# Capítulo 5

# Transformaciones de variables aleatorias

En prácticas anteriores hemos visto cómo simular valores aleatorios en distintas situaciones: variables aleatorias (unidimensionales) con distribuación conocida, variables aleatorias (unidimensionales) a través de la función cuantil, vectores aleatorios o distribución normal bivariante. A lo largo de esta práctica veremos cómo simular valores aleatorios de transformaciones de variables aleatorias.

## 5.1. Transformaciones de variables aleatorias unidimensionales

Consideremos una variable aleatoria (unidimensional) $X$, y consideremos una función medible $g : \mathbb{R} \rightarrow \mathbb{R}$. De esta forma podemos definir una nueva variable aleatoria (también unidimensional) dada por $Y = g(X)$.

Ahora bien, aunque la variable $X$ tenga una distribución conocida y fácil de manejar, es posible que la nueva variable aleatoria $Y$ no sea tan manejable. La pregunta que surge es entonces: ¿cómo podemos conocer su comportamiento? En esta situación, podemos utilizar la simulación para poder estimar la distribución de esta nueva variable aleatoria. Para ello, basta simular una muestra aleatoria de la variable $X$ (como hemos hecho en las prácticas 1 y 2) y posteriormente aplicar sobre dicha muestra la función $g$.

**Ejemplo 5.1.** Sea $X \equiv \mathcal{U}(0,1)$, y consideremos la nueva variable aleatoria $Y = X^2$. Básicamente esta nueva variable aleatoria considera el cuadrado de la variable aleatoria original $X$. Vamos a seguir los siguientes pasos:

(a) Simularemos $n = 5000$ valores aleatorios de $Y$.

(b) Estimaremos los siguientes parámetros: $E(Y)$, $DT(Y)$, $Med(Y)$ y el primer y tercer cuantil.

(c) Estimaremos las probabilidades $P(Y \leq 0.5)$ y $P(0.25 < Y \leq 0.75)$.

(d) Dibujaremos su función de distribución empírica.

Comencemos simulando los valores aleatorios de $Y$. Para ello, primero simulamos la muestra aleatoria de $X$:

```
n<-5000
x<-runif(n,0,1)
```

para posteriormente realizar la transformación de estos valores:

```
y<-x^2
```

Recordemos con esta última instrucción se eleva al cuadrado punto a punto cada uno de los elementos del vector.

Una vez construida la muestra aleatoria, podemos estimar los parámetros pedidos:

```
mean(y)                  # Estimación de la media
sd(y)                    # Estimación de la desviación típica
median(y)                # Estimación de la mediana
quantile(y,c(0.75,0.25)) # Estimación de los cuantiles 0.75 y 0.25
```

Por otra parte, podemos estimar las probabilidades:

```
sum(y<=0.5)/n
sum(y<=0.75 & y>0.25)/n
```

Por último, para dibujar la función de distribución empírica, procedemos como en clases anteriores:

```
library(ggplot2)
ggplot(data=NULL, aes(y))+
  stat_ecdf(col="red",size=1.2)+
  labs(x="",y="Función de distribución")
```

De forma similar se puede dibujar la estimación de su función de densidad:

```
library(ggplot2)
ggplot(data=NULL, aes(y))+
  geom_density(col="red",size=1.2)+
  labs(x="",y="Función de densidad")
```

Los gráficos obtenidos se muestran en la Figura 5.1.

## 5.2. Transformaciones de vectores aleatorios

Consideremos a continuación un vector aleatorio bidimensional $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, y una función $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ medible. Con estas herramientas, podemos definir una variable aleatoria (unidimensional) dada por $Y = g(X_1, X_2)$.

Al igual que hicimos en la sección anterior, podemos utilizar la simulación para estimar el comportamiento de $Y$. Para ello, bastaría simular valores aleatorios de $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ y posteriormente aplicar la transformación $g$.

**Ejemplo 5.2.** Vamos a continuar con el Ejemplo 4.1 de la práctica anterior, donde $X = (X_1, X_2)$ sigue una distribución normal bivariante con vcetor de medias $\mu = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$ y matriz

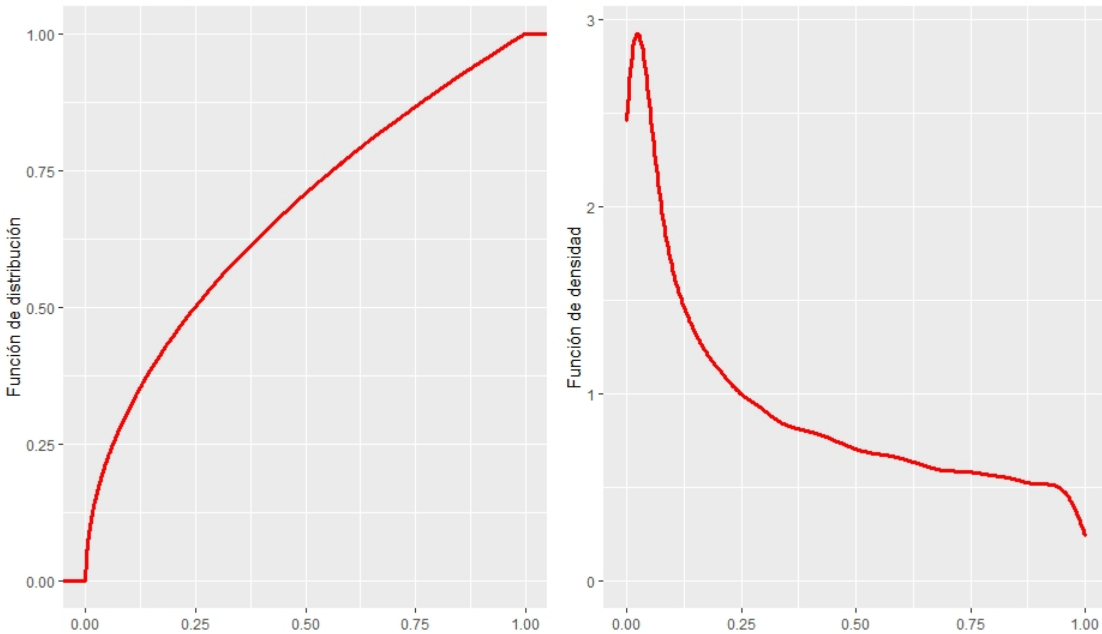


Figura 5.1: Función de distribución (izquierda) y de densidad (derecha) empírica construidas a partir de la muestra aleatoria.

de varianzas-covarianzas $\Sigma = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$. En dicho ejemplo habíamos visto simulado $n = 1000$ valores aleatorio utilizando las siguientes instrucciones:

```
library("MASS")
n<-1000
mu<-c(5,1)
sigma<-matrix(c(4,3,3,9),nrow=2)
x<-mvrnorm(n,mu,sigma)
```

En lo que sigue, nos interesará trabajar con la siguiente trasformación de $X = (X_1, X_2)$: $Y = X_1^2 + X_2^2$. Para esta nueva variable aleatoria, estamos interesados en hacer lo siguiente:

1. Estimar media, mediana y desviación típica.
2. Estimar $P(Y > 20)$.
3. Dibujar la función de distribución empírica.

Antes de comenzar a resolver los apartados anteriores, hemos de aplicar la trasformación a los valores simulados de $X = (X_1, X_2)$:

```
y<-x[,1]^2+x[,2]^2
```

El primer apartado lo resolvemos de manera sencilla, utilizando los comandos `mean`, `sd` y `median` para media, desviación típica y mediana, como ya hicimos en prácticas anteriores:

```
mean(y)          # Estimación de la media
[1] 40.0084
```

```
sd(y)            # Estimación de la desviación típica
[1] 29.43314
median(y)        # Estimación de la mediana
[1] 32.18269
```

En el segundo apartado nos piden estimar una probabilidad. Para estimar la probabilidad de $Y > 20$, hemos de ver cuántos elementos en el vector `y` cumplen la condición dada de entre los 1000 que hemos simulado:

```
sum(y>20)/n
[1] 0.74
```

Recordemos que con la expresión lógica `y>20` obtenemos un vector de ceros y unos donde aparecerán unos siempre que se cumpla la condición y ceros en otro caso. Al hacer la suma de la expressión lógica estamos contando el número de valores en los que se cumple la condición `y>20`, y al dividirlo entre 1000 obtenemos la frecuencia muestral.

Por último, vamos a dibujar la función de distribución empírica:

```
library(ggplot2)
ggplot(data=NULL, aes(y))+
  stat_ecdf(col="red",size=1.2)+
  labs(x="",y="Función de distribución")
```

## 5.3. Ejercicios propuestos

**Ejercicio 5.1.** En el Ejercicio 4.1 de la práctica anterior nos pedían simular valores aleatorios de una distribución normal bivariante donde $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ y ambas componentes son independientes. Simula 1000 valores aleatorios siguiendo alguno de los dos métodos indicados en dicho ejercicio y responde a las siguientes cuestiones:

1. Vamos a estudiar la variable $Y = X_1^2 + X_2^2$. Obtén los 1000 valores aleatorios de esta distribución.

2. Estima su media y su desviación típica. ¿Son estos valores cercanos a 2?

3. Estima la siguiente probabilidad: $F_Y(1.5) = P(Y \leq 1.5)$. A continuación, utilizando los comandos vistos en la primera práctica, calcula la probabilidad exacta $F_W(1.5) = P(W \leq 1.5)$, donde $W$ sigue una distribución chi-cuadrado con 2 grados de libertad.

4. Dibuja su función de distribución empírica, así como la función de distribución de $W$. ¿Son similares estas funciones?

**Nota:** Si $X_1, \ldots, X_k$ son variables independientes con distribución $\mathcal{N}(0,1)$, la variable $Y = X_1^2 + \ldots + X_k^2$ sigue una distribución chi-cuadrado con $k$ grados de libertad. En el caso de este ejercicio, $Y$ sigue una distribución chi-cuadrado con 2 grados de libertad.

**Ejercicio 5.2.** Continuemos con el Ejercicio 4.2 de la práctica anterior, donde habíamos simulado $n = 10000$ valores aleatorios de una distribución normal bivariante con vector de medias $\vec{\mu} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ y matriz de varianzas-covarianzas $\Sigma = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}$. A continuación estamos interesados en trabajar con la variable $Y = \frac{(X_1 - X_2)^2}{X_1^2 + 1}$.

1. Calcula $n = 10000$ valores aleatorios utilizando los valores aleatorios de $X = (X_1, X_2)$.

2. Estima media, mediana, primer y tercer cuantil así como su desviación típica.

3. Dibuja su función de distribución empírica.

4. Utiliza el comando `ecdf` para estimar las siguientes probabilidades: $P(Y \leq 0)$, $P(-1 < Y \leq 1)$ y $P(Y > -1)$.

**Ejercicio 5.3.** Estamos interesados en conocer la distribución de la variable $X^2$, donde $X$ sigue una distribución normal estándar. Vamos a estimar el comportamiento de esta nueva variable mediante simulación. Para ello, genera un vector con una muestra aleatoria de tamaño 10000 de una distribución normal estándar, y a continuación eleva al cuadrado cada uno de los elementos del vector. Una vez que tengas los valores aleatorios de $X^2$, responde a los siguientes apartados:

1. Estima la media, desviación típica y cuartiles de $X^2$.

2. Estima las siguientes probabilidades: $P\big(X^2 \in (0,1)\big)$, $P(X^2 > 1.5)$ y $P(X^2 > 0.75)$.

3. Dibuja la función de distribución empírica de $X^2$. ¿Cuál es la estimación de $F_{X^2}$ en el punto 1?

**Ejercicio 5.4.** Considera el vector aleatorio $\vec{X} = (X_1, X_2)'$ donde $X_1 \equiv \mathcal{N}(0,1)$ y $X_2 = \mathcal{U}(0,10)$, siendo ambas componentes independientes.

1. Genera $n = 1000$ valores aleatorios de $\vec{X} = (X_1, X_2)'$.

2. Queremos conocer la distribución de la variable aleatoria $W = \frac{X_1^2}{X_2}$. Dibuja su función de distribución empírica y estima su media y mediana. ¿Cuál es la estimación de $P(1 < W \leq 4)$?