Contents

1 ¿Qué es esto? 1

2 2. Random variable generation 2

2.1 Introduction . . . 2

2.1.1 The inverse transformation . . . 2

2.1.2 2.2: General transformation methods . . . 6

2.1.3 2.2.1 A normal generator . . . 7

2.1.4 2.2.2 Discrete distributions . . . 9

3 Notas sobre R 12

3.1 Sobre los histogramas . . . 12

4 Notación 14

Chapter 1

¿Qué es esto?

Los ejercicios, desarrollos de argumentos y códigos realizados por mí mientras estudio el libro *Introducing Monte Carlo Methods with R* de Christian P. Robert y George Casella.

Chapter 2

2. Random variable generation

2.1 Introduction

2.1.1 The inverse transformation

$$
\begin{tikzcd}
\Omega \arrow[r, "X"] \arrow[dr, "F_X \circ X"'] & \mathbb{R} \arrow[d, "F_X"] \\
& (0,1)
\end{tikzcd}
$$

$F_X \circ X \sim Unif(0,1)$

Teorema 2.1.1

*(**p.44**) Sea X una variable aletoria continua, con funciones de densidad f_X y F_X, respectivamente, con $F_X : \mathbb{R} \longrightarrow (0,1)$ invertible. La variable aleatoria*

$$U = F_X \circ X$$

tiene distribución $\mathcal{U}(0,1)$.

Demostración. En efecto, sea $u \in (0,1)$. Puesto que por hipótesis F_X es invertible (o, equivalentemente, biyectiva), existe un único $x \in \mathbb{R}$ tal que $F_X(x) = u$. Se tiene entonces que

$$
\begin{aligned}
P(U \leq u) = P(F_X(X) \leq F_X(x)) = P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(F_X(X))) \\
= P(X \leq x) = F_X(x) = u.
\end{aligned}
$$

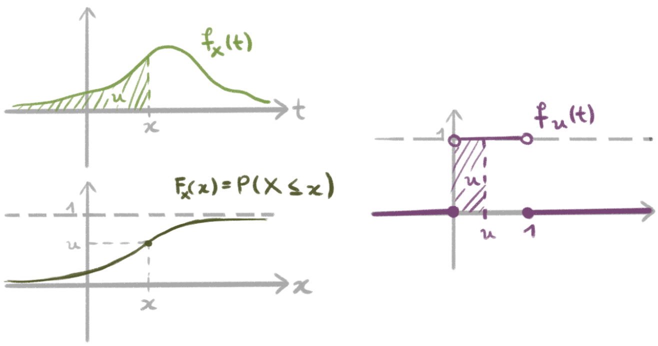

Figure 2.1: a

$\square$

El resultado dual se da a continuación.

Ejercicio 1

*(**2.1, p.44**) Si X es una variable aleatoria con función de acumulación F_X, defina a la **inversa generalizada de** F_X como sigue:*

$$\forall u \in (0,1): \quad F_X^-(u) := inf\{x \;:\; F_X(x) \geq u\}. \tag{2.1}$$

Demuestre que si U es una variable aleatoria con distribución $U(0,1)$, entonces

$$F_X^- \circ U \sim X.$$

Demostración.

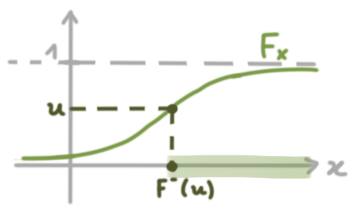

Figure 2.2: Definición de F_X^-.

Sea $Y : \Omega \longrightarrow \mathbb{R}$ la composición $F_X^- \circ U$. Debemos demostrar que

$$\forall z \in \mathbb{R}: \quad P(Y \leq z) = P(X \leq z).$$

Sea pues $z \in \mathbb{R}$.

a) Por ser F_X una función de acumulación, es una función creciente, luego, los eventos

$$(X \leq z) \;\text{ y }\; (F_X(X) \leq F_X(z))$$

son iguales; además, según el teorema 2.1.1, $F \circ X \sim U$; así,

$$P(X \leq z) = P(F_X(X) \leq F_X(z)) = P(U \leq F_X(z)). \tag{2.2}$$

b) Por la definición de F^- en términos de un ínfimo, se sigue de inmediato que

$$\begin{aligned} P(Y \leq z) = P(F_X^-(U) \leq z) =& P(\{w \in \Omega \;:\; F_X^-(U(w)) \leq z\}) \\ =& P(\{w \in \Omega \;:\; U(w) \leq F_X(z)\}) = P(U \leq F_X(z)), \end{aligned}$$

o sea, que

$$P(Y \leq z) = P(U \leq F_X(z)); \tag{2.3}$$

de (2.2) y (2.3) se sigue, como queríamos, que $P(X \leq z) = P(Y \leq z)$.

□

Ejercicio 2

> *(**2.2, p. 45**) Se dan algunas funciones de probabilidad. Calcule las correspondientes funciones de acumulación, y en base a estas y los resultados anteriores, simule via una transformación de variables aleatorias uniformes la distribución dada. Comparar el resultado obtenido con las gráficas hechas llamando a la distribución en R.*

Solución

a) **Distribución logística**: la pdf es

$$f(x) = \frac{1}{\beta}\frac{e^{-(x-\mu)/\beta}}{(1+e^{-(x-\mu)/\beta})^2} \tag{2.4}$$

(al parámetro μ se le llama **localización**, y al β **escala**). Integrando, calculamos que

$$\forall x \in \mathbb{R}: \quad F(x) = \frac{1}{1+e^{-(x-\mu)/\beta}};$$

la inversa de esta función es

$$F^{-1}(u) = ln\left(\left(\frac{u}{1-u}\right)^{\beta}\right) + \mu,$$

luego, según el ejercicio 1, si $U \sim U(0,1)$, la variable aleatoria

$$Y = ln\left(\left(\frac{U}{1-U}\right)^{\beta}\right) + \mu$$

tiene a la función (2.4) como función de distribución.

Los valores de default de los parámetros de la distribución logística en `rstudio` son

$$\mu = 0, \quad \beta = 1;$$

a continuación comprobamos con un experimento numérico en `rstudio` que las variables aleatorias

$$X = logistic(\mu = 0, \beta = 1) \;\text{ y }\; Y = ln\left(\frac{U}{1-U}\right) \tag{2.5}$$

parecen tener la misma distribución.

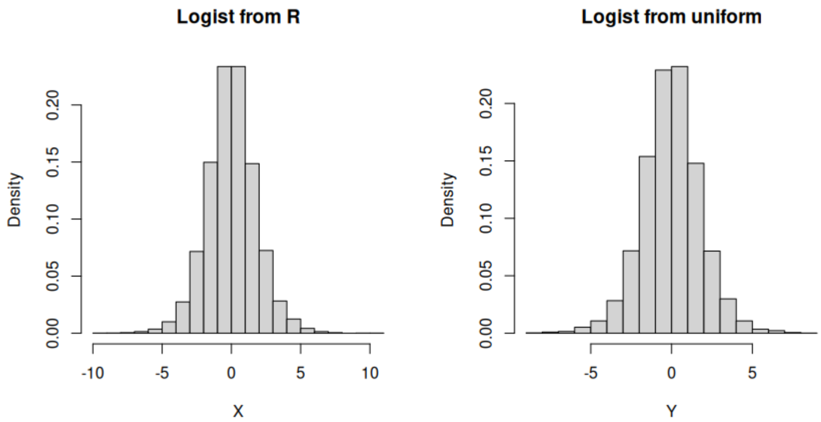

Figure 2.3: Histogramas de distribuciones logísticas (tamaño de muestra: 10^4) usando las variables aleatorias (2.5).

Código de R empleado para generar la imagen:

```
> Nsim=10^4
> U=runif(Nsim)
> X=rlogis(Nsim)
> Y=log(U/(1-U))
> par(mfrow=c(1,2))
> hist(X, freq=FALSE, main='Logist from R')
> hist(Y, freq=FALSE, main='Logist from uniform')
```

b) **Distribución de Cauchy**: la pdf es

$$f(x) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Después de integrar se llega a que

$$\forall x \in \mathbb{R}: \quad F(x) = \frac{1}{2}\frac{1}{\pi}arctan\left(\frac{x-\mu}{\sigma}\right).$$

La inversa de esta transformación es

$$F^{-1}(u) = \mu + \sigma tan\left(\pi u - \frac{\pi}{2}\right), \quad 0 < u < 1,$$

luego, según el ejercicio 1, si $U \sim U(0,1)$, la variable aleatoria

$$F^{-1}(U) = \mu + \sigma tan\left(\pi U - \frac{\pi}{2}\right)$$

tiene distribución de Cauchy con parámetros μ y σ.

Haciendo $\mu = 0$ y $\sigma = 1$, podemos graficar los histogramas de 10^4 variables aleatorias

$$X \sim Cauchy(\mu, sigma), \quad \text{y} \quad Z = \mu + \sigma tan\left(\pi U - \frac{\pi}{2}\right) \tag{2.6}$$

y compararlos; como se ve en la figura, resultado de correr el código de más abajo, estos histogramas son casi idénticos.

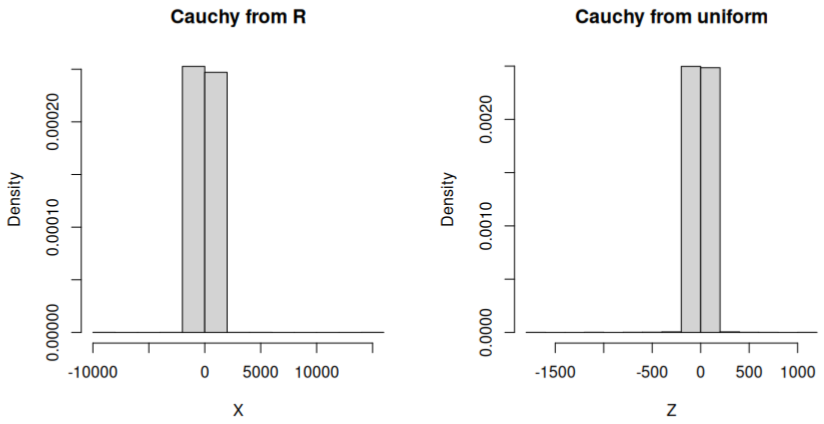

Figure 2.4: Histogramas de distribuciones Cauchy (tamaño de muestra: 10^4) usando las variables aleatorias (2.6).

Código de R empleado para generar la imagen:

```
> Nsim=10^4
> U=runif(Nsim)
> X=rcauchy(Nsim)
> Z=tan(3.1416*(U-0.5))
> par(mfrow=c(1,2))
> hist(X, freq=FALSE, main='Cauchy from R')
> hist(Z, freq=FALSE, main='Cauchy from uniform')
```

2.1.2 2.2: General transformation methods

Según un ejemplo del libro, si $U \sim U(0,1)$, entonces $-ln(U) \sim Exp(1)$. Se establecen además las siguientes relaciones:

Teorema 2.1.2

Sea $\{X_i\}_{i\in\mathbb{N}}$ iid $Exp(1)$ variables aleatorias. Para cualesquiera $a, b, v \in \mathbb{N}2^*$,

$$Y = 2\sum_{j=1}^{v} X_j \sim \chi^2_{2v}, \tag{2.7}$$

$$Y = \beta\sum_{j=1}^{a} X_j \sim \mathcal{G}(a,\beta), \tag{2.8}$$

$$Y = \frac{\sum_{j=1}^{a} X_j}{\sum_{j=1}^{a+b} X_j} \sim Beta(a,b). \tag{2.9}$$

Comprobemos las fórmulas (2.7), (2.8) y (2.9) con algunas simulaciones en R.

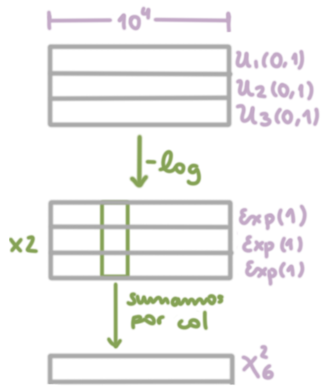

Figure 2.5: Explicación gráfica del proceso de simulación.

```
> U=runif(3*10^4)
> U=matrix(data=U, nrow=3)
```

```
> X=-log(U)
> Y=2*apply(X, 2, sum)
> Z=rchisq(10^4, df=6)
> par(mfrow=c(1,2))
> hist(Z, freq=FALSE, main='Chi df=6 from R')
> hist(Y, freq=FALSE, main='Chi df=6 from Exp')
```

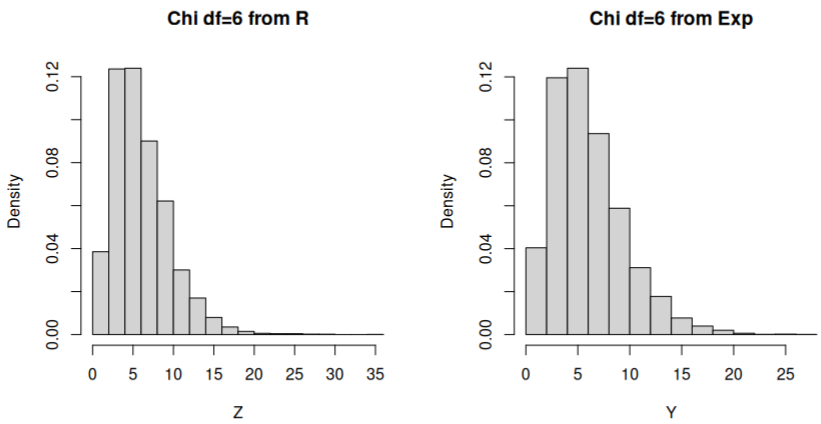

Figure 2.6: Histogramas de $Y = 2(X_1 + X_2 + X_3)$, con X_i, $1 \leq i \leq 3$ iid $Exp(1)$ y de $Z \sim \chi^2_6$.

No pude simular correctamente las otras dos distribuciones!

2.1.3 2.2.1 A normal generator

Teorema 2.1.3

(Box-Muller) *if U_1 and U_2 are iid uniforms in $]0,1[$, then the variables*

$$X_1 := \sqrt{-2log(U_1)}cos(2\pi U_2)$$

and

$$X_2 := \sqrt{-2log(U_1)}sen(2\pi U_2)$$

are iid N(0,1).

```
> n = 10**4
> U1 = runif(n)
> U2 = runif(n)
> X1 = sqrt(-2*log(U1))*cos(2*pi*U2)
> X2 = sqrt(-2*log(U1))*sin(2*pi*U2)
> N = rnorm(n,0,1)
> par(mfrow=c(1,3))
> hist(X1, freq=FALSE, main='Primera normal de BM')
> hist(X2, freq=FALSE, main='Segunda normal de BM')
> hist(N, freq=FALSE, main='Normal de R')
```

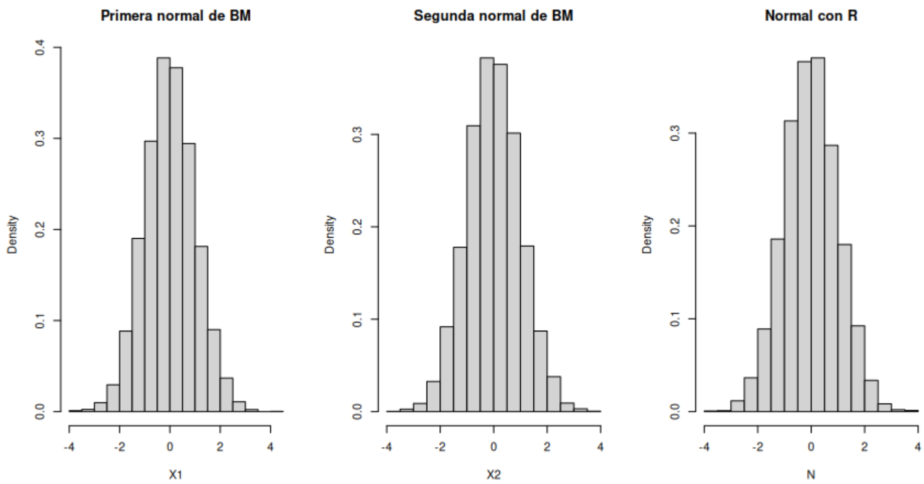

Figure 2.7: Usamos el algoritmo de Box-Muller para generar dos normales estandar. Comparamos sus histogramas con el de una normal generada por R.

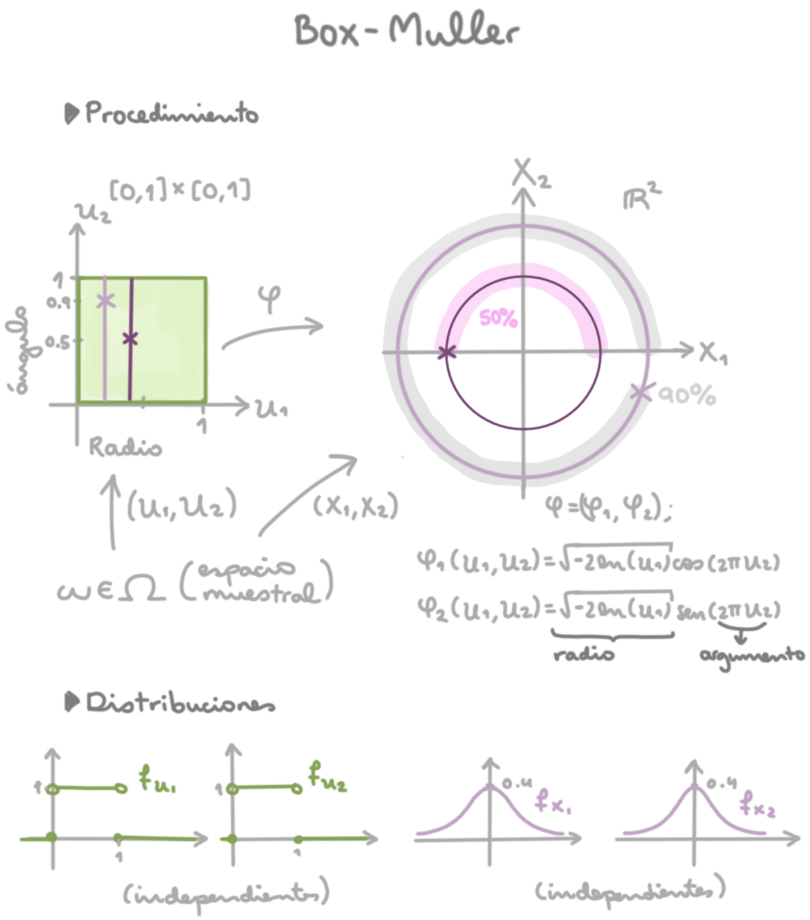

Figure 2.8

Ejercicio 3

(2.3) *Podemos usar el* ***teorema del límite central*** *para generar una distribución normal a partir de variables aleatorias uniformes e independientes.*

Sean $U_i \sim U[-1/2, 1/2]$, $0 \leq i \leq 12$. Se calcula que, para toda i, $E[U_i] = 0$ y

$Var[U_i] = 1/12$. Si $Z := \sum_{i=1}^{12} U_1$, entonces

$$E(Z) = \sum_{i=1}^{12} E(U_i) = 0$$

y

$$Var(Z) = \sum_{i=1}^{12} Var(U_i) = 1;$$

ahora bien, según el TLC, la variable aleatoira

$$\frac{Z - E(Z)}{\sqrt{Var(Z)}} = Z$$

tiende en probabilidad a $N(0,1)$

```
> n = 10**4
> Z1 = rnorm(n, 0, 1)
> U =runif(12*10**4)
> U = matrix(data = U, nrow=12)
> Z2 = apply(U, 2, sum)
> U1 = runif(n)
> U2 = runif(n)
> Z3 = sqrt(-2*log(U1))*cos(2*pi*U2)
> par(mfrow=c(1,3))
> hist(Z1, freq = FALSE, main = "Normal con R")
> hist(Z2, freq = FALSE, main = "Normal con TLC")
> hist(Z3, freq = FALSE, main = "Normal con BM")
```

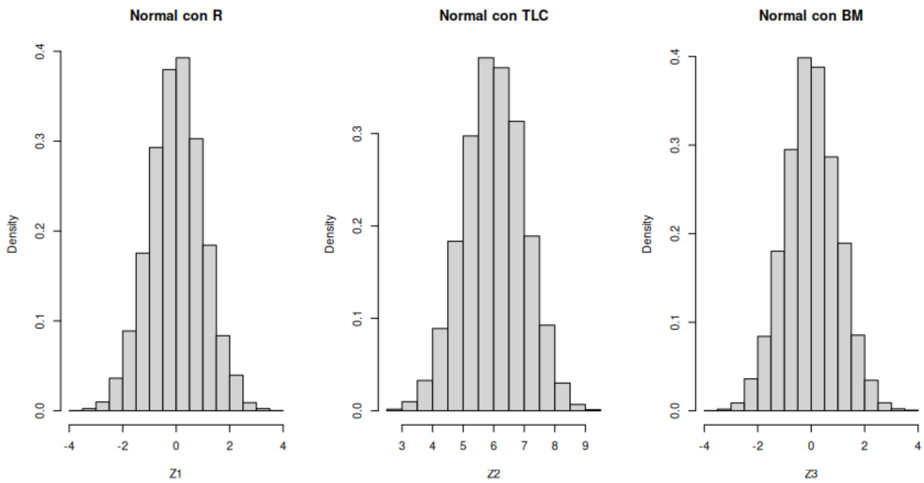

Figure 2.9: Resultados de las simulaciones en R. Por lo general se prefire usar BM a TLC, pues el primero da lugar a v.a. que en efecto tienen distribuciones $N(0,1)$, mientras que el segundo método siempre da aproximaciones.

2.1.4 2.2.2 Discrete distributions

A partir de una variable uniforme $U \sim U(0,1)$, también es posible generar distribuciones discretas. Sea P_θ una distribución cuyo soporte está contenido en $\mathbb{Z}$, y sea $X \sim P_\theta$.

Calcular a F_θ (la función de acumulación de la distribución) se reduce a calcular (**y, en la práctica, guardar en la memoria**) los siguientes números;

$$p_k := P(X \leq k) \tag{2.10}$$

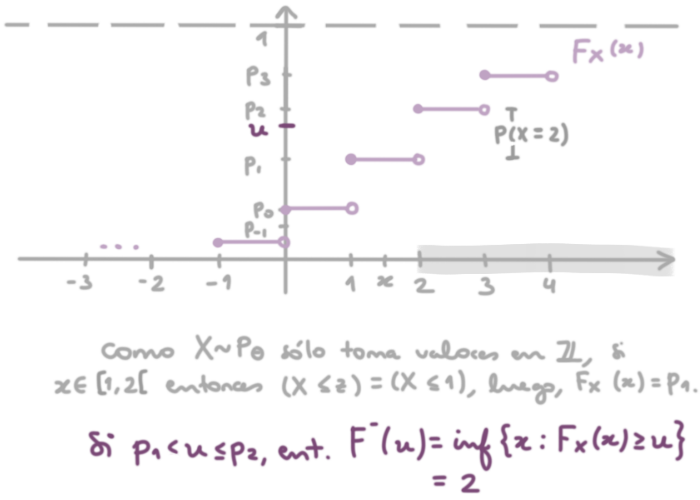

Figure 2.10: Recuerde cómo se define la inversa generalizada de la función de acumulación F en (2.1).

La inversa generalizada F_θ^- de la función de acumulación dada es

$$\begin{array}{rcccl} F_\theta^- : &]0,1[& \longrightarrow & \{p_k\}_{k\in\mathbb{Z}} \\ & u & \longmapsto & k \;\; si \; p_{k-1} < u \leq p_k. \end{array}$$

Según el ejercicio 1, $X := F_\theta^- \circ U \sim P_\theta$.

checa cómo tienes los environments de ejemplo en tu tesis.

Ejemplo 2.1.4

Este es el resultado de usar las ideas de esta sección para, a partir de números aleatorios de una distribución $U(0,1)$*, obtener una distribución binomial.*

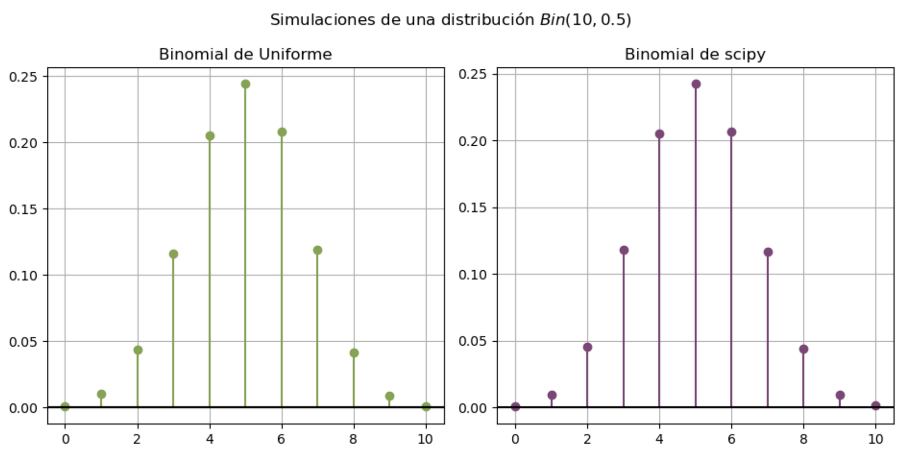

Figure 2.11: Figura generada usando el código de `generar_binomial.py`

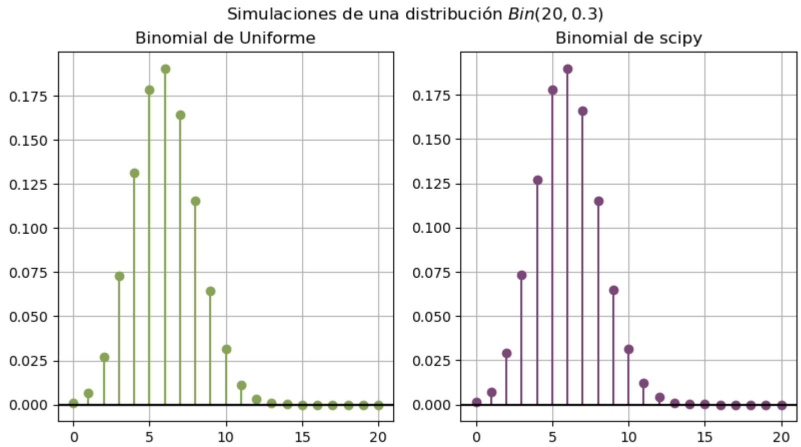

Figure 2.12: Figura generada usando el código de `generar_binomial.py`

Vamos a hacer lo mismo pero ahora con `R`.

```
> Nsim = 10^4; n = 10; p = 0.3
> soporte = seq(0, n, 1)
> vector_pk = pbinom(soporte, n, p)
> X = rep(0, Nsim)
> for (i in 1:Nsim){
+ u = runif(1)
+ X[i] = sum(vector$\_$pk<u)
+ }
```

Note que `vector_pk` es el vector en el que se almacenan los valores p_k para $0 \leq k \leq n$ (que, en el caso de la distribución binomial, son todos los valores de k que basta considerar), y que `vector_pk` es un vector de booleanos, cuya $i-$ésima entrada es `TRUE` si la i-ésima entrada de `vector_pk` es menor a u, y es `FALSE` en caso contrario. Nota entonces que `sum(vector_pk<u)` $= F^{-}(u)$.

Chapter 3

Notas sobre R

Aquí escribo algunas notas sobre las funciones y objetos de `R` que creo son útiles tener en cuenta.

3.1 Sobre los histogramas

Si `X` es un vector de números, puedes usar la función `hist` para crear un histograma a partir de los datos numéricos almacenados en `X`. Pon por ejemplo

$$\mathtt{X} = c(1, 1, 1, 1, 2, 3).$$

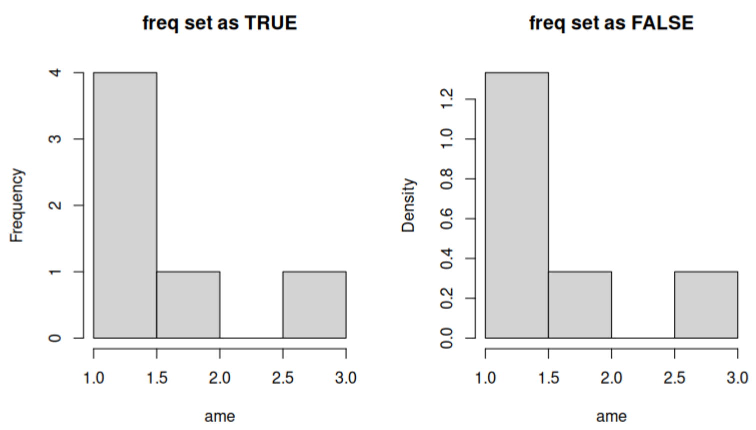

Figure 3.1: Se grafican los histogramas que resultaron de ejecutar, respectivamente, las instrucciones `hist(ame, main = 'freq set as TRUE')` y `hist(ame, main = 'freq set as FALSE', freq = FALSE)`.

Como se explica en [His], poniendo `freq = FALSE` no da lugar a un histograma de los *porcentajes* de los valores almacenados en los datos graficados, sino que normaliza el histograma (obviamente respetando proporciones) de tal forma que el área total de los rectángulos graficados sea 1.

Usar `freq = FALSE` parece ser la opción preferida por los autores para graficar histogramas. No me parece mal, pues se preservan las proporciones de las ocurrencias, pero personalmente prefiero graficar las probabilidades de ocurrencia de los elementos de `X`. Para hacer esto, tienes que hacer guardar el histograma en una variable (pidiendo no graficarlo) y modificar la propiedad `density` de este.

```
> h = hist(x, plot = FALSE)
> h\$density = h\$counts/sum(h\$counts)
> plot(h,freq=FALSE)
```

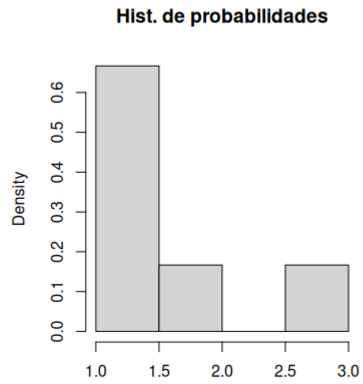

Figure 3.2: Histograma que resulta de hacer el cambio sugerido en el código de arriba.

Chapter 4

Notación

$\mathbb{N}^* = \{1, 2, \ldots\}$

Bibliography

[His] *Use hist() function in R to get percentages as opposed to raw frequencies.* https://stackoverflow.com/questions/7324683/use-hist-function-in-r-to-get-percentages-as-opposed-to-raw-frequencies. Visto el 23 de mayo 2023.