)$ es una variable aleatoria cuya distribución coincide con la distribución de X.*

¿Cuál es el significado del resultado anterior? Básicamente nos indica que si conocemos la función cuantil de la variable X, $F_X^{(-1)}$, podemos conocer su distribución sin más que hacer uso de la transformación $F_X^{(-1)}(U)$, donde U sigue una distribución uniforme en el intervalo $(0,1)$.

2.2. Simulación de valores utilizando la distribución cuantil

Partimos a continuación de una variable aleatoria X con función de distribución F_X. Nuestro objetivo es simular valores aleatorios de la variable X. Teniendo en cuenta el teorema enunciado anteriormente, hemos de seguir los siguientes pasos:

(1) Calcular la función cuantil de la variable X, F_X^{-1}.

(2) Simular un valor de una distribución uniforme en el intervalo $(0,1)$.

(3) Transformar los valores aleatorios obtenidos en el paso (2) mediante F_X^{-1}.

Veamos algunos ejemplos para ver cómo dar estos pasos:

Ejemplo 2.4. Consideremos una variable aleatoria X cuya probabilidad está distribuída de la siguiente manera: con probabilidad 0.5 toma el valor 0, y con probabilidad 0.5 se distribuye de manera uniforme en el intervalo $(0,1)$. Eso significa que su función de distribución F_X viene dada por:

$$F_X(t) = P(X \leq t) = \begin{cases} 0 & \text{si } t < 0. \\ 0.5 & \text{si } t = 0. \\ 0.5 + \frac{t}{2} & \text{si } t \in (0,1). \\ 1 & \text{si } t \geq 1. \end{cases} \tag{2.1}$$

A continuación, vamos a hacer lo siguiente:

1. Vamos a simular $n = 1000$ valores aleatorios de esta variable.
2. Calcularemos la función de distribución empírica y la dibujaremos.
3. Estimaremos la probabilidad $P(X \in (0, 0.5))$.
4. Estimaremos $E(X)$, $DT(X)$ y $Med(X)$.

Comencemos generando $n = 1000$ valores aleatorios de esta variable. Para ello, primero calculamos su función cuantil asociada, que viene dada por:

$$F_X^{-1}(p) = \begin{cases} 0 & \text{si } 0 < p \leq 0.5. \\ 2p - 1 & \text{si } 0.5 < p < 1. \end{cases}$$

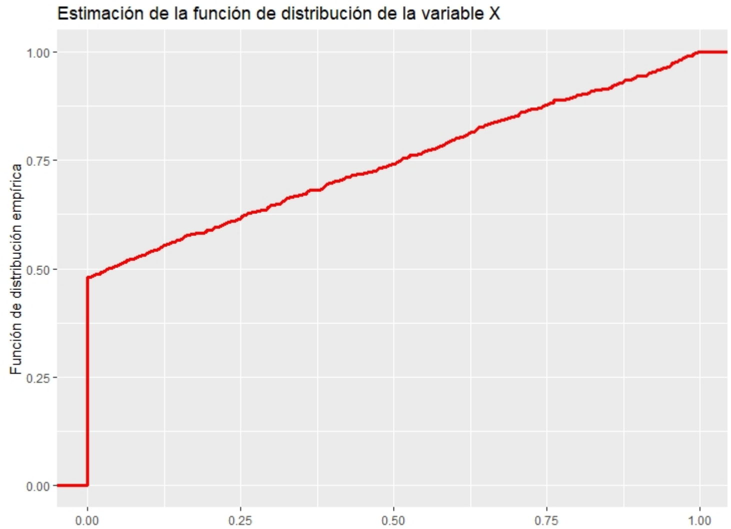

Figura 2.1: Función de distribución empírica construida a partir de los datos simulados en el Ejemplo 2.4.

A continuación, generamos $n = 1000$ valores aleatorios de una distribución uniforme $(0, 1)$. Para ello, hacemos uso de las instrucciones vistas en la práctica anterior:

```
n<-1000
u<-runif(n,0,1)
```

A continuación hacemos la transformación por la función cuantil. Para ello podemos utilizar, por ejemplo, la siguiente instrucción:

```
x<-(u>0.5)*(2*u-1)
```

Pasamos a continuación a representar gráficamente la función de distribución empírica obtenida con la simulación. Para ello podemos utilizar las siguientes instrucciones ya vistas en la práctica anterior:

```
library(ggplot2)  # Instrucción para cargar el paquete ggplot2
ggplot(data=NULL, aes(x))+
  stat_ecdf(col="red",size=1.2)+  # Instrucción para dibujar la ECDF
  labs(x="", y="Función de distribución empírica")+
    # Añadir etiquetas a los ejes
  ggtitle("Estimación de la función de distribución de la variable X")
    # Añadir un título al gráfico
```

La gráfica obtenida se ha representado en la Figura 2.1, donde se puede ver que la función de distribución empírica se parece la función de distribución de la Ecuación (2.1).

A continuación vamos a utilizar los valores aleatorios simulados de la variable X para estimar probabilidades. Si queremos estimar la probabilidad del intervalo $(0, 0.5)$, podemos utilizar las siguientes instrucciones:

```
sum(x>0 & x<0.5)/n
[1] 0.256
```

Por último, podemos estimar la esperanza y la desviación típica utilizando las instrucciones `mean` y `std`:

```
mean(x)
[1] 0.2556101
sd(x)
[1] 0.3181544
median(x)
[1] 0.02681938
```

Ejemplo 2.5. Consideremos una variable aleatoria X discreta tomando los siguientes valores:

X	0	1	2	3
p_X	0.2	0.3	0.1	0.4

Con esta variable, haremos lo siguiente:

1. Vamos a simular $n = 200$ valores aleatorios.
2. Vamos a estimar la media y la desviación típica.
3. Estimaremos las probabilidades de cada uno de los valores.
4. Calcularemos la función de distribución empírica.

Comencemos simulando $n = 200$ valores que provengan de esta variable. Para ello, primero calculamos la función de distribución:

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0. \\ 0.2 & \text{si } t \in [0, 1). \\ 0.5 & \text{si } t \in [1, 2). \\ 0.6 & \text{si } t \in [2, 3). \\ 1 & \text{si } t \geq 3. \end{cases}$$

Su función cuantil viene dada por:

$$F_X^{-1}(p) = \begin{cases} 0 & \text{si } 0 < p \leq 0.2. \\ 1 & \text{si } 0.2 < p \leq 0.5. \\ 2 & \text{si } 0.5 < p \leq 0.6. \\ 3 & \text{si } 0.6 < p < 1. \end{cases}$$

A continuación, generamos valores aleatorios de una uniforme en $(0, 1)$:

```
n<-200;
u<-runif(n,0,1)
```

Por último, transformamos los valores anteriores mediante la función cuantil. La manera más rápida de hacerlo, aunque no la única, es utilizando la siguiente instrucción:

```
x<-1*(u>0.2&u<=0.5)+2*(u>0.5&u<=0.6)+3*(u>0.6)
```

Para estimar la media y la desviación típica, usamos las siguientes instrucciones:

```
mean(x)
[1] 1.69
sd(x)
[1] 1.166492
```

Para hacer las estimaciones de las probabilidades de cada punto, podemos utilizar las siguientes expresiones:

```
p<-rep(0,4)  # Se inicializa un vector de ceros donde almacenar las
             # estimaciones de la probabilidad de cada punto.
for(i in 0:3){
        p[i+1]<-sum(x==i)/n  # Se estima la probabilidad del valor i.
        }
p
[1] 0.19 0.31 0.12 0.38
```

Por último, para calcular la función de distribución empírica, podemos utilizar las siguientes instrucciones:

```
xcdf<-ecdf(x)
```

Si en particular queremos estimar los valores de la función de distribución empírica en el punto 2, utilizaremos la instrucción:

```
xcdf(2)
```

Y para dibujar la función de distribución empírica, utilizamos el paquete `ggplot2` como ya se ha hecho en otras ocasiones:

```
library(ggplot2)  # Instrucción para cargar el paquete ggplot2
ggplot(data=NULL, aes(x))+
  stat_ecdf(col="red",size=1.2)+  # Instrucción para dibujar la ECDF
  labs(x="", y="Función de distribución empírica")+
    # Añadir etiquetas a los ejes
  ggtitle("Estimación de la función de distribución de la variable X")
    # Añadir un título al gráfico
```

2.3. Ejercicios propuestos

Ejercicio 2.1. Considera la variable aleatoria X cuya función de distribución aparece en el Ejemplo 2.1. Se pide:

1. Simula $n = 100$ valores aleatorios de X.

2. Utiliza la muestra anterior para estimar las probabilidades $P(X < 0.5)$, $P(X \leq 0.5)$, $P(X \in (0.25, 0.75))$.

3. Estima el valor de la mediana, $E(X)$ y $Var(X)$.

Ejercicio 2.2. Considera la variable aleatoria X cuya función de distribución aparece en el Ejemplo 2.2. Se pide:

1. Genera $n = 1000$ valores aleatorios utilizando las instrucciones vistas en la práctica anterior.

2. Calcula la función de distribución empírica y dibújala.

3. Repite los apartados anteriores pero simulando los valores aleatorios mediante la función cuantil.

Ejercicio 2.3. Considera la variable aleatoria X cuya función de distribución aparece en el Ejemplo 2.3. Se pide:

1. Simula $n = 2000$ valores aleatorios de X.

2. Estima $E(X), DT(X)$ y la mediana.

3. Estima las siguientes probabilidades: $P(X < 0.5)$, $P(X \in (0.5, 1))$ y $P(X \in (0.5, 1])$.

4. Calcula la función de distribución empírica y utilízala para estimar $P(X \leq 1.5)$.

Ejercicio 2.4. Considera una variable aleatoria X cuya función de distribución viene dada por:

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0. \\ \frac{1}{4} & \text{si } t = 0. \\ \frac{1+2x}{4} & \text{si } 0 < t < 1. \\ 1 & \text{si } t \geq 1. \end{cases}$$

Se pide:

1. Simula $n = 500$ valores aleatorios que provengan de esta distribución.

2. Dibuja la función de distribución empírica y la función de distribución F_X.

3. Estima $E(X)$, $DT(X)$, $Med(X)$, $P(X \in (0.4, 1))$ y $P(X \in (0.4, 1])$.

4. Repite los apartados anteriores utilizando una muestra de tamaño $n = 5000$.

Ejercicio 2.5. Considera una variable aleatoria X cuya función de distribución es:

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq 0. \\ t & \text{si } 0 < t < \frac{1}{2}. \\ \frac{1}{2} & \text{si } \frac{1}{2} \leq t < 1. \\ 1 & \text{si } t \geq 1. \end{cases}$$

Se pide:

1. Simula $n = 1000$ valores aleatorios de X.

2. Estima $P(X = 0.5)$ y $P(X \in (0.5, 0.8))$.

3. Estima $E(X)$, $DT(X)$ y los cuartiles de X.

4. Calcula y dibuja la función de distribución empírica.

Ejercicio 2.6. Considera una variable aleatoria X cuya función de distribución viene dada por:

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0. \\ \frac{1}{2}x + \frac{1}{4} & \text{si } t \in [0, 1). \\ 1 & \text{si } t \geq 1. \end{cases}$$

Se pide:

1. Simular $n = 200$ valores aleatorios de X.
2. Estima $E(X), DT(X)$ y el primer cuartil.
3. Calcula la función de distribución empírica y dibújala.
4. ¿Se te ocurre alguna manera de simular valores aleatorios de X sin necesidad de utilizar la función cuantil (solamente utilizando las instrucciones de la primera práctica)?