

Simulación Estadística

Rubén Fernández Casal (ruben.fcасal@udc.es), Ricardo Cao (rcao@udc.es)

2020-01-28

Índice general

Prólogo	5
1 Introducción a la simulación	7
1.1 Introducción	7
1.2 Generación de números (pseudo)aleatorios	8
1.3 Números aleatorios puros	9
1.4 Números pseudoaleatorios	10
2 Números aleatorios en R	13
2.1 Opciones	13
2.2 Paquetes de R	14
2.3 Ejercicios	14
2.4 Tiempo de CPU	18
3 Generación de números pseudoaleatorios	23
3.1 Generadores congruenciales	23
3.2 Análisis de la calidad de un generador	27
3.3 Ejercicios de fin de práctica	37
4 Análisis de resultados de simulación	39
4.1 Convergencia	39
4.2 Estimación de la precisión	42
4.3 Teorema central del límite	42
4.4 Determinación del número de generaciones	43
4.5 El problema de la dependencia	44
4.6 Observaciones	51
5 Simulación de variables continuas	53
5.1 Método de inversión	53
5.2 Método de aceptación rechazo	57
5.3 Modificaciones del método de AR	65
5.4 Método de composición	67
5.5 Métodos específicos para generación de distribuciones notables	68
6 Simulación de variables discretas	69
6.1 Método de la transformación cuantil	69
6.2 Método de la tabla guía	74
6.3 Método de Alias	76
6.4 Simulación de una variable discreta con dominio infinito	78
6.5 Cálculo directo de la función cuantil	79
6.6 Otros métodos	80
6.7 Métodos específicos para generación de distribuciones notables	83

7 Simulación de Distribuciones Multidimensionales	85
7.1 Introducción	85
7.2 El método de aceptación/rechazo	86
7.3 Factorización de la matriz de covarianzas	88
7.4 Método de las distribuciones condicionadas	91
7.5 Simulación condicional e incondicional	93
7.6 Simulación basada en cónulas	99
7.7 Otros métodos para variables continuas	106
7.8 Simulación de distribuciones multidimensionales discretas	106
7.9 Ejercicios propuestos	111
8 Aplicaciones de la simulación en Inferencia Estadística	113
8.1 Aplicaciones de la simulación	113
8.2 Distribución en el muestreo	114
8.3 Intervalos de confianza	117
8.4 Contrastos de hipótesis	125
8.5 Comparación de estimadores	134
8.6 Remuestreo Bootstrap	137
9 Integración y Optimización Montecarlo	147
9.1 Integración Monte Carlo (clásica)	147
9.2 Muestreo por importancia	152
9.3 Optimización Monte Carlo	157
9.4 Temple simulado	161
9.5 Algoritmos genéticos	164
10 Técnicas de reducción de la varianza	167
10.1 Reducción de la varianza	167
10.2 Variables antitéticas	167
10.3 Estratificación	172
10.4 Variables de control	175
10.5 Números aleatorios comunes	176
10.6 Ejercicios fin de práctica	177
Referencias	179
Bibliografía básica	179
Bibliografía complementaria	179
A Enlaces	181
B Integración numérica	185
B.1 Introducción	185
B.2 Integración numérica unidimensional	185
B.3 Integración numérica bidimensional	189

Prólogo

Este libro contiene los apuntes de la asignatura de Simulación Estadística del Máster en Técnicas Estadísticas.

Este libro ha sido escrito en R-Markdown empleando el paquete `bookdown` y está disponible en el repositorio Github: rubenfcasal/simbook. Se puede acceder a la versión en línea a través del siguiente enlace:

<https://rubenfcasal.github.io/simbook/index.html>.

donde puede descargarse en formato pdf.

Para instalar los paquetes necesarios para poder ejecutar los ejemplos mostrados en el libro se puede emplear el siguiente comando:

```
pkgs <- c('boot', 'MASS', 'DEoptim', 'nortest', 'geoR', 'copula', 'sm',
         'tseries', 'forecast', 'plot3D', 'rgl')
install.packages(setdiff(pkgs, installed.packages()[, "Package"]),
                 dependencies = TRUE)

# Si aparecen errores debidos a incompatibilidades entre las versiones de los paquetes,
# probar a ejecutar en lugar de lo anterior:
# install.packages(pkgs, dependencies = TRUE) # Instala todos...
```

Para generar el libro (compilar) serán necesarios paquetes adicionales, para lo que se recomendaría consultar el libro de “Escritura de libros con bookdown” en castellano.



Este obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional (esperamos poder liberarlo bajo una licencia menos restrictiva más adelante...).

Capítulo 1

Introducción a la simulación

1.1 Introducción

Problemas de la experimentación directa sobre la realidad:

- Coste elevado.
- En ocasiones las pruebas son destructivas.
- Lentitud.
- Puede no ser ética.
- experimentación sobre seres humanos.
- Puede resultar imposible.
- Acontecimientos futuros, ...
- ...

Puede ser preferible trabajar con un modelo del sistema real. La realidad puede ser muy compleja por lo que es habitual emplear un modelo para tratar de explicarla:

- Modelos deterministas.
- Modelos estocásticos (con componente aleatoria).
 - Cuando no se dispone de la suficiente información sobre las variables que influyen en el fenómeno en estudio.
 - Se tienen en cuenta esta incertidumbre.

La inferencia estadística proporciona herramientas para estimar los parámetros y contrastar la validez de un modelo estocástico a partir de los datos observados.

La idea es emplear el modelo para resolver el problema de interés. Cuando la solución no se puede obtener de modo analítico (teórico) se puede recurrir a la simulación.

Simulación: realizar experimentos con un modelo con el objetivo de recopilar información bajo determinadas condiciones. Nos centraremos en el caso de la **simulación estocástica**:

- Las conclusiones se obtienen habitualmente generando repetidamente simulaciones del modelo aleatorio.

1.1.1 Ventajas de la simulación (Shannon, 1975):

- Cuando la resolución analítica no puede llevarse a cabo.
- Cuando existen medios de resolver analíticamente el problema pero dicha resolución es complicada y costosa (o solo proporciona una solución aproximada).
- Si se desea experimentar antes de que exista el sistema (pruebas para la construcción de un sistema).
- Cuando es imposible experimentar sobre el sistema real por ser dicha experimentación destructiva.
- En ocasiones en las que la experimentación sobre el sistema es posible pero no ética.
- En sistemas que evolucionan muy lentamente en el tiempo.

1.1.2 Inconvenientes de la simulación:

- La construcción de un buen modelo puede ser una tarea muy costosa (compleja, laboriosa y requerir mucho tiempo; e.g. modelos climáticos).
- Frecuentemente el modelo omite variables o relaciones importantes entre ellas (los resultados pueden no ser válidos para el sistema real).
- Resulta difícil conocer la precisión de la simulación (especialmente en lo relativo a la precisión del modelo formulado).
- Problemas de extrapolación de las conclusiones.
- Tiempo de computación.

1.2 Generación de números (pseudo)aleatorios

Aplicaciones:

- Estadística:
- Muestreo, remuestreo, ...
- Aproximación de distribuciones (de estadísticos, estimadores, ...)
- Realización de contrastes, intervalos de confianza, ...
- Comparación de estimadores, contrastes, ...
- Validación teoría (distribución asintótica,...)
- Inferencia Bayesiana
- Optimización: Algoritmos genéticos, ...
- Computación: Diseño, verificación y validación de algoritmos,...
- Criptografía: Protocolos de comunicación segura, ...
- Física: Simulación de fenómenos naturales, ...
- Análisis numérico: Evaluación de expresiones, ...
- ...

En el Capítulo XX nos centraremos en algunas de las aplicaciones de utilidad en Estadística.

1.3 Números aleatorios puros

Una sucesión de números aleatorios puros (true random), se caracteriza por que no existe ninguna regla o plan que nos permita conocer sus valores.

Se almacena(ban) en *tablas de dígitos aleatorios (true random)* y normalmente son obtenidos por procesos físicos (loterías, ruletas, ruidos,...)

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

Se emplean para seleccionar números aleatorios en un rango de 1 a m :

- Se selecciona al azar un pto de inicio en la tabla y una dirección.
- Se agrupan los dígitos de forma que “cubran” el valor de m .
- Se seleccionan los valores menores o iguales que m (se descarta el resto).

Algunos enlaces:

- A Million Random Digits with 100,000 Normal Deviates . RAND Corporation. 1955.
- Generadores de números aleatorios “online”:
 - <http://www.random.org/integers> (ver paquete **random** en R).
 - <http://www.fourmilab.ch/hotbits>
- Generadores mediante hardware:
 - <http://software.intel.com>.
 - <http://spectrum.ieee.org>

Sus principales aplicaciones hoy en día son en criptografía (impredictibilidad).

1.3.1 Inconvenientes:

- Es necesario/recomendable conocer su distribución.
- Los datos generados deberían ser i.i.d.
- Reproductividad.
- Pueden requerir del almacenamiento en tablas.

1.3.2 Alternativas:

- **números pseudo-aleatorios**: simulan realizaciones de una variable aleatoria (uniforme).
- números quasi-aleatorios: secuencias determinísticas con una distribución más uniforme en el rango considerado (se podría pensar que son una única generación de una variable aleatoria).

1.4 Números pseudoaleatorios

1.4.1 Generación de números pseudoaleatorios mediante software

La mayoría de los métodos de simulación se basan en la posibilidad de generar números pseudoaleatorios con distribución $\mathcal{U}(0, 1)$. Se obtienen mediante un algoritmo recursivo denominado **generador**:

$$x_i = f(x_{i-1}, x_{i-2}, \dots, x_{i-k})$$

- k orden del generador.
- $(x_0, x_1, \dots, x_{k-1})$ **semilla** (estado inicial).

El **periodo** o *longitud del ciclo* es la longitud de la secuencia antes de que vuelva a repetirse. Lo denotaremos por p .

Los números de la sucesión serán predecibles, conociendo el algoritmo y la semilla.

- Sin embargo, si no se conociesen, **no se debería poder distinguir** una serie de números pseudoaleatorios **de una sucesión de números verdaderamente aleatoria** (utilizando recursos computacionales razonables).
- En caso contrario esta predecibilidad puede dar lugar a serios problemas (e.g. <http://eprint.iacr.org/2007/419>).

Como regla general, por lo menos mientras se está desarrollando un programa, interesa **fijar la semilla de aleatorización**.

- Permite la reproducibilidad de los resultados.
- Facilita la depuración del código.

Todo generador de números pseudoaleatorios mínimamente aceptable debe comportarse como si se tratase de una muestra genuina de datos independientes de una $\mathcal{U}(0, 1)$. Otras propiedades de interés serían:

- Reproducibilidad a partir de la semilla.
- Periodo suficientemente largo.
- Eficiencia (rapidez y requerimientos de memoria).
- Portabilidad.

- Generación de sub-secuencias (computación en paralelo).
- Parsimonia, ...

Gran cantidad de algoritmos:

- Cuadrados medios, Lehmer,...
- Congruenciales
- Registros desfasados
- Combinaciones
- ...

Código fuente disponible en múltiples librerías:

- GNU Scientific Library (GSL): <http://www.gnu.org/software/gsl/manual>
- StatLib: <http://lib.stat.cmu.edu>
- Numerical recipes: <http://www.nrbook.com/nr3>
- <http://random.mat.sbg.ac.at/software>
- KISS (Keep It Simple Stupid / Small and Simple): <http://www.fortran.com/kiss.f90>
- UNU.RAN (paquete **Runuran**): <http://statmath.wu.ac.at/unuran>
- ...

Nos centraremos en los generadores congruenciales, descritos en la Sección 3.1.

Capítulo 2

Números aleatorios en R

La generación de números pseudoaleatorios en R es una de las mejores disponibles en paquetes estadísticos. Entre las herramientas en el paquete base de R estarían:

- `set.seed(entero)`: permite establecer la semilla (y el generador).
- `RNGkind`: selecciona el generador.
- `rdistribución(n,...)`: genera valores aleatorios de la correspondiente distribución. Por ejemplo: `runif(n, min = 0, max = 1)`, generaría `n` valores de una uniforme.
- `sample`: genera muestras aleatorias (v.a. discretas) y permutaciones.

La semilla se almacena (en `globalenv`) en `.Random.seed`; es un vector de enteros cuya dimensión depende del tipo de generador:

- No debe ser modificado manualmente; se guarda con el entorno de trabajo.
- Si no se especifica con `set.seed` (o no existe) se genera a partir del reloj del sistema.

Nota: . Puede ser recomendable (para depurar) almacenarla antes de generar simulaciones, e.g. `semilla <- .Random.seed`.

2.1 Opciones

La función `RNGkind(kind = NULL, normal.kind = NULL)` permite seleccionar el tipo de generador (en negrita los valores por defecto):

- `kind` especifica el generador aleatorio:
 - “Wichmann-Hill”: Ciclo 6.9536×10^{12}
 - “Marsaglia-Multicarry”: Ciclo mayor de 2^{60}
 - “Super-Duper”: Ciclo aprox. 4.6×10^{18} (S-PLUS)
 - “**Mersenne-Twister**”: Ciclo $2^{19937} - 1$ y equidistribution en 623 dim.
 - “Knuth-TAOCP-2002”: Ciclo aprox. 2^{129} .
 - “Knuth-TAOCP”
 - “user-supplied”

- `normal.kind` selecciona el método de generación de normales (se tratará más adelante). “Kinderman-Ramage”, “Buggy Kinderman-Ramage”, “Ahrens-Dieter”, “Box-Muller”, “Inversion” , o “user-supplied”

2.2 Paquetes de R

Otros paquetes de R que pueden ser de interés:

- `setRNG` contiene herramientas que facilitan operar con la semilla (dentro de funciones,...).
- `random` generación de numeros “true random”.
- `randtoolbox` implementa generadores más recientes (`rngWELL`) y generación de secuencias cuasi-aleatorias.
- `RDieHarder` implementa diversos contrastes para el análisis de la calidad de un generador y varios generadores.
- `Runuran` interfaz para la librería UNU.RAN para la generación (automática) de variables aleatorias no uniformes.
- `rsprng` y `rstream` implementan la generación de múltiples secuencias (e.g. para programación paralela).
- `gls`, `rngwell19937`, `randaes`, `SuppDists`, `lhs`, `mc2d`, `f0ptions`, ...

2.3 Ejercicios

Ejercicio 2.1.

Sea (X, Y) es un vector aleatorio con distribución uniforme en el cuadrado $[-1, 1] \times [-1, 1]$ de área 4.

- Aproximar mediante simulación $P(X + Y \leq 0)$ y compararla con la probabilidad teórica (obtenida aplicando la regla de Laplace $\frac{\text{área favorable}}{\text{área posible}}$).

Probabilidad teórica $1/2$ (area favorable / área total)

```
set.seed(1)
n <- 10000
x <- runif(n, -1, 1)
y <- runif(n, -1, 1)
indice <- (x+y < 0)
# Aproximación por simulación
sum(indice)/n
```

```
## [1] 0.4996
```

```
mean(indice) # Alternativa
```

```
## [1] 0.4996
```

Nota: . R maneja internamente los valores lógicos como 1 (TRUE) y 0 (FALSE).

- Aproximar el valor de π mediante simulación a partir de $P(X^2 + Y^2 \leq 1)$.

```
set.seed(1)
n <- 10000
x <- runif(n, -1, 1)
y <- runif(n, -1, 1)
```

```

indice <- (x^2+y^2 < 1)
mean(indice)

## [1] 0.7806
pi/4

## [1] 0.7853982
pi_aprox <- 4*mean(indice)
pi_aprox

## [1] 3.1224

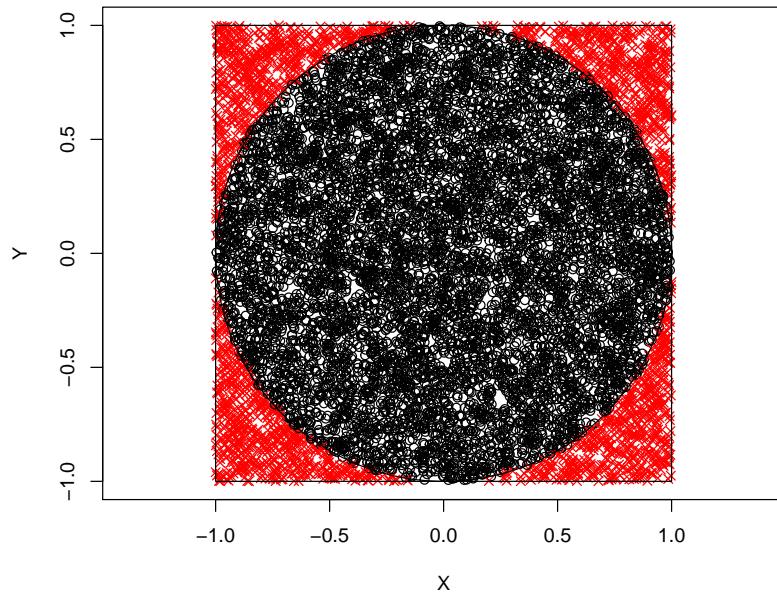
```

Gráfico

```

# Colores y símbolos dependiendo de si el índice correspondiente es verdadero:
color <- ifelse(indice, "black", "red")
simbolo <- ifelse(indice, 1, 4)
plot(x, y, pch = simbolo, col = color,
      xlim = c(-1, 1), ylim = c(-1, 1), xlab="X", ylab="Y", asp = 1)
# asp = 1 para dibujar circulo
symbols(0, 0, circles = 1, inches = FALSE, add = TRUE)
symbols(0, 0, squares = 2, inches = FALSE, add = TRUE)

```



Ejercicio 2.2.

Consideramos el experimento de Bernoulli consistente en el lanzamiento de una moneda.

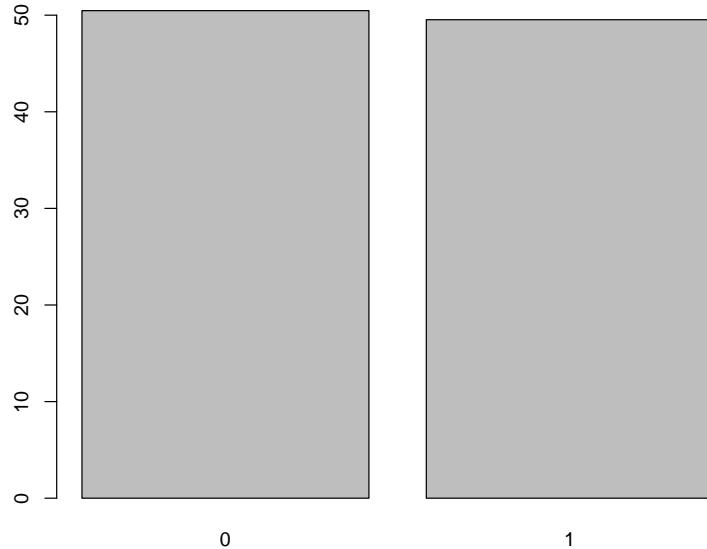
- Empleando la función `sample`, obtener 1000 simulaciones del lanzamiento de una moneda (`0 = cruz`, `1 = cara`), suponiendo que no está trucada. Aproximar la probabilidad de cara a partir de las simulaciones.

```

set.seed(1)
nsim <- 10000
x <- sample(c(cara = 1, cruz = 0), nsim, replace = TRUE, prob = c(0.5,0.5))
mean(x)

## [1] 0.4953
barplot(100*table(x)/nsim) # Representar porcentajes

```



- b) En R pueden generarse valores de la distribución de Bernoulli mediante la función `rbinom(nsim, size=1, prob)`. Generar un gráfico de líneas considerando en el eje X el número de lanzamientos (de 1 a 10000) y en el eje Y la frecuencia relativa del suceso cara (puede ser recomendable emplear la función `cumsum`).

```

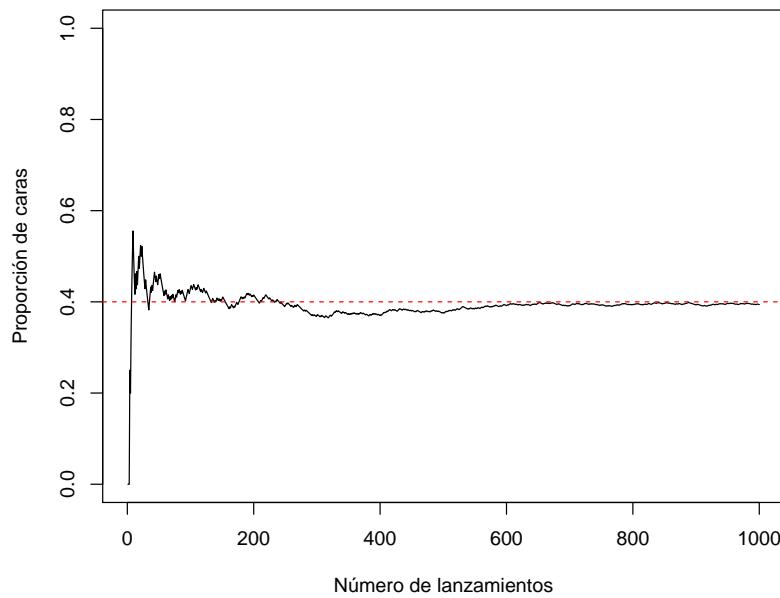
set.seed(1)
nsim <- 1000
p <- 0.4
x <- rbinom(nsim, size = 1, prob = p) # Simulamos una Bernouilli
n <- 1:nsim
# Alternativa programación: x <- runif(nsim) < p
mean(x)

```

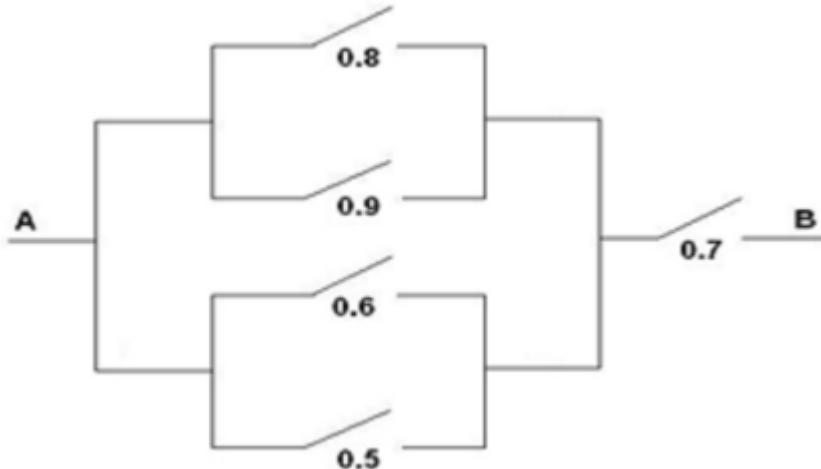
```

## [1] 0.394
plot(n, cumsum(x)/n, type="l", ylab="Proporción de caras",
      xlab="Número de lanzamientos", ylim=c(0,1))
abline(h=p, lty=2, col="red")

```

**Ejercicio 2.3.**

Simular el paso de corriente a través del siguiente circuito, donde figuran las probabilidades de que pase corriente por cada uno de los interruptores:



Considerar que cada interruptor es una v.a. de Bernoulli independiente para simular 1000 valores de cada una de ellas.

Nota: R maneja internamente los valores lógicos como 1 (TRUE) y 0 (FALSE). Recíprocamente, cualquier nº puede ser tratado como lógico (al estilo de C). El entero 0 es equivalente a FALSE y cualquier entero distinto de 0 a TRUE.

```
set.seed(1)
nsim <- 10000
x1 <- rbinom(nsim, size=1, prob=0.9)
x2 <- rbinom(nsim, size=1, prob=0.8)
z1 <- x1 | x2 # Operador lógico "O"
```

```
x3 <- rbinom(nsim, size=1, prob=0.6)
x4 <- rbinom(nsim, size=1, prob=0.5)
z2 <- x3 | x4
z3 <- z1 | z2
x5 <- rbinom(nsim, size=1, prob=0.7)
fin <- z3 & x5 # Operador lógico "Y"
mean(fin)

## [1] 0.6918
```

Ejercicio 2.4.

En 1651, el Caballero de Méré le planteó a Pascal una pregunta relacionada con las apuestas y los juegos de azar: ¿es ventajoso apostar a que en cuatro lanzamientos de un dado se obtiene al menos un seis? Este problema generó una fructífera correspondencia entre Pascal y Fermat que se considera, simbólicamente, como el nacimiento del Cálculo de Probabilidades.

- a) Escribir una función que simule el lanzamiento de n dados. El parámetro de entrada es el número de lanzamientos n , que toma el valor 4 por defecto, y la salida debe ser TRUE si se obtiene al menos un 6 y FALSE en caso contrario.

```
deMere <- function(n = 4){
  lanz <- sample(1:6, replace=TRUE, size=n)
  return(6 %in% lanz)
}

n <- 4
lanz <- sample(1:6, replace=TRUE, size=n)
lanz

## [1] 4 5 3 1
6 %in% lanz

## [1] FALSE
```

- b) Utilizar la función anterior para simular $nsim = 10000$ jugadas de este juego y calcular la proporción de veces que se gana la apuesta (obtener al menos un 6 en n lanzamientos), usando $n = 4$. Comparar el resultado con la probabilidad teórica $1 - (5/6)^n$.

```
set.seed(1)
n <- 4
nsim <- 10000
mean(replicate(nsim, deMere(n)))

## [1] 0.5195
1-(5/6)^n

## [1] 0.5177469
```

2.4 Tiempo de CPU

La velocidad del generador suele ser una característica importante. Para evaluar el rendimiento están disponibles en R distintas herramientas:

- `proc.time()`: permite obtener tiempo de computación real y de CPU.

```
tini <- proc.time()
# Código a evaluar
tiempo <- proc.time() - tini
```

- `system.time(expresión)`: muestra el tiempo de computación (real y de CPU) de expresión.
- `Rprof(fichero)`: permite evaluar el rendimiento muestreando la pila en intervalos para determinar en que funciones se emplea el tiempo de computación (de utilidad para optimizar la velocidad). `Rprof(NULL)`: desactiva el muestreo. `summaryRprof(fichero)`: muestra los resultados.

2.4.1 Utilidades tiempo de CPU

Por ejemplo, podríamos emplear las siguientes funciones para ir midiendo los tiempos de CPU durante una simulación:

```
CPUtimeini <- function() {
  .tiempo.ini <- proc.time()
  .tiempo.last <- .tiempo.ini
}

CPUtimeprint <- function() {
  tmp <- proc.time()
  cat("\nTiempo última operación:\n")
  print(tmp-.tiempo.last)
  cat("Tiempo total operación:\n")
  print(tmp-.tiempo.ini)
  .tiempo.last <- tmp
}
```

Llamando a `CPUtimeini()` donde se quiere empezar a contar, y a `CPUtimeprint()` para imprimir el tiempo total y el tiempo desde la última llamada a una de estas funciones. Ejemplo:

```
funtest <- function(n) median(runif(n)) # Función de prueba...
CPUtimeini()
funtest(1000000)

## [1] 0.5004536
CPUtimeprint()

##
## Tiempo última operación:
##   user   system elapsed
## 0.08    0.00    0.08
## Tiempo total operación:
##   user   system elapsed
## 0.08    0.00    0.08

funtest(1000)

## [1] 0.4954613
CPUtimeprint()

##
## Tiempo última operación:
##   user   system elapsed
## 0       0       0
```

```
## Tiempo total operación:
##      user    system elapsed
##      0.08     0.00     0.08
```

2.4.2 Paquetes de R

Por ejemplo, se puede emplear el paquete `npsp` (fichero `cpu.time.R`):

- Call `cpu.time(restart = TRUE)` where you want to start counting.
- Call `cpu.time()` to print/get total and/or partial (since the last call to this function) real and CPU times.

```
# CPU time utilities
# ----

#' @rdname npsp-internals
#' @keywords internal
#' @export
.cpu.time.ini <- function() {
  time.ini <- structure(rep(0, 5), .Names = c("user.self", "sys.self", "elapsed",
    "user.child", "sys.child"), class = "proc_time")# proc.time()
  time.last <- time.ini
  function(..., reset = FALSE, total = TRUE, last = TRUE, flush = FALSE) {
    res <- list(time = proc.time())
    if (reset) {
      time.ini <-> res$time
      time.last <-> time.ini
      res$last <- res$total <- 0
      if (total | last) cat("CPU time has been initialized.\n")
    } else {
      res$last <- res$time - time.last
      res$total <- res$time - time.ini
      if (last) {
        cat("Time of last operation:", ..., "\n")
        print(res$last)
      }
      if (total) {
        cat("Total time:\n")
        print(res$total)
      }
      if (flush) flush.console()
      time.last <-> res$time
    }
    return(invisible(res))
  }
}

#' Total and partial CPU time used
#'
#' Returns and (optionally) prints the total and/or partial
#' (since the last call to this function)
#' real and CPU times.
#' @param ... objects (describing the last operation) to be printed
#' (using \code{\link{cat}}),
```

```

#' if \code{last == TRUE}.
#' @param reset logical; if \code{TRUE}, time counters are initialized.
#' @param total logical; if \code{TRUE}, the total time used is printed.
#' @param last logical; if \code{TRUE}, the partial time used is printed.
#' @param flush logical; if \code{TRUE}, \code{\link{flush.console}} is called.
#' @return Invisibly returns a list with the following 3 components
#' (objects of class \code{"proc_time"}):
#' \item{time}{user, system, and total elapsed times for the currently running R process
#' (result of a call to \code{\link{proc.time}}). }
#' \item{last, total}{differences between the corresponding \code{\link{proc.time}} calls.}
#' @seealso
#' \code{\link{proc.time}}, \code{\link{system.time}}, \code{\link{flush.console}}.
#' @export
cpu.time <- .cpu.time.ini()

```

Ejemplo:

```

cpu.time(reset = TRUE)

## CPU time has been initialized.

res <- funtest(1000000)
cpu.time('`nSample median of', 1000000, 'values =', res, total = FALSE)

## Time of last operation:
## Sample median of 1e+06 values = 0.4993793
##     user   system elapsed
##     0.05    0.02    0.06

res <- funtest(1000)
cpu.time('`nSample median of', 1000, 'values =', res)

## Time of last operation:
## Sample median of 1000 values = 0.4936829
##     user   system elapsed
##     0       0       0
## Total time:
##     user   system elapsed
##     0.05    0.02    0.06

```

Otro paquete que puede ser de utilidad es **microbenchmark** (si se quieren estudiar con más detalle los tiempos de computación; aunque en este libro no será el caso...). Hay que tener en cuenta que, por construcción, aunque se realicen en la mismas condiciones (en el mismo equipo), los tiempos de CPU en R pueden variar “ligeramente” entre ejecuciones.

Capítulo 3

Generación de números pseudoaleatorios

3.1 Generadores congruenciales

3.1.1 Generador congruencial lineal (simple)

Partiendo de una semilla inicial x_0 :

$$\begin{aligned}x_i &= (ax_{i-1} + c) \bmod m \\u_i &= \frac{x_i}{m} \\i &= 1, 2, \dots\end{aligned}$$

donde a es el **multiplicador**, c el **incremento** y m el **módulo** del generador (parámetros enteros fijados de antemano).

- Si $c = 0$ se denomina congruencial **multiplicativo**
- Si $c \neq 0$ se denomina congruencial **mixto**

Implementado en el fichero *RANDC.R* (tratando de imitar el funcionamiento en R, aunque de forma no muy eficiente...):

```
# Generador congruencial de números pseudoaleatorios
# -----
#
# -----
# initRANDC(semilla, a, c, m)
#   Selecciona el generador congruencial
#   Por defecto RANDU de IBM con semilla del reloj
#   OJO: No se hace ninguna verificación de los parámetros
initRANDC <- function(semilla=as.numeric(Sys.time()), a=2^16+3, c=0, m=2^31) {
  .semilla <- as.double(semilla) %% m #Cálculos en doble precisión
  .a <- a
  .c <- c
  .m <- m
  return(invisible(list(semilla=.semilla,a=.a,c=.c,m=.m))) #print(initRANDC())
}

# -----
```

```

# RANDC()
#   Genera un valor pseudoaleatorio con el generador congruencial
#   Actualiza la semilla (si no existe llama a initRANDC)
RANDC <- function() {
  if (!exists(".semilla", envir=globalenv())) initRANDC()
  .semilla <- (.a * .semilla + .c) %% .m
  return(.semilla/.m)
}

# -----
# RANDCN(n)
#   Genera un vector de valores pseudoaleatorios con el generador congruencial
#   (por defecto de dimensión 1000)
#   Actualiza la semilla (si no existe llama a initRANDC)
RANDCN <- function(n=1000) {
  x <- numeric(n)
  for(i in 1:n) x[i] <- RANDC()
  return(x)
  # return(replicate(n,RANDC())) # Alternativa más rápida
}

initRANDC(543210)      # Fijar semilla 543210 para reproductibilidad

```

Nota: Para evitar problemas computacionales, se recomienda emplear un algoritmo como el descrito en L'Ecuyer (1988).

Ejemplos:

- $c = 0$, $a = 2^{16} + 3 = 65539$ y $m = 2^{31}$, generador RANDU de IBM (**no recomendable**).
- $c = 0$, $a = 7^5 = 16807$ y $m = 2^{31} - 1$, Park y Miller (1988) “minimal standar”, empleado por las librerías IMSL y NAG.

Los parámetros y la semilla determinan los valores generados:

$$x_i = \left(a^i x_0 + c \frac{a^i - 1}{a - 1} \right) \bmod m$$

A pesar de su simplicidad, una adecuada elección de los parámetros permite obtener de manera eficiente secuencias de números “aparentemente” i.i.d. $\mathcal{U}(0, 1)$.

3.1.2 Periodo

Teorema 3.1. (Hull y Dobell, 1962)

Un generador congruencial tiene período máximo ($p = m$) si y solo si:

1. c y m son primos relativos (i.e. $m.c.d.(c, m) = 1$).
2. $a - 1$ es múltiplo de todos los factores primos de m (i.e. $a \equiv 1 \pmod{q}$, para todo q factor primo de m).
3. Si m es múltiplo de 4, entonces $a - 1$ también lo ha de ser (i.e. $m \equiv 0 \pmod{4} \Rightarrow a \equiv 1 \pmod{4}$).

Algunas consecuencias:

- Si m primo, $p = m \Leftrightarrow a = 1$

- Un generador multiplicativo no cumple la condición 1.

Teorema 3.2. *Un generador multiplicativo tiene período máximo ($p = m - 1$) si:*

1. *m es primo.*
2. *a es una raíz primitiva de m (i.e. el menor entero q tal que $a^q \equiv 1 \pmod{m}$ es $q = m - 1$).*

Además de preocuparse de la longitud del ciclo, las secuencias generadas deben aparecer muestras i.i.d. $\mathcal{U}(0, 1)$. Por ejemplo, los valores generados pueden mostrar una estructura reticular.

- Marsaglia (1968): k -uplas de generadores multiplicativos contenidas en a lo sumo $(k!m)^{1/k}$ hiperplanos paralelos.
- Generador RANDU de IBM (70's):

```
system.time(u <- RANDCN(9999)) # Generar

##      user    system elapsed
##     0.03     0.01    0.05

xyz <- matrix(u, ncol = 3, byrow = TRUE)

library(plot3D)
points3D(xyz[,1], xyz[,2], xyz[,3], colvar = NULL, phi = 60, theta = -50, pch = 21, cex = 0.2)
```

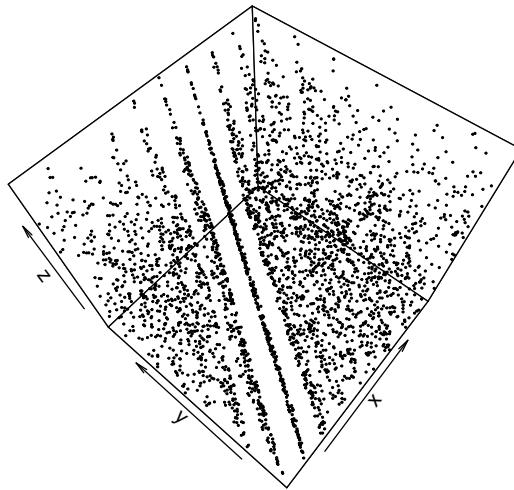


Figura 3.1: Gráfico de dispersión de tripletas del generador RANDU de IBM (contenidas en 15 planos)

NOTA: Alternativamente se podría utilizar la función `plot3d` del paquete `rgl`, y pulsando con el ratón se podría rotar la figura para ver los hiperplanos:

```
library(rgl)
plot3d(xyz)
```

Se han propuesto diversas pruebas (ver sección siguiente) para determinar si un generador tiene problemas de este tipo y se han realizado numerosos estudios para determinadas familias (e.g. Park y Miller, 1988, estudiaron que parámetros son adecuados para $m = 2^{31} - 1$).

- En cualquier caso, se recomienda considerar un “periodo de seguridad” $\approx \sqrt{p}$ para evitar este tipo de problemas.
- Aunque estos generadores tienen limitaciones en su capacidad para producir secuencias muy largas de números i.i.d. $\mathcal{U}(0, 1)$, es un elemento básico en generadores más avanzados.

3.1.3 Otros generadores

Generalizaciones del generador congruencial lineal simple:

$$x_i = f(x_{i-1}, x_{i-2}, \dots, x_{i-k}) \bmod m$$

- No lineal: e.g. $x_i = a_1 x_{i-1} + a_2 x_{i-1}^2 \bmod m$
- Lineal múltiple: $x_i = a_1 x_{i-1} + a_2 x_{i-2} + \dots + a_k x_{i-k} \bmod m$, con $p \leq m^k - 1$)
- Matricial: $\mathbf{x}_i = A_1 \mathbf{x}_{i-1} + A_2 \mathbf{x}_{i-2} + \dots + A_k \mathbf{x}_{i-k} \bmod m$ (por ejemplo el generador por defecto de R).

Un caso particular del generador lineal múltiple son los denominados *generadores de registros desfasados*. Se generan bits de forma secuencial considerando $m = 2$ y $a_i \in \{0, 1\}$, y los cálculos se pueden realizar rápidamente mediante operaciones lógicas (se pueden emplear directamente los registros del procesador).

Otras alternativas bastante utilizadas consisten en la combinación de varios generadores:

- Combinación de salidas: $u_i = \sum u_i^{(l)} \bmod 1$, donde $u_i^{(l)}$ es el i -ésimo valor obtenido con el generador l .
- Barajar salidas...

Ejercicio 3.1.

Considera el generador congruencial definido por:

$$\begin{aligned} x_{n+1} &= (5x_n + 1) \bmod 512, \\ u_{n+1} &= \frac{x_{n+1}}{512}, \quad n = 0, 1, \dots \end{aligned}$$

(de ciclo máximo).

NOTA: El algoritmo está implementado en el fichero *RANDC.R* y se muestra en la Sección 3.1.1.

- a) Generar 500 valores de este generador, obtener el tiempo de CPU, representar su distribución mediante un histograma (en escala de densidades) y compararla con la densidad teórica.

```
initRANDC(321, 5, 1, 512)          # Fijar semilla para reproductibilidad
nsim <- 500
system.time(u <- RANDCN(nsim))  # Generar

##      user  system elapsed
##        0       0       0

hist(u, freq = FALSE)            # Equivalente a curve(dunif(x, 0, 1), add = TRUE)
```

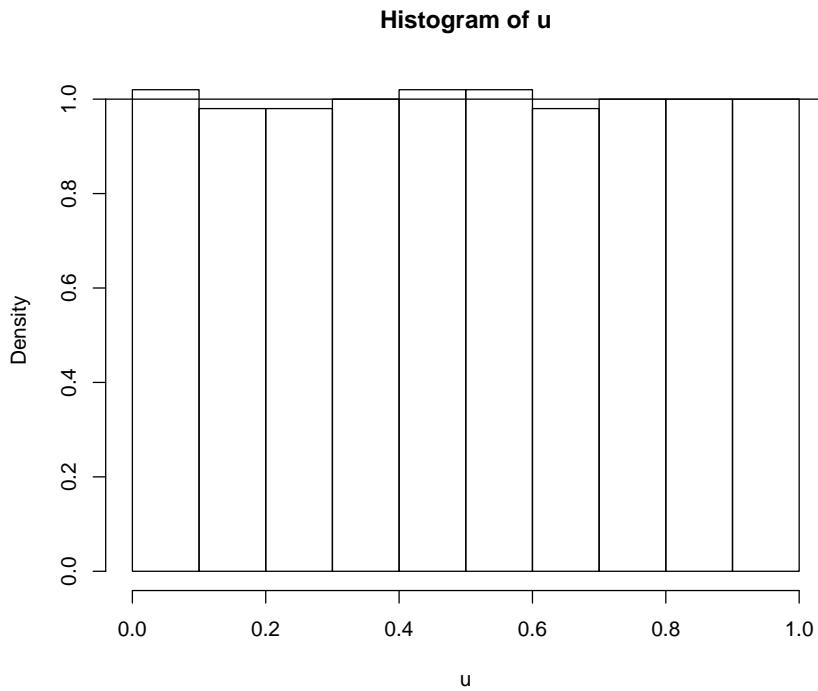


Figura 3.2: Histograma de los valores generados

- b) Calcular la media de las simulaciones (`mean`) y compararla con la teórica.

La aproximación por simulación de la media teórica es:

```
mean(u)
```

```
## [1] 0.4999609
```

La media teórica es 0.5. Error absoluto 3.90625×10^{-5} .

- c) Aproximar (mediante simulación) la probabilidad del intervalo $(0.4; 0.8)$ y compararla con la teórica.

La probabilidad teórica es $0.8 - 0.4 = 0.4$

La aproximación mediante simulación:

```
sum((0.4 < u) & (u < 0.8))/nsim
```

```
## [1] 0.402
```

```
mean((0.4 < u) & (u < 0.8))      # Alternativa
```

```
## [1] 0.402
```

3.2 Análisis de la calidad de un generador

Para verificar si un generador tiene las propiedades estadísticas deseadas hay disponibles una gran cantidad de test de hipótesis (baterías de contrastes) y métodos gráficos:

- Contrastos genéricos de bondad de ajuste y aleatoriedad.

- Contrastos específicos para generadores aleatorios.

Se trata principalmente de contrastar si las muestras generadas son i.i.d. $\mathcal{U}(0, 1)$ (análisis univariante). Aunque los métodos más avanzados tratan normalmente de contrastar si las k -uplas:

$$(U_{t+1}, U_{t+2}, \dots, U_{t+k-1}); t = (i-1)k, i = 1, \dots, m$$

son i.i.d. $\mathcal{U}(0, 1)^k$ (uniformes independientes en el hipercubo; análisis multivariante).

Nos centraremos en los métodos genéricos. Pueden usarse en:

- Evaluación de generadores aleatorios
- Evaluación de generadores de variables aleatorias
- Modelado de entradas de modelos de simulación

Uno de los contrastes más conocidos es el test ji-cuadrado de bondad de ajuste (`chisq.test` para el caso discreto). Aunque si la variable de interés es continua, habría que discretizarla (con la correspondiente perdida de información). Por ejemplo, se podría emplear la siguiente función (que imita a las incluidas en R):

```
#-----
# chisq.test.cont(x, distribution, nclasses, output, nestpar,...)
#-----
# Realiza el test ji-cuadrado de bondad de ajuste para una distribución continua
# discretizando en intervalos equiprobables.
# Parámetros:
#   distribution = "norm", "unif", etc
#   nclasses = floor(length(x)/5)
#   output = TRUE
#   nestpar = 0= nº de parámetros estimados
#   ... = parámetros distribución
# Ejemplo:
#   chisq.test.cont(x, distribution="norm", nestpar=2, mean=mean(x), sd=sqrt((nx-1)/nx)*sd(x))
#-----
chisq.test.cont <- function(x, distribution = "norm", nclasses = floor(length(x)/5),
  output = TRUE, nestpar = 0, ...) {
  # Funciones distribución
  q.distrib <- eval(parse(text = paste("q", distribution, sep = "")))
  d.distrib <- eval(parse(text = paste("d", distribution, sep = "")))
  # Puntos de corte
  q <- q.distrib((1:(nclasses - 1))/nclasses, ...)
  tol <- sqrt(.Machine$double.eps)
  xbreaks <- c(min(x) - tol, q, max(x) + tol)
  # Gráficos y frecuencias
  if (output) {
    xhist <- hist(x, breaks = xbreaks, freq = FALSE, lty = 2, border = "grey50")
    curve(d.distrib(x, ...), add = TRUE)
  } else {
    xhist <- hist(x, breaks = xbreaks, plot = FALSE)
  }
  # Cálculo estadístico y p-valor
  O <- xhist$counts # Equivalente a table(cut(x, xbreaks)) pero más eficiente
  E <- length(x)/nclasses
  DNAME <- deparse(substitute(x))
  METHOD <- "Pearson's Chi-squared test"
  STATISTIC <- sum((O - E)^2/E)
```

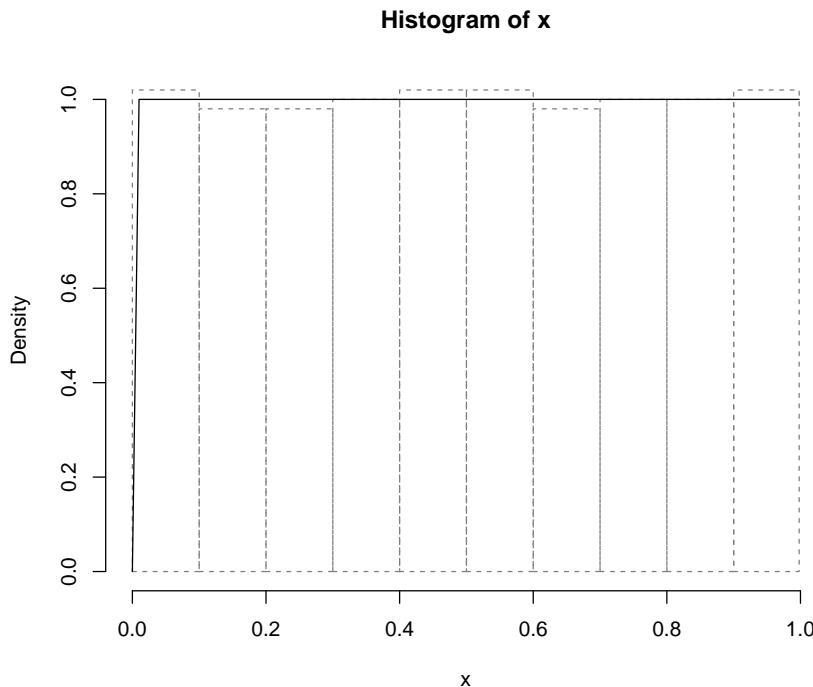
```

names(STATISTIC) <- "X-squared"
PARAMETER <- nclasses - nestpar - 1
names(PARAMETER) <- "df"
PVAL <- pchisq(STATISTIC, PARAMETER, lower.tail = FALSE)
# Preparar resultados
classes <- format(xbreaks)
classes <- paste("(", classes[-(nclasses + 1)], ",",
                 classes[-1], "]", ,
                 sep = ""))
RESULTS <- list(classes = classes, observed = 0, expected = E, residuals = (0 -
                           E)/sqrt(E))
if (output) {
  cat("\nPearson's Chi-squared test table\n")
  print(as.data.frame(RESULTS))
}
if (any(E < 5))
  warning("Chi-squared approximation may be incorrect")
structure(c(list(statistic = STATISTIC, parameter = PARAMETER, p.value = PVAL,
             method = METHOD, data.name = DNAME), RESULTS), class = "htest")
}

```

Por ejemplo, continuando con el generador congruencial anterior, obtendríamos:

```
chisq.test.cont(u, distribution = "unif",
                 nclasses = 10, nestpar = 0, min = 0, max = 1)
```



```
##
## Pearson's Chi-squared test table
##          classes observed expected residuals
## 1  (-1.490116e-08, 1.000000e-01]      51      50  0.1414214
## 2  ( 1.000000e-01, 2.000000e-01]      49      50 -0.1414214
```

```

## 3  ( 2.000000e-01, 3.000000e-01]      49      50 -0.1414214
## 4  ( 3.000000e-01, 4.000000e-01]      50      50  0.0000000
## 5  ( 4.000000e-01, 5.000000e-01]      51      50  0.1414214
## 6  ( 5.000000e-01, 6.000000e-01]      51      50  0.1414214
## 7  ( 6.000000e-01, 7.000000e-01]      49      50 -0.1414214
## 8  ( 7.000000e-01, 8.000000e-01]      50      50  0.0000000
## 9  ( 8.000000e-01, 9.000000e-01]      50      50  0.0000000
## 10 ( 9.000000e-01, 9.980469e-01]     50      50  0.0000000

##
## Pearson's Chi-squared test
##
## data: u
## X-squared = 0.12, df = 9, p-value = 1

```

Importante:

Empleando los métodos genéricos del modo habitual, desconfiamos del generador si la muestra/secuencia no se ajusta a la distribución teórica (p -valor $\leq \alpha$). En este caso además, **también se sospecha si se ajusta demasiado bien** a la distribución teórica (p -valor $\geq 1 - \alpha$).

Otro contraste de bondad de ajuste muy conocido es el test de Kolmogorov-Smirnov, implementado en `ks.test`.

Ejercicio 3.2.

Continuando con el generador congruencial anterior:

```

initRANDC(321, 5, 1, 512)      # Fijar semilla para reproductibilidad
nsim <- 500
system.time(u <- RANDCN(nsim)) # Generar

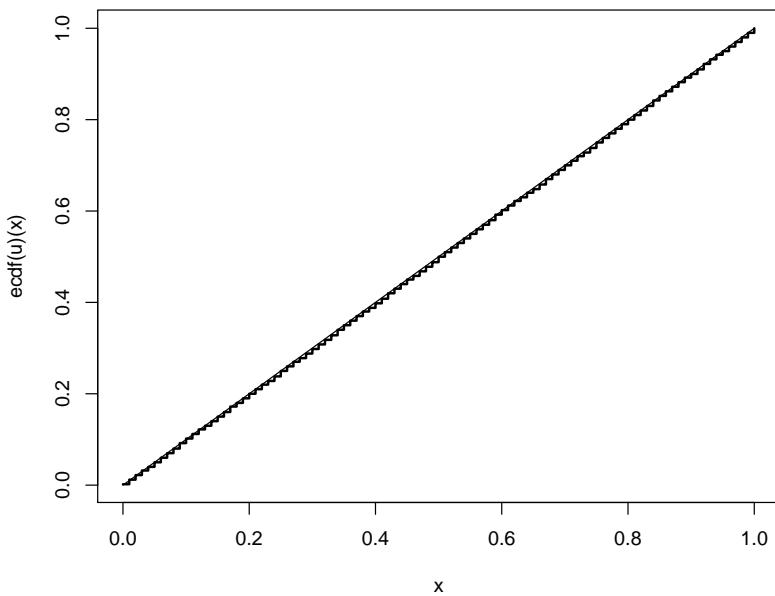
```

- a) Realizar el contraste de Kolmogorov-Smirnov para estudiar el ajuste a una $\mathcal{U}(0, 1)$.

```

# Distribución empírica
curve(ecdf(u)(x), type = "s", lwd = 2)
curve(punif(x, 0, 1), add = TRUE)

```



```
# Test de Kolmogorov-Smirnov
ks.test(u, "punif", 0, 1)

##
##  One-sample Kolmogorov-Smirnov test
##
## data: u
## D = 0.0033281, p-value = 1
## alternative hypothesis: two-sided
```

- b) Obtener el gráfico secuencial y el de dispersión retardado, ¿se observa algún problema?

Gráfico secuencial:

```
plot(as.ts(u))
```

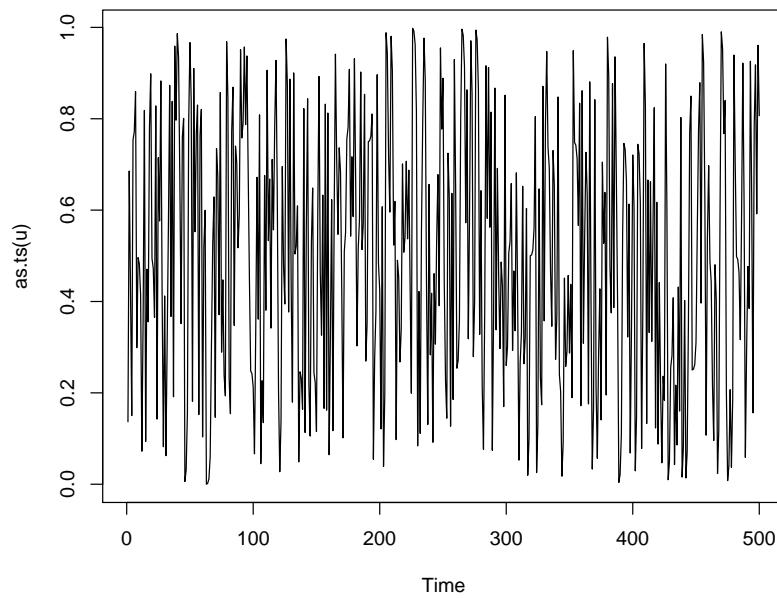
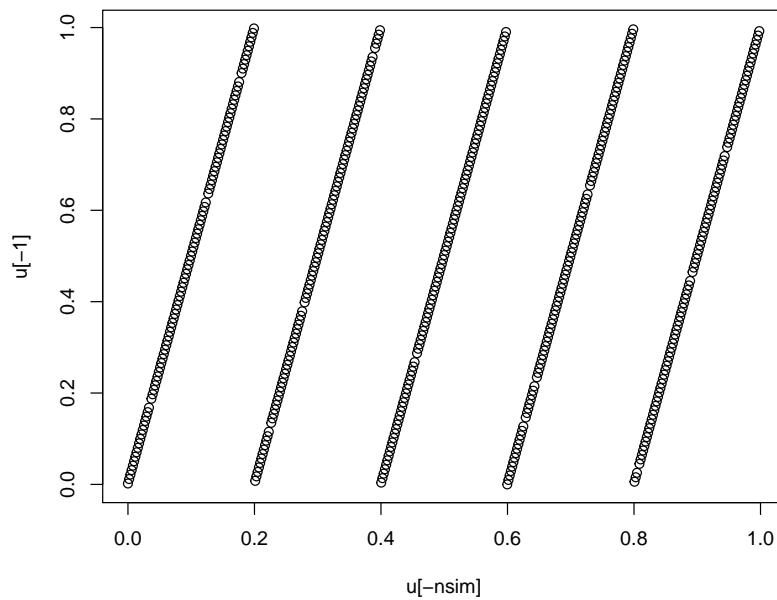


Gráfico de dispersión retardado:

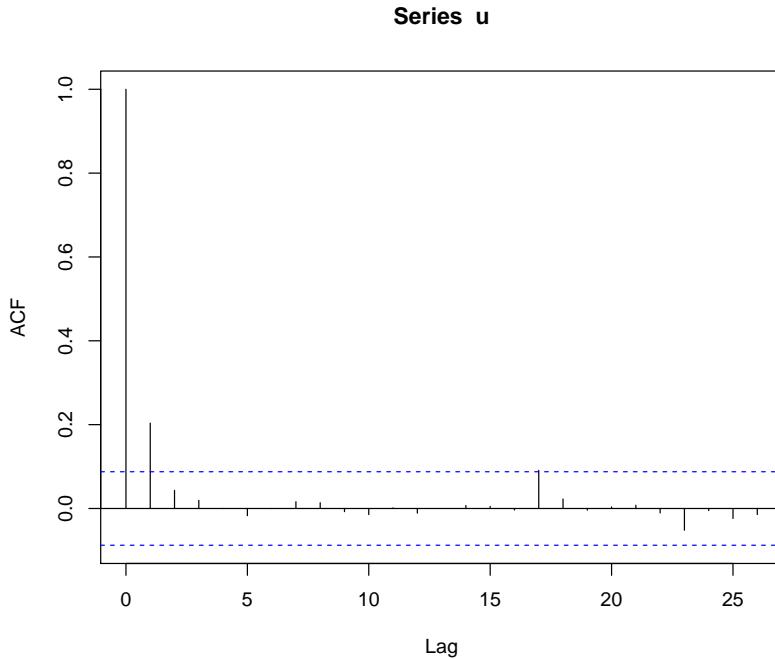
```
plot(u[-nsim], u[-1])
```



- c) Estudiar las correlaciones del vector (u_i, u_{i+k}) , con $k = 1, \dots, 10$. Contrastar si son nulas.

Correlaciones:

```
acf(u)
```



Test de Ljung-Box:

```
Box.test(u, lag = 10, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data: u
## X-squared = 22.533, df = 10, p-value = 0.01261
```

3.2.1 Repetición de contrastes

Los contrastes se plantean habitualmente desde el punto de vista de la inferencia estadística en la práctica: se realiza una prueba sobre la única muestra disponible. Si se realiza una única prueba, en las condiciones de H_0 hay una probabilidad α de rechazarla. En simulación tiene mucho más sentido realizar un gran número de pruebas:

- La proporción de rechazos debería aproximarse al valor de α (se puede comprobar para distintos valores de α).
- La distribución del estadístico debería ajustarse a la teórica bajo H_0 (se podría realizar un nuevo contraste de bondad de ajuste).
- Los p -valores obtenidos deberían ajustarse a una $\mathcal{U}(0, 1)$ (se podría realizar también un contraste de bondad de ajuste).

Este procedimiento es también el habitual para validar un método de contraste de hipótesis por simulación.

Ejemplo 3.1.

Consideramos el generador congruencial RANDU:

```
# Valores iniciales
initRANDC(543210)    # Fijar semilla para reproducibilidad
# set.seed(543210)
n <- 500
nsim <- 1000
estadistico <- numeric(nsim)
pvalor <- numeric(nsim)

# Realizar contrastes
for(isim in 1:nsim) {
  u <- RANDCN(n)      # Generar
  # u <- runif(n)
  tmp <- chisq.test.cont(u, distribution="unif",
                         nclasses=100, output=FALSE, nestpar=0, min=0, max=1)
  estadistico[isim] <- tmp$statistic
  pvalor[isim] <- tmp$p.value
}
```

Proporción de rechazos:

```
# cat("\nProporción de rechazos al 1% =", sum(pvalor < 0.01)/nsim, "\n")
cat("\nProporción de rechazos al 1% =", mean(pvalor < 0.01), "\n")
```

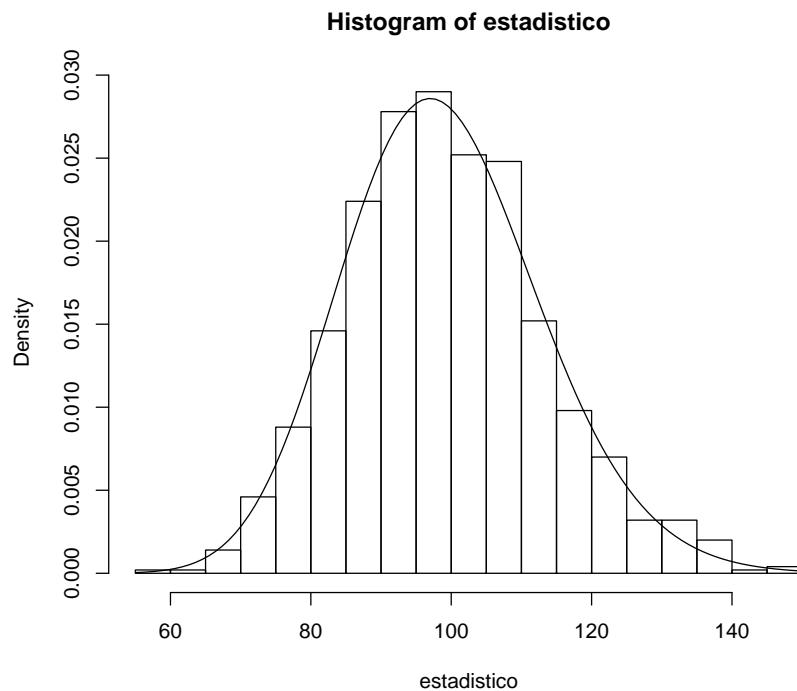
```
##
## Proporción de rechazos al 1% = 0.014
# cat("Proporción de rechazos al 5% =", sum(pvalor < 0.05)/nsim, "\n")
cat("Proporción de rechazos al 5% =", mean(pvalor < 0.05), "\n")
```

```
## Proporción de rechazos al 5% = 0.051
# cat("Proporción de rechazos al 10% =", sum(pvalor < 0.1)/nsim, "\n")
cat("Proporción de rechazos al 10% =", mean(pvalor < 0.1), "\n")
```

```
## Proporción de rechazos al 10% = 0.112
```

Análisis del estadístico contraste:

```
# Histograma
hist(estadistico, breaks = "FD", freq=FALSE)
curve(dchisq(x,99), add=TRUE)
```

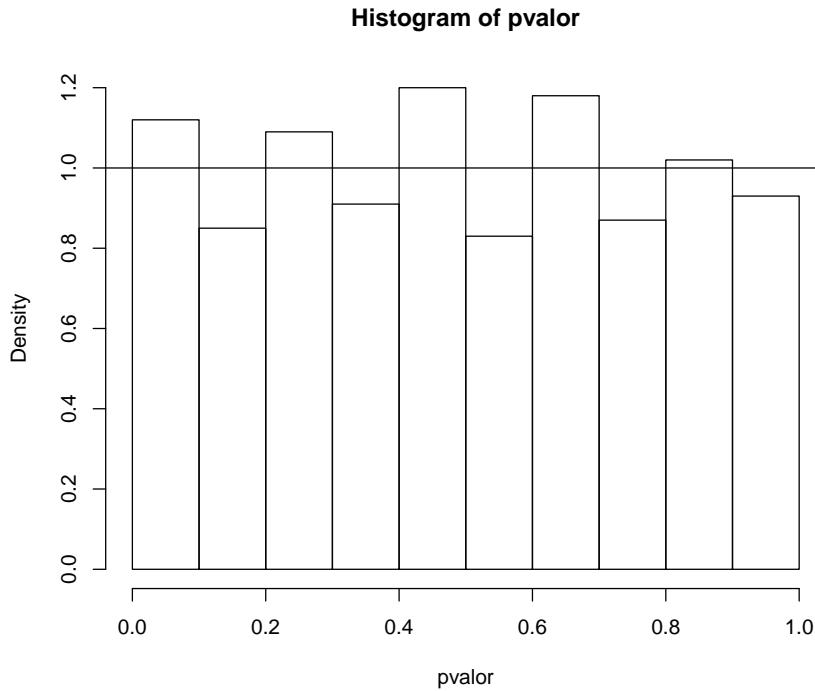


```
# Test ji-cuadrado
# chisq.test.estadistico, distribution="chisq", nclasses=20, nestpar=0, df=99)
# Test de Kolmogorov-Smirnov
ks.test(estadistico, "pchisq", df=99)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data: estadistico
## D = 0.023499, p-value = 0.6388
## alternative hypothesis: two-sided
```

Análisis de los p-valores:

```
# Histograma
hist(pvalor, freq=FALSE)
abline(h=1) # curve(dunif(x,0,1), add=TRUE)
```



```
# Test ji-cuadrado
# chisq.test.cont(pvalor, distribution="unif", nclasses=20, nestpar=0, min=0, max=1)
# Test de Kolmogorov-Smirnov
ks.test(pvalor, "punif", min=0, max=1)

##
## One-sample Kolmogorov-Smirnov test
##
## data: pvalor
## D = 0.023499, p-value = 0.6388
## alternative hypothesis: two-sided
```

3.2.2 Baterías de contrastes

Contrastes específicos para generadores aleatorios:

- Diehard tests (The Marsaglia Random Number CDROM): <http://www.stat.fsu.edu/pub/diehard>.
- TestU01: <http://www.iro.umontreal.ca/simardr/testu01/tu01.html>.
- NIST test suite: <http://csrc.nist.gov/groups/ST/toolkit/rng>.
- Dieharder (paquete RDieHarder): <http://www.phy.duke.edu/rgb/General/dieharder.php>
- Entidad Certificadora (gratuita): CAcert.

Documentación adicional:

- Randomness Tests: A Literature Survey <http://www.ciphersbyritter.com/RES/RANDTEST.HTM>
- Marsaglia, Tsang (2002). Some Difficult-to-pass Tests of Randomness: <http://www.jstatsoft.org/v07/i03>
<http://www.csis.hku.hk/cisc/download/ideetect>

3.3 Ejercicios de fin de práctica

Ejercicio 3.3.

Uno de los primeros generadores fue el denominado método de los cuadrados medios propuesto por Von Neumann (1946). Con este procedimiento se generan números pseudoaleatorios de 4 dígitos de la siguiente forma:

- i. Se escoge un número de cuatro dígitos x_0 (semilla).
- ii. Se eleva al cuadrado (x_0^2) y se toman los cuatro dígitos centrales (x_1).
- iii. Se genera el número pseudo-aleatorio como

$$u_1 = \frac{x_1}{10^4}.$$

- iv. Volver al paso ii y repetir el proceso.

Para obtener los k (número par) dígitos centrales de x_i^2 se puede utilizar que:

$$x_{i+1} = \left\lfloor \left(x_i^2 - \left\lfloor \frac{x_i^2}{10^{(2k-\frac{k}{2})}} \right\rfloor 10^{(2k-\frac{k}{2})} \right) / 10^{\frac{k}{2}} \right\rfloor$$

El algoritmo está implementado en el fichero *RANDVN.R*:

```
# Generador Von Neumann de números pseudoaleatorios
# =====

# -----
# initRANDVN(semilla,n)
#   Inicia el generador
#   n número de dígitos centrales, 4 por defecto (debe ser un nº par)
#   Por defecto semilla del reloj
#   OJO: No se hace ninguna verificación de los parámetros
initRANDVN <- function(semilla = as.numeric(Sys.time()), n = 4) {
  .semilla <- as.double(semilla) %% 10^n # Cálculos en doble precisión
  .n <- n
  .aux <- 10^(2*n-n/2)
  .aux2 <- 10^(n/2)
  return(invisible(list(semilla=.semilla,n=.n)))
}

# -----
# RANDVN()
#   Genera un valor pseudoaleatorio con el generador de Von Neumann
#   Actualiza la semilla (si no existe llama a initRANDVN)
RANDVN <- function() {
  if (!exists(".semilla", envir=globalenv())) initRANDVN()
  z <- .semilla^2
  .semilla <- trunc((z-trunc(z/.aux)*.aux)/.aux2)
  return(.semilla/10^.n)
}

# -----
# RANDVNN(n)
#   Genera un vector de valores pseudoaleatorios con el generador congruencial
```

```
#   (por defecto de dimensión 1000)
# Actualiza la semilla (si no existe llama a initRANDVN)
RANDVNN <- function(n = 1000) {
  x <- numeric(n)
  for(i in 1:n) x[i] <- RANDVN()
  return(x)
  # return(replicate(n,RANDVN())) # Alternativa más rápida
}
```

Estudiar las características del generador de cuadrados medios a partir de una secuencia de 500 valores. Emplear únicamente métodos gráficos.

Ejercicio 3.4.

Considerando el generador congruencial multiplicativo de parámetros $a = 7^5 = 16807$, $c = 0$ y $m = 2^{31} - 1$. ¿Se observan los mismos problemas que con el algoritmo RANDU al considerar las tripletas (x_k, x_{k+1}, x_{k+2}) ?

Capítulo 4

Análisis de resultados de simulación

4.1 Convergencia

Supongamos que estamos interesados en aproximar la media teórica $E(X)$ a partir de una secuencia i.i.d. X_1, X_2, \dots, X_n mediante la media muestral \bar{X}_n (aproximar una probabilidad sería un caso particular considerando una distribución de Bernoulli). Una justificación teórica de la validez de la aproximación obtenida mediante simulación es *la ley (débil) de los grandes números*:

- Si X_1, X_2, \dots es una secuencia de v.a.'s independientes con:

$$E(X_i) = \mu \text{ y } Var(X_i) = \sigma^2 < \infty,$$

entonces $\bar{X}_n = (X_1 + \dots + X_n)/n$ converge en probabilidad a μ . i.e. para cualquier $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

- *La ley fuerte* establece la convergencia casi segura.

Ejemplo 4.1. Aproximación de una probabilidad

Simulamos una distribución de Bernoulli de parámetro $p = 0.5$:

```
p <- 0.5
set.seed(1)
nsim <- 10000
rx <- runif(nsim) <= p
```

La aproximación por simulación de p será:

```
mean(rx)
```

```
## [1] 0.5047
```

Podemos generar un gráfico con la evolución de la aproximación con el siguiente código:

```
plot(cumsum(rx)/1:nsim, type="l", lwd=2, xlab="Número de generaciones",
      ylab="Proporción muestral", ylim=c(0,1))
abline(h = mean(rx), lty = 2)
# valor teórico
abline(h = p)
```

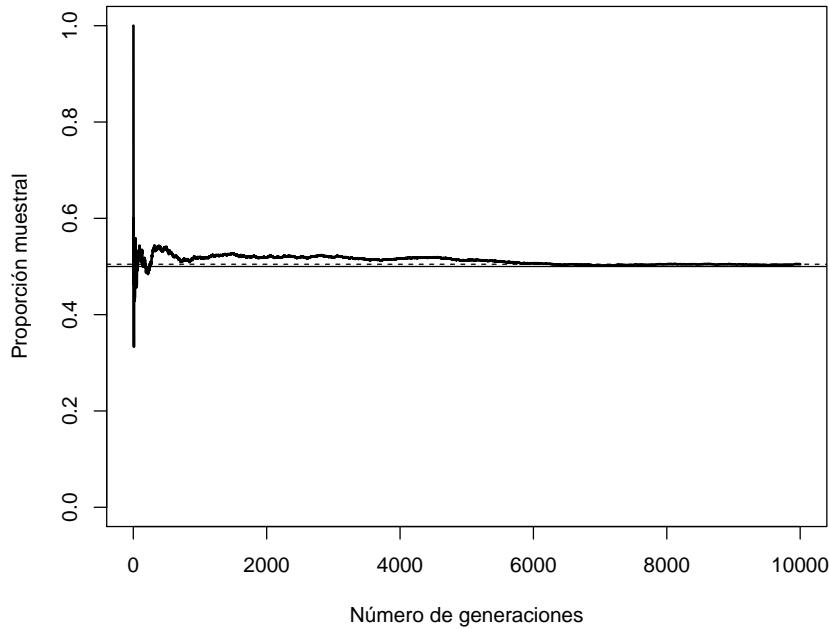


Figura 4.1: Aproximación de la proporción en función del número de generaciones.

4.1.1 Detección de problemas de convergencia

Una suposición crucial es que las variables X_i deben tener varianza finita (realmente esta suposición puede relajarse: $E(|X_i|) < \infty$). En caso contrario la media muestral puede no converger a una constante. Un ejemplo conocido es la distribución de Cauchy:

```
set.seed(1)
nsim <- 10000
rx <- rcauchy(nsim)
plot(cumsum(rx)/1:nsim, type="l", lwd=2,
     xlab="Número de generaciones", ylab="Media muestral")
```

Para detectar problemas de convergencia es recomendable representar la evolución de la aproximación de la característica de interés (sobre el número de generaciones). Además del análisis descriptivo de las simulaciones. Por ejemplo, en el siguiente gráfico de cajas observamos los valores que producen estos saltos:

```
boxplot(rx)
```

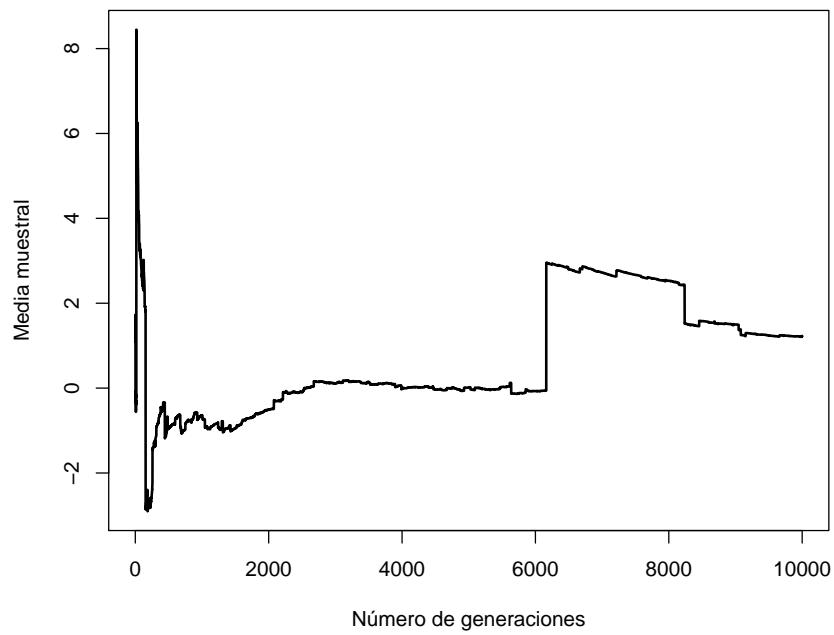
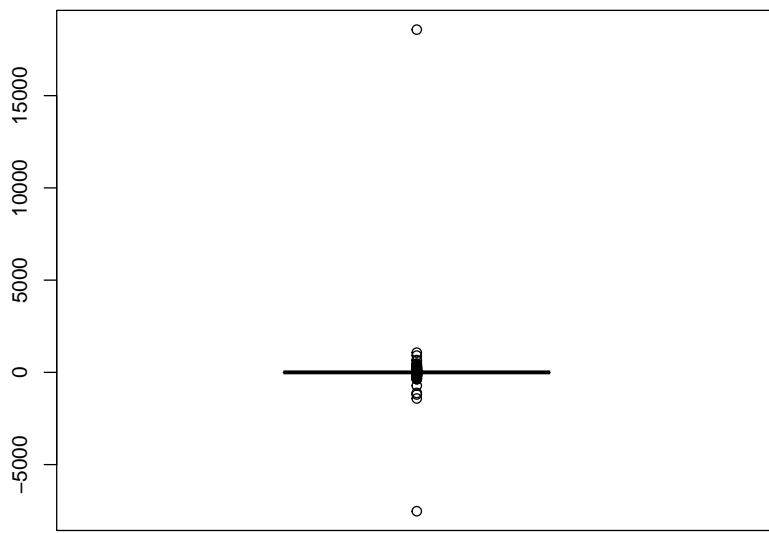


Figura 4.2: Evolución de la media muestral de una distribución de Cauchy en función del número de generaciones.



4.2 Estimación de la precisión

En el caso de la media muestral \bar{X}_n , un estimador insesgado de $Var(\bar{X}_n) = \sigma^2/n$ es:

$$\widehat{Var}(\bar{X}_n) = \frac{\hat{S}^2}{n}$$

con:

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

En el caso de una proporción \hat{p}_n :

$$\widehat{Var}(\hat{p}_n) = \frac{\hat{p}_n(1-\hat{p}_n)}{n-1}.$$

Los valores obtenidos servirían como medidas básicas de la precisión de la aproximación, aunque su principal aplicación es la construcción de intervalos de confianza.

4.3 Teorema central del límite

Si X_1, X_2, \dots es una secuencia de v.a.'s independientes con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2 < \infty$, entonces:

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1)$$

i.e. $\lim_{n \rightarrow \infty} F_{Z_n}(z) = \Phi(z)$. Por tanto, un intervalo de confianza asintótico para μ es:

$$IC_{1-\alpha}(\mu) = \left(\bar{X}_n - z_{1-\alpha/2} \frac{\hat{S}_n}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha/2} \frac{\hat{S}_n}{\sqrt{n}} \right).$$

Podemos considerar que $z_{1-\alpha/2} \frac{\hat{S}_n}{\sqrt{n}}$ es la precisión obtenida (con nivel de confianza $1 - \alpha$).

Ejemplo 4.2. Aproximación de la media de una distribución normal

```
xsd <- 1
xmed <- 0
set.seed(1)
nsim <- 1000
rx <- rnorm(nsim, xmed, xsd)
```

La aproximación por simulación de la media será:

```
mean(rx)
```

```
## [1] -0.01164814
```

Como medida de la precisión de la aproximación podemos considerar:

```
2*sd(rx)/sqrt(nsim)
```

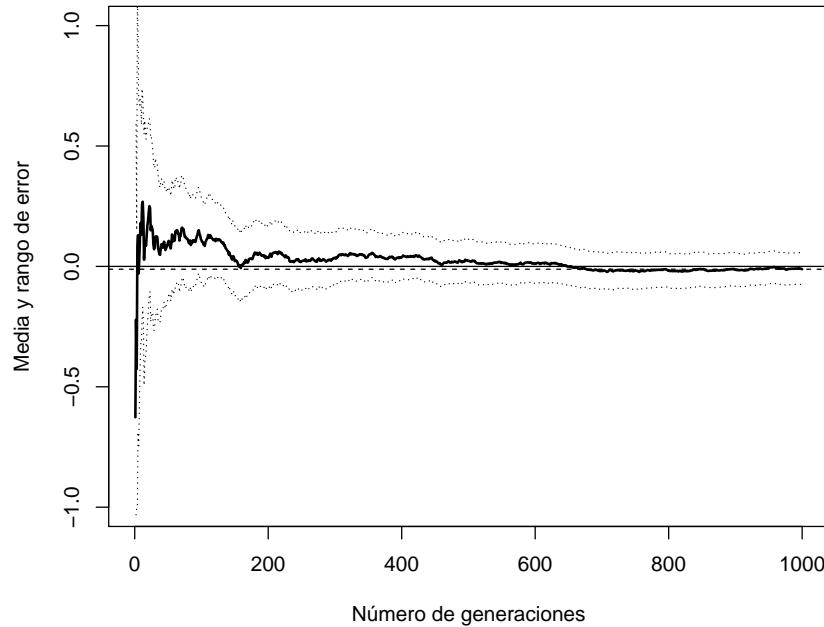
```
## [1] 0.06545382
```

(es habitual emplear 2 en lugar de 1.96, lo que se correspondería con $1 - \alpha = 0.9545$ en el caso de normalidad). Podemos añadir también los correspondientes intervalos de confianza al gráfico de convergencia:

```

n <- 1:n sim
est <- cumsum(rx)/n
esterr <- sqrt(cumsum((rx-est)^2))/(n - 1) # Error standar
plot(est, type = "l", lwd = 2, xlab = "Número de generaciones",
     ylab = "Media y rango de error", ylim = c(-1, 1))
abline(h = est[n sim], lty=2)
lines(est + 2*esterr, lty=3)
lines(est - 2*esterr, lty=3)
abline(h = xmed)

```



4.4 Determinación del número de generaciones

En muchas ocasiones puede interesar obtener una aproximación con un nivel de precisión fijado.

Para una precisión absoluta ε , se trata de determinar n de forma que:

$$z_{1-\alpha/2} \frac{\hat{S}_n}{\sqrt{n}} < \varepsilon$$

Un algoritmo podría ser el siguiente:

1. Hacer $j = 0$ y fijar un tamaño inicial n_0 (e.g. 30 ó 60).
2. Generar $\{X_i\}_{i=1}^{n_0}$ y calcular \hat{S}_{n_0} .
3. Mientras $z_{1-\alpha/2} \hat{S}_{n_{j-1}} / \sqrt{n_j} > \varepsilon$ hacer:
 4. $j = j + 1$.
 5. $n_j = \left\lceil \left(z_{1-\alpha/2} \hat{S}_{n_{j-1}} / \varepsilon \right)^2 \right\rceil$.

6. Generar $\{X_i\}_{i=n_{j-1}+1}^{n_j}$ y calcular \widehat{S}_{n_j} .

Para una precisión relativa $\varepsilon |\mu|$ se procede análogamente de forma que:

$$z_{1-\alpha/2} \frac{\widehat{S}_n}{\sqrt{n}} < \varepsilon |\bar{X}_n|.$$

4.5 El problema de la dependencia

En el caso de dependencia, la estimación de la precisión se complica:

$$Var(\bar{X}) = \frac{1}{n^2} \left(\sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j) \right).$$

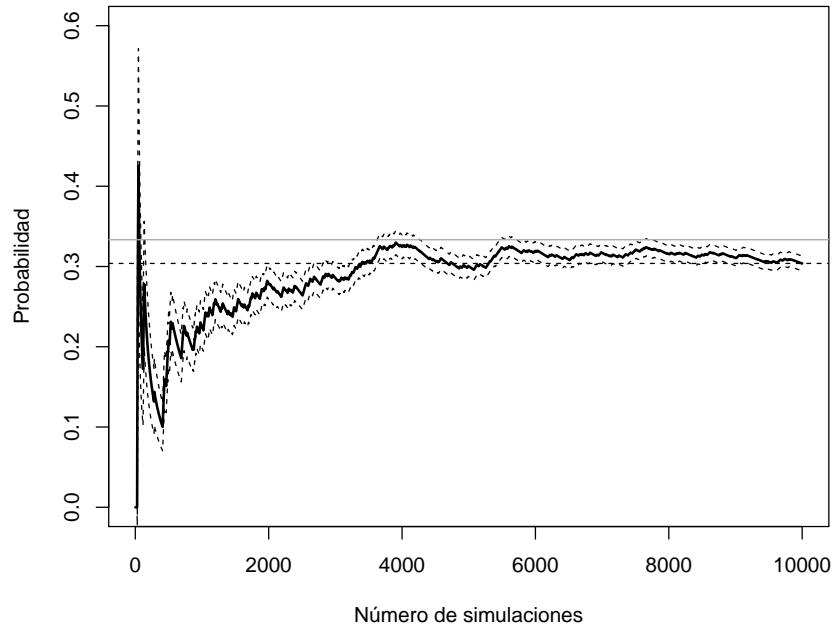
Ejemplo 4.3. Aproximación de una proporción bajo dependencia (cadena de Markov)

Supongamos que en A Coruña llueve de media $1/3$ días al año, y que la probabilidad de que un día llueva solo depende de lo que ocurrió el día anterior, siendo 0.94 si el día anterior llovió y 0.03 si no. Podemos generar valores de la variable indicadora de día lluvioso con el siguiente código:

```
# Variable dicotómica 0/1 (FALSE/TRUE)
set.seed(1)
nsim <- 10000
alpha <- 0.03 # prob de cambio si seco
beta <- 0.06 # prob de cambio si lluvia
rx <- logical(nsim) # x == "llueve"
rx[1] <- FALSE # El primer día no llueve
for (i in 2:nsim)
  rx[i] <- if (rx[i-1]) runif(1) > beta else runif(1) < alpha
```

Se podría pensar en emplear las expresiones anteriores:

```
n <- 1:nsim
est <- cumsum(rx)/n
esterr <- sqrt(est*(1-est)/(n-1)) # OJO! Supone independencia
plot(est, type="l", lwd=2, ylab="Probabilidad",
     xlab="Número de simulaciones", ylim=c(0,0.6))
abline(h = est[nsim], lty=2)
lines(est + 2*esterr, lty=2)
lines(est - 2*esterr, lty=2)
abline(h = 1/3, col="darkgray")      # Prob. teor. cadenas Markov
```



La aproximación de la proporción sería correcta (es consistente):

```
est[nsim]
```

```
## [1] 0.3038
```

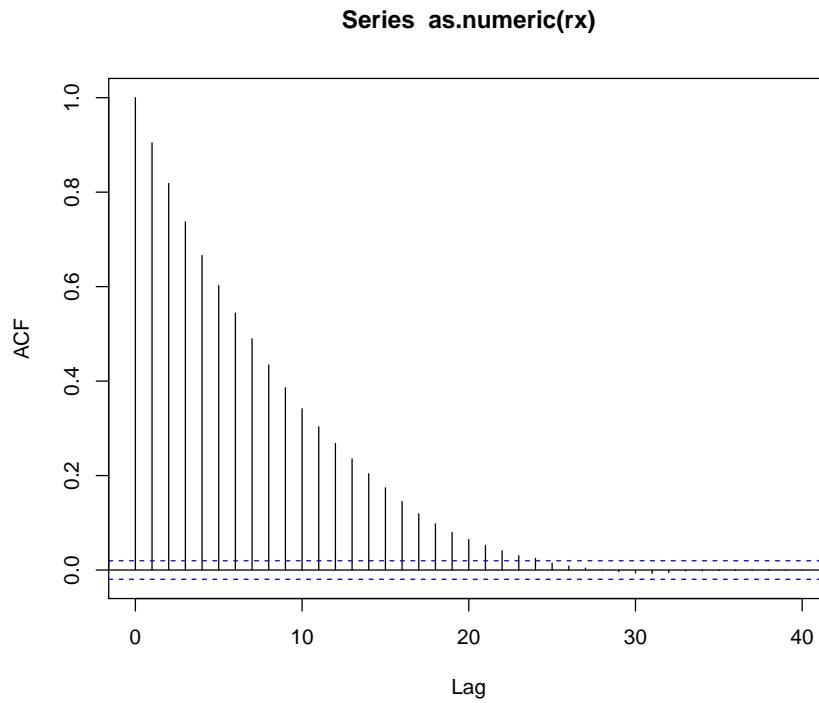
Sin embargo, al ser datos dependientes **esta aproximación del error estandar no es adecuada:**

```
esterr[nsim]
```

```
## [1] 0.004599203
```

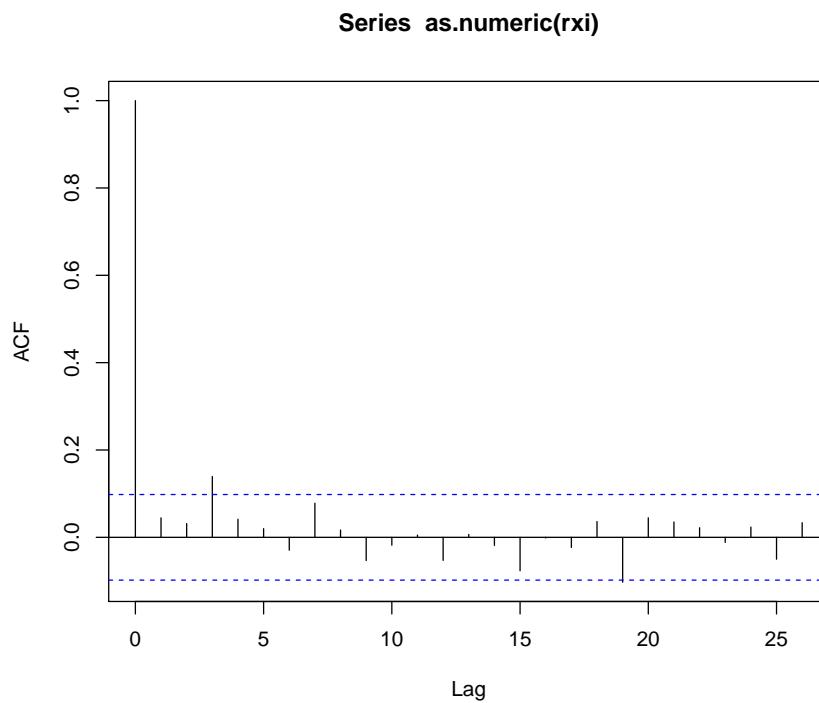
En este caso al haber dependencia positiva se produce una subestimación del verdadero error estandar.

```
acf(as.numeric(rx))
```



El gráfico de autocorrelaciones sugiere que si tomamos 1 de cada 25 podemos suponer independencia.

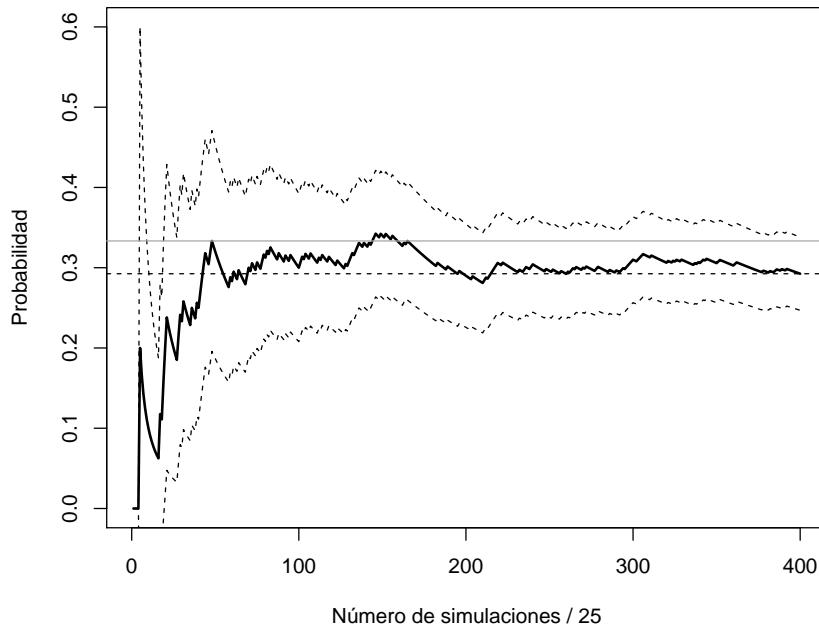
```
lag <- 24
xlag <- c(rep(FALSE, lag), TRUE)
rxi <- rx[xlag]
acf(as.numeric(rxi))
```



```

n <- 1:length(rxi)
est <- cumsum(rxi)/n
esterr <- sqrt(est*(1-est)/(n-1))
plot(est, type="l", lwd=2, ylab="Probabilidad",
     xlab=paste("Número de simulaciones /", lag + 1), ylim=c(0,0.6))
abline(h = est[length(rxi)], lty=2)
lines(est + 2*esterr, lty=2) # Supone independencia
lines(est - 2*esterr, lty=2)
abline(h = 1/3, col="darkgray")      # Prob. teor. cadenas Markov

```



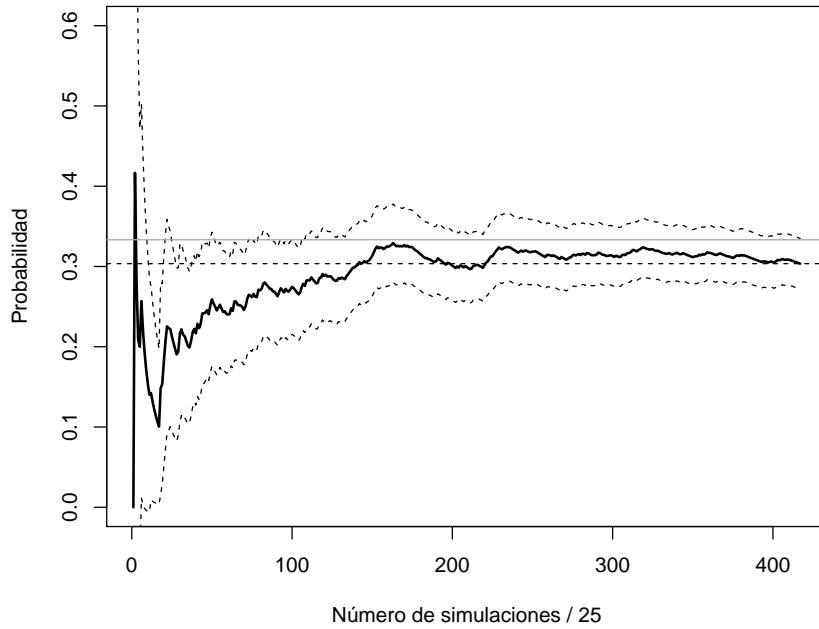
Esta forma de proceder podría ser adecuada para tratar de aproximar la precisión pero no sería eficiente para aproximar la media. Siempre será preferible emplear todas las observaciones.

Por ejemplo, se podría pensar en considerar las medias de grupos de 24 valores consecutivos y suponer que hay independencia entre ellas:

```

rxm <- rowMeans(matrix(rx, ncol = lag, byrow = TRUE))
n <- 1:length(rxm)
est <- cumsum(rxm)/n
esterr <- sqrt(cumsum((rxm-est)^2))/(n - 1) # Error estándar
plot(est, type="l", lwd=2, ylab="Probabilidad",
     xlab=paste("Número de simulaciones /", lag + 1), ylim=c(0,0.6))
abline(h = est[length(rxm)], lty=2)
lines(est + 2*esterr, lty=2) # OJO! Supone independencia
lines(est - 2*esterr, lty=2)
abline(h = 1/3, col="darkgray")      # Prob. teor. cadenas Markov

```



Esta es la idea del método de medias por lotes (*batch means; macro-micro replicaciones*) para la estimación de la varianza. Alternativamente se podría recurrir a la generación de múltiples secuencias independientes.

Trataremos este tipo de problemas en la diagnosis de algoritmos de simulación Monte Carlo de Cadenas de Markov (MCMC). Aparecen también en la simulación dinámica (por eventos o cuantos).

Nota: En el caso anterior se calcula el error estándar de la aproximación por simulación de la proporción, pero si el objetivo es la aproximación de la varianza (de la variable y no de las medias por lotes), habrá que reescalarlo adecuadamente:

```
var.aprox <- nsim * esterr[length(rxm)]^2
var.aprox
```

```
## [1] 2.473664
```

Obtenida asumiendo independencia entre las medias por lotes, y que será una mejor aproximación que asumir independencia entre las generaciones de la variable:

```
var(rx)
```

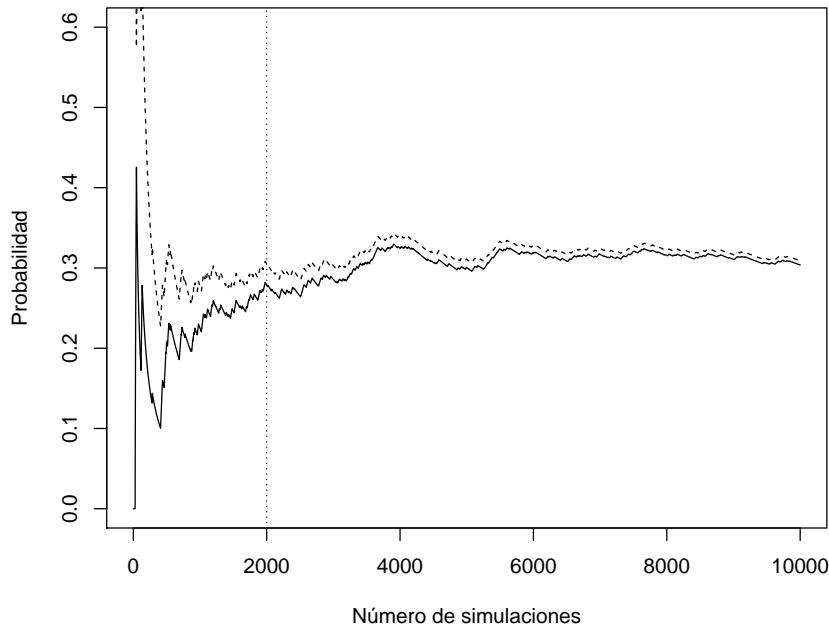
```
## [1] 0.2115267
```

4.5.1 Período de calentamiento

En el caso de simulación de datos dependientes (simulación dinámica) pueden aparecer problemas de estabilización. Puede ocurrir que el sistema evolucione lentamente en el tiempo hasta alcanzar su distribución estacionaria, siendo muy sensible a las condiciones iniciales con las que se comenzó la simulación. En tal caso resulta conveniente ignorar los resultados obtenidos durante un cierto período inicial de tiempo (denominado período de calentamiento o estabilización), cuyo único objeto es conseguir que se establezca la distribución de probabilidad.

Como ejemplo comparamos la simulación anterior con la obtenida considerando como punto de partida un día lluvioso.

```
set.seed(1)
rx2 <- logical(nsim)
rx2[1] <- TRUE # El primer día llueve
for (i in 2:nsim)
  rx2[i] <- if (rx2[i-1]) runif(1) > beta else runif(1) < alpha
n <- 1:nsim
est <- cumsum(rx)/n
est2 <- cumsum(rx2)/n
plot(est, type="l", ylab="Probabilidad",
     xlab="Número de simulaciones", ylim=c(0,0.6))
lines(est2, lty = 2)
abline(v = 2000, lty = 3)
```



En estos casos puede ser recomendable ignorar los primeros valores generados (por ejemplo los primeros 2000) y recalcular los estadísticos deseados.

También trataremos este tipo de problemas en la diagnosis de algoritmos MCMC.

Ejemplo 4.4. Simulación de un proceso autorregresivo (serie de tiempo)

$$X_t = \mu + \rho * (X_{t-1} - \mu) + \varepsilon_t$$

Podemos tener en cuenta que en este caso la varianza es:

$$\text{var}(X_t) = E(X_t^2) - \mu^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}.$$

Establecer parámetros

```

nsim <- 200 # Número de simulaciones
xvar <- 1 # Varianza
xmed <- 0 # Media
rho <- 0.5 # Coeficiente AR
nburn <- 10 # Periodo de calentamiento (burn-in)
# Varianza del error
evar <- xvar*(1 - rho^2)

```

Alternativamente se podría fijar la varianza del error:

```

# evar <- 1
# xvar <- evar / (1 - rho^2)

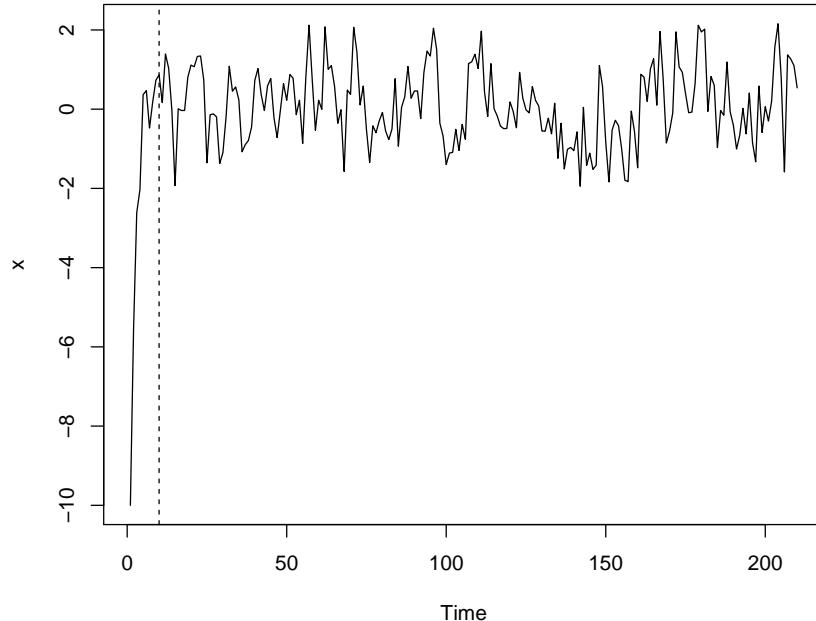
```

Simular

```

set.seed(1)
x <- numeric(nsim + nburn)
# Normalmente simularíamos el primer valor
# rx[1] <- rnorm(1, mean = xmed, sd = sqrt(xvar))
# Aunque en este caso lo fijamos para alejarnos un poco de la distribución estacionaria
# (lo que requerirá de un mayor periodo de calentamiento)
x[1] <- -10
# Simular el resto de la secuencia
for (i in 2:length(x))
  x[i] <- xmed + rho*(x[i-1] - xmed) + rnorm(1, sd=sqrt(evar))
x <- as.ts(x)
plot(x)
abline(v = nburn, lty = 2)

```



```

# Eliminar periodo de calentamiento
rx <- x[-seq_len(nburn)]

```

Para simular una serie de tiempo en R se puede emplear la función `arima.sim` del paquete base `stats`. En este caso el periodo de calentamiento se establece mediante el parámetro `n.start` (que se fija automáticamente a un valor adecuado).

Por ejemplo, podemos generar este serie autoregresiva con:

```
rx2 <- arima.sim(list(order = c(1,0,0), ar = rho), n = nsim, n.start = nburn, sd = sqrt(evar))
```

La recomendación es fijar la varianza de las series simuladas si se quieren comparar resultados considerando distintos parámetros de dependencia.

4.6 Observaciones

- En el caso de que la característica de interés de la distribución de X no sea la media, los resultados anteriores no serían en principio aplicables.
- Incluso en el caso de la media, las bandas de confianza obtenidas con el TCL son puntuales (si generamos nuevas secuencias de simulación es muy probable que no estén contenidas).
- En muchos casos (p.e. la generación de múltiples secuencias de simulación puede suponer un coste computacional importante), puede ser preferible emplear un método de remuestreo.

Capítulo 5

Simulación de variables continuas

En este capítulo se expondrán métodos generales para simular distribuciones continuas: el método de inversión y los basados en aceptación/rechazo. En todos los casos como punto de partida es necesario disponer de un método de generación de números pseudoaleatorios uniformes en (0,1).

5.1 Método de inversión

Se trataría del método preferible para la simulación de una variable continua (siempre que se disponga de la función cuantil). Está basado en los siguientes resultados:

Si X es una variable aleatoria con función de distribución F continua y estrictamente monótona (invertible), entonces:

$$U = F(X) \sim \mathcal{U}(0, 1),$$

ya que:

$$G(u) = P(Y \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u.$$

El recíproco también es cierto, si $U \sim \mathcal{U}(0, 1)$ entonces:

$$F^{-1}(U) \sim X$$

Algoritmo (método de inversión):

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Devolver $X = F^{-1}(U)$.

5.1.1 Ejemplo

Algoritmo para simular el método de inversión la distribución exponencial $\exp(\lambda)$:

- $f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$
- $F(x) = 1 - e^{-\lambda x}$ si $x \geq 0$
- $1 - e^{-\lambda x} = u \Leftrightarrow x = -\frac{\ln(1-u)}{\lambda}$

Algoritmo:

1. Generar $U \sim \mathcal{U}(0, 1)$.

$$2. \text{ Devolver } X = -\frac{\ln(1-U)}{\lambda}.$$

NOTA: en el último paso podemos emplear directamente U en lugar de $1-U$, ya que $1-U \sim \mathcal{U}(0,1)$. Esta última expresión para acelerar los cálculos es la que denominaremos forma simplificada.

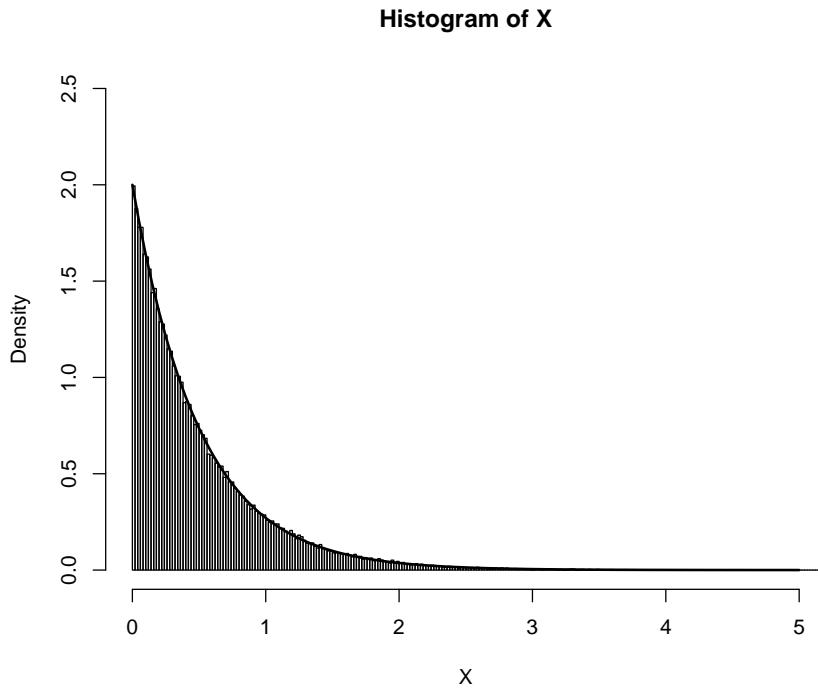
```
# Simular vector exp(lambda)
tini <- proc.time()

lambda <- 2
nsim <- 10^5
set.seed(1)
U <- runif(nsim)
X <- -log(U)/lambda # -log(1-U)/lambda

tiempo <- proc.time() - tini
tiempo

##      user    system elapsed
##      0.02     0.00    0.02

hist(X,breaks="FD",freq=FALSE,xlim=c(0,5),ylim=c(0,2.5))
curve(dexp(x,lambda),lwd=2,add=TRUE)
```



5.1.2 Algunas distribuciones que pueden simularse por el método de inversión

A continuación se incluyen algunas distribuciones que se pueden simular fácilmente mediante el método de inversión. Se adjunta una forma simplificada del método que tiene por objeto evitar cálculos innecesarios (tal y como se hizo en el ejemplo de la exponencial).

Nombre	Densidad	$F(x)$	$F^{-1}(U)$	Forma simplificada
$\exp(\lambda) (\lambda > 0)$	$\lambda e^{-\lambda x}, \text{ si } x \geq 0$	$1 - e^{-\lambda x}$	$-\frac{\ln(1-U)}{\lambda}$	$-\frac{\ln U}{\lambda}$
Cauchy	$\frac{1}{\pi(1+x^2)}$	$\frac{1}{2} + \frac{\arctan x}{\pi}$	$\tan\left(\pi\left(U - \frac{1}{2}\right)\right)$	$\tan \pi U$
Triangular en $(0, a)$	$\frac{2}{a} \left(1 - \frac{x}{a}\right), \text{ si } 0 \leq x \leq a$	$\frac{2}{a} \left(x - \frac{x^2}{2a}\right)$	$a(1 - \sqrt{1-U})$	$a(1 - \sqrt{U})$
Pareto ($a, b > 0$)	$\frac{ab^a}{x^{a+1}}, \text{ si } x \geq b$	$1 - \left(\frac{b}{x}\right)^a$	$\frac{b}{(1-U)^{1/a}}$	$\frac{b}{U^{1/a}}$
Weibull ($\lambda, \alpha > 0$)	$\alpha \lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha}, \text{ si } x \geq 0$	$1 - e^{-(\lambda x)^\alpha}$	$\frac{(-\ln(1-U))^{1/\alpha}}{\lambda}$	$\frac{(-\ln U)^{1/\alpha}}{\lambda}$

Ejercicio 5.1.

La distribución doble exponencial (o distribución de Laplace) de parámetro λ tiene función de densidad:

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}, x \in \mathbb{R}$$

y función de distribución:

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} \frac{1}{2} e^{\lambda x} & x < 0 \\ 1 - \frac{1}{2} e^{-\lambda x} & x \geq 0 \end{cases}$$

- a) Escribir una función que permita generar, por el método de inversión, una muestra de n observaciones de esta distribución (NOTA: esta distribución puede generarse fácilmente simulando una distribución exponencial y otorgarle un signo positivo o negativo con equiprobabilidad; ver método de composición).

```
ddexp <- function(x, lambda = 1){
  # Densidad doble exponencial
  lambda*exp(-lambda*abs(x))/2
}

rdexp <- function(lambda = 1){
  # Simulación por inversión
  # Doble exponencial
  U <- runif(1)
  if (U<0.5) {
    return(log(2*U)/lambda)
  } else {
    return(-log(2*(1-U))/lambda)
  }
}

rdexpn <- function(n = 1000, lambda = 1) {
  # Simulación n valores de doble exponencial
  x <- numeric(n)
  for(i in 1:n) x[i] <- rdexp(lambda)
  return(x)
}
```

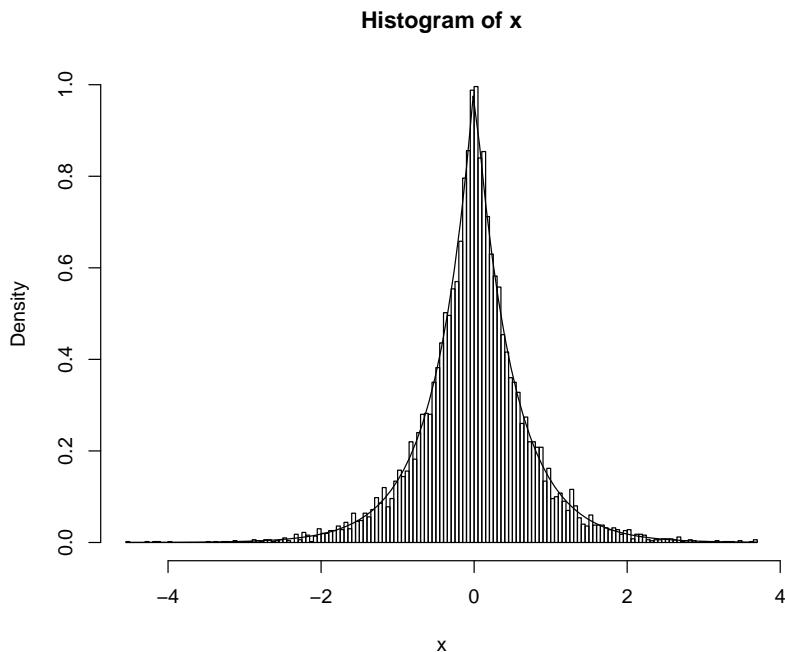
- b) Generar 10^4 valores de la distribución doble exponencial de parámetro $\lambda = 2$ y obtener el tiempo de CPU que tarda en generar la secuencia.

```
set.seed(54321)
system.time(x <- rdexp(10^4, 2))
```

```
##    user  system elapsed
##    0.03    0.00    0.04
```

- c) Representar el histograma y compararlo con la densidad teórica.

```
hist(x, breaks = "FD", freq = FALSE)
# lines(density(x), col = 'blue')
curve(ddexp(x, 2), add = TRUE)
```



5.1.3 Ventajas e inconvenientes

Ventajas:

- Aplicable, en principio, a cualquier distribución continua.

Inconvenientes:

- Puede no ser posible encontrar una expresión explícita para $F^{-1}(u)$.
- Aún disponiendo de una expresión explícita para $F^{-1}(u)$, su evaluación directa puede requerir mucho tiempo de computación.

Alternativas:

- Emplear métodos numéricos para resolver $F(x) - u = 0$ (requeriría resolver numéricamente esta ecuación para cada valor aleatorio que se desee generar).
- Utilizar una aproximación a $F^{-1}(u)$ (inversión aproximada).

5.1.4 Inversión aproximada

En muchos casos en los que no se puede emplear la expresión exacta de la función cuantil $F^{-1}(u)$, se dispone de una aproximación suficientemente buena que se puede emplear en el algoritmo anterior (se obtendrían simulaciones con una distribución aproximada a la deseada).

Por ejemplo, para aproximar la función cuantil de la normal estándar, Odeh y Evans consideraron la siguiente función auxiliar :

$$g(v) = \sqrt{-2 \ln v} \frac{A(\sqrt{-2 \ln v})}{B(\sqrt{-2 \ln v})},$$

siendo $A(x) = \sum_{i=0}^4 a_i x^i$ y $B(x) = \sum_{i=0}^4 b_i x^i$ con:

$$\begin{array}{ll} a_0 = -0.322232431088 & b_0 = 0.0993484626060 \\ a_1 = -1 & b_1 = 0.588581570495 \\ a_2 = -0.342242088547 & b_2 = 0.531103462366 \\ a_3 = -0.0204231210245 & b_3 = 0.103537752850 \\ a_4 = -0.0000453642210148 & b_4 = 0.0038560700634 \end{array}$$

La aproximación consiste en utilizar $g(1-u)$ en lugar de $F^{-1}(u)$ para los valores de $u \in [10^{-20}, \frac{1}{2}]$ y $-g(u)$ si $u \in [\frac{1}{2}, 1 - 10^{-20}]$. Para $u \notin [10^{-20}, 1 - 10^{-20}]$ (que sólo ocurre con una probabilidad de $2 \cdot 10^{-20}$) la aproximación no es recomendable.

Algoritmo de Odeh y Evans

1. Generar $U \sim U(0, 1)$.
2. Si $U < 10^{-20}$ ó $U > 1 - 10^{-20}$ entonces volver a 1.
3. Si $U < 0.5$ entonces hacer $X = g(1-U)$ en caso contrario hacer $X = -g(U)$.
4. Devolver X .

En manuales de funciones matemáticas, como Abramowitz y Stegun (1964), se tienen aproximaciones de la función cuantil de las principales distribuciones (por ejemplo en la página 993 las correspondientes a la normal estándar).

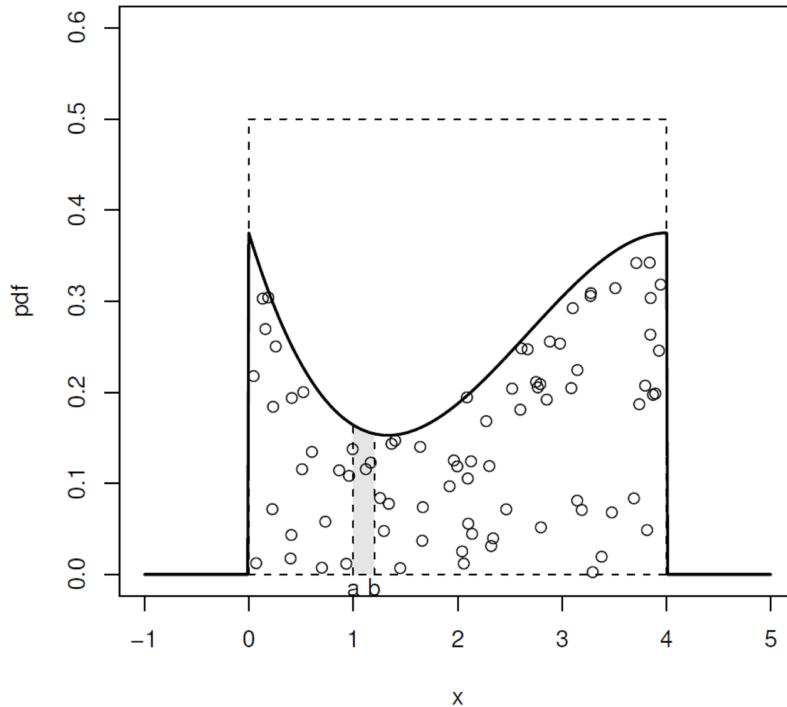
5.2 Método de aceptación rechazo

Se trata de un método universal alternativo al de inversión para el caso de que no se pueda emplear la función cuantil, pero se dispone de una expresión (preferiblemente sencilla) para la función de densidad $f(x)$.

Si f es la densidad objetivo, la idea es simular una variable aleatoria bidimensional (X, Y) con distribución uniforme en el hipógrafo de f (el conjunto de puntos del plano comprendidos entre el eje OX y f):

$$A_f = \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq f(x)\}.$$

De esta forma la primera componente tendrá la distribución deseada:



$$P(a < X < b) = \frac{\text{Área de } \{(x, y) \in \mathbb{R}^2 : a < x < b; 0 \leq y \leq f(x)\}}{\text{Área de } A_f} = \int_a^b f(x) dx$$

El resultado anterior es también válido para una cuasi-densidad f^* (no depende de la constante normalizadora). El resultado general sería en siguiente:

- Si X es una variable aleatoria con función de densidad f y $U \sim \mathcal{U}(0, 1)$ entonces

$$(X, c \cdot U \cdot f(X)) \sim \mathcal{U}(A_{cf})$$

siendo $A_{cf} = \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq cf(x)\}$.

- Recíprocamente si $(X, Y) \sim \mathcal{U}(A_{cf}) \Rightarrow X \sim f$.

Para generar valores de una variable aleatoria bidimensional con distribución uniforme en A_f (o en A_{f^*}), se emplea el resultado anterior para generar valores en $A_{cg} \supset A_f$, siendo g una densidad auxiliar (preferiblemente fácil de simular y similar a f). Teniendo en cuenta además que:

- Si $(X, Y) \sim \mathcal{U}(A)$ y $B \subset A \Rightarrow (X, Y)|_B \sim \mathcal{U}(B)$

Por tanto, si (T, Y) sigue una distribución uniforme en A_{cg} , aceptando los valores de (T, Y) que pertenezcan a A_f (o a A_{f^*}) se obtendrán generaciones con distribución uniforme sobre A_f (o A_{f^*}) y la densidad de la primera componente T será f .

5.2.1 Algoritmo

Supongamos que f es la densidad objetivo y g es una densidad auxiliar (fácil de simular y similar a f), de forma que existe una constante $c > 0$ tal que:

$$f(x) \leq c \cdot g(x), \forall x \in \mathbb{R}.$$

Algoritmo (Von Neuman, 1951):

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Generar $T \sim g$.
3. Si $c \cdot U \cdot g(T) \leq f(T)$ devolver $X = T$,
en caso contrario volver al paso 1.

5.2.2 Densidades acotadas en un intervalo cerrado

Sea f una función de densidad cualquiera con soporte en un intervalo cerrado $[a, b]$ (es decir, $\{x/f(x) \neq 0\} = [a, b]$) de tal forma que existe una constante $M > 0$ tal que $f(x) \leq M \forall x$ (es decir, f es acotada superiormente). En este caso puede tomarse como densidad auxiliar g , la de una $\mathcal{U}(a, b)$. En efecto, tomando $c = M(b - a)$ y teniendo en cuenta que

$$g(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{en caso contrario} \end{cases}$$

se tiene que $f(x) \leq M = \frac{c}{b-a} = c \cdot g(x)$, $\forall x \in [a, b]$. Así pues, el algoritmo quedaría como sigue:

1. Generar $U, V \sim \mathcal{U}(0, 1)$.
2. Hacer $T = a + (b - a)V$.
3. Si $M \cdot U \leq f(T)$ devolver $X = T$,
en caso contrario volver al paso 1.

Nota: no confundir M con $c = M(b - a)$.

Ejercicio 5.2.

Desarrollar el código necesario para generar, por el método de aceptación-rechazo, una muestra de n observaciones de una distribución normal estándar:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R},$$

empleando como distribución auxiliar una doble exponencial con $\lambda = 1$ (más adelante veremos que esta es la elección óptima para el parámetro de la densidad auxiliar) y que la cota optima es:

$$c_{\text{opt}} = \sqrt{\frac{2e}{\pi}} \simeq 1.3155.$$

Para establecer la condición de aceptación o rechazo se puede tener en cuenta que:

$$c \cdot U \cdot \frac{g(T)}{f(T)} = \sqrt{\frac{2e}{\pi}} U \sqrt{\frac{\pi}{2}} \exp\left(\frac{T^2}{2} - |T|\right) = U \cdot \exp\left(\frac{T^2}{2} - |T| + \frac{1}{2}\right),$$

aunque en general puede ser recomendable emplear $c \cdot U \cdot g(T) \leq f(T)$.

```
# densidad objetivo: dnorm
# densidad auxiliar: ddexp

# EJECUTAR CÓDIGO DEL APARTADO A DEL EJERCICIO 1
c.opt <- sqrt(2*exp(1)/pi)
lambda.opt <- 1
ngen <- 0

rnormAR <- function() {
```

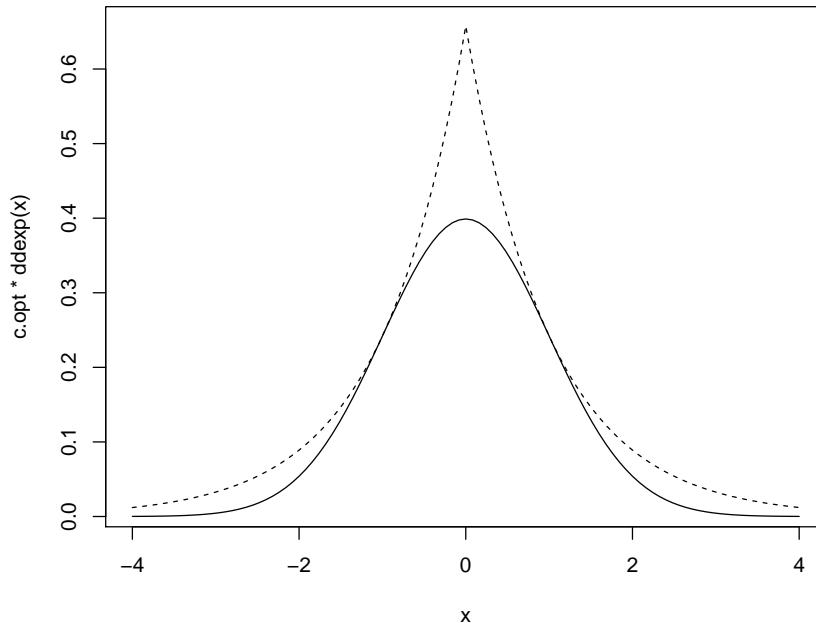
```

# Simulación por aceptación-rechazo
# Normal estandar a partir de doble exponencial
while (TRUE) {
  U <- runif(1)
  X <- rdexp(1) # lambda = 1
  ngen <- ngen+1 # Comentar esta linea para uso normal
  # if (U*exp((X^2+1)*0.5-abs(X)) <= 1) return(X)
  if (c.opt * U * ddexp(X, lambda.opt) <= dnorm(X)) return(X)
}
}

rnormARn <- function(n=1000) {
# Simulación n valores N(0,1)
  x <- numeric(n)
  for(i in 1:n) x[i]<-rnormAR()
  return(x)
}

# Grafico
curve(c.opt * ddexp(x), xlim = c(-4, 4), lty = 2)
curve(dnorm(x), add = TRUE)

```



- a) Generar una muestra de 10^4 observaciones empleando este algoritmo. Obtener el tiempo de CPU y calcular el número medio de generaciones de la distribución auxiliar.

```

set.seed(54321)
nsim <- 10^4

ngen <- 0
system.time(x <- rnormARn(nsim))

```

```

##      user    system elapsed
##      0.08     0.00    0.08

# N° generaciones
{
  cat("\nNº de generaciones = ", ngen)
  cat("\nNº medio de generaciones = ", ngen/nsim)
  cat("\nProporción de rechazos = ", 1-nsim/ngen, "\n")
}

##
## Nº de generaciones = 13163
## Nº medio de generaciones = 1.3163
## Proporción de rechazos = 0.2402948

```

- b) Representar el histograma y compararlo con la densidad teórica.

```

hist(x, breaks="FD", freq=FALSE)
curve(dnorm(x), add=TRUE)

```



5.2.3 Eficiencia del algoritmo

Como medida de la eficiencia del algoritmo de aceptación-rechazo podríamos considerar el número de iteraciones del algoritmo, es decir, el número de generaciones de la densidad auxiliar y de comparaciones para aceptar un valor de la densidad objetivo. Este número N es aleatorio y sigue una distribución geométrica (número de pruebas necesarias hasta obtener el primer éxito) con parámetro p (probabilidad de éxito) la probabilidad de aceptación en el paso 3:

$$p = \frac{\text{area}(A_f)}{\text{area}(A_{cg})} = \frac{1}{c}.$$

Por tanto:

$$E(N) = \frac{1}{p} = c$$

es el número medio de iteraciones del algoritmo (el número medio de pares de variables (T, U) que se necesitan generar, y de comparaciones, para obtener una simulación de la densidad objetivo).

Es obvio, por tanto, que cuanto más cercano a 1 sea el valor de c más eficiente será el algoritmo.

El principal problema con este método es encontrar una densidad auxiliar g de forma que:

$$c_{\text{opt}} = \max_{x/g(x)>0} \frac{f(x)}{g(x)}.$$

sea próximo a 1.

Ejercicio 5.3.

Continuando con el ejercicio anterior del método de aceptación-rechazo para generar observaciones de una distribución normal estándar, empleando como distribución auxiliar una doble exponencial:

- c) Aproximar la cota óptima numéricamente.

```
# Obtención de un valor c óptimo aproximado
optimize(f=function(x){dnorm(x)/ddexp(x)}, maximum=TRUE, interval=c(-1,1))
```

```
## $maximum
## [1] -0.999959
##
## $objective
## [1] 1.315489

# NOTA: Cuidado con los límites
# optimize(f=function(x){dnorm(x)/ddexp(x)}, maximum=TRUE, interval=c(0,2))

# Valor óptimo real
# sqrt(2*exp(1)/pi)
c.opt
```

```
## [1] 1.315489
```

- d) Aproximar el parámetro óptimo de la densidad auxiliar numéricamente (normalmente comenzaríamos por este paso).

```
# Obtención de valores c y lambda óptimos aproximados
fopt <- function(lambda) {
  # Obtiene c fijado lambda
  optimize(f = function(x){dnorm(x)/ddexp(x,lambda)},
            maximum=TRUE, interval=c(0,2))$objective
}

# Encontrar lambda que minimiza
res <- optimize(f=function(x){fopt(x)}, interval=c(0.5,2))
lambda.opt2 <- res$minimum
c.opt2 <- res$objective
```

5.2.4 Ejemplo

El algoritmo de Aceptación-Rechazo se emplea habitualmente en Inferencia Bayesiana:

- $f(x|\theta)$ densidad muestral.
- $\pi(\theta)$ densidad a priori.

- $\mathbf{x} = (x_1, \dots, x_n)'$ muestra observada.
- La distribución a posteriori de θ es:

$$\pi(\theta|\mathbf{x}) = \frac{L(\mathbf{x}|\theta)\pi(\theta)}{\int L(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

siendo $L(\mathbf{x}|\theta)$ la función de verosimilitud ($L(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$ suponiendo i.i.d.). Es decir:

$$\pi(\theta|\mathbf{x}) \propto L(\mathbf{x}|\theta)\pi(\theta).$$

Para simular valores de la densidad a posteriori $\pi(\theta|\mathbf{x})$ a partir de la densidad a priori $\pi(\theta)$

- $\pi(\theta|\mathbf{x})/\pi(\theta) \propto L(\mathbf{x}|\theta)$
- $L(\mathbf{x}|\theta) \leq c' = L(\mathbf{x}|\hat{\theta})$ siendo $\hat{\theta}$ el estimador MV de θ .

Algoritmo:

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Generar $\tilde{\theta} \sim \pi(\theta)$.
3. Si $L(\mathbf{x}|\hat{\theta}) \cdot U \leq L(\mathbf{x}|\tilde{\theta})$ devolver $\tilde{\theta}$,
en caso contrario volver al paso 1.

Ejercicio 5.4.

Para la estimación Bayes de la media de una normal se suele utilizar como distribución a priori una Cauchy.

- a) Generar una muestra i.i.d. $X_i \sim N(\theta_0, 1)$ de tamaño $n = 10$ con $\theta_0 = 1$. Utilizar una $Cauchy(0, 1)$ (`rcauchy`) como distribución a priori y como densidad auxiliar para simular por aceptación-rechazo una muestra de la densidad a posteriori (emplear `dnorm` para construir la verosimilitud). Obtener el intervalo de probabilidad al 95%.

```

mu0 <- 1
n <- 10
nsim <- 10^3
set.seed(54321)
x <- rnorm(n, mean = mu0)

# Función de verosimilitud
lik <- function(mu){prod(dnorm(x, mean = mu))}

# Cota óptima
# Estimación por máxima verosimilitud
env <- optimize(f = lik, int = range(x), maximum = TRUE)
env

## $maximum
## [1] 0.7353805
##
## $objective
## [1] 3.303574e-08
c <- env$objective

mean(x)

```

```

## [1] 0.7353958
c <- lik(mean(x))
c

## [1] 3.303574e-08

# Simulación distribución a posteriori
# Simulación por aceptación-rechazo a partir de Cauchy(0,1)
ngen <- nsim
Y <- rcauchy(nsim)
ind <- (c*runif(nsim) > sapply(Y, lik)) # TRUE si no verifica condición
# Volver a generar si no verifica condición
while (sum(ind)>0){
  le <- sum(ind)
  ngen <- ngen + le
  Y[ind] <- rcauchy(le)
  ind[ind] <- (c*runif(le) > sapply(Y[ind], lik)) # TRUE si no verifica condición
}

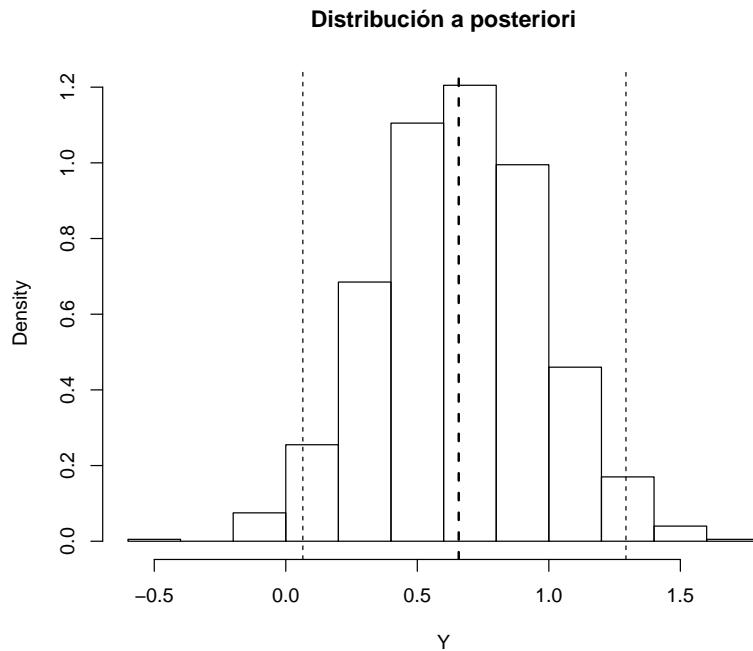
{ # N° generaciones
  cat("\nNº de generaciones = ", ngen)
  cat("\nNº medio de generaciones = ", ngen/nsim)
  cat("\nProporción de rechazos = ", 1-nsim/ngen, "\n")
}

## 
## Nº de generaciones = 5898
## Nº medio de generaciones = 5.898
## Proporción de rechazos = 0.830451

# Intervalo de probabilidad al 95% (IC Bayes)
q <- quantile(Y, c(0.025, 0.975))

# Representar estimador e IC Bayes
hist(Y, freq=FALSE, main="Distribución a posteriori")
# abline(v = mean(x), lty = 3) # Estimación frecuentista
abline(v = mean(Y), lty = 2, lwd = 2) # Estimación Bayesiana
abline(v = q, lty = 2)

```



- b) Repetir el apartado anterior con $n = 100$.

5.3 Modificaciones del método de AR

En el tiempo de computación influye:

- La proporción de aceptación (debería ser grande).
- La dificultad de simular con la densidad auxiliar.
- El tiempo necesario para hacer la comparación en el paso 4.

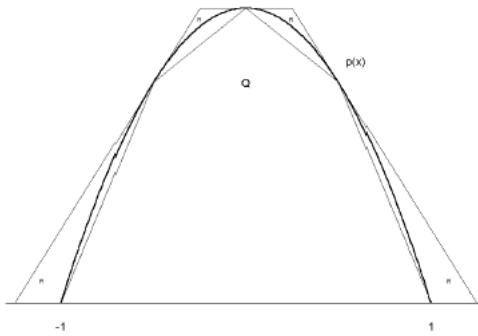
En ciertos casos el tiempo de computación necesario para evaluar $f(x)$ puede ser alto.

Para evitar evaluaciones de la densidad se puede emplear una función “squeeze” (Marsaglia, 1977) que aproxime la densidad por abajo:

$$s(x) \leq f(x).$$

Algoritmo:

1. Generar $U \sim \mathcal{U}(0, 1)$ y $T \sim g$.
2. Si $c \cdot U \cdot g(T) \leq s(T)$ devolver $X = T$,
en caso contrario
2.a. si $c \cdot U \cdot g(T) \leq f(T)$ devolver $X = T$,
2.b. en caso contrario volver al paso 1.



Cuanto mayor sea el área bajo $s(x)$ (más próxima a 1) más efectivo será el algoritmo.

Se han desarrollado métodos generales para la construcción de las funciones g y s de forma automática (cada vez que se evalúa la densidad se mejoran las aproximaciones). Estos métodos se basan principalmente en que una transformación de la densidad objetivo es cóncava o convexa.

5.3.1 Muestreo por rechazo adaptativo (ARS)

Supongamos que f es una cuasi-densidad log-cóncava (i.e. $\frac{\partial^2}{\partial x^2} \log f(x) < 0, \forall x$).

Sea $S_n = \{x_i : i = 0, \dots, n+1\}$ con $f(x_i)$ conocidos.

Denotamos por $L_{i,i+1}(x)$ la recta pasando por $(x_i, \log f(x_i))$ y $(x_{i+1}, \log f(x_{i+1}))$

- $L_{i,i+1}(x) \leq \log f(x)$ en el intervalo $I_i = (x_i, x_{i+1}]$
- $L_{i,i+1}(x) \geq \log f(x)$ fuera de I_i

En el intervalo I_i se definen las envolventes de $\log f(x)$:

- $\underline{\phi}_n(x) = L_{i,i+1}(x)$
- $\bar{\phi}_n(x) = \min\{L_{i-1,i}(x), L_{i+1,i+2}(x)\}$

Las envolventes de $f(x)$ en I_i serán:

- $s_n(x) = \exp(\underline{\phi}_n(x))$
- $G_n(x) = \exp(\bar{\phi}_n(x))$

Tenemos entonces que:

$$s_n(x) \leq f(x) \leq G_n(x) = c \cdot g_n(x)$$

donde $g_n(x)$ es una mixtura discreta de distribuciones tipo exponencial truncadas (las tasas pueden ser negativas), que se puede simular fácilmente mediante el método de inversión.

Algoritmo (Gilks, 1992):

1. Inicializar n y s_n .
2. Generar $U \sim \mathcal{U}(0, 1)$ y $T \sim g_n$.
3. Si $U \cdot G_n(T) \leq s_n(T)$ devolver $X = T$,
en caso contrario,
 - 3.a Si $U \cdot G_n(T) \leq f(T)$ devolver $X = T$.
 - 3.b Hacer $n = n + 1$, añadir T a S_n y actualizar s_n y G_n .
4. Volver al paso 2.

Gilks y Wild (1992) propusieron un método similar pero empleando tangentes para construir la cota superior.

La mayoría de las densidades de la familia de distribuciones exponencial son log-cónicas. Hörmann (1995) extendió esta aproximación al caso de densidades T_c -cónicas:

$$T_c(x) = \text{signo}(c)x^c T_0(x) = \log(x).$$

A parte de la transformación logarítmica, la transformación $T_{-1/2}(x) = -1/\sqrt{x}$ es habitualmente la más empleada.

5.3.2 Método del cociente de uniformes

Se puede ver como una modificación del método de aceptación-rechazo, de especial interés cuando el soporte no es acotado.

Si (U, V) se distribuye uniformemente sobre:

$$C_h = \left\{ (u, v) \in \mathbb{R}^2 : 0 < u \leq \sqrt{h(v/u)} \right\},$$

siendo h una función no negativa integrable (cuasi-densidad), entonces $X = V/U$ tiene función de densidad proporcional a h (Kinderman y Monahan, 1977). Además C_h tiene área finita.

De modo análogo al método de aceptación-rechazo, hay modificaciones para acelerar los cálculos y automatizar el proceso, construyendo regiones mediante polígonos:

$$C_i \subset C_h \subset C_s.$$

5.4 Método de composición

En ocasiones la densidad de interés se puede expresar como una mixtura discreta de densidades:

$$f(x) = \sum_{j=1}^k p_j f_j(x)$$

con $\sum_{j=1}^k p_j = 1$, $p_j \geq 0$ y f_j densidades (sería también válido para funciones de distribución o variables aleatorias).

Algoritmo:

1. Generar J con distribución $P(J = j) = p_j$.
2. Generar $X \sim f_J$.

En el caso continuo tendríamos:

$$f(x) = \int g(x|y)h(y)dy.$$

Algoritmo:

1. Generar $Y \sim h$.
2. Generar $X \sim g(\cdot|Y)$.

5.4.1 Ejemplo

Simulación de una densidad doble exponencial:

$$f(x) = \frac{\lambda}{2}e^{-\lambda|x|}, \forall x \in \mathbb{R}.$$

Se tiene que:

$$f(x) = \frac{1}{2}f_1(x) + \frac{1}{2}f_2(x)$$

siendo:

$$f_1(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad f_2(x) = \begin{cases} \lambda e^{\lambda x} & \text{si } x < 0 \\ 0 & \text{si } x \geq 0 \end{cases}$$

Algoritmo:

1. Generar $U, V \sim \mathcal{U}(0, 1)$.
2. Si $U < 0.5$ devolver $X = -\ln(1 - V)$.
3. En caso contrario devolver $X = \ln V$.

Observaciones:

- En ocasiones se hace un reciclado de los números aleatorios (solo se genera una uniforme, e.g. $V = 2(U - 0.5)$ si $U \in (0.5, 1)$).
- En ciertas ocasiones por comodidad, para simular una muestra de tamaño n , se simulan muestras de tamaño np_i con densidad f_i y se combinan aleatoriamente.

5.5 Métodos específicos para generación de distribuciones notables

Ver ejemplos en el libro de Ricardo y ayuda de las funciones implementadas en R

Capítulo 6

Simulación de variables discretas

Se trata de simular una variable aleatoria discreta X con función de masa de probabilidad (f.m.p.):

x_i		x_1	x_2	\cdots	x_n	\cdots
$P(X = x_i)$		p_1	p_2	\cdots	p_n	\cdots

Considerando como partida una $\mathcal{U}(0, 1)$, la idea general consiste en asociar a cada posible valor x_i de X un subintervalo de $(0, 1)$ de longitud igual a la correspondiente probabilidad. Por ejemplo, como ya se mostró en capítulos anteriores, es habitual emplear código de la forma:

```
x <- runif(nsim) < p
```

para simular una distribución $Bernoulli(p)$.

Para generar variables discretas con dominio finito en R, si no se dispone de un algoritmo específico más eficiente, es recomendable emplear:

```
sample(valores, nsim, replace = TRUE, prob)
```

Esta función del paquete base implementa eficientemente el método “alias” que describiremos más adelante en la sección 6.3.

6.1 Método de la transformación cuantil

Este método es una adaptación del método de inversión (válido para el caso continuo) a distribuciones discretas. En este caso, la función de distribución es:

$$F(x) = \sum_{x_j \leq x} p_j,$$

y la distribución de la variable aleatoria $F(X)$ no es uniforme (es una variable aleatoria discreta que toma los valores $F(x_i)$ con probabilidad p_i , $i = 1, 2, \dots$). Sin embargo, se puede generalizar el método de inversión a situaciones en las que F no es invertible considerando la función cuantil.

Se define la función cuantil o inversa generalizada de una función de distribución F como:

$$Q(u) = \inf \{x \in \mathbb{R} : F(x) \geq u\}, \quad \forall u \in (0, 1).$$

Si F es invertible $Q = F^{-1}$.

Teorema 6.1 (de inversión generalizada). *Si $U \sim \mathcal{U}(0,1)$, la variable aleatoria $Q(U)$ tiene función de distribución F .*

Prueba. Bastaría ver que:

$$Q(u) \leq x \iff u \leq F(x).$$

Como F es monótona y por la definición de Q :

$$Q(u) \leq x \Rightarrow u \leq F(Q(u)) \leq F(x).$$

Por otro lado como Q también es monótona:

$$u \leq F(x) \Rightarrow Q(u) \leq Q(F(x)) \leq x$$

□

A partir de este resultado se deduce el siguiente algoritmo general para simular una distribución de probabilidad discreta.

Algoritmo 6.1 (de transformación cuantil). 1. Generar $U \sim \mathcal{U}(0,1)$.

2. Devolver $X = Q(U)$.

El principal problema es el cálculo de $Q(U)$. En este caso, suponiendo por comodidad que los valores que toma la variable están ordenados ($x_1 < x_2 < \dots$), la función cuantil será:

$$\begin{aligned} Q(U) &= \inf \left\{ x_j : \sum_{i=1}^j p_i \geq U \right\} \\ &= x_k, \text{ tal que } \sum_{i=1}^{k-1} p_i < U \leq \sum_{i=1}^k p_i \end{aligned}$$

Para encontrar este valor se puede emplear el siguiente algoritmo:

Algoritmo 6.2 (de transformación cuantil con búsqueda secuencial). 1. Generar $U \sim \mathcal{U}(0,1)$.

2. Hacer $I = 1$ y $S = p_1$.

3. Mientras $U > S$ hacer $I = I + 1$ y $S = S + p_I$

4. Devolver $X = x_I$.

Este algoritmo no es muy eficiente, especialmente si el número de posibles valores de la variable es grande.

Nota: . El algoritmo anterior es válido independientemente de que los valores que tome la variable estén ordenados.

Si la variable toma un número finito de valores, se podría implementar en R de la siguiente forma:

```
rfmp0 <- function(x, prob = 1/length(x), nsim = 1000) {
  # Simulación nsim v.a. discreta a partir de fmp
  # por inversión generalizada (transformación cuantil)
  X <- numeric(nsim)
  U <- runif(nsim)
  for(j in 1:nsim) {
    i <- 1
    Fx <- prob[1]
    while (Fx < U[j]) {
      i <- i + 1
    }
    X[j] <- i
  }
}
```

```

    Fx <- Fx + prob[i]
  }
  X[j] <- x[i]
}
return(X)
}

```

Adicionalmente, para disminuir el tiempo de computación, se puede almacenar las probabilidades acumuladas en una tabla. Si también se quiere obtener el número de comparaciones se puede considerar una variable global `ncomp`:

```

ncomp <- 0

rfmp <- function(x, prob = 1/length(x), nsim = 1000) {
  # Simulación nsim v.a. discreta a partir de fmp
  # por inversión generalizada (transformación cuantil)
  # Inicializar FD
  Fx <- cumsum(prob)
  # Simular
  X <- numeric(nsim)
  U <- runif(nsim)
  for(j in 1:nsim) {
    i <- 1
    while (Fx[i] < U[j]) i <- i + 1
    X[j] <- x[i]
    ncomp <- ncomp + i
  }
  return(X)
}

```

Ejercicio 6.1.

Generar, por el método de la transformación cuantil usando búsqueda secuencial, una muestra de $nsim = 10^5$ observaciones de una variable $\mathcal{B}(10, 0.5)$. Obtener el tiempo de CPU empleado. Aproximar por simulación la función de masa de probabilidad, representarla gráficamente y compararla con la teórica. Calcular también la media muestral (compararla con la teórica np) y el número medio de comparaciones para generar cada observación.

Empleamos la rutina anterior para generar las simulaciones:

```

set.seed(54321)
n <- 10
p <- 0.5
nsim <- 10^5

x <- 0:n
fmp <- dbinom(x, n, p)

ncomp <- 0
system.time( rx <- rfmp(x, fmp, nsim) )

##   user  system elapsed
## 0.08    0.00    0.08

```

Aproximación de la media:

```
mean(rx)
## [1] 5.00322
```

El valor teórico es $n \cdot p = 5$.

Número medio de comparaciones:

```
ncomp/nsim
## [1] 6.00322
```

*# Se verá más adelante que el valor teórico es $\sum(1:length(x)) * fmp$*

Análisis de los resultados:

```
res <- table(rx)/nsim
plot(res, ylab = "frecuencia relativa", xlab = "valores")
points(x, fmp, pch = 4) # Comparación teórica
```

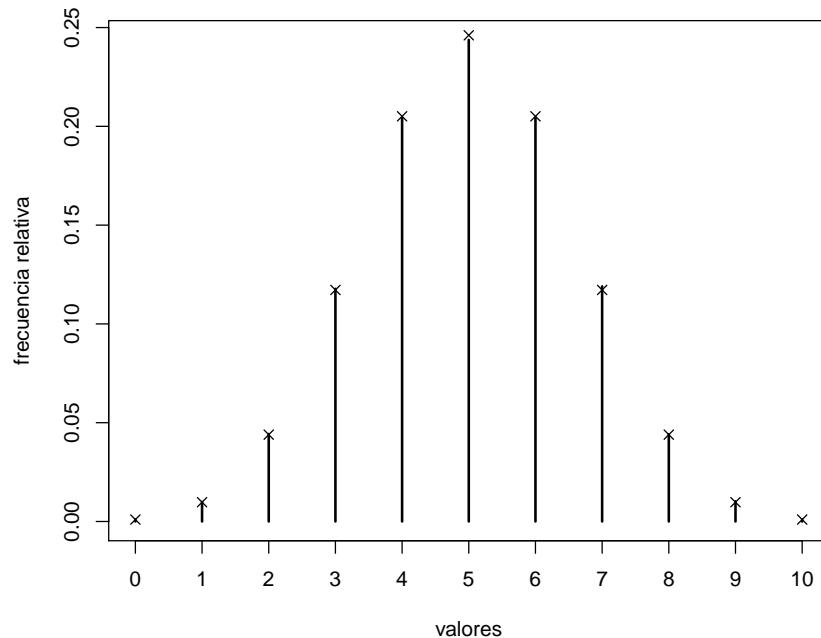


Figura 6.1: Comparación de las frecuencias relativas de los valores generados con las probabilidades teóricas.

```
res <- as.data.frame(res)
names(res) <- c("x", "psim")
res$pteor <- fmp
res

##      x     psim      pteor
## 1    0 0.00107 0.0009765625
## 2    1 0.00990 0.0097656250
## 3    2 0.04432 0.0439453125
## 4    3 0.11778 0.1171875000
## 5    4 0.20425 0.2050781250
```

```

## 6 5 0.24375 0.2460937500
## 7 6 0.20454 0.2050781250
## 8 7 0.11898 0.1171875000
## 9 8 0.04419 0.0439453125
## 10 9 0.01023 0.0097656250
## 11 10 0.00099 0.0009765625

# Errores
max(abs(res$psim - res$pteor))

## [1] 0.00234375

max(abs(res$psim - res$pteor) / res$pteor)

## [1] 0.09568

```

Nota: . Puede ocurrir que no todos los valores sean generados en la simulación. En el código anterior si `length(x) > length(psim)`, la sentencia `res$pteor <- fmp` gerará un error. Alternativamente se podría emplear por ejemplo:

```

res <- data.frame(x = x, pteor = fmp, psim = 0)
res.sim <- table(rx)/nsim
index <- match(names(res.sim), x)
res$psim[index] <- res.sim

```

6.1.1 Eficiencia del algoritmo

Si consideramos la variable aleatoria \mathcal{I} correspondiente a las etiquetas, su función de masa de probabilidad sería:

$$\frac{i}{P(\mathcal{I} = i)} \begin{array}{c|ccccc} & 1 & 2 & \cdots & n & \cdots \\ \hline P(\mathcal{I} = i) & p_1 & p_2 & \cdots & p_n & \cdots \end{array}$$

y el número de comparaciones en el paso 3 sería un valor aleatorio de esta variable. Una medida de la eficiencia del algoritmo de la transformación cuantil es el número medio de comparaciones:

$$E(\mathcal{I}) = \sum_i ip_i.$$

Realmente, cuando la variable toma un número finito de valores: x_1, x_2, \dots, x_n , no sería necesario hacer la última comprobación $U > \sum_{i=1}^n p_i = 1$ y se generaría directamente x_n , por lo que el número medio de comparaciones sería:

$$\sum_{i=1}^{n-1} ip_i + (n-1)p_n.$$

Para disminuir el número esperado de comparaciones podemos reordenar los valores x_i de forma que las probabilidades correspondientes sean decrecientes. Esto equivale a considerar un etiquetado l de forma que:

$$p_{l(1)} \geq p_{l(2)} \geq \cdots \geq p_{l(n)} \geq \cdots$$

Ejercicio 6.2.

Repetir el ejercicio anterior ordenando previamente las probabilidades en orden decreciente y también empleando la función `sample` de R.

```

tini <- proc.time()

ncomp <- 0

```

```

ind <- order(fmp, decreasing=TRUE)
rx <- rfmp(x[ind], fmp[ind], nsim)

tiempo <- proc.time() - tini
tiempo

##      user    system elapsed
##      0.04    0.02    0.04

# Número de comparaciones
ncomp/nsim

## [1] 3.08969
sum((1:length(x))*fmp[ind]) # Valor teórico

## [1] 3.083984

```

Como ya se comentó, en R se recomienda emplear la función `sample` (implementa eficientemente el método de Alias):

```

system.time( rx <- sample(x, nsim, replace = TRUE, prob = fmp) )

##      user    system elapsed
##          0        0        0

```

6.2 Método de la tabla guía

La idea consiste en construir m subintervalos equiespaciados contenidos en $[0, 1]$ de la forma

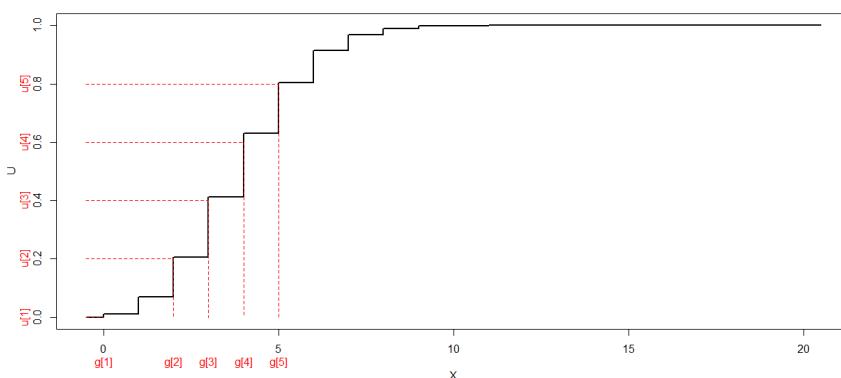
$$I_j = [u_j, u_{j+1}) = \left[\frac{j-1}{m}, \frac{j}{m} \right) \text{ para } j = 1, 2, \dots, m$$

y utilizarlos como punto de partida para la búsqueda. En una tabla guía se almacenan los índices de los cuantiles correspondientes a los extremos inferiores de los intervalos:

$$g_j = Q_{\mathcal{I}}(u_j) = \inf \left\{ i : F_i \geq u_j = \frac{j-1}{m} \right\}$$

El punto de partida para un valor U será g_{j_0} siendo:

$$j_0 = \lfloor mU \rfloor + 1$$



En este caso, puede verse que una cota del número medio de comparaciones es:

$$E(N) \leq 1 + \frac{n}{m}$$

Algoritmo 6.3 (de simulación mediante una tabla guía). *Inicialización:*

1. Hacer $F_1 = p_1$.
2. Desde $i = 2$ hasta n hacer $F_i = F_{i-1} + p_i$.

Cálculo de la tabla guía:

1. Hacer $g_1 = 1$ e $i = 1$.
2. Desde $j = 2$ hasta m hacer
 - 2.a Mientras $(j - 1)/m > F_i$ hacer $i = i + 1$.
 - 2.b $g_j = i$

Simulación mediante tabla guía:

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Hacer $j = \lfloor mU \rfloor + 1$.
3. Hacer $i = g_j$.
4. Mientras $U > F_i$ hacer $i = i + 1$.
5. Devolver $X = x_i$.

Ejercicio 6.3.

Diseñar una rutina que permita generar $nsim$ valores de una distribución discreta usando una tabla guía. Repetir el ejercicio anterior empleando esta rutina con $m = n - 1$.

```
rfmp.tabla <- function(x, prob = 1/length(x), m, nsim = 1000) {
  # Simulación v.a. discreta a partir de función de masa de probabilidad
  # por tabla guía de tamaño m
  # Inicializar tabla y FD
  Fx <- cumsum(prob)
  g <- rep(1,m)
  i <- 1
  for(j in 2:m) {
    while (Fx[i] < (j-1)/m) i <- i+1
    g[j] <- i
  }
  # Generar valores
  X <- numeric(nsim)
  U <- runif(nsim)
  for(j in 1:nsim) {
    i <- g[floor(U[j]*m)+1]
    while (Fx[i] < U[j]) i <- i + 1
    X[j] <- x[i]
  }
  return(X)
}

system.time( rx <- rfmp.tabla(x, fmp, n-1, nsim) )
```

```
##    user  system elapsed
##    0.07    0.00    0.06
```

Análisis de los resultados:

```
res <- table(rx)/nsim
plot(res, ylab = "frecuencia relativa", xlab = "valores")
points(x, fmp, pch = 4) # Comparación teórica
```

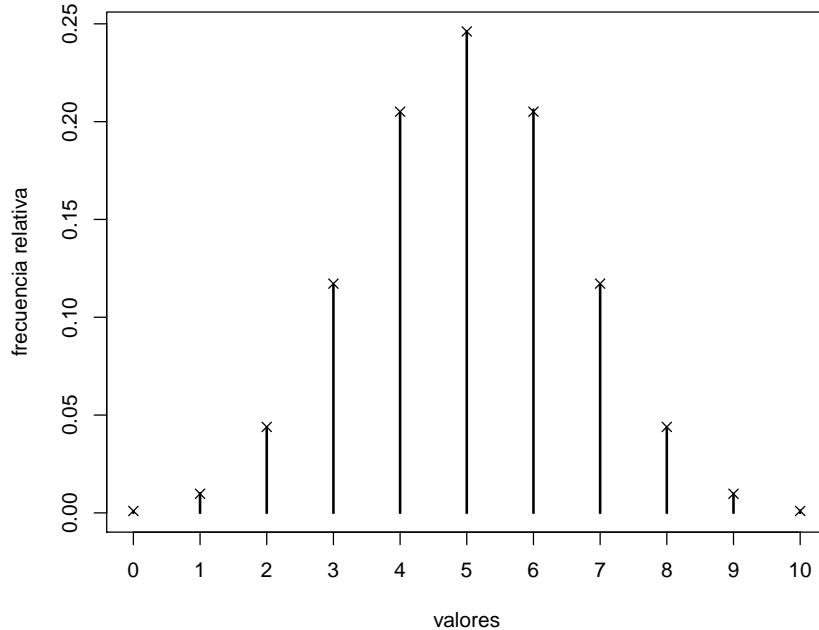


Figura 6.2: Comparación de las frecuencias relativas de los valores generados con las probabilidades teóricas.

6.3 Método de Alias

Se basa en representar la distribución de X como una mixtura (uniforme) de variables dicotómicas (Walker, 1977):

$$Q^{(i)} = \begin{cases} x_i & \text{con prob. } q_i \\ x_{a_i} & \text{con prob. } 1 - q_i \end{cases}$$

La idea para construir estas variables es “tomar prestada” parte de la probabilidad de los valores más probables (ricos) para asignársela a los valores menos probables (pobres), almacenando en a_i el índice del valor de donde procede. El algoritmo “Robin Hood” de inicialización (Kronmal y Peterson, 1979) es el siguiente:

1. Desde $i = 1$ hasta n hacer $q_i = np_i$.
2. Establecer $L = \{l : q_l < 1\}$ y $H = \{h : q_h \geq 1\}$.
3. Si L ó H vacíos terminar.
4. Seleccionar $l \in L$ y $h \in H$.

5. Hacer $a_l = h$.
6. Eliminar l de L .
7. Hacer $q_h = q_h - (1 - q_l)$.
8. Si $q_h < 1$ mover h de H a L .
9. Ir al paso 3.

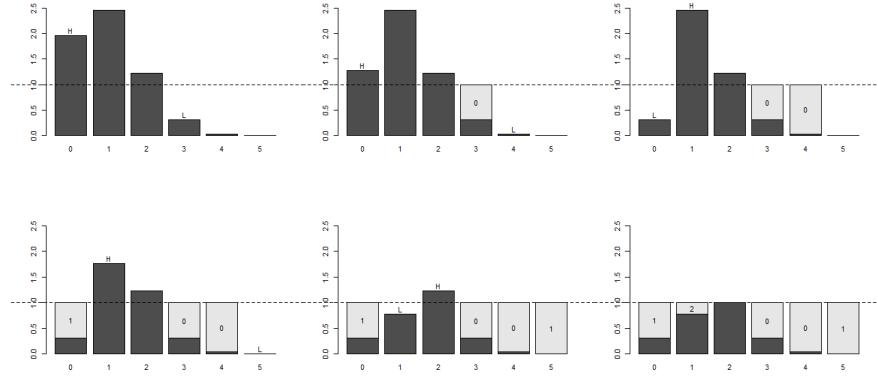


Figura 6.3: Pasos del algoritmo de inicialización del método Alias.

El algoritmo para generar las simulaciones es el estándar del método de composición:

1. Generar $U, V \sim \mathcal{U}(0, 1)$.
2. Hacer $i = \lfloor nU \rfloor + 1$.
3. Si $V < q_i$ devolver $X = x_i$.
4. En caso contrario devolver $X = x_{a_i}$.

Este algoritmo es muy eficiente y es el implementado en la función `sample` de R.

Ejercicio 6.4.

Diseñar una rutina que permita generar $nsim$ valores de una distribución discreta usando el método de Alias. Repetir el ejercicio anterior empleando esta rutina.

```
rfmp.alias <- function(x, prob = 1/length(x), nsim = 1000) {
  # Inicializar tablas
  a <- numeric(length(x))
  q <- prob*length(x)
  low <- q < 1
  high <- which(!low)
  low <- which(low)
  while (length(high) && length(low)) {
    l <- low[1]
    h <- high[1]
    a[l] <- h
    q[h] <- q[h] - (1 - q[l])
    if (q[h] < 1) {
      high <- high[-1]
      low[1] <- h
    } else low <- low[-1]
```

```

} # while
# Generar valores
V <- runif(nsim)
i <- floor(runif(nsim)*length(x)) + 1
return( x[ ifelse( V < q[i] , i, a[i]) ] )
}

system.time( rx <- rfmp.alias(x,fmp,nsim) )

```

```

##      user    system elapsed
##  0.03     0.00    0.03

```

Análisis de los resultados:

```

res <- table(rx)/nsim
plot(res, ylab = "frecuencia relativa", xlab = "valores")
points(x, fmp, pch = 4)  # Comparación teórica

```

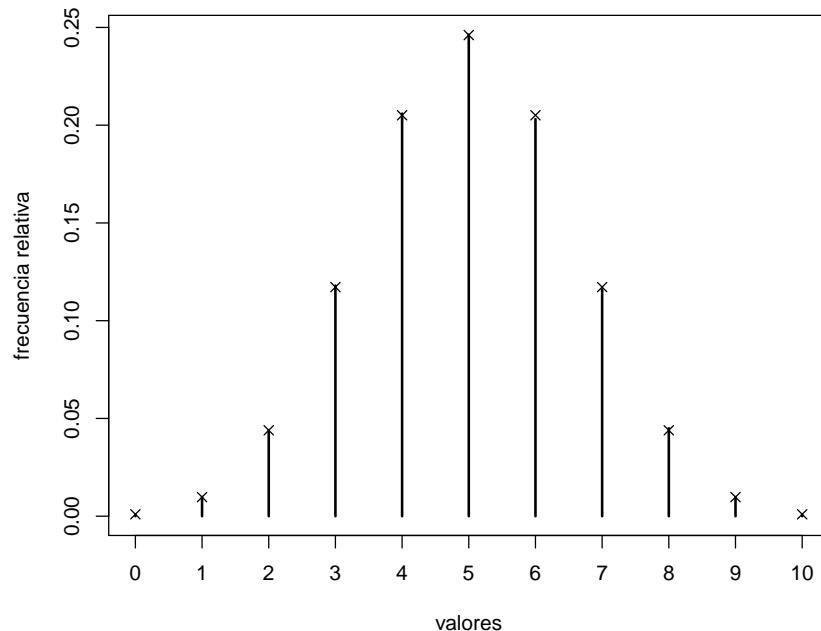


Figura 6.4: Comparación de las frecuencias relativas de los valores generados con las probabilidades teóricas.

6.4 Simulación de una variable discreta con dominio infinito

Los métodos anteriores están pensados para variables que toman un número finito de valores. Si la variable discreta tiene dominio infinito no se podrían almacenar las probabilidades acumuladas, aunque en algunos casos podrían calcularse de forma recursiva.

Ejemplo 6.1 (distribución de Poisson).

Una variable X con distribución de Poisson de parámetro λ , toma los valores $x_1 = 0, x_2 = 1, \dots$ con probabilidades:

$$p_j = P(X = x_j) = P(X = j - 1) = \frac{e^{-\lambda} \lambda^{j-1}}{(j-1)!}, \quad j = 1, 2, \dots$$

En este caso, como:

$$P(X = j) = \frac{e^{-\lambda} \lambda^j}{j!} = \frac{\lambda}{j} \frac{e^{-\lambda} \lambda^{j-1}}{(j-1)!} = \frac{\lambda}{j} P(X = j - 1),$$

el algoritmo de inversión con búsqueda secuencial sería:

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Hacer $I = 0, p = e^{-\lambda}$ y $S = p$.
3. Mientras $U > S$ hacer $I = I + 1, p = \frac{\lambda}{I} p$ y $S = S + p$.
4. Devolver $X = I$.

Hay modificaciones de los algoritmos anteriores, e.g. incluyendo una búsqueda secuencial en la cola de la distribución, para estos casos.

Como alternativa, siempre se puede pensar en truncar la distribución, eliminando los valores muy poco probables (teniendo en cuenta el número de generaciones que se pretenden realizar), de esta forma la distribución de las simulaciones sería aproximada.

6.5 Cálculo directo de la función cuantil

En ocasiones el método de la transformación cuantil puede acelerarse computacionalmente porque, mediante cálculos directos, es posible encontrar el valor de la función cuantil en cualquier U , evitando el bucle de búsqueda. Normalmente se realiza mediante truncamiento de una distribución continua.

Ejemplo 6.2 (distribución uniforme discreta).

La función de masa de probabilidad de una distribución uniforme discreta en $\{1, 2, \dots, n\}$ viene dada por

$$p_j = \frac{1}{n}, \text{ para } j = 1, 2, \dots, n.$$

Pueden generarse valores de esta distribución de forma muy eficiente truncando la distribución uniforme:

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Devolver $X = \lfloor nU \rfloor + 1$.

Ejemplo 6.3 (distribución geométrica).

La función de masa de probabilidad de una distribución geométrica es:

$$P(X = j) = P(I = j + 1) = p(1 - p)^j, \quad j = 0, 1, \dots$$

Si se considera como variable aleatoria continua auxiliar una exponencial, con función de distribución $G(x) = 1 - e^{-\lambda x}$ si $x \geq 0$, se tiene que:

$$\begin{aligned} G(i) - G(i-1) &= 1 - e^{-\lambda i} - (1 - e^{-\lambda(i-1)}) = e^{-\lambda(i-1)} - e^{-\lambda i} \\ &= e^{-\lambda(i-1)} (1 - e^{-\lambda}) = (1 - e^{-\lambda}) (e^{-\lambda})^{i-1} \\ &= p(1 - p)^{i-1}, \end{aligned}$$

tomando $p = 1 - e^{-\lambda}$. De donde se obtendría el algoritmo:

0. Hacer $\lambda = -\ln(1-p)$.

1. Generar $U \sim \mathcal{U}(0,1)$.

2. Hacer $T = -\frac{\ln U}{\lambda}$.

3. Devolver $X = \lfloor T \rfloor$.

6.6 Otros métodos

- Aceptación-Rechazo: Este método también sería directamente aplicable al caso discreto. En principio habría que considerar una variable auxiliar discreta con el mismo soporte, pero también hay modificaciones para variables auxiliares continuas.
- Método de composición: este es otro método directamente aplicable y que se emplea en el método de Alias
- Hay otros métodos que tratan de reducir el número medio de comparaciones de la búsqueda secuencial, por ejemplo los árboles (binarios) de Huffman (e.g. Cao, 2002, Sección 4.2).

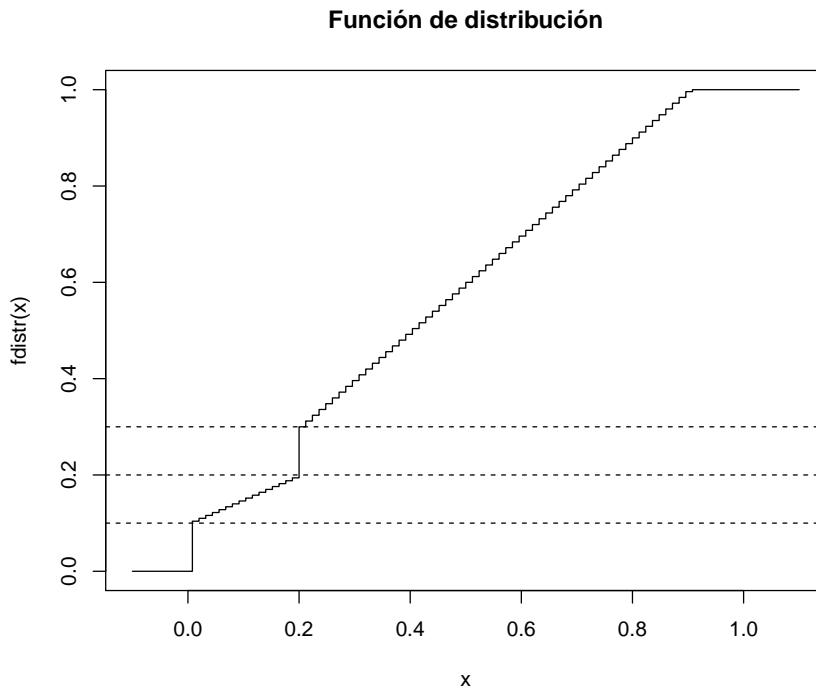
Ejercicio 6.5.

Considera la variable aleatoria con función de distribución dada por:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{2} + \frac{1}{10} & \text{si } x \in [0, \frac{1}{5}) \\ x + \frac{1}{10} & \text{si } x \in [\frac{1}{5}, \frac{9}{10}] \\ 1 & \text{en otro caso} \end{cases}$$

Función de distribución:

```
fdistr <- function(x) {
  ifelse(x < 0, 0,
    ifelse(x < 1/5, x/2 + 1/10,
      ifelse(x <= 9/10, x + 1/10, 1) )
  }
# Empleando ifelse se complica un poco más pero el resultado es una función vectorial.
curve(fdistr(x), from = -0.1, to = 1.1, type = 's',
      main = 'Función de distribución')
# Discontinuidades en 0 y 1/5
abline(h = c(1/10, 2/10, 3/10), lty = 2)
```



Nota: Esta variable toma los valores 0 y 1/5 con probabilidad 1/10.

- a) Diseña un algoritmo basándose en el método de inversión generalizado para generar observaciones de esta variable.

El algoritmo general es siempre el mismo. Empleando la función cuantil:

$$Q(u) = \inf \{x \in \mathbb{R} : F(x) \geq u\},$$

el algoritmo sería:

1. Generar $U \sim \mathcal{U}(0, 1)$
2. Devolver $X = Q(U)$

En este caso concreto:

1. Generar $U \sim \mathcal{U}(0, 1)$
2. Si $U < \frac{1}{10}$ devolver $X = 0$
3. Si $U < \frac{2}{10}$ devolver $X = 2(U - \frac{1}{10})$
4. Si $U < \frac{3}{10}$ devolver $X = \frac{2}{10}$
5. En caso contrario devolver $X = U - \frac{1}{10}$

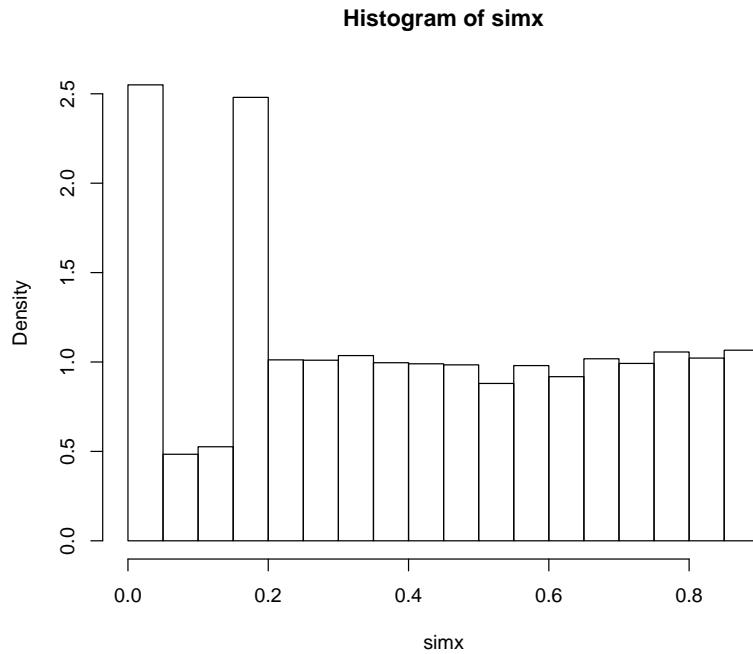
- b) Implementa el algoritmo en una función que permita generar $nsim$ valores de la variable.

```
# Función cuantil:
fquant <- function(u) {
  ifelse(u < 1/10, 0,
    ifelse(u < 2/10, 2*(u - 1/10),
      ifelse(u < 3/10, 1/5, u - 1/10) )
  }
# Función para generar nsim valores:
rx <- function(nsim) fquant(runif(nsim))
```

Ejemplo:

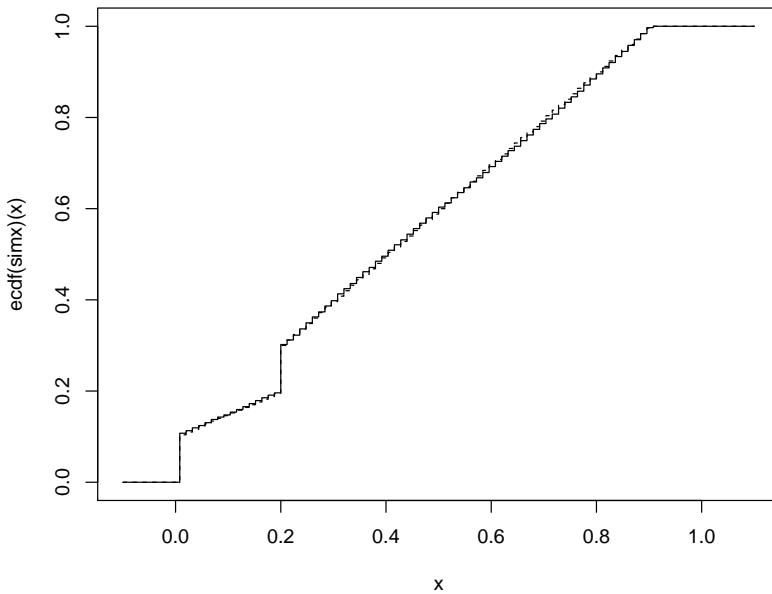
```
set.seed(1)
nsim <- 10^4
system.time(simx <- rx(nsim))

##      user    system   elapsed
##        0         0         0
hist(simx, breaks = "FD", freq = FALSE)
```



En este caso como no es una variable absolutamente continua mejor emplear la función de distribución para compararla con la teórica:

```
curve(ecdf(simx)(x), from = -0.1, to = 1.1, type = "s")
curve(fdistr(x), type = "s", lty = 2, add = TRUE)
```



Ejercicio 6.6. (propuesto)

Se pretende simular $nsim = 10^4$ observaciones de una variable hipergeométrica (`dhyper(x, m, n, k)`) de parámetros m = el nº de grupo multiplicado por 10, $n = 100 - m$ y $k = 20$.

- Comprobar que el rango de posibles valores de esta variable es `max(0, k-n):min(m, k)`. Generar los valores empleando el método de la transformación cuantil usando búsqueda secuencial. Obtener el tiempo de CPU empleado. Aproximar por simulación la función de masa de probabilidad, representarla gráficamente y compararla con la teórica. Calcular también la media muestral (compararla con la teórica $km/(m + n)$) y el número medio de comparaciones para generar cada observación.
- Repetir el apartado anterior ordenando previamente las probabilidades en orden decreciente y también: empleando la función `sample` de R, mediante una tabla guía (con $k-1$ subintervalos) y usando el método de Alias.

6.7 Métodos específicos para generación de distribuciones notables

Ver ejemplos del libro de Ricardo y de las funciones implementadas en R.

Capítulo 7

Simulación de Distribuciones Multidimensionales

7.1 Introducción

Supongamos que se pretende simular un vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_d)$ con función de densidad conjunta $f(x_1, x_2, \dots, x_d)$ (análogo para una función de densidad discreta con función de masa de probabilidad).

Si las componentes son independientes y f_i son las correspondientes densidades marginales, bastará con generar $X_i \sim f_i$. Las dificultades aparecen cuando se quieren simular componentes con una determinada estructura de dependencia.

Ejemplo 7.1 (simulación de normales independientes).

Si:

$$\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$$

con $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)^t$ y $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_d^2)$, entonces $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$.

Algoritmo 7.1 (de simulación de normales independientes). 1. Simular $Z_1, Z_2, \dots, Z_d \sim \mathcal{N}(0, 1)$ independientes.

2. Para $i = 1, 2, \dots, d$ hacer $X_i = \mu_i + \sigma_i Z_i$.

Las funciones implementadas en el paquete base de R permiten simular fácilmente en el caso independiente. Por ejemplo en el caso bidimensional con $X_1 \sim \mathcal{N}(0, 1)$ y $X_2 \sim \mathcal{N}(-1, 0.5^2)$:

```
f1 <- function(x) dnorm(x)
f2 <- function(x) dnorm(x, -1, 0.5)
curve(f1, -3, 3, ylim = c(0, f2(-1)), ylab = "f_i(x)")
curve(f2, add = TRUE, lty = 2)
```

Para simular una generación bastaría con:

```
set.seed(1)
rnorm(2, c(0, -1), c(1, 0.5))
## [1] -0.6264538 -0.9081783
```

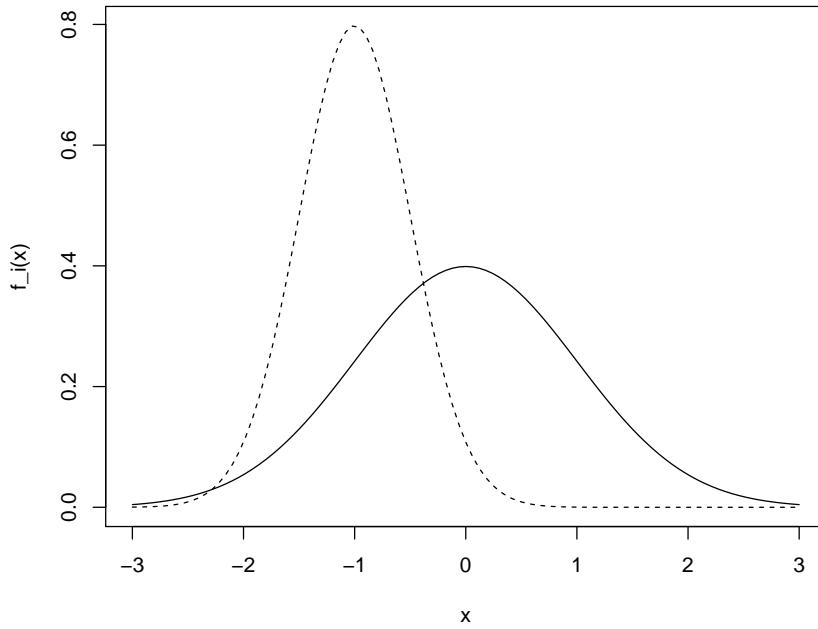


Figura 7.1: Densidades marginales de las componentes.

y para simular `nsim`:

```
set.seed(1)
nsim <- 5
rx <- matrix(rnorm(2*nsim, c(0, -1), c(1, 0.5)), nrow = nsim, byrow = TRUE)
colnames(rx) <- paste0("X", 1:ncol(rx))
rx

##           X1          X2
## [1,] -0.6264538 -0.9081783
## [2,] -0.8356286 -0.2023596
## [3,]  0.3295078 -1.4102342
## [4,]  0.4874291 -0.6308376
## [5,]  0.5757814 -1.1526942
```

7.2 El método de aceptación/rechazo

Muchos de los algoritmos descritos en los temas anteriores son válidos para el caso multivariante (aceptación-rechazo, cociente de uniformes, composición, ...). En este caso sin embargo, puede ser mucho más difícil cumplir los requerimientos (e.g. encontrar una densidad auxiliar adecuada) y los algoritmos obtenidos pueden ser computacionalmente poco eficientes (especialmente si el número de dimensiones es grande).

El algoritmo de aceptación-rechazo sería exactamente el mismo, lo único que cambiaría serían que las densidades son multidimensionales. Supongamos que la densidad objetivo f y la densidad auxiliar g verifican:

$$f(x_1, x_2, \dots, x_d) \leq c \cdot g(x_1, x_2, \dots, x_d), \forall \mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d.$$

para una constante $c > 0$. El algoritmo sería:

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Generar $\mathbf{T} = (T_1, T_2, \dots, T_d) \sim g$.
3. Si $c \cdot U \cdot g(T_1, T_2, \dots, T_d) \leq f(T_1, T_2, \dots, T_d)$ devolver $\mathbf{X} = \mathbf{T}$.
4. En caso contrario volver al paso 1.

Por ejemplo, de forma análoga al caso unidimensional, en el caso de una densidad acotada en un hipercubo (intervalo cerrado multidimensional) siempre podríamos considerar una uniforme como densidad auxiliar.

Ejemplo 7.2 (distribución bidimensional acotada).

Supongamos que estamos interesados en generar valores de una variable aleatoria bidimensional (X, Y) con función de densidad:

$$f(x, y) = \begin{cases} \frac{3}{16} (2 - (x^2 + y^2)) & \text{si } x \in [-1, 1] \text{ e } y \in [-1, 1] \\ 0 & \text{en otro caso} \end{cases}$$

Podríamos considerar como densidad auxiliar la uniforme en $[-1, 1] \times [-1, 1]$:

$$g(x, y) = \begin{cases} \frac{1}{4} & \text{si } x, y \in [-1, 1] \\ 0 & \text{en otro caso} \end{cases}$$

Como $f(x, y) \leq M = f(0, 0) = \frac{3}{8}$, tomando $c = \frac{M}{g(x, y)} = \frac{3}{2}$ tendríamos que $f(x, y) \leq cg(x, y) = M$ y el algoritmo sería:

1. Generar $U \sim \mathcal{U}(0, 1)$.
2. Generar $T_1, T_2 \sim \mathcal{U}(-1, 1)$.
3. Si $M \cdot U \leq f(T_1, T_2)$ devolver $\mathbf{X} = (T_1, T_2)$.
4. En caso contrario volver al paso 1.

En este caso, la condición de aceptación del paso 3 simplificada sería: $U \leq 1 - (T_1^2 + T_2^2) / 2$.

Ejemplo 7.3 (distribución uniforme en la esfera).

Supongamos que el objetivo es simular puntos uniformemente distribuidos sobre la “esfera” unitaria d -dimensional:

$$C_d = \{(x_1, x_2, \dots, x_d) : x_1^2 + x_2^2 + \dots + x_d^2 \leq 1\}.$$

Denotando por $V_d(1)$, el “volumen” (la medida) de la esfera d -dimensional de radio 1 (en general, la de radio r verifica $V_d(r) = r^d V_d(1)$), se tiene:

$$f(x_1, x_2, \dots, x_d) = \begin{cases} \frac{1}{V_d(1)} & \text{si } (x_1, x_2, \dots, x_d) \in C_d \\ 0 & \text{si } (x_1, x_2, \dots, x_d) \notin C_d \end{cases}$$

Para simular valores en \mathbb{R}^d , con densidad f , podemos utilizar como distribución auxiliar una $\mathcal{U}([-1, 1] \times [-1, 1] \times \dots \times [-1, 1]) = \mathcal{U}([-1, 1]^d)$, dada por:

$$g(x_1, x_2, \dots, x_d) = \begin{cases} \frac{1}{2^d} & \text{si } x_i \in [-1, 1], \text{ para todo } i = 1, 2, \dots, d \\ 0 & \text{en otro caso} \end{cases}$$

La constante c óptima para la utilización del método de aceptación/rechazo es:

$$c = \max_{\mathbf{x}/g(\mathbf{x}) > 0} \frac{f(\mathbf{x})}{g(\mathbf{x})} = \frac{\frac{1}{V_d(1)}}{\frac{1}{2^d}} = \frac{2^d}{V_d(1)}$$

y la condición de aceptación $cUg(\mathbf{T}) \leq f(\mathbf{T})$ se convierte en:

$$\frac{2^d}{V_d(1)} U \frac{1}{2^d} 1_{[-1,1]^d}(\mathbf{T}) \leq \frac{1}{V_d(1)} 1_{C_d}(\mathbf{T}),$$

o, lo que es lo mismo, $U 1_{[-1,1]^d}(\mathbf{T}) \leq 1_{C_d}(\mathbf{T})$. Dado que el número aleatorio U está en el intervalo $[0, 1)$ y que las funciones indicadoras valen 0 ó 1, esta condición equivale a que $1_{[-1,1]^d}(\mathbf{T}) = 1_{C_d}(\mathbf{T})$, es decir, a que $\mathbf{T} \in C_d$, es decir, que se verifique:

$$T_1^2 + T_2^2 + \cdots + T_d^2 \leq 1.$$

Por otra parte, la simulación de $T \sim \mathcal{U}([-1, 1]^d)$ puede hacerse trivialmente mediante $T_i \sim \mathcal{U}(-1, 1)$ para cada $i = 1, 2, \dots, d$, ya que las componentes son independientes. Como el valor de U es superfluo en este caso, el algoritmo queda:

1. Simular $V_1, V_2, \dots, V_d \sim \mathcal{U}(0, 1)$ independientes.
2. Para $i = 1, 2, \dots, d$ hacer $T_i = 2V_i - 1$.
3. Si $T_1^2 + T_2^2 + \cdots + T_d^2 > 1$ entonces volver al paso 1.
4. Devolver $\mathbf{X} = (T_1, T_2, \dots, T_d)^t$.

Usando las fórmulas del “volumen” de una “esfera” d -dimensional:

$$V_d(r) = \begin{cases} \frac{\pi^{d/2} r^d}{(d/2)!} & \text{si } d \text{ es par} \\ \frac{2^{\lfloor \frac{d}{2} \rfloor + 1} \pi^{\lfloor \frac{d}{2} \rfloor} r^d}{1 \cdot 3 \cdot 5 \cdots d} & \text{si } d \text{ es impar} \end{cases}$$

puede verse que el número medio de iteraciones del algoritmo, dado por la constante $c = \frac{2^d}{V_d(1)}$, puede llegar a ser enórmemente grande. Así, si $d = 2$ se tiene $c = 1.27$, si $d = 3$ se tiene $c = 1.91$, si $d = 4$ entonces $c = 3.24$ y para $d = 10$ resulta $c = 401.5$ que es un valor que hace que el algoritmo sea tremadamente lento en dimensión 10. Esto está relacionado con la “maldición de la dimensionalidad” (curse of dimensionality), a medida que aumenta el número de dimensiones el volumen de la “frontera” crece exponencialmente.

7.3 Factorización de la matriz de covarianzas

Teniendo en cuenta que si $Cov(\mathbf{X}) = \Sigma$, entonces:

$$Cov(A\mathbf{X}) = A\Sigma A^t.$$

La idea de este tipo de métodos es simular datos independientes y transformarlos linealmente de modo que el resultado tenga la covarianza deseada.

Este método se emplea principalmente para la simulación de una normal multivariante, aunque también es válido para muchas otras distribuciones como la t -multivariante.

7.3.1 Simulación de la normal multivariante

Si $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ y A matriz $p \times d$, entonces:

$$A\mathbf{X} \sim \mathcal{N}_p(A\boldsymbol{\mu}, A\Sigma A^t).$$

Partiendo de $\mathbf{Z} \sim \mathcal{N}_d(\mathbf{0}, I_d)$, se podrían considerar distintas factorizaciones de la matriz de covarianzas:

- Factorización espectral: $\Sigma = H\Lambda H^t = H\Lambda^{1/2}(H\Lambda^{1/2})^t$

$$\mathbf{Y} = \mu + H\Lambda^{1/2}\mathbf{Z} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma).$$

- Factorización de Cholesky: $\Sigma = LL^t$

$$\mathbf{Y} = \mu + L\mathbf{Z} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma).$$

Desde el punto de vista de la eficiencia computacional la factorización de Cholesky sería la preferible. Pero en ocasiones, para evitar problemas numéricos (por ejemplo, en el caso de matrices no definidas positivas) puede ser más adecuado emplear la factorización espectral. En el primer caso el algoritmo sería el siguiente:

Algoritmo 7.2 (de simulación de una normal multivariante). 1. *Obtener la factorización de Cholesky*
 $\Sigma = LL^t$.

2. *Simular* $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)$ i.i.d. $\mathcal{N}(0, 1)$.
3. *Hacer* $\mathbf{X} = \mu + L\mathbf{Z}$.
4. *Repetir los pasos 2 y 3 las veces necesarias.*

Nota: Hay que tener en cuenta el resultado del algoritmo empleado para la factorización de Cholesky. Por ejemplo si se obtiene $\Sigma = U^tU$, hará que emplear $L = U^t$.

Ejemplo 7.4 (simulación de datos funcionales o temporales).

Supongamos que el objetivo es generar una muestra de tamaño `nsim` de la variable funcional:

$$Y(x) = \sin(2\pi x) + \varepsilon(x)$$

con $0 \leq x \leq 1$ y $Cov(\varepsilon(x), \varepsilon(y)) = e^{-\|x-y\|}$, considerando 100 puntos de discretización (se puede pensar también que es un proceso temporal).

```
nsim <- 20
n <- 100
x <- seq(0, 1, length = n)
# Media
mu <- sin(2*pi*x)
# Covarianzas
x.dist <- as.matrix(dist(x))
x.cov <- exp(-x.dist)
```

Para la factorización de la matriz de covarianzas emplearemos la función `chol` del paquete base de R (si las dimensiones fueran muy grandes podría ser preferible emplear otros paquetes, e.g. `spam::chol.spam`), pero al devolver la matriz triangular superior habrá que transponer el resultado:

```
U <- chol(x.cov)
L <- t(U)
```

Si queremos simular una realización:

```
set.seed(1)
head(mu + L %*% rnorm(n))

##      [,1]
## 1 -0.6264538
## 2 -0.5307633
## 3 -0.5797968
```

```

## 4 -0.2844357
## 5 -0.1711797
## 6 -0.2220796

y para simular nsim:
z <- matrix(rnorm(nsim * n), nrow = n)
y <- mu + L %*% z

matplot(x, y, type = "l", ylim = c(-3.5, 3.5))
lines(x, mu, lwd = 2)

```

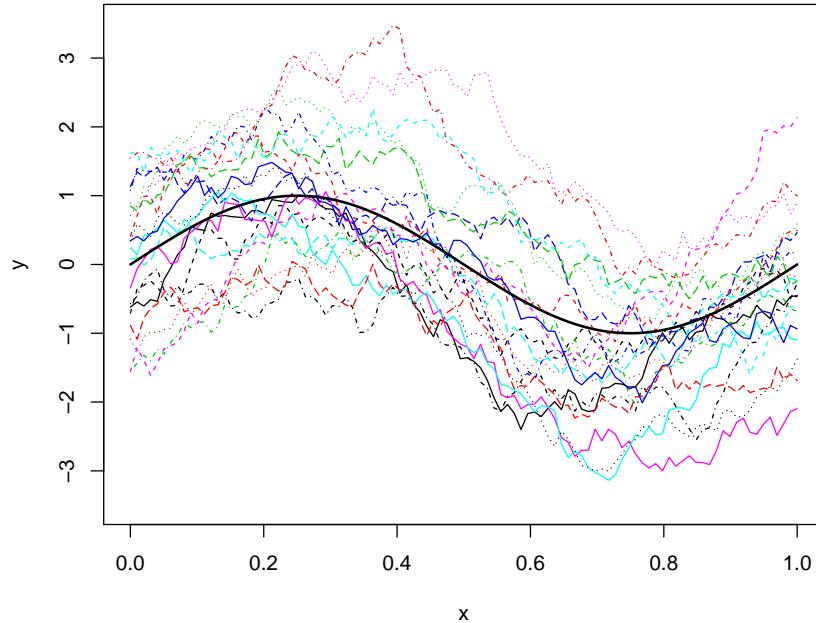


Figura 7.2: Realizaciones del proceso funcional.

Alternativamente se podría emplear, por ejemplo, la función `mvrnorm` del paquete `MASS` que emplea la factorización espectral (`eigen`):

```

library(MASS)
mvrnorm

## function (n = 1, mu, Sigma, tol = 1e-06, empirical = FALSE, EISPACK = FALSE)
## {
##     p <- length(mu)
##     if (!all(dim(Sigma) == c(p, p)))
##         stop("incompatible arguments")
##     if (EISPACK)
##         stop("'EISPACK' is no longer supported by R", domain = NA)
##     eS <- eigen(Sigma, symmetric = TRUE)
##     ev <- eS$values
##     if (!all(ev >= -tol * abs(ev[1L])))
##         stop("'Sigma' is not positive definite")

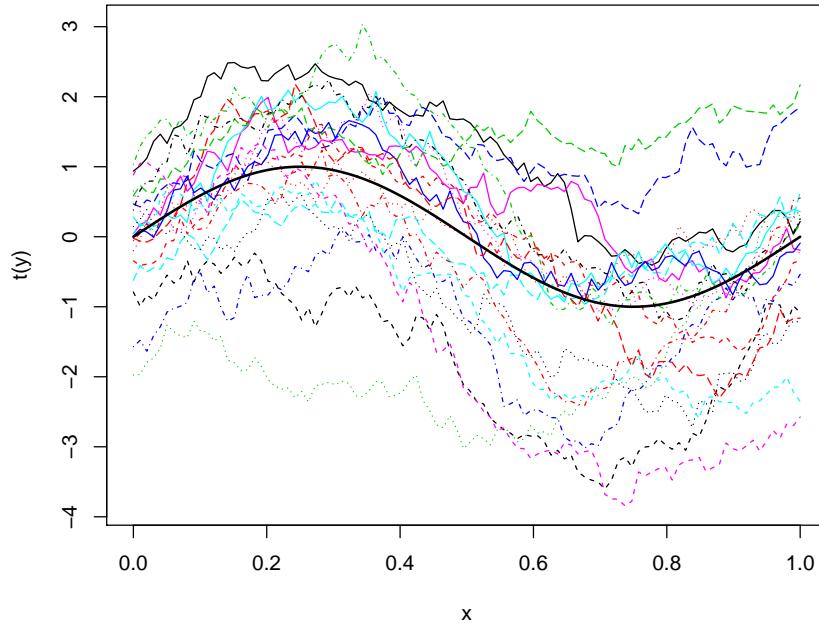
```

```

##      X <- matrix(rnorm(p * n), n)
##      if (empirical) {
##          X <- scale(X, TRUE, FALSE)
##          X <- X %*% svd(X, nu = 0)$v
##          X <- scale(X, FALSE, TRUE)
##      }
##      X <- drop(mu) + eS$vectors %*% diag(sqrt(pmax(ev, 0)), p) %*%
##              t(X)
##      nm <- names(mu)
##      if (is.null(nm) && !is.null(dn <- dimnames(Sigma)))
##          nm <- dn[[1L]]
##      dimnames(X) <- list(nm, NULL)
##      if (n == 1)
##          drop(X)
##      else t(X)
##  }
## <bytecode: 0x000000002f6b3860>
## <environment: namespace:MASS>
y <- mvtnorm(nsim, mu, x.cov)

matplot(x, t(y), type = "l")
lines(x, mu, lwd = 2)

```



7.4 Método de las distribuciones condicionadas

Teniendo en cuenta que:

$$f(x_1, x_2, \dots, x_d) = f_1(x_1) \cdot f_2(x_2|x_1) \cdots f_d(x_d|x_1, x_2, \dots, x_{d-1}),$$

donde las densidades condicionales pueden obtenerse a partir de las correspondientes marginales:

$$f_i(x_i|x_1, x_2, \dots, x_{i-1}) = \frac{f_{1,\dots,i}(x_1, x_2, \dots, x_i)}{f_{1,\dots,i-1}(x_1, x_2, \dots, x_{i-1})},$$

se obtiene el siguiente algoritmo general:

Algoritmo 7.3 (de simulación mediante distribuciones condicionadas). 1. Generar $X_1 \sim f_1$.

2. Desde $i = 2$ hasta d generar $X_i \sim f_i(\cdot|X_1, X_2, \dots, X_{i-1})$.
3. Devolver $\mathbf{X} = (X_1, X_2, \dots, X_d)$.

Nota: En las simulaciones unidimensionales se puede tener en cuenta que $f_i(x_i|x_1, x_2, \dots, x_{i-1}) \propto f_{1,\dots,i}(x_1, x_2, \dots, x_i)$.

Ejemplo 7.5 (distribución uniforme en el círculo unitario).

Se trata de la distribución bidimensional continua con densidad constante en el círculo:

$$C = \{(x_1, x_2) \in \mathbb{R} : x_1^2 + x_2^2 \leq 1\}.$$

Su función de densidad viene dada por:

$$f(x_1, x_2) = \begin{cases} \frac{1}{\pi} & \text{si } (x_1, x_2) \in C \\ 0 & \text{si } (x_1, x_2) \notin C \end{cases}$$

La densidad marginal de la primera variable resulta:

$$f_1(x_1) = \int_{-\sqrt{1-x_1^2}}^{+\sqrt{1-x_1^2}} \frac{1}{\pi} dx_2 = \frac{2\sqrt{1-x_1^2}}{\pi} \text{ si } x_1 \in [-1, 1],$$

es decir:

$$f_1(x_1) = \begin{cases} \frac{2}{\pi}\sqrt{1-x_1^2} & \text{si } x_1 \in [-1, 1] \\ 0 & \text{si } x_1 \notin [-1, 1] \end{cases}$$

Además:

$$f_2(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = \frac{\frac{1}{\pi}}{\frac{2\sqrt{1-x_1^2}}{\pi}} = \frac{1}{2\sqrt{1-x_1^2}}, \text{ si } x_2 \in \left[-\sqrt{1-x_1^2}, \sqrt{1-x_1^2}\right]$$

valiendo cero en otro caso. Se tiene entonces que:

$$X_2|X_1 \sim \mathcal{U}\left(-\sqrt{1-X_1^2}, \sqrt{1-X_1^2}\right),$$

siempre que $X_1 \in [-1, 1]$.

Finalmente, el algoritmo resulta:

1. Simular X_1 con densidad $f_1(x_1) = \frac{2}{\pi}\sqrt{1-x_1^2}1_{\{|x_1|\leq 1\}}$.
2. Simular X_2 con densidad $\mathcal{U}\left(-\sqrt{1-X_1^2}, \sqrt{1-X_1^2}\right)$.
3. Devolver $\mathbf{X} = (X_1, X_2)^t$.

Para el paso 1 puede utilizarse, por ejemplo, el método de aceptación/rechazo, pues se trata de una densidad acotada definida en un intervalo acotado.

Ejemplo 7.6 (distribución normal bidimensional).

En el caso de una distribución normal bidimensional:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix} \right)$$

tendríamos que:

- $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
- $X_2|X_1 \sim \mathcal{N}\left(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho(X_1 - \mu_1), \sigma_2^2(1 - \rho^2)\right)$.

Por lo que el algoritmo sería el siguiente:

Algoritmo 7.4 (de simulación de una normal bidimensional). 1. Simular $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ independientes.

2. Hacer $X_1 = \mu_1 + \sigma_1 Z_1$.
3. Hacer $X_2 = \mu_2 + \sigma_2 \rho Z_1 + \sigma_2 \sqrt{1 - \rho^2} Z_2$.

Este algoritmo es el mismo que obtendríamos con la factorización de la matriz de covarianzas ya que $\Sigma = LL^t$ con:

$$L = \begin{pmatrix} \sigma_1^2 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1 - \rho^2} \end{pmatrix}$$

Además, esta aproximación puede generalizarse al caso multidimensional, ver Sección 7.5.1.

7.5 Simulación condicional e incondicional

En ocasiones en inferencia estadística interesa la simulación condicional de nuevos valores de forma que se preserve la muestra observada.

En este caso se suele emplear el algoritmo anterior:

Algoritmo 7.5 (de simulación condicional). 1. Obtener la distribución condicional (correspondiente al punto que se quiere simular) dada la muestra y los valores simulados anteriormente.

2. Simular un valor de la distribución condicional.
3. Agregar este valor al conjunto de datos y volver al paso 1.

En el caso de normalidad, en lugar de simular punto a punto, podemos obtener fácilmente la distribución condicionada para simular los valores de forma conjunta.

7.5.1 Simulación condicional de una normal multivariante

Si $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ es tal que \mathbf{X} , $\boldsymbol{\mu}$ y Σ se particionan de la forma:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

suponiendo que \mathbf{X}_1 se corresponde con los valores observados y \mathbf{X}_2 con los que se pretende simular, entonces puede verse (e.g. Ripley, 1987) que la distribución de $\mathbf{X}_2|\mathbf{X}_1$ es normal con:

$$E(\mathbf{X}_2|\mathbf{X}_1) = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1), \quad (7.1)$$

$$Cov(\mathbf{X}_2|\mathbf{X}_1) = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}. \quad (7.2)$$

Nota: La ecuación (7.1) coincide con la expresión de la predicción lineal óptima de \mathbf{X}_2 a partir de \mathbf{X}_1 con media y varianzas conocidas (denominado predictor del kriging simple en estadística espacial, y la diagonal de (7.2) son las correspondientes varianzas kriging).

Ejemplo 7.7 (simulación condicional de datos funcionales o temporales).

Continuando con el Ejemplo 7.4 anterior, consideramos los primeros valores de una simulación incondicional como los datos:

```
idata <- x < 0.5
# idata <- x < 0.2 / x > 0.8
mu1 <- mu[idata]
n1 <- length(mu1)
cov.data <- x.cov[idata, idata]
U <- chol(cov.data)
# Simulación (incondicional):
set.seed(1)
data <- drop(mu1 + t(U) %*% rnorm(n1))
```

Para obtener la simulación condicional en los puntos de predicción, calculamos la correspondiente media y matriz de covarianzas condicionadas:

```
mu2 <- mu[!idata]
n2 <- length(mu2)
cov.pred <- x.cov[!idata, !idata]
cov.preddat <- x.cov[!idata, idata]
# Cálculo de los pesos kriging:
cov.data.inv <- chol2inv(U)
lambda <- cov.preddat %*% cov.data.inv
# La media serán las predicciones del kriging simple:
kpred <- mu2 + drop(lambda %*% (data - mu1))
# Varianza de la distribución condicional
kcov <- cov.pred - lambda %*% t(cov.preddat)
# (La diagonal serán las varianzas kriging).
```

Las simulaciones condicionales se obtendrán de forma análoga:

```
z <- matrix(rnorm(nsim * n2), nrow = n2)
y <- kpred + t(chol(kcov)) %*% z
# Representación gráfica:
plot(x, mu, type = "l", lwd = 2, ylab = "y", ylim = c(-3.5, 3.5)) # media teórica
lines(x[idata], data) # datos
# y <- rep(NA, n)
# y[idata] <- data
# lines(x, y)
matplot(x[!idata], y, type = "l", add = TRUE) # simulaciones condicionales
lines(x[!idata], kpred, lwd = 2, lty = 2) # media condicional (predicción kriging)
```

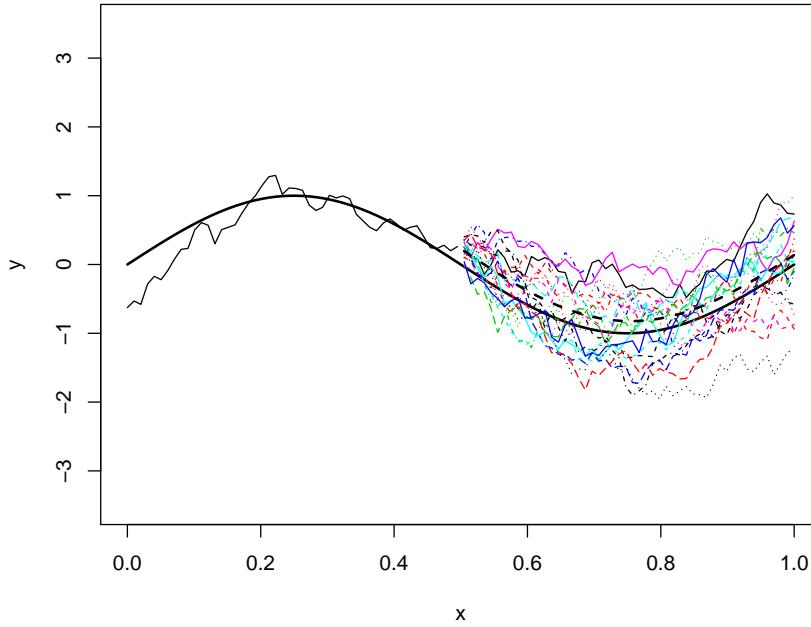


Figura 7.3: Realizaciones condicionales del proceso funcional.

Ejemplo 7.8 (simulación condicional de datos espaciales).

Consideramos un proceso espacial bidimensional normal $Z(\mathbf{s}) \equiv Z(x, y)$ de media 0 y covariograma exponencial:

$$\text{Cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\|\mathbf{s}_1 - \mathbf{s}_2\|) = e^{-\|\mathbf{s}_1 - \mathbf{s}_2\|}.$$

En primer lugar, obtendremos una simulación del proceso en las posiciones $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ que será considerada posteriormente como los datos observados.

Empleando las herramientas del paquete `geoR`, resulta muy fácil obtener una simulación incondicional en una rejilla en el cuadrado unidad mediante la función `grf`:

```
library(geoR)
n <- 4
set.seed(1)
z <- grf(n, grid="reg", cov.pars=c(1,1))

## grf: generating grid 2 * 2 with 4 points
## grf: process with 1 covariance structure(s)
## grf: nugget effect is: tausq= 0
## grf: covariance model 1 is: exponential(sigmasq=1, phi=1)
## grf: decomposition algorithm used is: cholesky
## grf: End of simulation procedure. Number of realizations: 1

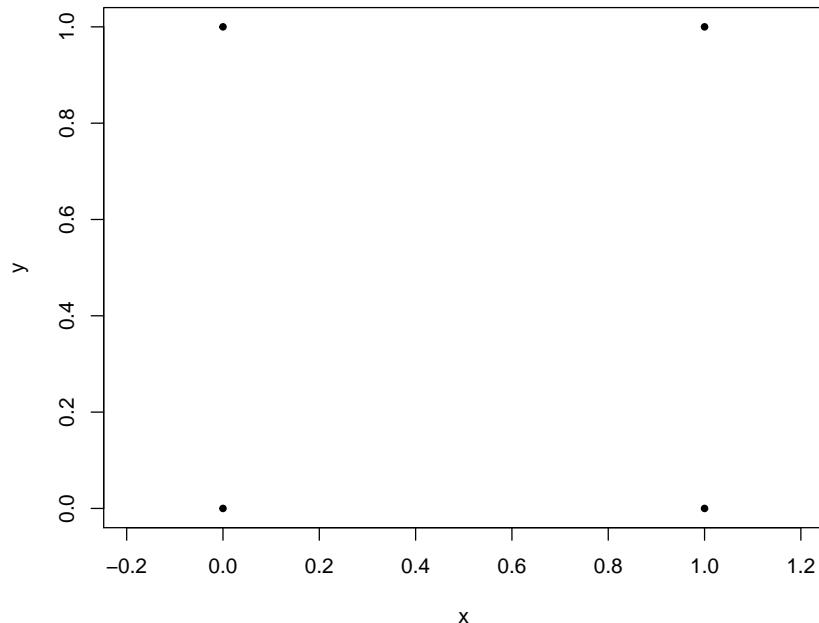
# names(z)
z$coords

##      x y
## [1,] 0 0
```

```
## [2,] 1 0
## [3,] 0 1
## [4,] 1 1
z$data
## [1] -0.62645381 -0.05969442 -0.98014198  1.09215113
```

La `grf` función emplea por defecto el método de la factorización de la matriz de covarianzas, sin embargo, si se desean obtener múltiples realizaciones, en lugar de llamar repetidamente a esta función, puede ser preferible emplear la siguiente aproximación:

```
# Posiciones datos
nx <- c(2, 2)
data.s <- expand.grid(x = seq(0, 1, len = nx[1]), y = seq(0, 1, len = nx[2]))
plot(data.s, type = "p", pch = 20, asp = 1) # Representar posiciones
```



```
# Matriz de varianzas covarianzas
cov.matrix <- varcov.spatial(coords=data.s, cov.pars=c(1,1))$varcov
cov.matrix

##          [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.3678794 0.3678794 0.2431167
## [2,] 0.3678794 1.0000000 0.2431167 0.3678794
## [3,] 0.3678794 0.2431167 1.0000000 0.3678794
## [4,] 0.2431167 0.3678794 0.3678794 1.0000000

# Simular valores
set.seed(1)
L <- t(chol(cov.matrix))

# Bucle simulación
nsim <- 1 # 1000
```

```

for (i in 1:nSIM) {
  y <- L %*% rnorm(n)
  # calcular estadísticos, errores, ...
}
y

##          [,1]
## [1,] -0.62645381
## [2,] -0.05969442
## [3,] -0.98014198
## [4,]  1.09215113

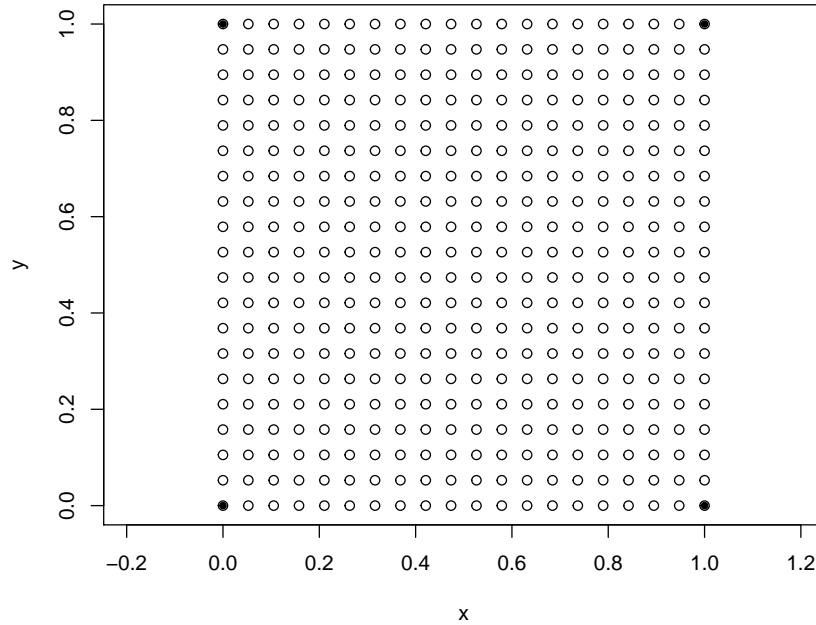
```

Para generar simulaciones condicionales podemos emplear la función `krige.conv`. Por ejemplo, para generar 4 simulaciones en la rejilla regular 10×10 en el cuadrado unidad $[0, 1] \times [0, 1]$ condicionadas a los valores generados en el apartado anterior podríamos emplear el siguiente código:

```

# Posiciones simulación condicional
nnx <- c(20, 20)
nn <- prod(nnx)
ndata.s <- expand.grid(x = seq(0, 1, len = nnx[1]), y = seq(0, 1, len = nnx[2]))
plot(ndata.s, type = "p", pch = 20, asp = 1)
points(ndata.s)

```



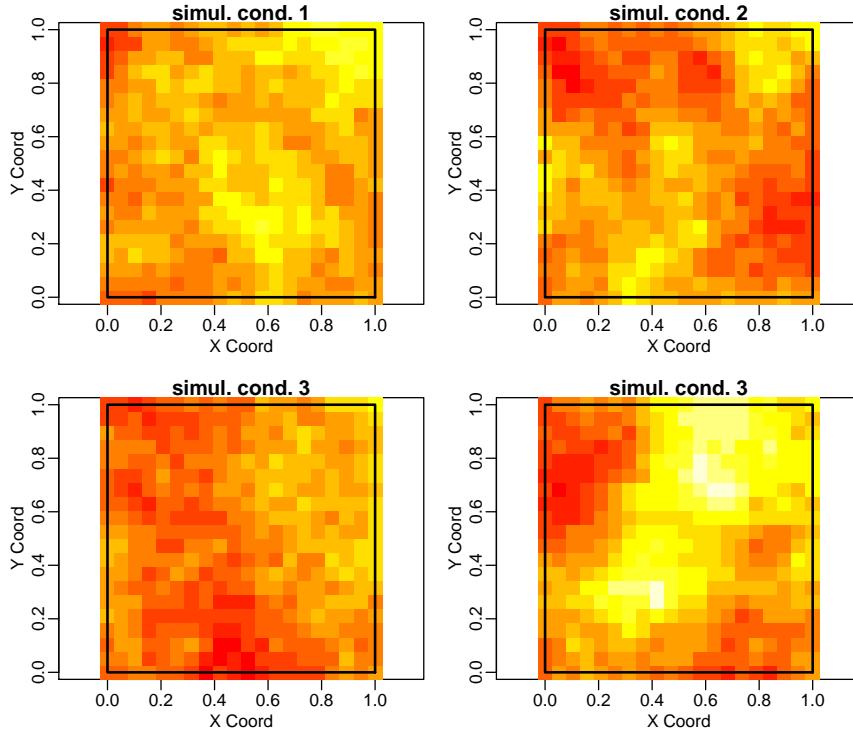
```

# Simulación condicional
set.seed(1)
s.out <- output.control(n.predictive = 4)
kc <- krige.conv(z, loc = ndata.s, output = s.out,
                  krige = krige.control(type.krige = "SK", beta = 0, cov.pars = c(1, 1)))

## krige.conv: results will be returned only for prediction locations inside the borders
## krige.conv: model with constant mean

```

```
## krige.conv: sampling from the predictive distribution (conditional simulations)
## krige.conv: Kriging performed using global neighbourhood
# Generar gráficos
par.old <- par(mfrow = c(2, 2), mar = c(3.5, 3.5, 1, 0), mgp = c(1.5, .5, 0))
zlim <- range(kc$simul) # Escala común
image(kc, val=kc$simul[,1], main="simul. cond. 1", zlim=zlim)
image(kc, val=kc$simul[,2], main="simul. cond. 2", zlim=zlim)
image(kc, val=kc$simul[,3], main="simul. cond. 3", zlim=zlim)
image(kc, val=kc$simul[,4], main="simul. cond. 3", zlim=zlim)
```



```
par(par.old)
```

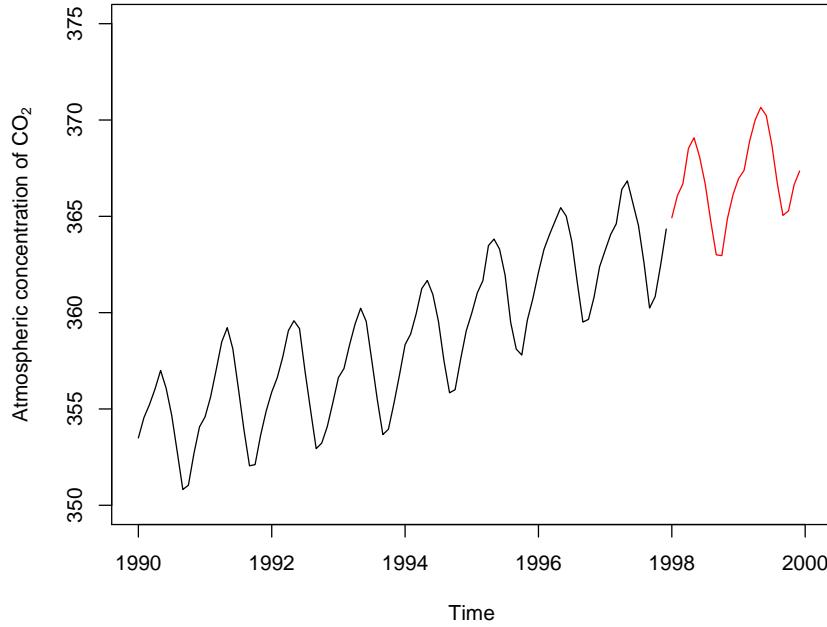
Los valores en las posiciones $\{(0,0), (0,1), (1,0), (1,1)\}$ coinciden con los generados anteriormente.

7.5.2 Simulación condicional a partir de un modelo ajustado

En la práctica normalmente se ajusta un modelo a los datos observados y posteriormente se obtienen las simulaciones condicionadas empleando el modelo ajustado.

Por ejemplo, en el caso de series de tiempo, se puede emplear la función `simulate` del paquete `forecast`:

```
library(forecast)
data <- window(co2, 1990) # datos de co2 desde 1990
plot(data, ylab = expression("Atmospheric concentration of CO"[2]),
      xlim=c(1990,2000), ylim=c(350, 375))
fit <- ets(data)
# plot(forecast(fit, h=24))
set.seed(1)
lines(simulate(fit, 24), col="red")
```



7.6 Simulación basada en cópulas

Una cópula es una función de distribución multidimensional con distribuciones marginales uniformes (e.g. Nelsen, 2006). Se emplean principalmente para la construcción de distribuciones multivariantes a partir de distribuciones marginales (también en análisis de dependencia y medidas de asociación).

Nelsen, R.B. (2006). *An introduction to copulas*, 2^a ed., Springer

Hofert, M. (2018). *Elements of Copula Modeling with R*, Springer.

Por simplicidad nos centraremos en el caso bidimensional. El teorema central en la teoría de cópulas es el teorema de Sklar (1959), que en este caso es:

Teorema 7.1 (de Sklar, caso bidimensional). *Si (X, Y) es una variable aleatoria bidimensional con función de distribución conjunta $F(\cdot, \cdot)$ y distribuciones marginales $F_1(\cdot)$ y $F_2(\cdot)$ respectivamente, entonces existe una cópula $C(\cdot, \cdot)$ tal que:*

$$F(x, y) = C(F_1(x), F_2(y)), \quad \forall x, y \in \mathbb{R}.$$

Además, si $F_1(\cdot)$ y $F_2(\cdot)$ son continuas $C(\cdot, \cdot)$ es única.

El recíproco también es cierto.

7.6.1 Cópulas Arquimediana

Además de las cópulas Gausianas, es una de las familias de cópulas más utilizadas. Son de la forma:

$$C(x_1, x_2, \dots, x_n) = \Psi^{-1} \left(\sum_{i=1}^n \Psi(F_i(x_i)) \right),$$

siendo Ψ su función generadora.

Una condición suficiente para que sea una cópula multidimensional válida es que $\Psi(1) = 0$, $\lim_{x \rightarrow 0} \Psi(x) = \infty$, $\Psi'(x) < 0$ y $\Psi''(x) > 0$.

Ejemplos:

- Cópula producto o independiente: $\Psi(x) = -\ln(x)$,

$$F(x, y) = F_1(x)F_2(y).$$

- Cópula de Clayton: $\Psi(x) = \frac{1}{\alpha}(x^{-\alpha} - 1)$; $\alpha > 0$,

$$F(x, y) = (F_1(x)^{-\alpha} + F_2(y)^{-\alpha} - 1)^{-1/\alpha}.$$

- Cópula de Gumbel: $\Psi(x) = (-\ln(x))^\alpha$; $\alpha \geq 1$

7.6.2 Simulación

Las cópulas pueden facilitar notablemente la simulación de la distribución conjunta. Si $(U, V) \sim C(\cdot, \cdot)$ (marginales uniformes):

$$(F_1^{-1}(U), F_2^{-1}(V)) \sim F(\cdot, \cdot)$$

En la mayoría de los casos se dispone de expresiones explícitas de $C_u(v) \equiv C_2(v|u)$ y de su inversa $C_u^{-1}(w)$, por lo que se puede generar (U, V) fácilmente mediante el **método secuencial de distribuciones condicionadas**.

Algoritmo 7.6 (de simulación bidimensional mediante cópulas). 1. Generar $U, W \sim \mathcal{U}(0, 1)$

2. Obtener $V = C_U^{-1}(W)$
3. Devolver $(F_1^{-1}(U), F_2^{-1}(V))$

Ejercicio 7.1.

Consideramos una v.a. bidimensional con distribuciones marginales uniformes y distribución bidimensional determinada por la cópula de Clayton.

- a) Teniendo en cuenta que en este caso:

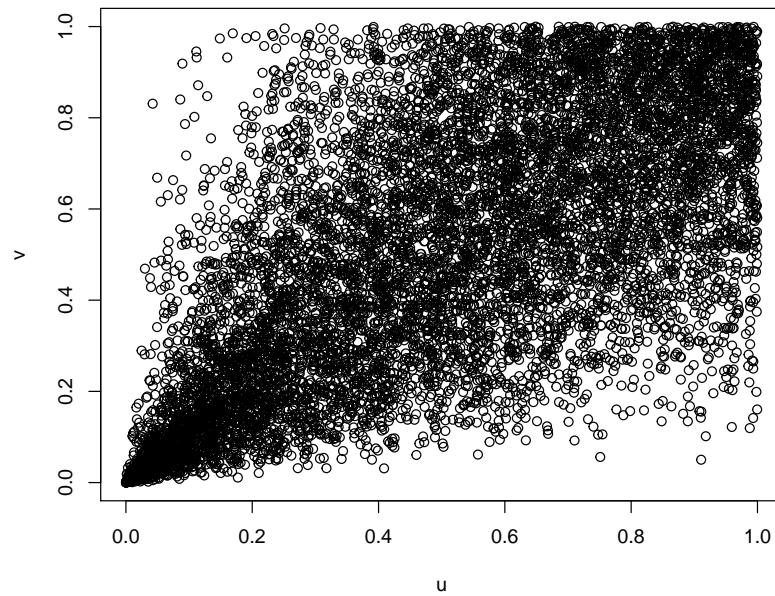
$$C_u^{-1}(w) \equiv \left(w^{-\alpha} \left(w^{-\frac{\alpha}{\alpha+1}} - 1 \right) + 1 \right)^{-\frac{1}{\alpha}},$$

diseñar una rutina que permita generar una muestra de tamaño n de esta distribución.

```
rcclayton <- function(alpha, n) {
  val <- cbind(runif(n), runif(n))
  val[, 2] <- (val[, 1]^(−alpha) *
    (val[, 2]^(−alpha/(alpha + 1)) − 1) + 1)^^(−1/alpha)
  return(val)
}
```

- b) Utilizando la rutina anterior generar una muestra de tamaño 10000 y representar gráficamente los valores obtenidos y sus distribuciones marginales.

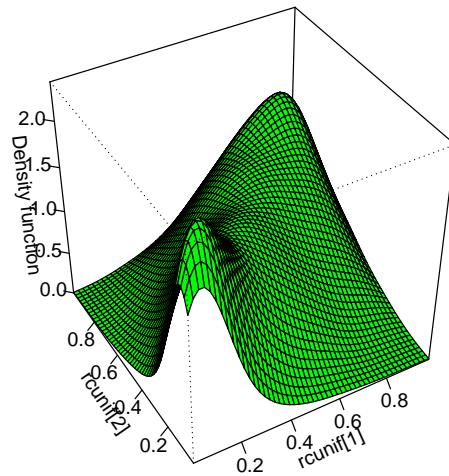
```
set.seed(54321)
rcunif <- rcclayton(2,10000)
plot(rcunif, xlab = "u", ylab = "v")
```



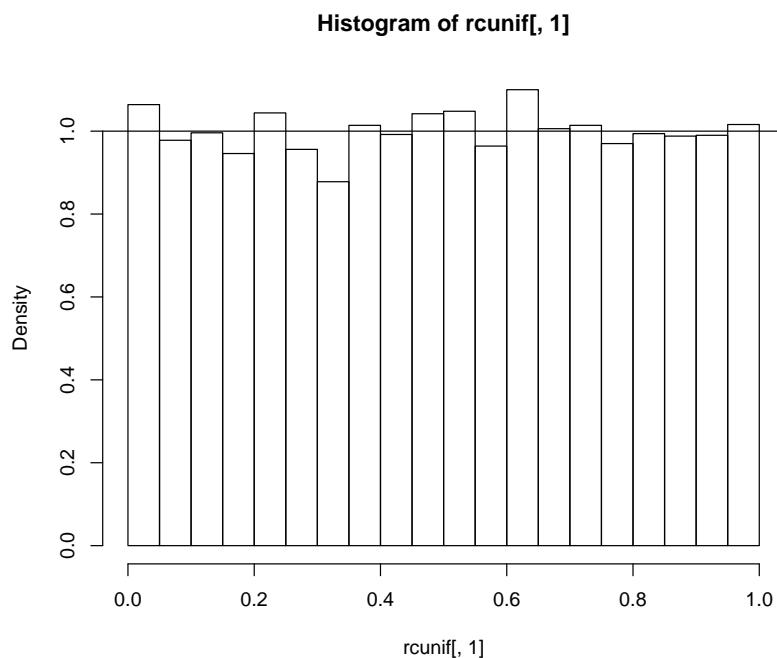
Representar la densidad conjunta (con `sm::sm.density()`) y las marginales:

```
# Densidad conjunta
# if(!require(sm)) stop('Required package `sm` not installed.')
sm::sm.density(rcunif)
```

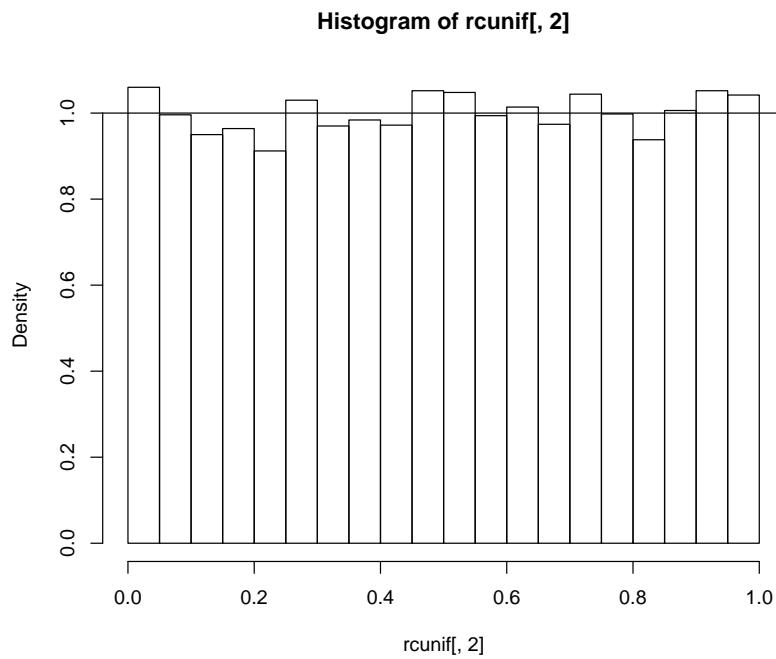
```
## Warning: weights overwritten by binning
```



```
# Distribuciones marginales  
hist(rcunif[,1], freq = FALSE)  
abline(h = 1)
```

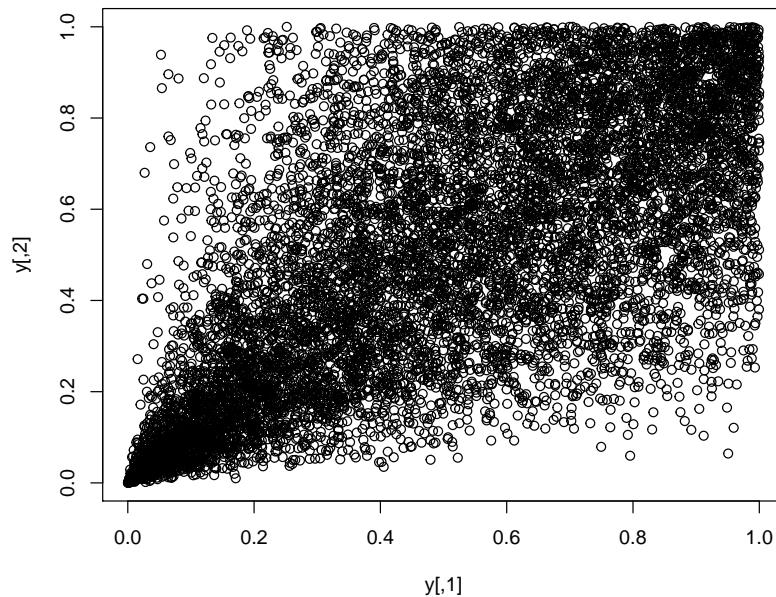


```
hist(rcunif[,2], freq = FALSE)  
abline(h = 1)
```



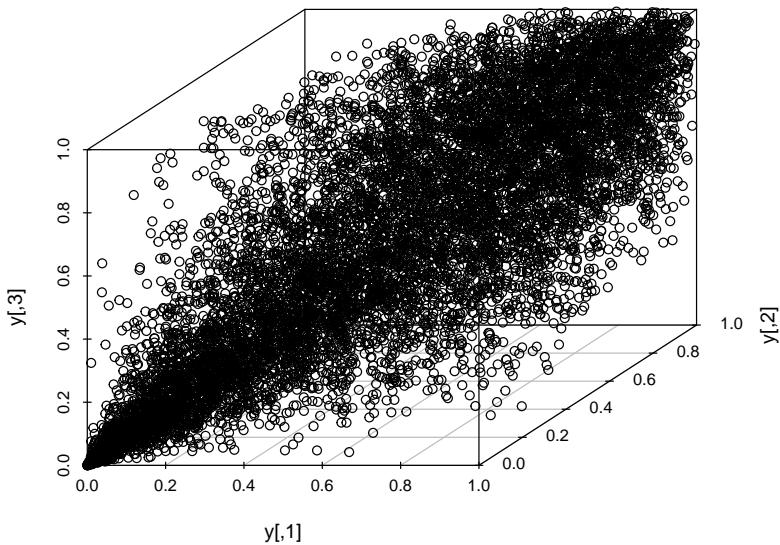
Empleando el paquete `copula`:

```
if(!require(copula)) stop('Required package `copula` not installed.')
clayton.cop <- claytonCopula(2, dim = 2) # caso bidimensional
y <- rCopula(10000, clayton.cop)
plot(y)
```



```
clayton.cop <- claytonCopula(2, dim = 3) # caso tridimensional
y <- rCopula(10000, clayton.cop)
```

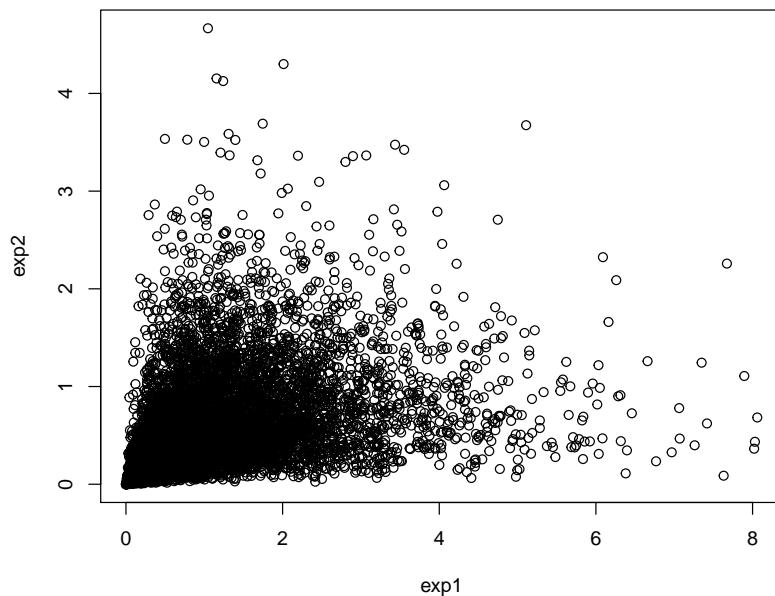
```
scatterplot3d::scatterplot3d(y)
```



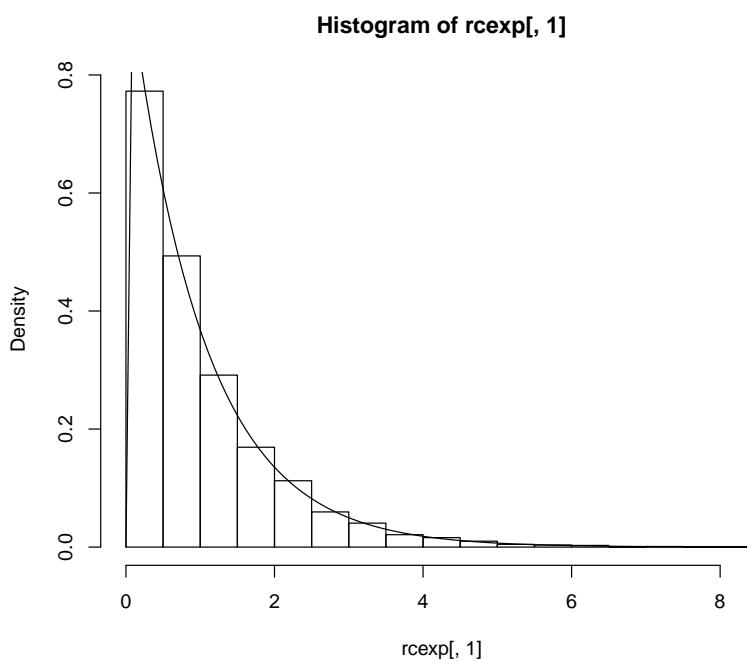
```
# plot3D:::points3D(y[,1], y[,2], y[, 3], colvar = NULL)
```

- c) A partir de la muestra anterior generar una muestra de una v.a. bidimensional con distribuciones marginales exponenciales de parámetros 1 y 2 respectivamente (y distribución bidimensional determinada por la cópula de Clayton).

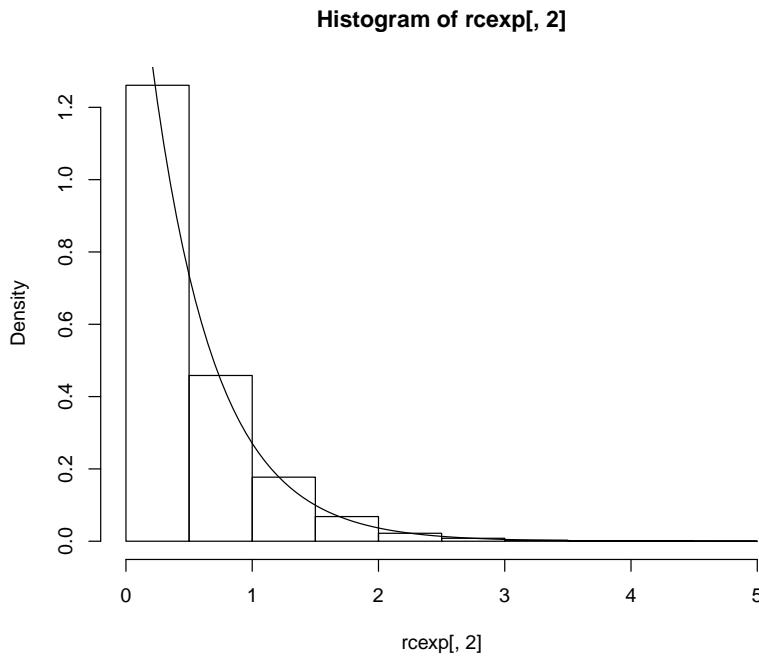
```
rcexp <- cbind(qexp(rcunif[,1], 1), qexp(rcunif[,2], 2))
plot(rcexp, xlab = "exp1", ylab = "exp2")
```



```
# Distribuciones marginales  
hist(rcexp[,1], freq = FALSE)  
curve(dexp(x,1), add = TRUE)
```



```
hist(rcexp[,2], freq = FALSE)  
curve(dexp(x,2), add = TRUE)
```



```
# ...  
z <- 1:10  
xy <- matrix(z, ncol = 2)  
xy  
  
##      [,1] [,2]  
## [1,]     1    6  
## [2,]     2    7  
## [3,]     3    8  
## [4,]     4    9  
## [5,]     5   10  
as.vector(xy)  
  
##  [1] 1 2 3 4 5 6 7 8 9 10  
# ...
```

7.7 Otros métodos para variables continuas

- Método de la transformada de Fourier (FFT) (e.g. Davies y Harte, 1987).
- Circular embedding (Dietrich and Newsam, 1997).

7.8 Simulación de distribuciones multidimensionales discretas

7.8.1 Métodos de codificación o etiquetado para variables discretas

En el caso de una distribución d -dimensional discreta el procedimiento habitual es simular una variable aleatoria discreta unidimensional equivalente. Este tipo de procedimientos son conocidos como métodos de

etiquetado o codificación y la idea básica consistiría en construir un índice unidimensional equivalente al índice multidimensional.

Si la variable discreta multidimensional tiene soporte finito, este tipo de recodificación se puede hacer de forma automática en R cambiando simplemente el indexado (internamente un objeto `matrix` o `array` ya está almacenado como un vector y R permite un indexado multidimensional a partir del atributo `dim`).

```
z <- 1:10
xy <- matrix(z, ncol = 2)
xy

##      [,1] [,2]
## [1,]    1    6
## [2,]    2    7
## [3,]    3    8
## [4,]    4    9
## [5,]    5   10
as.vector(xy)

##  [1] 1 2 3 4 5 6 7 8 9 10
```

Si la variable discreta multidimensional no tiene soporte finito (no se podría guardar la f.m.p. en una tabla), se podrían emplear métodos de codificación más avanzados (ver sección 6.3 del libro de R. Cao).

7.8.2 Simulación de una variable discreta bidimensional

Consideramos datos recogidos en un estudio de mejora de calidad en una fábrica de semiconductores. Se obtuvo una muestra de obleas que se clasificaron dependiendo de si se encontraron partículas en la matriz que producía la oblea y de si la calidad de oblea era buena (para más detalles Hall, 1994. Analysis of defectivity of semiconductor wafers by contingency table. Proceedings of the Institute of Environmental Sciences 1, 177-183).

```
n <- c(320, 14, 80, 36)
particulas <- gl(2, 1, 4, labels = c("no", "si"))
calidad <- gl(2, 2, labels = c("buena", "mala"))
df <- data.frame(n, particulas, calidad)
df

##      n particulas calidad
## 1 320        no    buena
## 2  14        si    buena
## 3  80        no     mala
## 4  36        si     mala
```

En lugar de estar en el formato de un conjunto de datos (`data.frame`) puede que los datos estén en formato de tabla (`table`, `matrix`):

```
tabla <- xtabs(n ~ calidad + particulas)
tabla

##      particulas
## calidad no si
##    buena 320 14
##    mala   80 36
```

Lo podemos convertir directamente a `data.frame`:

```
as.data.frame(tabla)

##   calidad particulas Freq
## 1 buena            no  320
## 2 mala             no   80
## 3 buena            si   14
## 4 mala             si   36
```

En este caso definimos las probabilidades a partir de las frecuencias:

```
df$p <- df$n/sum(df$n)
df

##      n particulas calidad      p
## 1 320      no    buena 0.71111111
## 2  14      si    buena 0.03111111
## 3  80      no    mala  0.17777778
## 4  36      si    mala  0.08000000
```

En formato tabla:

```
pij <- tabla/sum(tabla)
pij

##      particulas
## calidad      no      si
##  buena 0.71111111 0.03111111
##  mala  0.17777778 0.08000000
```

Para simular la variable bidimensional consideramos una variable unidimensional de índices:

```
z <- 1:nrow(df)
z

## [1] 1 2 3 4

Con probabilidades:
```

```
pz <- df$p
pz

## [1] 0.71111111 0.03111111 0.17777778 0.08000000
```

Si las probabilidades estuviesen en una matriz, las convertiríamos a un vector con:

```
as.vector(pij)

## [1] 0.71111111 0.17777778 0.03111111 0.08000000
```

Si simulamos la variable unidimensional:

```
set.seed(1)
nsim <- 20
rz <- sample(z, nsim, replace = TRUE, prob = pz)
```

Podríamos obtener simulaciones bidimensionales, por ejemplo:

```
etiquetas <- as.matrix(df[c('particulas', 'calidad')])
rxy <- etiquetas[rz, ]
head(rxy)
```

```
##      particulas calidad
```

```
## [1,] "no"      "buena"
## [2,] "no"      "buena"
## [3,] "no"      "buena"
## [4,] "si"      "mala"
## [5,] "no"      "buena"
## [6,] "si"      "mala"
```

Alternativamente, si queremos trabajar con data.frames:

```
etiquetas <- df[c('particulas', 'calidad')]
rxy <- etiquetas[rz, ]
head(rxy)
```

```
##   particulas calidad
## 1      no    buena
## 1.1    no    buena
## 1.2    no    buena
## 4      si     mala
## 1.3    no    buena
## 4.1    si     mala

# Si se quieren eliminar las etiquetas de las filas:
row.names(rxy) <- NULL
head(rxy)
```

```
##   particulas calidad
## 1      no    buena
## 2      no    buena
## 3      no    buena
## 4      si     mala
## 5      no    buena
## 6      si     mala
```

7.8.3 Simulación de tablas de contingencia

El código anterior puede ser empleado para simular tablas de contingencia. Aunque en estos casos se suele fijar el total de la tabla (o incluso las frecuencias marginales). En este caso, sólo habría que fijar el nº de simulaciones al total de la tabla:

```
nsim <- sum(n)
set.seed(1)
rz <- sample(z, nsim, replace = TRUE, prob = pz)
rtable <- table(rz) # Tabla de frecuencias unidimensional
matrix(rtable, ncol = 2) # Tabla de frecuencias bidimensional

##      [,1] [,2]
## [1,] 321   78
## [2,] 15    36
```

Aunque puede ser preferible emplear directamente `rmultinom` si se van a generar muchas:

```
ntsim <- 1000
rtablas <- rmultinom(ntsim, sum(n), pz)
rtablas[ , 1:5] # Las cinco primeras simulaciones

##      [,1] [,2] [,3] [,4] [,5]
## [1,] 298  329  323  323  307
```

```
## [2,] 15 21 5 15 15
## [3,] 92 68 91 77 92
## [4,] 45 32 31 35 36
```

Por ejemplo, si se quiere simular bajo independencia, estimando las probabilidades a partir de la tabla:

```
tabla
```

```
##      particulas
## calidad no si
## buena 320 14
## mala  80 36
```

Consideraríamos como probabilidades:

```
pind <- (rowSums(tabla) %o% colSums(tabla))/(sum(tabla)^2)
matrix(pind, nrow = nrow(tabla))
```

```
## [,1]      [,2]
## [1,] 0.6597531 0.08246914
## [2,] 0.2291358 0.02864198
rtablas <- rmultinom(ntsim, sum(n), pind)
rtablas[, 1:5] # Las cinco primeras simulaciones
```

```
## [,1] [,2] [,3] [,4] [,5]
## [1,] 292 285 309 303 290
## [2,] 96 105 97 84 113
## [3,] 48 48 36 49 39
## [4,] 14 12 8 14 8
```

Para realizar el contraste de independencia:

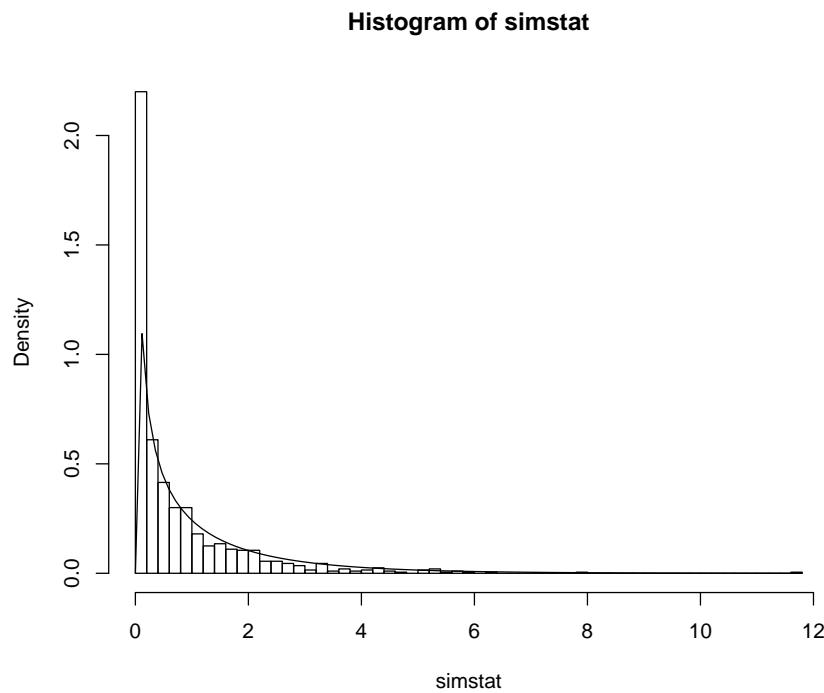
```
res <- chisq.test(tabla)
res
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: tabla
## X-squared = 60.124, df = 1, p-value = 8.907e-15
```

Ejercicio 7.2.

Aproximar por simulación la distribución (exacta) del estadístico ji-cuadrado bajo independencia.

```
simstat <- apply(rtablas, 2, function(x){chisq.test(matrix(x,nrow=nrow(tabla)))$statistic})
hist(simstat, freq = FALSE, breaks = 'FD')
# Distribución asintótica (aproximación ji-cuadrado)
curve(dchisq(x, res$parameter), add = TRUE)
```



7.9 Ejercicios propuestos

Ejercicio 7.3.

Dar un algoritmo basado en el método de aceptación-rechazo (considerando como densidad auxiliar una uniforme) que permita generar observaciones de una variable aleatoria bidimensional (X, Y) con función de densidad:

$$f(x, y) = \begin{cases} \frac{3}{16} (2 - (x^2 + y^2)) & \text{si } x \in [-1, 1] \text{ e } y \in [-1, 1] \\ 0 & \text{en otro caso} \end{cases}$$

Ejercicio 7.4.

Considerando la variable aleatoria bidimensional del ejercicio anterior y teniendo en cuenta que la densidad marginal de la variable X es:

$$f_X(x) = \begin{cases} \frac{1}{8} (5 - 3x^2) & \text{si } x \in [-1, 1] \\ 0 & \text{en otro caso} \end{cases}$$

Describir brevemente un algoritmo para la simulación del vector aleatorio basado en el método de las distribuciones condicionadas (asumir que se dispone de un algoritmo para generar observaciones de las distribuciones unidimensionales de interés).

Capítulo 8

Aplicaciones de la simulación en Inferencia Estadística

8.1 Aplicaciones de la simulación

Aplicaciones:

- Estadística: Muestreo, aproximación de distribuciones, remuestreo, ...
- Optimización: Algoritmos genéticos, ...
- Análisis numérico: Evaluación de expresiones, ...
- ...

En este capítulo nos centraremos en algunas de las aplicaciones en inferencia estadística, aunque habría muchas:

- Distribución de estimadores puntuales/estadísticos:
 - Aproximación de la distribución.
 - * Aproximación de características de la distribución.
 - * Validación de la distribución asintótica.
 - Comparación de estimadores.
- Estimación por intervalo de confianza:
 - Obtención de intervalos/bandas de confianza (probabilidad).
 - Análisis de un estimador por intervalo de confianza.
- Contrastos de hipótesis:
 - Aproximación del p -valor.
 - Análisis de un contraste de hipótesis.
 - Validación teoría.
- Métodos de remuestro bootstrap.
- Inferencia Bayesiana
- ...

En el siguiente capítulo traremos la Integración y Optimización Monte Carlo...

Observación: En este capítulo se obtendrán simulaciones de estadísticos a partir de muestras (podemos pensar que se parte de generaciones de una variable multivariante). En la mayoría de los ejemplos se generan todas las muestras de una vez, se guardan y se procesan vectorialmente (normalmente empleando la función `apply`). En problemas mas complejos, en los que no es necesario almacenar todas las simulaciones, puede ser preferible emplear un bucle para generar y procesar las muestras iterativamente. Por ejemplo podríamos emplear el siguiente esquema:

```
# Fijar semilla
set.seed(1)
for (isim in 1:nSIM) {
  seed <- .Random.seed
  # Si se produce un error, podremos depurarlo ejecutando:
  # .Random.seed <- seed

  # Generar muestra
  ...
}
```

o alternativamente fijar la semilla en cada iteración, por ejemplo:

```
for (isim in 1:nSIM) {
  set.seed(isim)

  # Generar muestra
  ...
}
```

8.2 Distribución en el muestreo

Ejercicio 8.1.

Si X_1, \dots, X_n es una muestra aleatoria simple de una variable aleatoria $X \sim N(\mu, \sigma)$, la distribución en el muestreo de:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

es:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Confirmar este resultado mediante simulación, para ello:

- Crear un conjunto de datos `muestras` con 500 muestras de tamaño $n = 10$ de una $N(1, 2)$. Añadir al conjunto de datos las estimaciones de la media y desviación típica obtenidas con cada una de las muestras.

Valores iniciales:

```
set.seed(54321) # Fijar semilla para reproductibilidad
nsim <- 500
nx <- 10
```

Valores teóricos:

```
mux <- 1
sdx <- 2
```

Simulación de las muestras (al estilo Rcmdr):

```
muestras <- as.data.frame(matrix(rnorm(nsim*nx, mean=mux, sd=sdx), ncol=nx))
rownames(muestras) <- paste("muestra", 1:nsim, sep="")
colnames(muestras) <- paste("obs", 1:nx, sep="")
str(muestras)

## 'data.frame': 500 obs. of 10 variables:
## $ obs1 : num 0.642 -0.856 -0.568 -2.301 0.184 ...
## $ obs2 : num 3.483 2.216 1.1 4.305 0.677 ...
## $ obs3 : num 1.24 -1.51 -3.98 2.29 2.46 ...
## $ obs4 : num 3.286 0.947 0.953 -1.663 2.623 ...
## $ obs5 : num 3.77 -1.34 1.61 -2.46 1.11 ...
## $ obs6 : num -2.044 0.32 3.046 0.136 3.555 ...
## $ obs7 : num 0.6186 -1.8614 4.3386 0.0996 0.8334 ...
## $ obs8 : num -0.829 2.202 -1.688 1.534 -0.114 ...
## $ obs9 : num 0.4904 -0.6713 0.5451 -0.6517 0.0168 ...
## $ obs10: num 2.79 2.84 1.27 3.93 2.17 ...
```

Estimaciones:

```
muestras$mean <- rowMeans(muestras[,1:nx])
muestras$sd <- apply(muestras[,1:nx], 1, sd)
```

Nota: . La fila `muestras[i,]` contiene las observaciones de la i-ésima muestra y la correspondiente media y desviación típica.

```
muestras[1,]
```

```
##          obs1     obs2     obs3     obs4     obs5     obs6     obs7
## muestra1 0.6421985 3.482661 1.242483 3.28559 3.766896 -2.04443 0.6186323
##          obs8     obs9    obs10      mean      sd
## muestra1 -0.8293636 0.4903819 2.790091 1.344514 1.951292
```

Normalmente emplearemos sin embargo una ordenación por columnas (cada fila se corresponderá con una generación).

- b) Generar el histograma (en escala de densidades) de las medias muestrales y compararlo con la densidad teórica.

Distribución de la media muestral:

```
hist(muestras$mean, freq = FALSE, breaks = "FD",
      xlab = "Medias", ylab = "Densidad")
# Densidad observada (estimación)
lines(density(muestras$mean))
# Densidad teórica (bajo normalidad)
curve(dnorm(x, mux, sdx/sqrt(nx)), lwd = 2, col = "blue", add = TRUE)
# Aproximación del valor esperado de la media muestral mediante simulación
abline(v = mean(muestras$mean), lty = 2)
# Valor esperado de la media muestral (teórico)
abline(v = mux, col = "blue")
```

Ejercicio 8.2.

Si X_1, \dots, X_n es una m.a.s. de una variable aleatoria X (cualquiera) con $E(X) = \mu$ y $Var(X) = \sigma^2$, por el

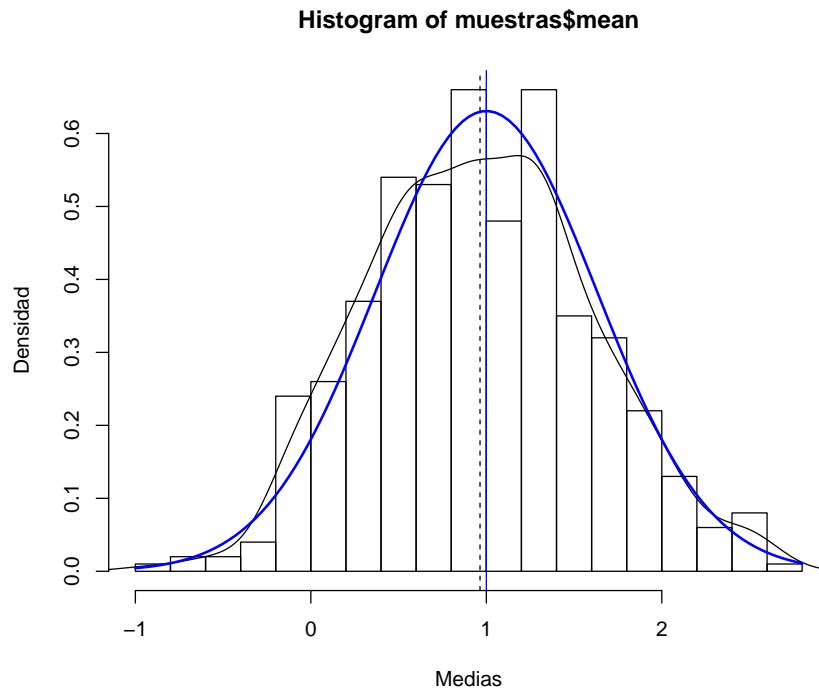


Figura 8.1: Distribución de la media muestral de una distribución normal.

Teorema Central del Límite, la distribución en el muestreo de $\hat{\mu} = \bar{X}$ se aproxima a la normalidad:

$$\bar{X} \xrightarrow{n \rightarrow \infty} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Típicamente se suele considerar que esta aproximación es buena para tamaños muestrales $n > 30$, aunque dependerá de las características de la distribución de X .

- a) Repetir el ejercicio anterior considerando muestras de una $Exp(1)$ (tener en cuenta que $X \sim Exp(\lambda) \Rightarrow \mu_X = \sigma_X = 1/\lambda$). ¿Qué ocurre con la distribución de la media muestral?

```
set.seed(54321) # Fijar semilla para reproductibilidad
nsim <- 500
nx <- 10
# nx <- 50
```

Valores teóricos:

```
lambda <- 1
muexp <- 1/lambda
sdexp <- muexp
```

Simulación de las muestras:

```
muestras2 <- as.data.frame(matrix(rexp(nsim*nx, rate=lambda), ncol=nx))
rownames(muestras2) <- paste("muestra", 1:nsim, sep="")
colnames(muestras2) <- paste("obs", 1:nx, sep="")
```

Estimaciones:

```
muestras2$mean <- rowMeans(muestras2[,1:nx])
muestras2$sd <- apply(muestras2[,1:nx], 1, sd)
```

Distribución de la media muestral:

```
hist(muestras2$mean, xlim = c(-0.1, 2.5), freq = FALSE, breaks = "FD",
      xlab = "Medias", ylab = "Densidad")
# Densidad observada (estimación)
lines(density(muestras2$mean))
# Distribución asintótica (TCL)
curve(dnorm(x,muexp,sdexp/sqrt(nx)), lwd=2, col="blue", add=TRUE)
# Aproximación del valor esperado de la media muestral mediante simulación
abline(v=mean(muestras2$mean),lty=2)
# Valor esperado de la media muestral (teórico)
abline(v=muexp, col="blue")
```

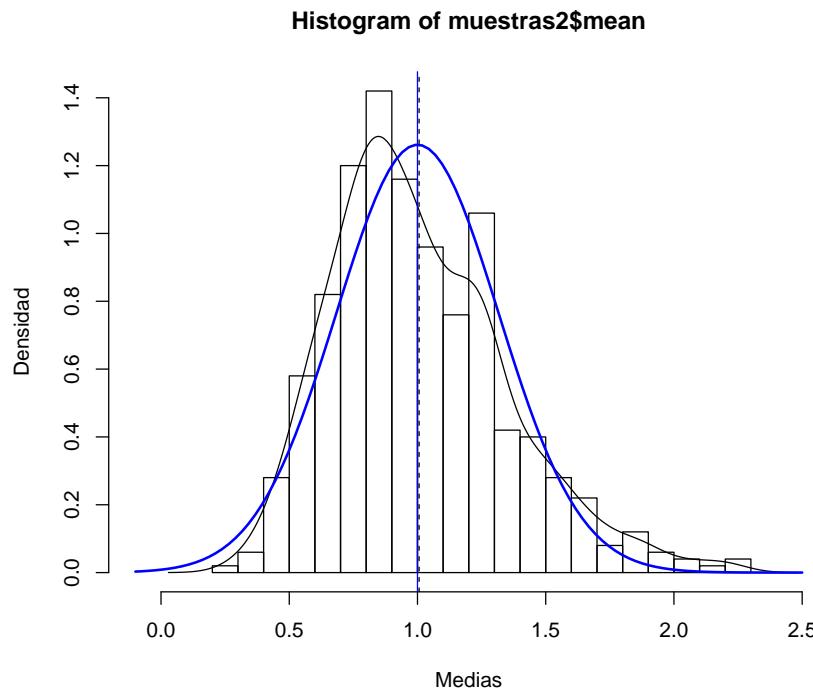


Figura 8.2: Distribución de la media muestral de una distribución exponencial y distribución asintótica.

- b) Aumentar el tamaño muestral a 50. ¿Se aproxima más la distribución de las medias muestrales a la teórica bajo normalidad?

Ejecutar el código del apartado anterior fijando `nx <- 50`.

8.3 Intervalos de confianza

Ejercicio 8.3.

Siguiendo el enunciado del ejercicio 1, se deduce que el intervalo de confianza (de nivel $1 - \alpha$) para la media μ de una población normal con varianza conocida es:

$$IC_{1-\alpha}(\mu) = \left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right).$$

La idea es que el $100(1 - \alpha)\%$ de los intervalos así construidos contendrán el verdadero valor del parámetro.

- a) Utilizando el conjunto de datos `muestras` del ejercicio 1 (500 muestras de tamaño $n = 10$ de una $N(1, 2)$), añadir en dos nuevas variables los extremos del intervalo de confianza para la media con varianza conocida al conjunto de datos. Analizar la cobertura de estas estimaciones por IC.

IC para la media con varianza conocida (bajo normalidad):

```
alfa <- 0.05
z <- qnorm(1 - alfa/2)
muestras$ici <- muestras$mean - z*sdx/sqrt(nx)
muestras$ics <- muestras$mean + z*sdx/sqrt(nx)
```

Cobertura de las estimaciones por IC:

```
muestras$cob <- (muestras$ici < mux) & (mux < muestras$ics)
ncob <- sum(muestras$cob) # N° de intervalos que contienen la verdadera media
ncob
```

```
## [1] 480
100*ncob/nsim      # Proporción de intervalos
```

```
## [1] 96
100*(1 - alfa)    # Proporción teórica bajo normalidad
```

```
## [1] 95
```

Como ejemplo ilustrativo, generamos el gráfico de los primeros 50 intervalos:

```
m <- 50
tmp <- muestras[1:m,]
attach(tmp)
color <- ifelse(cob, "blue", "red")
plot(1:m, mean, col = color, ylim = c(min(ici),max(ics)),
     xlab = "Muestra", ylab = "IC")
arrows(1:m, ici, 1:m, ics, angle = 90, length = 0.05, code = 3, col = color)
abline(h = mux, lty = 3)

detach(tmp)
```

- b) Repetir el apartado anterior considerando muestras de una $Exp(1)$. ¿Qué ocurre con la cobertura del intervalo de confianza obtenido bajo normalidad?

Ejecutar el código del apartado a) del ejercicio 2.

IC para la media con varianza conocida (bajo normalidad)

```
alfa <- 0.05
z <- qnorm(1 - alfa/2)
muestras2$ici <- muestras2$mean - z*sdexp/sqrt(nx)
muestras2$ics <- muestras2$mean + z*sdexp/sqrt(nx)
```

Cobertura de las estimaciones por IC:

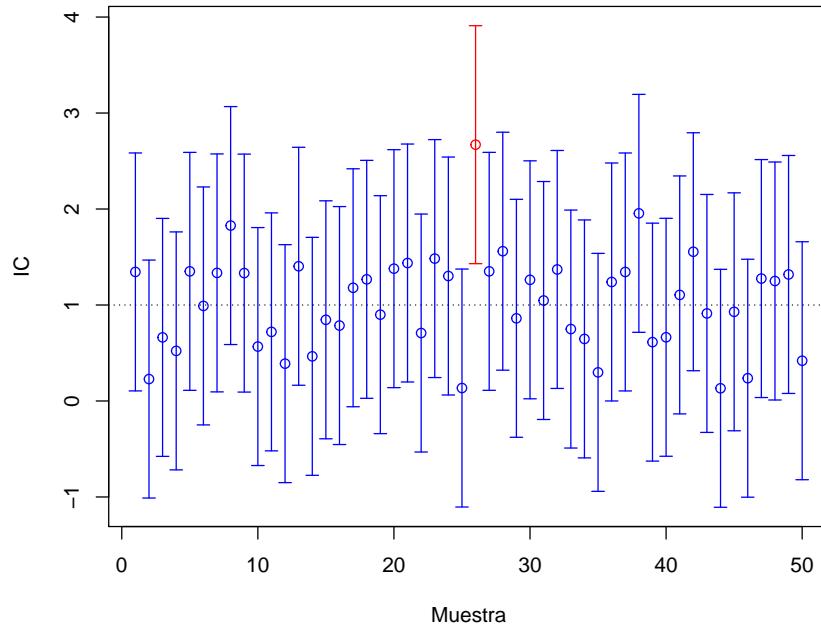


Figura 8.3: Cobertura de las estimaciones por IC.

```
muestras2$cob <- (muestras2$ici < muexp) & (muexp < muestras2$ics)
ncob <- sum(muestras2$cob) # N° de intervalos que contienen la verdadera media
ncob
```

```
## [1] 469
100*ncob/nsim      # Proporción de intervalos
## [1] 93.8
100*(1 - alfa)    # Proporción teórica bajo normalidad
## [1] 95
```

Como ejemplo ilustrativo, generamos el gráfico de los primeros 100 intervalos:

```
m <- 100
tmp <- muestras2[1:m,]
attach(tmp)
color <- ifelse(cob, "blue", "red")
plot(1:m, mean, col = color, ylim = c(min(ici),max(ics)),
     xlab = "Muestra", ylab = "IC")
arrows(1:m, ici, 1:m, ics, angle = 90, length = 0.05, code = 3, col = color)
abline(h = muexp, lty = 3)

detach(tmp)
```

c) ¿Qué ocurre si aumentamos el tamaño muestral a 50?

Ejecutar el código del ejercicio anterior fijando `nx <- 50` y el del apartado anterior.

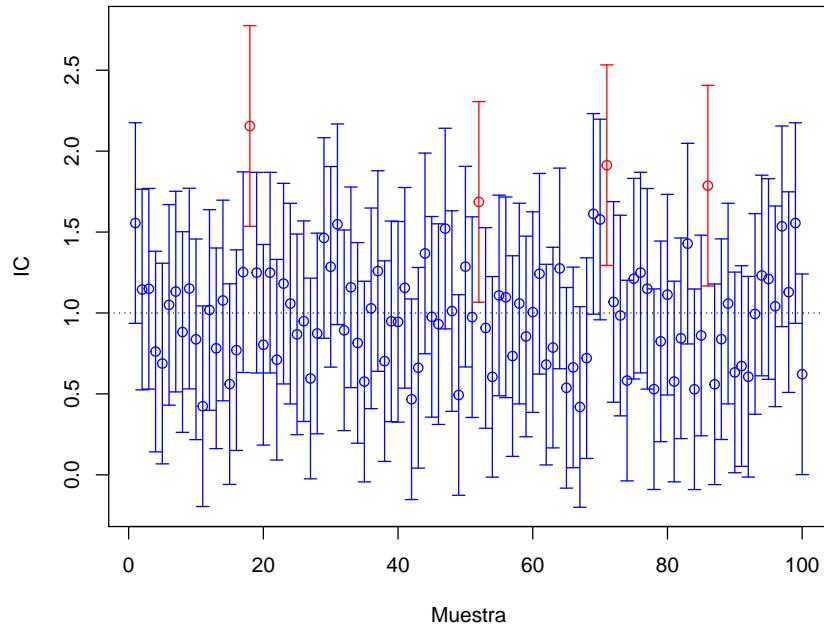


Figura 8.4: Cobertura de las estimaciones por IC (bajo normalidad).

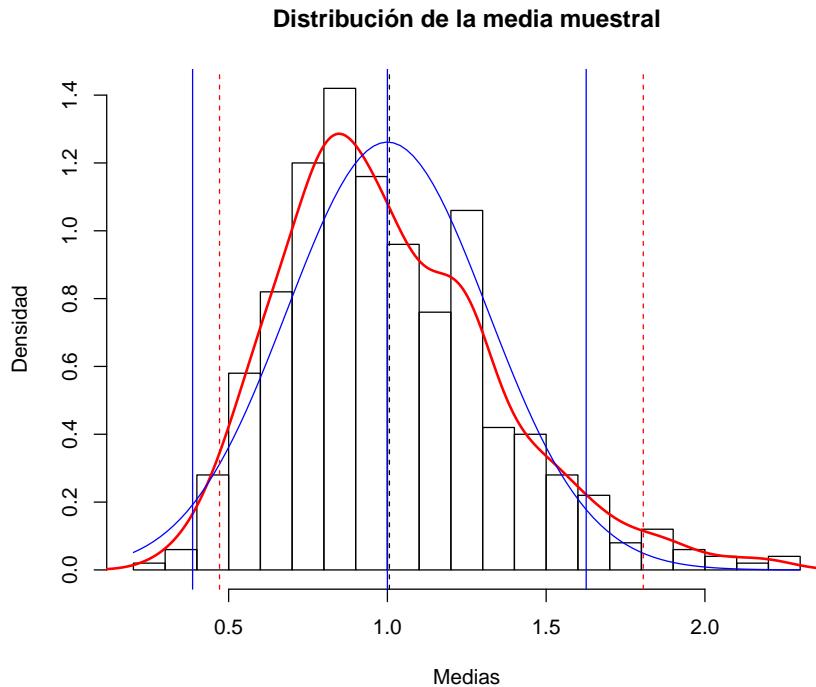
En los apartados b) y c) podíamos considerar bootstrap descrito al final de este capítulo.

Podemos aproximar por simulación los intervalos de probabilidad de la media muestral (tendríamos una idea del valor esperado de lo que obtendríamos con el bootstrap percentil; en este caso el estimador es insesgado...):

```
# Distribución de la media muestral
hist(muestras2$mean, freq=FALSE, breaks="FD",
      main="Distribución de la media muestral", xlab="Medias", ylab="Densidad")
# Densidad observada (estimación)
lines(density(muestras2$mean), lwd=2, col='red')
# Densidad teórica (bajo normalidad)
curve(dnorm(x,muexp,sdexp/sqrt(nx)), col="blue", add=TRUE)
# Aproximación por simulación del valor esperado de la media muestral
abline(v=mean(muestras2$mean), lty=2)
# Valor esperado de la media muestral (teórico)
abline(v=muexp, col="blue")
# IP bajo normalidad
ic.aprox <- apply(muestras2[,c('ici','ics')], 2, mean)
## ic.aprox
##      ici      ics
## 0.3865199 1.6261099
# Intervalo de probabilidad para la media muestral aproximado bajo normalidad
abline(v = ic.aprox, col='blue')

# Intervalo de probabilidad para la media muestral (aproximado por simulación)
ic.sim <- quantile(muestras2$mean, c(alfa/2, 1 - alfa/2))
## ic.sim
##      2.5%      97.5%
```

```
## 0.4714233 1.8059094
# IP (aprox.)
abline(v=ic.sim, lty=2, col='red')
```



Nota: . Estimaciones puntuales, por intervalo de confianza y contrastes de hipótesis para la media con varianza desconocida bajo normalidad se pueden obtener con la función `t.test`.

Ejercicio 8.4.

El Intervalo de confianza para una proporción construido usando la aproximación normal tiene un mal comportamiento cuando el tamaño de la muestra es pequeño. Una simple y efectiva mejora consiste en añadir a la muestra $2a$ elementos, a éxitos y a fracasos. Así el intervalo de confianza al $(1 - \alpha) 100\%$ para una proporción mejorado es:

$$IC_{1-\alpha}^a(p) = \left(\tilde{p} - z_{1-\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}, \tilde{p} + z_{1-\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \right),$$

siendo $\tilde{n} = n + 2a$, $\tilde{p} = \frac{np + a}{\tilde{n}}$.

En el caso de $a = 2$ se denomina IC Agresti-Coull.

- a) Teniendo en cuenta que la v.a. $X = n\hat{p} \sim \mathcal{B}(n, p)$, obtener y representar gráficamente la cobertura teórica del intervalo de confianza estándar ($a = 0$) de una proporción para una muestra de tamaño $n = 30$, $\alpha = 0.05$ y distintos valores de p (`p.teor <- seq(1/n, 1 - 1/n, length = 1000)`).

Parámetros:

```
n <- 30
alpha <- 0.05
adj <- 0 # (adj <- 2 para Agresti-Coull)
```

Probabilidades teóricas:

```
m <- 1000
p.teor <- seq(1/n, 1 - 1/n, length = m)
```

Posibles resultados:

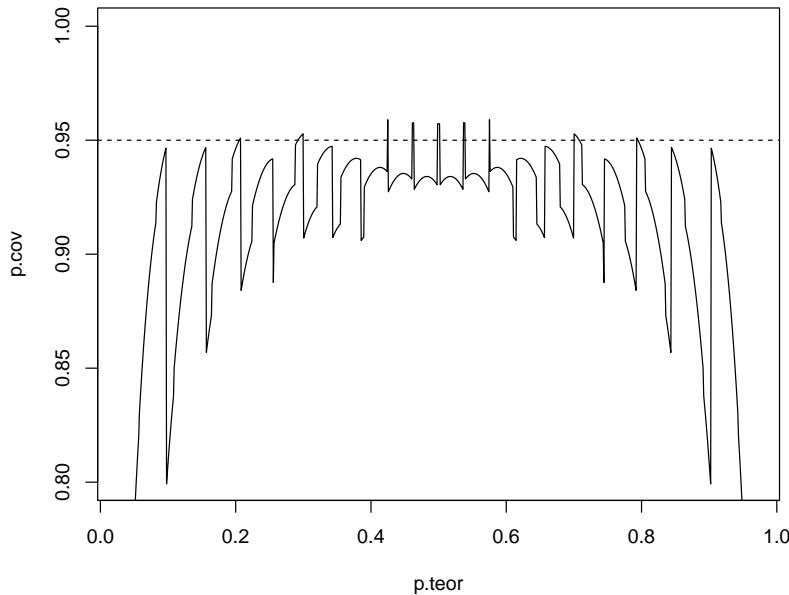
```
x <- 0:n
p.est <- (x + adj)/(n + 2 * adj)
ic.err <- qnorm(1 - alpha/2) * sqrt(p.est * (1 - p.est)/(n + 2 * adj))
lcl <- p.est - ic.err
ucl <- p.est + ic.err
```

Recorrer prob. teóricas:

```
p.cov <- numeric(m)
for (i in 1:m) {
  # cobertura de los posibles intervalos
  cover <- (p.teor[i] >= lcl) & (p.teor[i] <= ucl)
  # prob. de los posibles intervalos
  p.rel <- dbinom(x[cover], n, p.teor[i])
  # prob. total de cobertura
  p.cov[i] <- sum(p.rel)
}
```

Gráfico coberturas:

```
plot(p.teor, p.cov, type = "l", ylim = c(1 - 4 * alpha, 1))
abline(h = 1 - alpha, lty = 2)
```



Fuente Suess y Trumbo (2010).

- b) Repetir el apartado anterior considerando intervalos de confianza Agresti-Coull ($a = 2$).

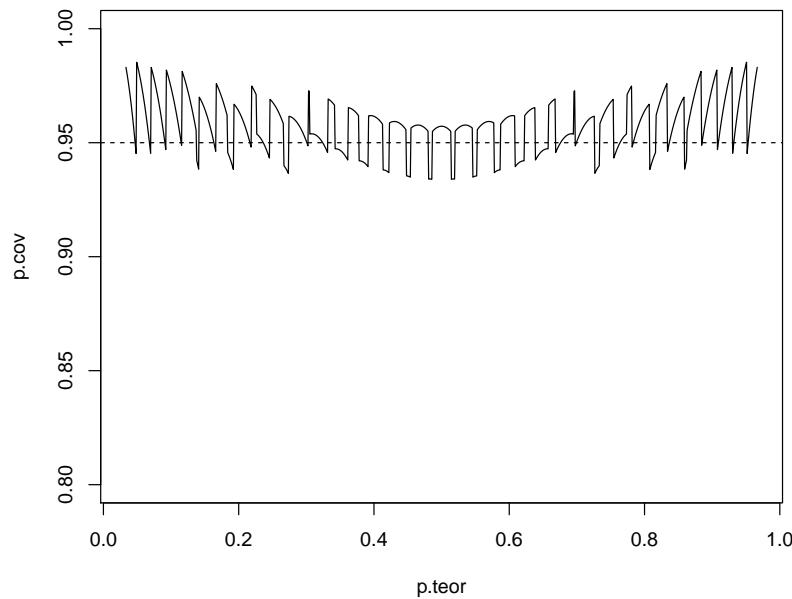
Parámetros:

```

n <- 30
alpha <- 0.05
adj <- 2 # Agresti-Coull

# Probabilidades teóricas:
m <- 1000
p.teor <- seq(1/n, 1 - 1/n, length = m)
# Posibles resultados:
x <- 0:n
p.est <- (x + adj)/(n + 2 * adj)
ic.err <- qnorm(1 - alpha/2) * sqrt(p.est * (1 - p.est)/(n + 2 * adj))
lcl <- p.est - ic.err
ucl <- p.est + ic.err
# Recorrer prob. teóricas:
p.cov <- numeric(m)
for (i in 1:m) {
  # cobertura de los posibles intervalos
  cover <- (p.teor[i] >= lcl) & (p.teor[i] <= ucl)
  # prob. de los posibles intervalos
  p.rel <- dbinom(x[cover], n, p.teor[i])
  # prob. total de cobertura
  p.cov[i] <- sum(p.rel)
}
# Gráfico coberturas:
plot(p.teor, p.cov, type = "l", ylim = c(1 - 4 * alpha, 1))
abline(h = 1 - alpha, lty = 2)

```



- c) Repetir el apartado anterior empleando simulación para aproximar la cobertura.

Parámetros:

```

n <- 30
alpha <- 0.05
adj <- 2 #' (2 para Agresti-Coull)

set.seed(54321)
nsim <- 500
# Probabilidades teóricas:
m <- 1000
p.teor <- seq(1/n, 1 - 1/n, length = m)

```

Recorrer prob. teóricas:

```

# m <- length(p.teor)
p.cov <- numeric(m)
for (i in 1:m) {
  # Equivalente a simular nsim muestras de tamaño n
  # ry <- matrix(rbinom(n*nsim, 1, p.teor[i]), ncol=n)
  # rx <- apply(ry, 1, sum)
  rx <- rbinom(nsim, n, p.teor[i])
  p.est <- (rx + adj)/(n + 2 * adj)
  ic.err <- qnorm(1 - alpha/2) * sqrt(p.est * (1 - p.est)/(n + 2 * adj))
  p.cov[i] <- mean( abs(p.est - p.teor[i]) < ic.err )
}

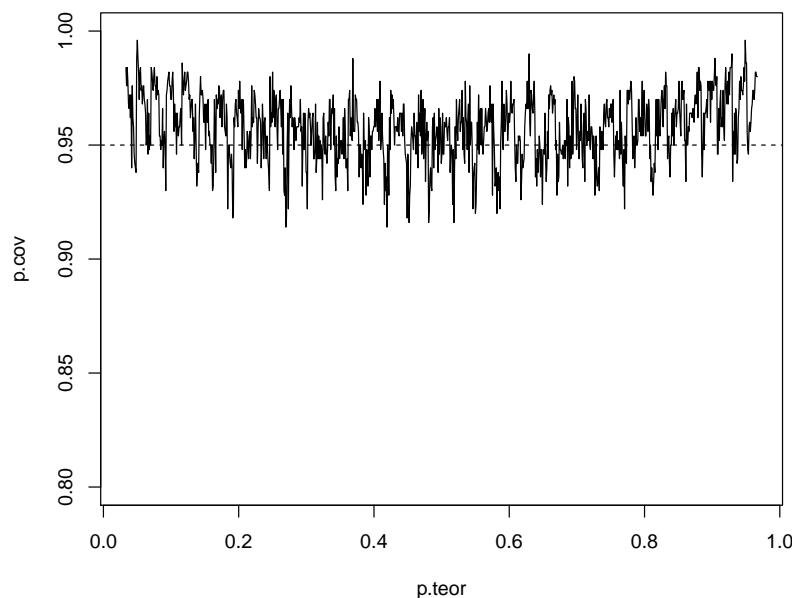
```

Representar:

```

plot(p.teor, p.cov, type = "l", ylim = c(1 - 4 * alpha, 1))
abline(h = 1 - alpha, lty = 2)

```



8.4 Contrastes de hipótesis

Ejercicio 8.5.

En el tema 2 se propuso el análisis de la bondad de ajuste de un generador de números pseudo-aleatorios mediante el test de Kolmogorov-Smirnov. Sin embargo, si H_0 es compuesta (los parámetros desconocidos se estiman por máxima verosimilitud y se trabaja con \hat{F}_0) los cuantiles de la distribución (asintótica) de D_n pueden ser demasiado conservativos y sería preferible utilizar la distribución exacta.

- Analizar el comportamiento del contraste de Kolmogorov-Smirnov para contrastar normalidad empleando repetidamente este test, considerando 1000 pruebas con muestras de tamaño 30 de una $\mathcal{N}(0, 1)$. Comparar gráficamente el ajuste de la distribución del p -valor a la de referencia (estudiar el tamaño del contraste).

Valores iniciales:

```
set.seed(54321)
nx <- 30
mx <- 0
sx <- 1
nsim <- 1000
estadistico <- numeric(nsim)
pvalor <- numeric(nsim)
```

Realizar contrastes

```
for(isim in 1:nsim) {
  rx <- rnorm(nx, mx, sx)
  tmp <- ks.test(rx, "pnorm", mean(rx), sd(rx))
  estadistico[isim] <- tmp$statistic
  pvalor[isim] <- tmp$p.value
}
```

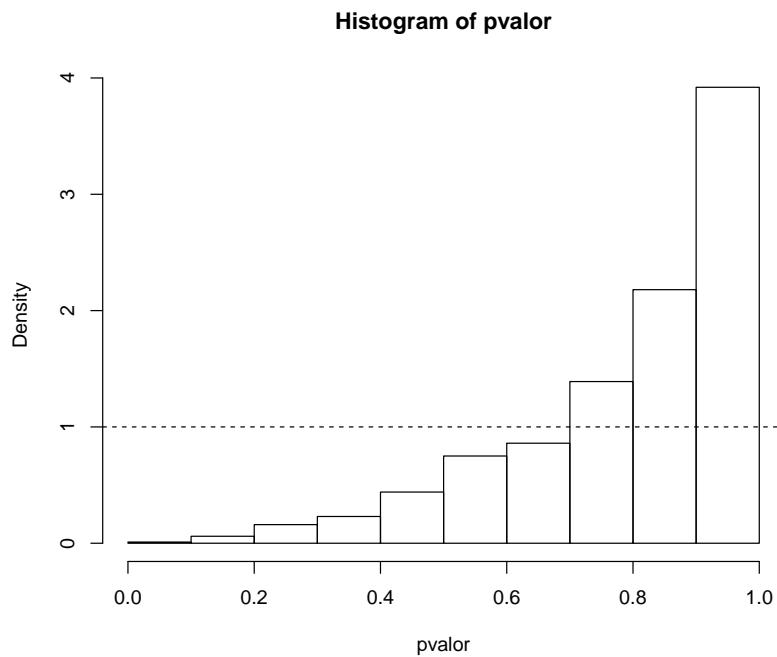
Proporción de rechazos:

```
{
  cat("\nProporción de rechazos al 1% =", mean(pvalor < 0.01), "\n")
  cat("Proporción de rechazos al 5% =", mean(pvalor < 0.05), "\n")
  cat("Proporción de rechazos al 10% =", mean(pvalor < 0.1), "\n")
}
```

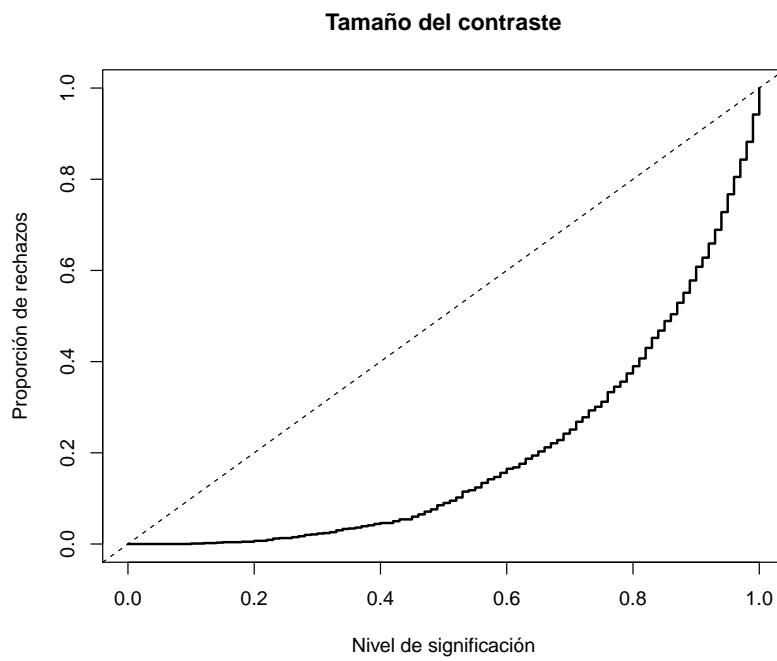
```
##
## Proporción de rechazos al 1% = 0
## Proporción de rechazos al 5% = 0
## Proporción de rechazos al 10% = 0.001
```

Análisis de los p -valores:

```
hist(pvalor, freq=FALSE)
abline(h=1, lty=2)  # curve(dunif(x, 0, 1), add=TRUE)
```



```
# Distribución empírica
curve(ecdf(pvalor))(x), type = "s", lwd = 2,
      main = 'Tamaño del contraste', ylab = 'Proporción de rechazos',
      xlab = 'Nivel de significación')
abline(a=0, b=1, lty=2) # curve(punif(x, 0, 1), add = TRUE)
```



- b) Repetir el apartado anterior considerando el test de Lilliefors (rutina `lillie.test` del paquete `nortest`).

```
library(nortest, quietly = TRUE)
```

Valores iniciales:

```
set.seed(54321)
nx <- 30
mx <- 0
sx <- 1
nsim <- 1000
estadistico <- numeric(nsim)
pvalor <- numeric(nsim)
```

Realizar contrastes

```
for(isim in 1:nsim) {
  rx <- rnorm(nx, mx, sx)
  # tmp <- ks.test(rx, "pnorm", mean(rx), sd(rx))
  tmp <- lillie.test(rx)
  estadistico[isim] <- tmp$statistic
  pvalor[isim] <- tmp$p.value
}
```

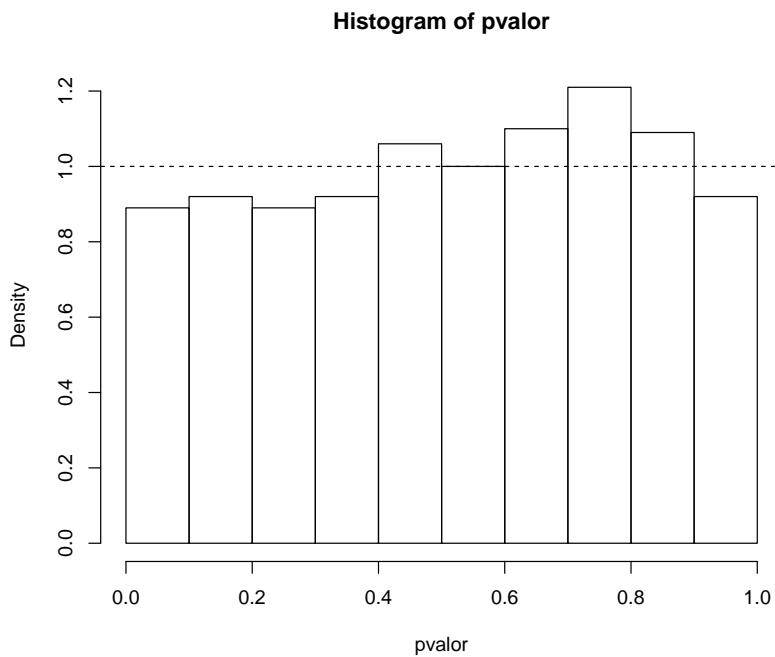
Proporción de rechazos:

```
{
  cat("\nProporción de rechazos al 1% =", mean(pvalor < 0.01), "\n")
  cat("Proporción de rechazos al 5% =", mean(pvalor < 0.05), "\n")
  cat("Proporción de rechazos al 10% =", mean(pvalor < 0.1), "\n")
}
```

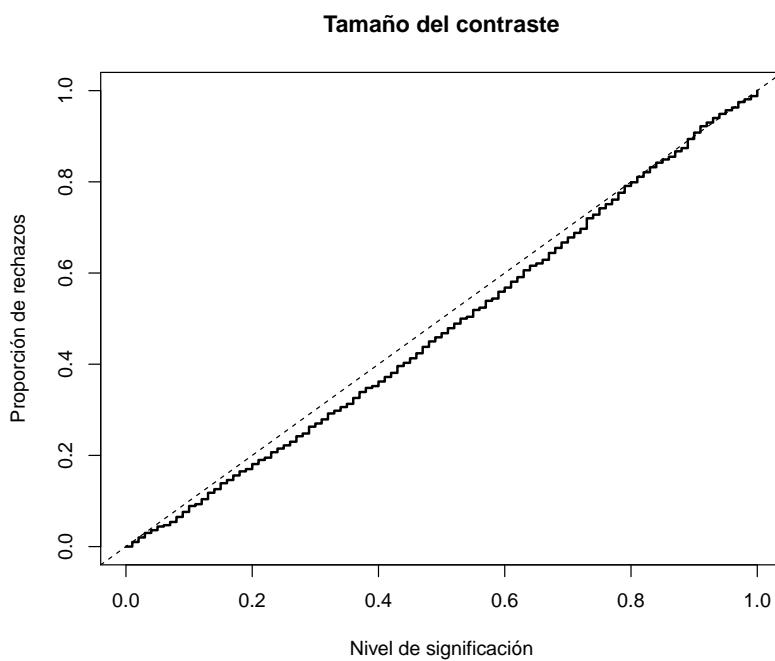
```
##
## Proporción de rechazos al 1% = 0.01
## Proporción de rechazos al 5% = 0.044
## Proporción de rechazos al 10% = 0.089
```

Análisis de los p-valores:

```
hist(pvalor, freq=FALSE)
abline(h=1, lty=2)  # curve(dunif(x,0,1), add=TRUE)
```



```
# Distribución empírica
curve(ecdf(pvalor))(x), type = "s", lwd = 2, main = 'Tamaño del contraste',
      ylab = 'Proporción de rechazos', xlab = 'Nivel de significación')
abline(a=0, b=1, lty=2) # curve(unif(x, 0, 1), add = TRUE)
```



- c) Repetir el apartado a) contrastando una distribución exponencial y considerando 500 pruebas con muestras de tamaño 30 de una $Exp(1)$.

Valores iniciales:

```
set.seed(54321)
nx <- 30
ratex <- 1
nsim <- 500
estadistico <- numeric(nsim)
pvalor <- numeric(nsim)
```

Realizar contrastes

```
for(isim in 1:nsim) {
  rx <- rexp(nx, ratex)
  tmp <- ks.test(rx, "pexp", 1/mean(rx))
  estadistico[isim] <- tmp$statistic
  pvalor[isim] <- tmp$p.value
}
```

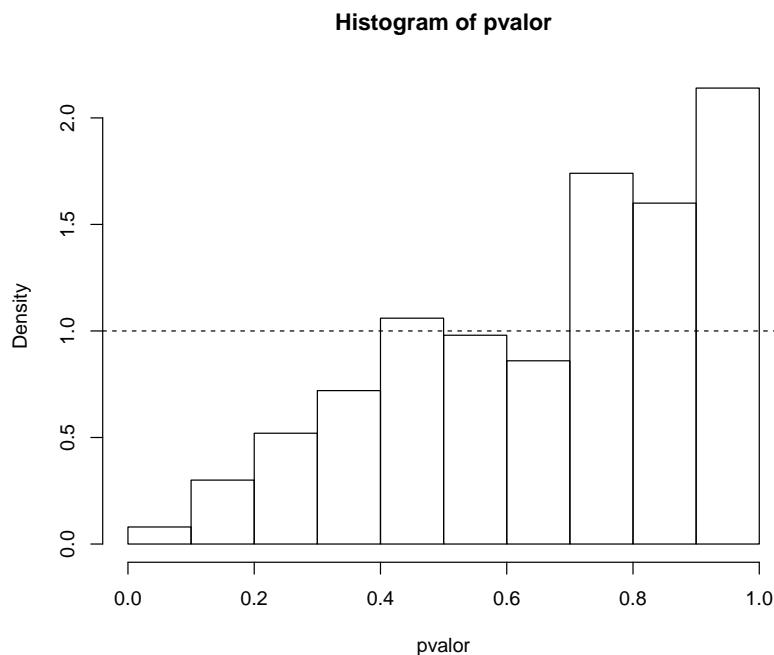
Proporción de rechazos:

```
{
  cat("\nProporción de rechazos al 1% =", mean(pvalor < 0.01), "\n")
  cat("Proporción de rechazos al 5% =", mean(pvalor < 0.05), "\n")
  cat("Proporción de rechazos al 10% =", mean(pvalor < 0.1), "\n")
}

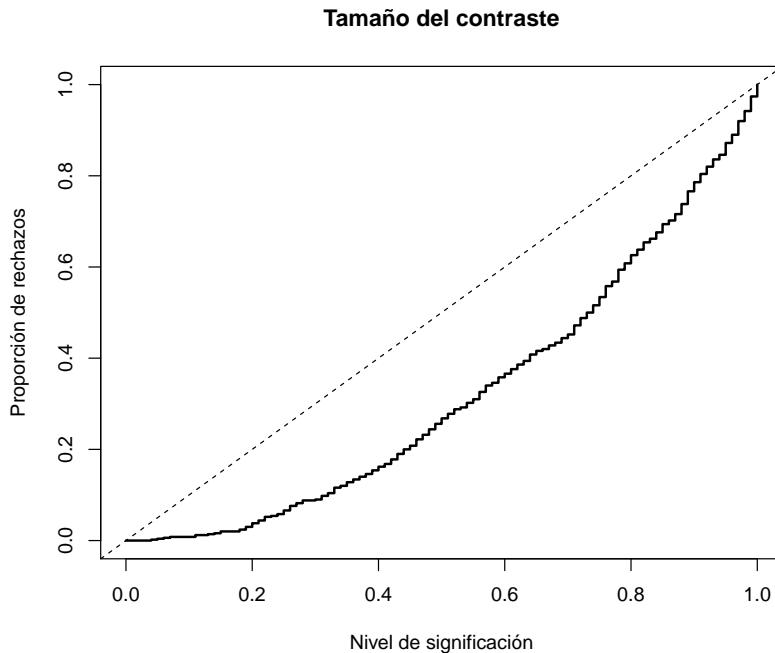
##
## Proporción de rechazos al 1% = 0
## Proporción de rechazos al 5% = 0.004
## Proporción de rechazos al 10% = 0.008
```

Análisis de los p-valores:

```
hist(pvalor, freq=FALSE)
abline(h=1, lty=2)  # curve(dunif(x,0,1), add=TRUE)
```



```
# Distribución empírica
curve(ecdf(pvalor)(x), type = "s", lwd = 2,
      main = 'Tamaño del contraste', ylab = 'Proporción de rechazos',
      xlab = 'Nivel de significación')
abline(a=0, b=1, lty=2) # curve(punif(x, 0, 1), add = TRUE)
```



- d) Diseñar una rutina que permita realizar el contraste KS de bondad de ajuste de una variable exponencial aproximando el p -valor por simulación y repetir el apartado anterior empleando esta rutina.

```
ks.exp.sim <- function(x, nsim = 10^3) {
  DNAME <- deparse(substitute(x))
  METHOD <- "Kolmogorov-Smirnov Test of pexp by simulation"
  n <- length(x)
  RATE <- 1/mean(x)
  ks.exp.stat <- function(x, rate=1/mean(x)) {
    DMinus <- pexp(sort(x), rate=rate) - (0:(n - 1))/n
    DPlus <- 1/n - DMinus
    Dn = max(c(DMinus, DPlus))
  }
  STATISTIC <- ks.exp.stat(x, rate = RATE)
  names(STATISTIC) <- "Dn"
  # PVAL <- 0
  # for(i in 1:nsim) {
  #   rx <- rexp(n, rate = RATE)
  #   if (STATISTIC <= ks.exp.stat(rx)) PVAL <- PVAL+1
  # }
  # PVAL <- PVAL/nsim
  # PVAL <- PVAL/(nsim + 1)
  # PVAL <- (PVAL + 1)/(nsim + 2)
  rx <- matrix(rexp(n*nsim, rate = RATE), ncol=n)
  PVAL <- mean(STATISTIC <= apply(rx, 1, ks.exp.stat))
```

```

    return(structure(list(statistic = STATISTIC, alternative = "two.sided",
                          p.value = PVAL, method = METHOD, data.name = DNAME),
                  class = "htest"))
}

```

Simulación:

```

set.seed(54321)
nx <- 30
ratex <- 1
nsim <- 500
estadistico <- numeric(nsim)
pvalor <- numeric(nsim)

```

Realizar contrastes

```

for(isim in 1:nsim) {
  rx <- rexp(nx, ratex)
  # tmp <- ks.test(rx, "pexp", 1/mean(rx))
  tmp <- ks.exp.sim(rx, nsim = 200)
  estadistico[isim] <- tmp$statistic
  pvalor[isim] <- tmp$p.value
}

```

Proporción de rechazos:

```

{
  cat("\nProporción de rechazos al 1% =", mean(pvalor < 0.01), "\n")
  cat("Proporción de rechazos al 5% =", mean(pvalor < 0.05), "\n")
  cat("Proporción de rechazos al 10% =", mean(pvalor < 0.1), "\n")
}

```

```

##
## Proporción de rechazos al 1% = 0.008
## Proporción de rechazos al 5% = 0.058
## Proporción de rechazos al 10% = 0.106

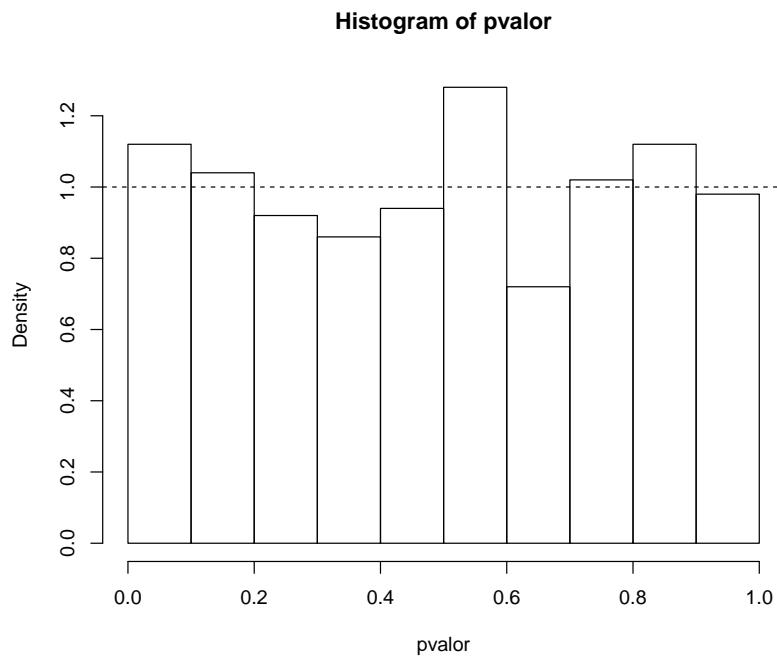
```

Análisis de los p-valores:

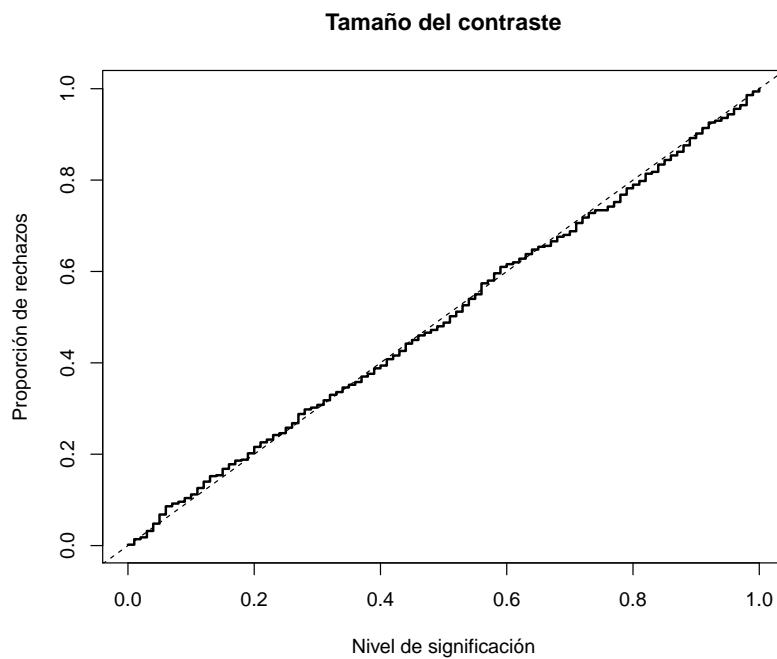
```

hist(pvalor, freq=FALSE)
abline(h=1, lty=2)  # curve(dunif(x,0,1), add=TRUE)

```



```
# Distribución empírica
curve(ecdf(pvalor))(x), type = "s", lwd = 2,
      main = 'Tamaño del contraste', ylab = 'Proporción de rechazos',
      xlab = 'Nivel de significación')
abline(a=0, b=1, lty=2) # curve(punif(x, 0, 1), add = TRUE)
```



- e) Estudiar la potencia de los contrastes de los apartados c) y d), considerando como alternativa una distribución Weibull.

La distribución exponencial es un caso particular de la Weibull: `dexp(x, ratex) == dweibull(x, 1, 1/ratex)`. Estudiamos lo que ocurre al desplazar `dweibull(x, shape, 1/ratex)` con $0 < \text{shape} < 2$.

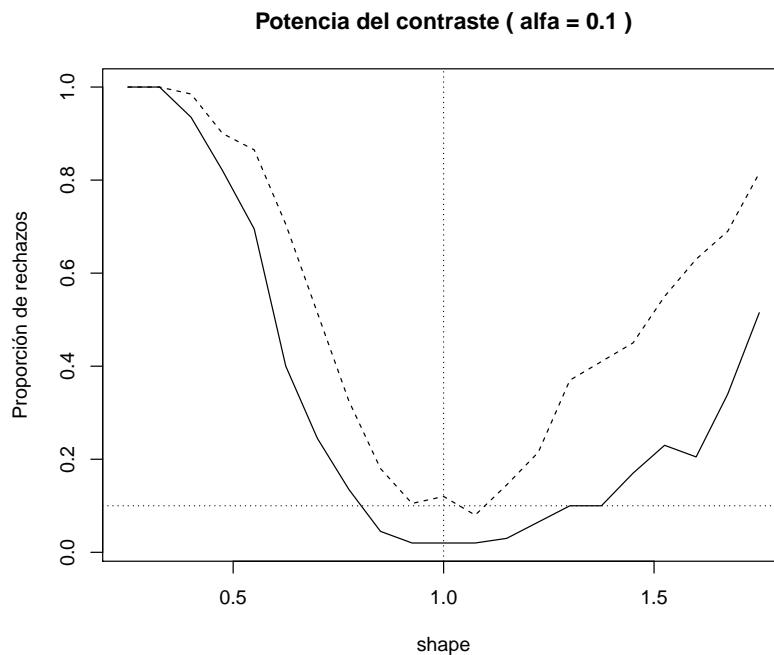
CUIDADO: las simulaciones pueden requerir de mucho tiempo de computación (consideramos valores pequeños de `nx` y `nsim` en datos y en `ks.exp.sim`).

```
set.seed(54321)
nx <- 20
ratex <- 1      # Puede ser interesante representarlo variando rate
nsim <- 200
alfa <- 0.1      # Puede ser interesante representarlo variando alfa

shapex <- seq(0.25, 1.75, len=21)
preject <- numeric(length(shapex)) # Porporciones de rechazos con ks.test
ks.test.p <- function(x) ks.test(x, "pexp", 1/mean(x))$p.value
preject2 <- preject # Porporciones de rechazos con ks.exp.sim
ks.exp.sim.p <- function(x) ks.exp.sim(x, 200)$p.value

for (i in seq_along(shapex)) {
  rx <- matrix(rweibull(nx*nsim, shape = shapex[i], scale = 1/ratex), ncol=nx)
  preject[i] <- mean(apply(rx, 1, ks.test.p) <= alfa )
  preject2[i] <- mean(apply(rx, 1, ks.exp.sim.p) <= alfa )
}

plot(shapex, preject, type="l", main = paste("Potencia del contraste ( alfa =", alfa, ")"),
     xlab = "shape", ylab = "Proporción de rechazos")
lines(shapex, preject2, lty = 2)
abline(h = alfa, v = 1, lty = 3)
```



El estadístico de Kolmogorov-Smirnov `Dn = max(c(DMinus, DPlus))` tiene ventajas desde el punto de vista

teórico, pero puede no ser muy potente para detectar diferencias entre la distribución bajo la hipótesis nula y la distribución de los datos. La ventaja de la aproximación por simulación es que no estamos atados a resultados teóricos y podemos emplear el estadístico que se considere oportuno (la principal desventaja es el tiempo de computación). Por ejemplo, podríamos pensar en utilizar como estadístico la suma de los errores en valor absoluto del correspondiente gráfico PP, y solo habría que cambiar el estadístico D_n en la función `ks.exp.sim` por $D_n = \text{sum}(\text{abs}(1:n - 0.5)/n - \text{pexp}(\text{sort}(x), \text{rate}=rate))$.

8.5 Comparación de estimadores

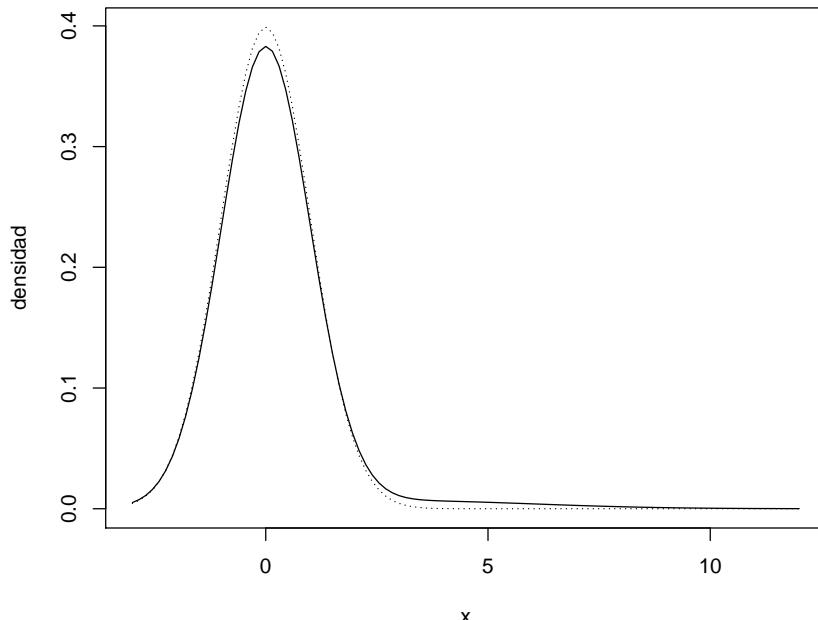
Ejercicio 8.6.

Supongamos que estamos interesados en estudiar el efecto de datos atípicos en la estimación de la media teórica mediante la media y la mediana muestrales. Consideraremos una v.a. con distribución normal contaminada, en la que una observación procede de una $N(0, 1)$ con probabilidad 0.95 y de una $N(3, 3^2)$ con probabilidad 0.05 (mixtura). Se puede generar una muestra de esta variable mediante los comandos:

```
p.sim <- rbinom(n, 1, 0.05)
dat.sim <- rnorm(n, 3*p.sim, 1+2*p.sim)
```

Función de densidad:

```
curve(dnorm(x, 0, 1), -3, 12, ylab = 'densidad', lty = 3)
curve(0.95*dnorm(x, 0, 1) + 0.05*dnorm(x, 3, 3), add = TRUE)
```



Nota: . También es habitual simular este tipo de datos generando un porcentaje alto de valores (en este caso un 95%) de la distribución base ($N(0, 1)$) y el resto (5%) de la distibución “contaminadora” ($N(3, 3^2)$), aunque se suele considerar un porcentaje de contaminación del 1% o inferior. En el tema 7 se describirá el método de composición para simular mixturas.

- a) Aproximar mediante simulación (500 generaciones) el sesgo y error estándar de la media y la mediana en el caso de una muestra de tamaño $n = 100$ (suponiendo que se pretende estimar la media no contaminada 0).

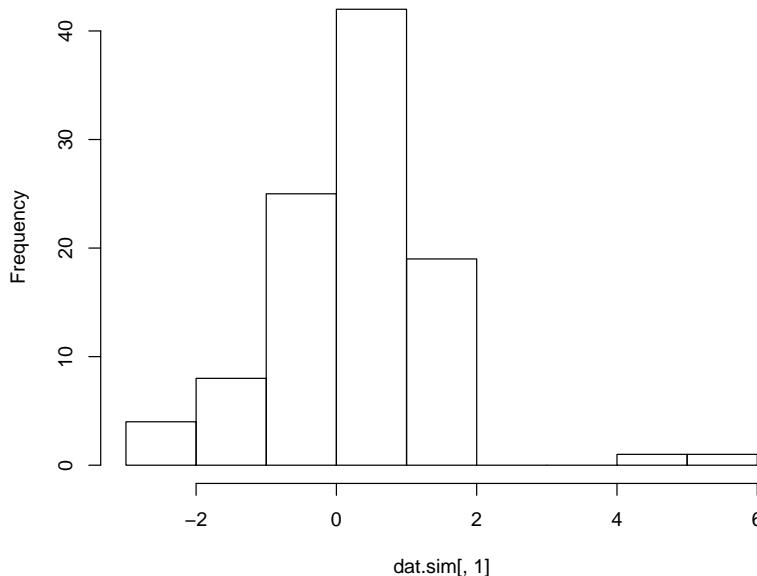
```
# media y mediana
xsd <- 1
xmed <- 0
ndat <- 100
nsim <- 500

# for (isim in 1:nsim) # evitar matrix y apply
set.seed(1)
ntsim <- ndat*nsim
p.sim <- rbinom(ntsim, 1, 0.05)
dat.sim <- rnorm(ntsim, 3*p.sim, 1+2*p.sim)
dat.sim <- matrix(dat.sim, ncol=nsim)
```

Cada columna es una muestra

```
str(dat.sim[,1])
## num [1:100] 0.197 -0.42 1.163 -0.406 0.744 ...
hist(dat.sim[,1])
```

Histogram of dat.sim[, 1]



Calculamos los estimadores:

```
mean.sim <- apply(dat.sim, 2, mean)
median.sim <- apply(dat.sim, 2, median)
```

Estimamos sus características:

```
mean(mean.sim) # Coincide con el sesgo (media teórica es 0)
```

```
## [1] 0.1459986
```

```

sd(mean.sim)
## [1] 0.1349537
mean(median.sim) # Coincide con el sesgo (media teórica es 0)

## [1] 0.04453509
sd(median.sim)
## [1] 0.1300611

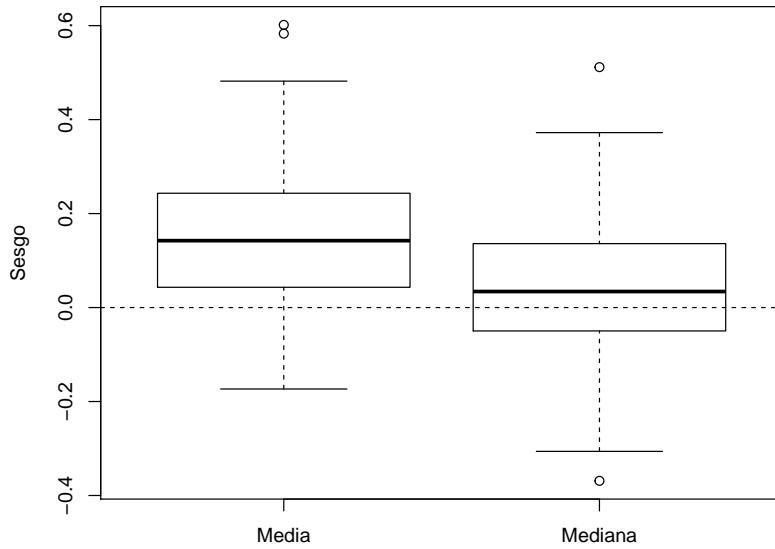
```

Sesgo:

```

boxplot(mean.sim-xmed, median.sim-xmed,
        names=c("Media", "Mediana"), ylab="Sesgo")
abline(h = 0, lty = 2)

```

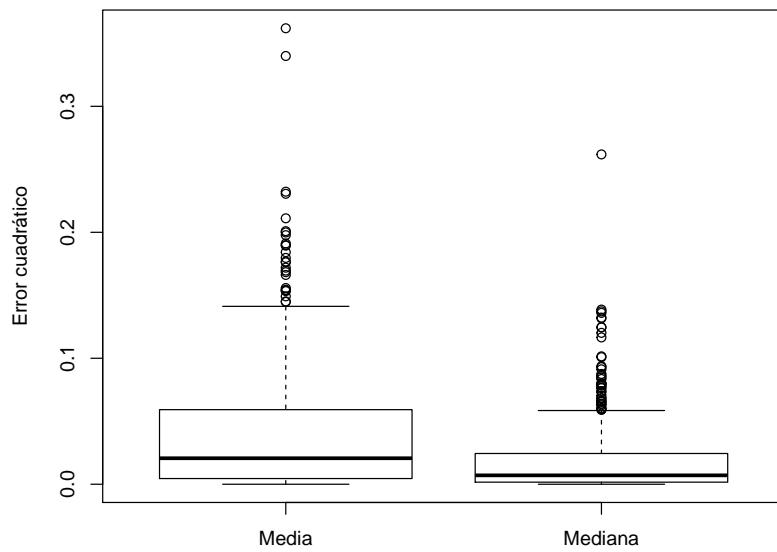


Error cuadrático:

```

boxplot((mean.sim-xmed)^2, (median.sim-xmed)^2,
        names=c("Media", "Mediana"), ylab="Error cuadrático")

```



Estadísticos error cuadrático:

```
# SE media
summary((mean.sim-xmed)^2)

##      Min.   1st Qu.    Median     Mean   3rd Qu.    Max.
## 0.0000005 0.0045072 0.0206272 0.0394917 0.0591531 0.3619587

# SE mediana
summary((median.sim-xmed)^2)

##      Min.   1st Qu.    Median     Mean   3rd Qu.    Max.
## 0.0000001 0.0016481 0.0070625 0.0188654 0.0243903 0.2618368
```

8.6 Remuestreo Bootstrap

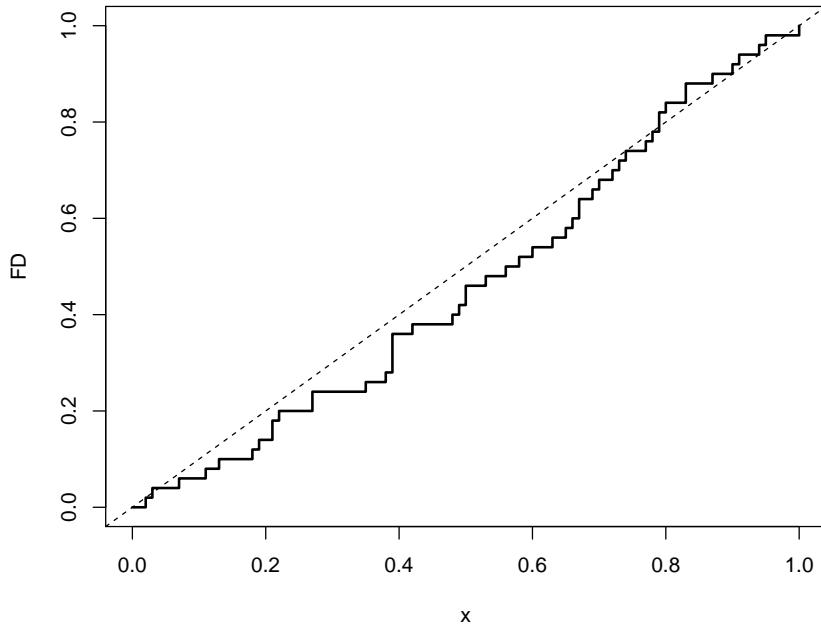
8.6.1 Idea:

Consideramos un conjunto de datos simulado:

```
set.seed(1)
data <- runif(50)
```

La idea es aproximar características poblacionales por las correspondientes de la distribución empírica de los datos observados:

```
# Distribución bootstrap
curve(ecdf(data)(x), ylab = "FD", type = "s", lwd = 2)
# Distribución teórica
abline(a = 0, b = 1, lty = 2)
```



Las características de la distribución empírica se pueden aproximar mediante simulación. En el caso i.i.d. esto puede ser implementado mediante remuestreo, realizando repetidamente muestreo aleatorio con reemplazamiento del conjunto de datos original (manteniendo el tamaño muestral):

```
# Muestra bootstrap
xboot <- sample(data, replace=TRUE)
```

8.6.2 Métodos de remuestreo Bootstrap

Método de remuestreo (Efron, 1979) utilizado para aproximar características de la distribución en el muestreo de un estadístico:

- Aproximación del sesgo y de la varianza.
- Construcción de intervalos de confianza
- Realizar contrastes de hipótesis.

De utilidad cuando no se dispone la distribución exacta, no es adecuado emplear la distribución asintótica, etc. La idea es aproximar características poblacionales por las correspondientes de la distribución empírica de los datos observados. En el caso i.i.d. esto puede ser implementado mediante remuestreo, realizando repetidamente **muestreo aleatorio con reemplazamiento del conjunto de datos original** (manteniendo el tamaño muestral).

Si $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$ es una muestra i.i.d. F_θ y $\hat{\theta} = T(\mathbf{x})$ es un estimador de θ . Para $b = 1, \dots, B$:

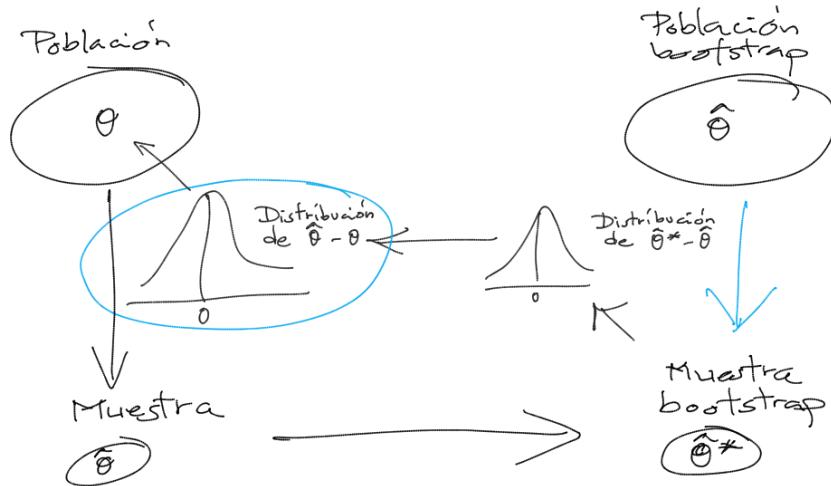
- $\mathbf{x}_b^* = (x_{b1}^*, x_{b2}^*, \dots, x_{bn}^*)^t$ muestra bootstrap obtenida mediante muestreo con reemplazamiento de $\{x_1, x_2, \dots, x_n\}$.
- $\hat{\theta}_b^* = T(\mathbf{x}_b^*)$ valor del estadístico en la muestra bootstrap.

La idea original (bootstrap natural, Efron) es que la variabilidad de $\hat{\theta}_b^*$ en torno a $\hat{\theta}$ aproxima la variabilidad de $\hat{\theta}$ en torno a θ :

- La distribución de $\hat{\theta}_b^* - \hat{\theta}$ aproxima la distribución de $\hat{\theta} - \theta$.

En general podríamos decir que:

- la muestra es a la población lo que la muestra bootstrap es a la muestra.



Para información adicional sobre bootstrap ver p.e.: Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press

8.6.3 Ejemplo

Como ejemplo consideraremos una muestra de tamaño $n = 100$ de una normal estándar para tratar de aproximar el sesgo y error estándar de la media y la mediana mediante bootstrap.

```
# dat <- dat.sim[, 1]
set.seed(54321)
ndat <- 100
datmed <- 0
datsd <- 1
dat <- rnorm(ndat, mean=datmed, sd=datsd)
```

Consideramos 1000 réplicas bootstrap:

```
nboot <- 1000
```

Es habitual tomar $nboot + 1$ entero múltiplo de 100 e.g. $nboot = 999$ ó 1999 (facilita el cálculo de percentiles, orden $(nboot + 1)*p$)

El valor del estadístico en la muestra es:

```
stat.dat <- mean(dat)
```

Generamos las réplicas bootstrap:

```
set.seed(1)
stat.boot <- numeric(nboot)
for (i in 1:nboot) {
  dat.boot <- sample(dat, replace=TRUE)
  stat.boot[i] <- mean(dat.boot)
}
# Valor esperado bootstrap del estadístico
```

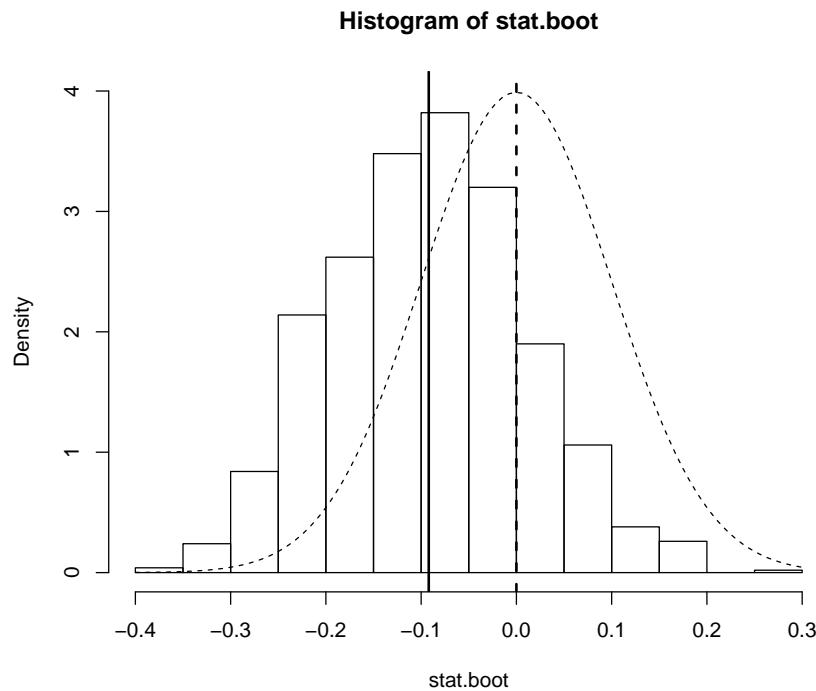
```
mean.boot <- mean(stat.boot)
mean.boot
```

```
## [1] -0.09202607
```

Bootstrap percentil:

```
hist(stat.boot, freq=FALSE, ylim = c(0,4))
abline(v=mean.boot, lwd=2)
# abline(v=stat.dat)

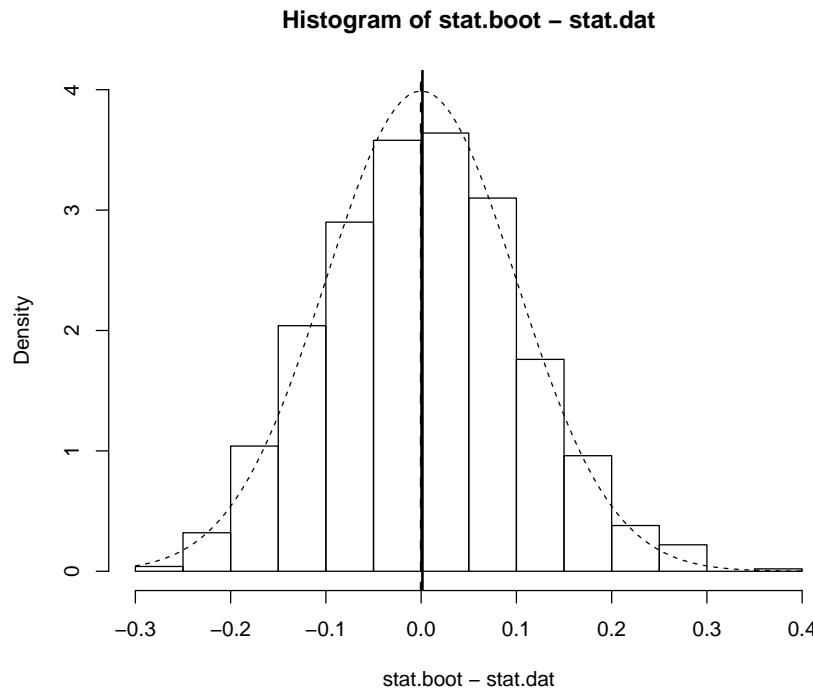
# Distribución poblacional
curve(dnorm(x, datmed, datsd/sqrt(ndat)), lty=2, add=TRUE)
abline(v=datmed, lwd=2, lty=2)
```



Bootstrap natural/básico:

```
hist(stat.boot-stat.dat, freq=FALSE, ylim = c(0,4))
abline(v=mean.boot-stat.dat, lwd=2)

# Distribución poblacional
# Distribución teórica de stat.dat - stat.teor
curve(dnorm(x, 0, datsd/sqrt(ndat)), lty=2, add=TRUE)
abline(v=0, lwd=2, lty=2)
```



Sesgo y error estándar bootstrap:

```
# sesgo (teor=0)
mean.boot = stat.dat

## [1] 0.00135433

# error estándar
sd(stat.boot)

## [1] 0.1027024

# error estándar teórico
datstd/sqrt(ndat)

## [1] 0.1
```

Versión “optimizada” para R:

```
boot.bootstrap <- function(dat, nboot=1000, statistic=mean)
{
  ndat <- length(dat)
  dat.boot <- sample(dat, ndat*nboot, replace=T)
  dat.boot <- matrix(dat.boot, ncol=nboot, nrow=ndat)
  stat.boot <- apply(dat.boot, 2, statistic)
}

fstatistic <- function(dat){
  # mean(dat)
  mean(dat, trim=0.2)
  # median(dat)
  # max(dat)
}
```

```

set.seed(1)
stat.dat <- fstatistic(dat)
stat.boot <- boot.strap(dat, nboot, fstatistic)

res.boot <- c(stat.dat, mean(stat.boot)-stat.dat, sd(stat.boot))
names(res.boot) <- c("Estadístico", "Sesgo", "Err.Std")
res.boot

##  Estadístico      Sesgo      Err.Std
## -0.144801929  0.006044904  0.117978183

```

8.6.4 Paquetes R: bootstrap, boot

```

library(boot)
# ?boot

```

Función estadístico:

```

boot.f <- function(data, indices){
  # data[indices] será la muestra bootstrap
  mean(data[indices])
}

```

Generación de las muestras

```

set.seed(1)
stat.boot <- boot(dat, boot.f, nboot)
stat.boot

##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = dat, statistic = boot.f, R = nboot)
##
##
## Bootstrap Statistics :
##      original     bias    std. error
## t1* -0.0933804 0.00135433   0.105385

names(stat.boot)

## [1] "t0"        "t"         "R"         "data"      "seed"
## [6] "statistic" "sim"       "call"      "stype"     "strata"
## [11] "weights"

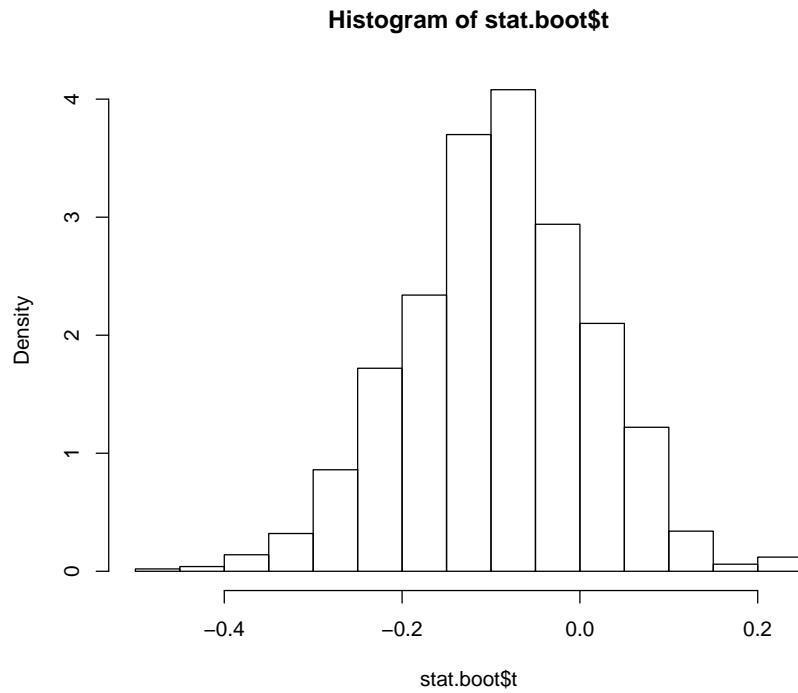
```

8.6.5 Gráficos

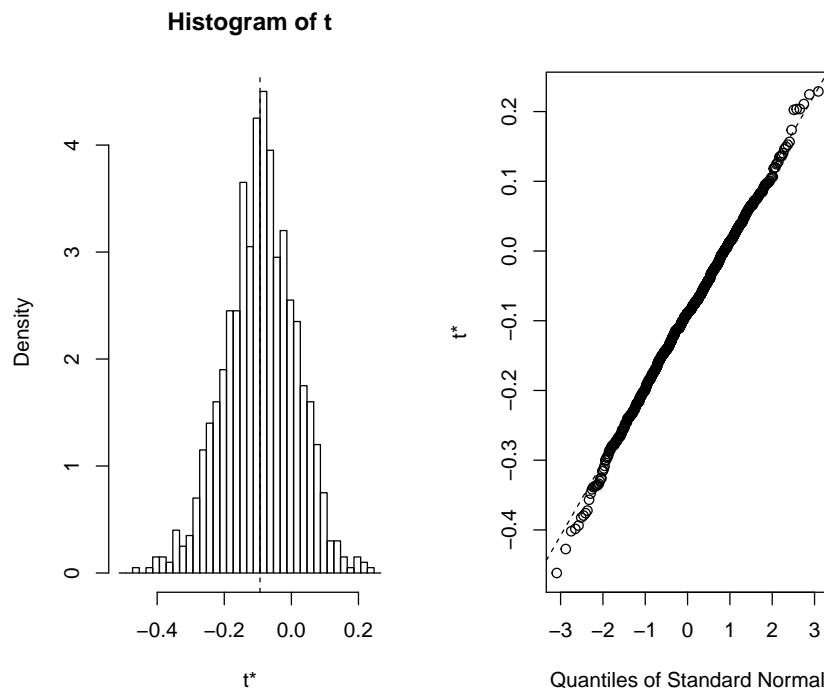
```

hist(stat.boot$t, freq=FALSE)

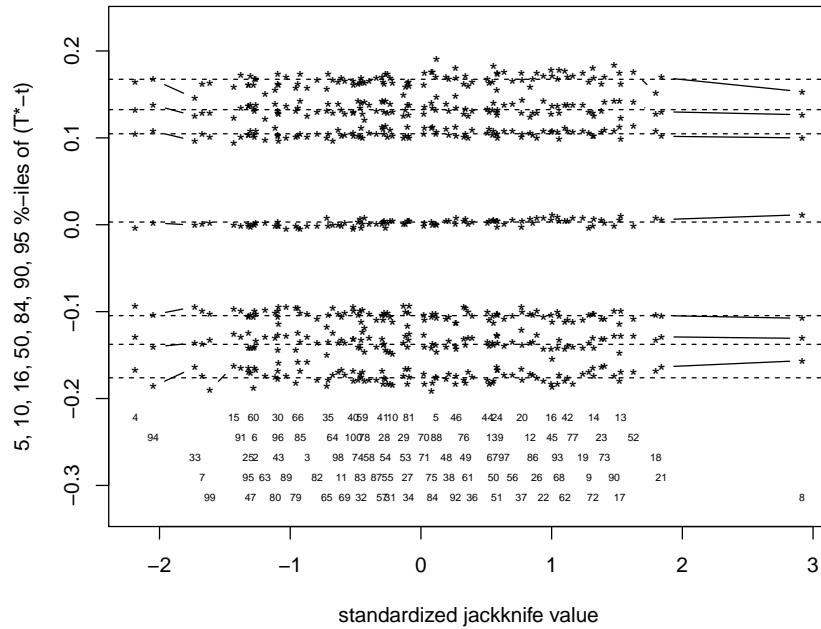
```



```
plot(stat.boot)
```



```
jack.after.boot(stat.boot)
```



8.6.6 Intervalos de confianza bootstrap

```
boot.ci(stat.boot, type=c("norm", "basic", "perc", "bca"))

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = stat.boot, type = c("norm", "basic", "perc",
##      "bca"))
##
## Intervals :
## Level      Normal          Basic
## 95%   (-0.3013,  0.1118 )  (-0.2900,  0.1212 )
##
## Level      Percentile        BCa
## 95%   (-0.3079,  0.1032 )  (-0.3258,  0.0973 )
## Calculations and Intervals on Original Scale
```

Ejercicio 8.7.

Repetir el ejemplo anterior considerando la media recortada al 10% (ejemplo con parámetros adicionales).

```
boot.f <- function(data, indices, trim=0.1){
  mean(data[indices], trim)
}

set.seed(1)
boot(dat, boot.f, nboot, trim=0.2)
```

```

## 
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = dat, statistic = boot.f, R = nboot, trim = 0.2)
##
##
## Bootstrap Statistics :
##      original     bias   std. error
## t1* -0.1448019 0.005146914 0.1226472

```

Ejercicio 8.8. (propuesto)

En el tema 2 se propuso el análisis de la aleatoriedad de un generador de números pseudo-aleatorios mediante el test de rachas, que se podría implementar repetidamente. Sin embargo, la aproximación asintótica empleada por la rutina `runs.test` de la librería `tseries` no es adecuada para tamaños muestrales pequeños ($n < 40$) y sería preferible utilizar la distribución exacta (o por lo menos utilizar una corrección por continuidad).

- Analizar el comportamiento del contraste empleando repetidamente el test de rachas, considerando 500 pruebas con muestras de tamaño 10 de una $Bernoulli(0.5)$. ¿Se observa algo extraño?
- Realiza un programa que permita aproximar por simulación la función de masa de probabilidad del estadístico número de rachas (a partir de valores de una $Bernoulli(0.5)$). Representarla gráficamente y compararla con la densidad normal. Obtener los puntos críticos para contrastar la hipótesis nula de aleatoriedad para $\alpha = 0.01, 0.05$ y 0.1 . ¿Es esta distribución adecuada para el contraste de aleatoriedad de variables continuas? ¿Cuál debería ser la probabilidad de obtener una única racha al aplicar el test a una variable continua?
- Diseñar una rutina que permita realizar el contraste de aleatoriedad de una variable continua approximando el p -valor por simulación. Asumir que la distribución del estadístico puede ser asimétrica, en cuyo caso el p -valor $p = 2 \min \left\{ P(R \leq \hat{R} | H_0), P(R \geq \hat{R} | H_0) \right\}$.
- Diseñar una rutina que permita realizar el contraste de aleatoriedad de una variable continua approximando el p -valor mediante bootstrap.

Capítulo 9

Integración y Optimización Montecarlo

9.1 Integración Monte Carlo (clásica)

La integración Monte Carlo se emplea principalmente para aproximar integrales multidimensionales:

$$I = \int \cdots \int h(x_1, \dots, x_d) dx_1 \cdots dx_d$$

donde puede presentar ventajas respecto a los métodos tradicionales de integración numérica.

Supongamos que nos interesa:

$$I = \int_0^1 h(x) dx$$

Si x_1, x_2, \dots, x_n i.i.d. $\mathcal{U}(0, 1)$ entonces:

$$I = E(h(\mathcal{U}(0, 1))) \approx \frac{1}{n} \sum_{i=1}^n h(x_i)$$

Si el intervalo de integración es genérico:

$$I = \int_a^b h(x) dx = (b - a) \int_a^b h(x) \frac{1}{(b - a)} dx = (b - a) E(h(\mathcal{U}(a, b))).$$

Si x_1, x_2, \dots, x_n i.i.d. $\mathcal{U}(a, b)$:

$$I \approx \frac{1}{n} \sum_{i=1}^n h(x_i) (b - a)$$

9.1.1 Ejercicio 1

Crear una función que implemente la integración Monte Carlo clásica para aproximar integrales del tipo:

$$I = \int_a^b h(x) dx.$$

```
mc.integral0 <- function(fun, a, b, n) {
  # Integración Monte Carlo de fun entre a y b utilizando una muestra de tamaño n
  # fun es una función de una sola variable (y que no es vectorial)
  # Se asume a < b y n entero positivo
  #
  x <- runif(n, a, b)
  fx <- sapply(x, fun) # Si fun fuese vectorial bastaría con: fx <- fun(x)
  return(mean(fx) * (b - a))
}
```

Emplearla para aproximar:

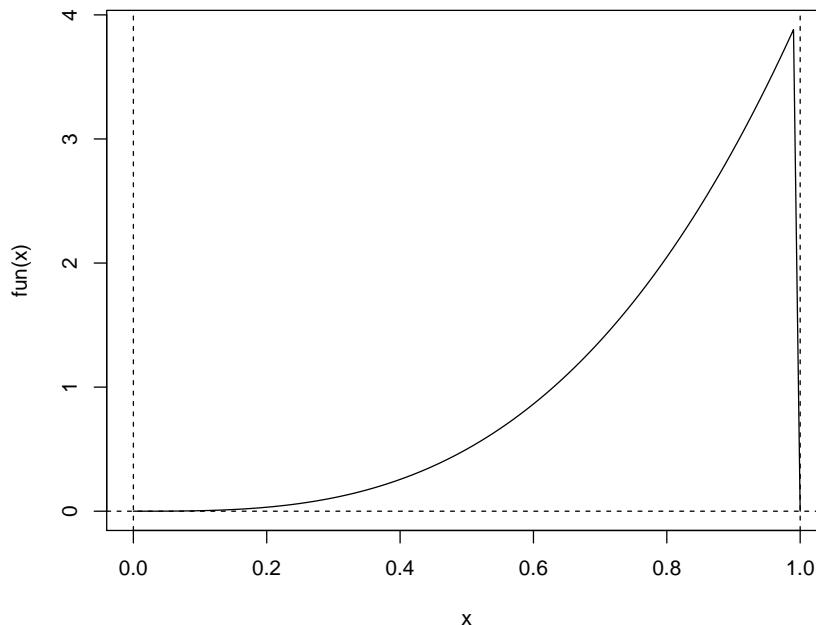
$$\int_0^1 4x^3 dx = 1,$$

y representar gráficamente la aproximación en función de n .

Función a integrar:

```
fun <- function(x) ifelse((x > 0) & (x < 1), 4 * x^3, 0)
# return(4 * x^3)

curve(fun, 0, 1)
abline(h=0,lty=2)
abline(v=c(0,1),lty=2)
```



```
set.seed(54321)
mc.integral0(fun, 0, 1, 20)

## [1] 0.8051635
mc.integral0(fun, 0, 1, 100)

## [1] 1.134583
```

```
mc.integral0(fun, 0, 1, 100)
```

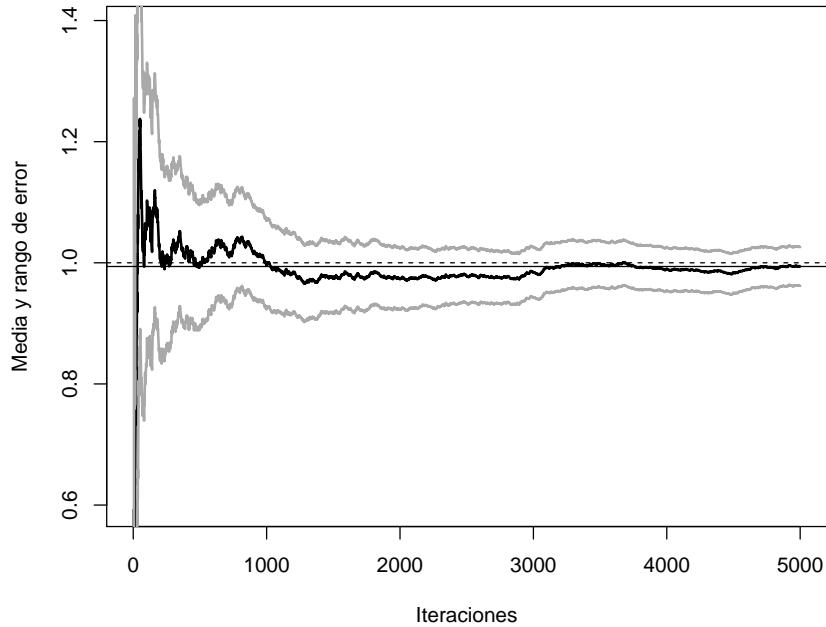
```
## [1] 0.935731
```

La función `mc.integral0` no es adecuada para analizar la convergencia de la aproximación por simulación. Una alternativa más eficiente para representar gráficamente la convergencia:

```
mc.integral <- function(fun, a, b, n, plot = TRUE) {
  fx <- sapply(runif(n, a, b), fun) * (b - a)
  if (plot) {
    estint <- cumsum(fx)/(1:n)
    esterr <- sqrt(cumsum((fx - estint)^2))/(1:n)
    plot(estint, ylab = "Media y rango de error", type = "l", lwd = 2, ylim = mean(fx) +
      2 * c(-esterr[1], esterr[1]), xlab = "Iteraciones")
    lines(estint + 2 * esterr, col = "darkgray", lwd = 2)
    lines(estint - 2 * esterr, col = "darkgray", lwd = 2)
    valor <- estint[n]
    abline(h = valor)
    return(list(valor = valor, error = 2 * esterr[n]))
  } else return(list(valor = mean(fx), error = 2 * sd(fx)/sqrt(n)))
}

set.seed(54321)
mc.integral(fun, 0, 1, 5000)
```

```
## $valor
## [1] 0.9939051
##
## $error
## [1] 0.03206515
abline(h = 1, lty = 2)
```



Si sólo interesa la aproximación:

```
set.seed(54321)
mc.integral(fun, 0, 1, 5000, plot = FALSE)

## $valor
## [1] 0.9939051
##
## $error
## [1] 0.03209283
```

Nota: Es importante tener en cuenta que la función `mc.integral` solo es válida para dominio finito.

9.1.2 Caso general

A partir a ahora consideraremos que interesa aproximar una integral de la forma:

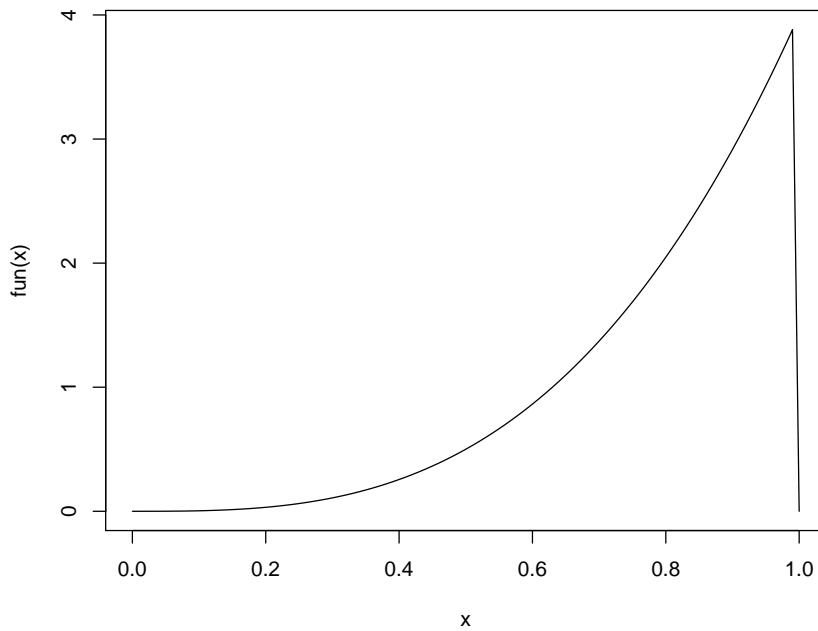
$$\theta = E(h(X)) = \int h(x) f(x) dx$$

siendo $X \sim f$, entonces, si x_1, x_2, \dots, x_n i.i.d. X :

$$\theta \approx \frac{1}{n} \sum_{i=1}^n h(x_i)$$

Por ejemplo, como en el ejercicio anterior se considera de una función de densidad, se correspondería con el caso general de $h(x) = 1$ y $f(x) = 4x^3$ para $0 < x < 1$. La idea es que, en lugar de considerar una distribución uniforme, es preferible generar más valores donde hay mayor “área”:

```
curve(fun(x), 0, 1)
```



Los pasos serían simular x con densidad `fun` y aproximar la integral por `mean(h(x))`:

```
x <- rfun(nsim) # Habría que implementarlo (por ejemplo por el método de inversión...)
res <- mean(h(x)) # Aproximación por simulación
error <- 2*sd(h(x))/sqrt(nsim)
```

En este caso como $h(x) = 1$ no tendría mucho sentido (`mean(1) = 1`).

9.1.3 Ejercicio 2

Aproximar:

$$\phi(t) = \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

para $t = 4.5$, empleando integración Monte Carlo (aproximación tradicional con dominio infinito).

```
# h <- function(x) x > 4.5
# f <- function(x) dnorm(x)
set.seed(54321)
nsim <- 10^3
x <- rnorm(nsim)
mean(x > 4.5) # mean(h(x))

## [1] 0
pnorm(-4.5) # valor teórico P(X > 4.5)

## [1] 3.397673e-06
```

De esta forma es difícil que se generen valores (en este caso ninguno) en la región que interesaría para la aproximación de la integral:

```
any(x > 4.5)
```

```
## [1] FALSE
```

Sería preferible generar más valores donde hay mayor “área”...

9.2 Muestreo por importancia

Para aproximar la integral:

$$\theta = E(h(X)) = \int h(x) f(x) dx,$$

puede ser preferible generar observaciones de una densidad g que tenga una forma similar al producto hf .

Si $Y \sim g$:

$$\theta = \int h(x) f(x) dx = \int \frac{h(x) f(x)}{g(x)} g(x) dx = E(q(Y)).$$

siendo $q(x) = \frac{h(x)f(x)}{g(x)}$.

Si y_1, y_2, \dots, y_n i.i.d. $Y \sim g$:

$$\theta \approx \frac{1}{n} \sum_{i=1}^n q(y_i) = \frac{1}{n} \sum_{i=1}^n w(y_i) h(y_i) = \hat{\theta}_g$$

con $w(x) = \frac{f(x)}{g(x)}$.

En este caso $Var(\hat{\theta}_g) = Var(q(Y))/n$, pudiendo reducirse significativamente respecto al método clásico si:

$$g(x) \underset{aprox.}{\propto} h(x) f(x).$$

Para aplicar el TCL, la varianza del estimador $\hat{\theta}_g$ es finita si:

$$E(q^2(Y)) = \int \frac{h^2(x) f^2(x)}{g(x)} dx = E\left(h^2(X) \frac{f(X)}{g(X)}\right) < \infty.$$

La idea básica es que si la densidad g tiene colas más pesadas que la densidad f con mayor facilidad puede dar lugar a “simulaciones” con varianza finita (podría emplearse en casos en los que no existe $E(h^2(X))$; ver Sección Convergencia en el Tema 4 de Análisis de resultados).

La distribución de los pesos $w(y_i)$ debería ser homogénea para evitar datos influyentes (inestabilidad).

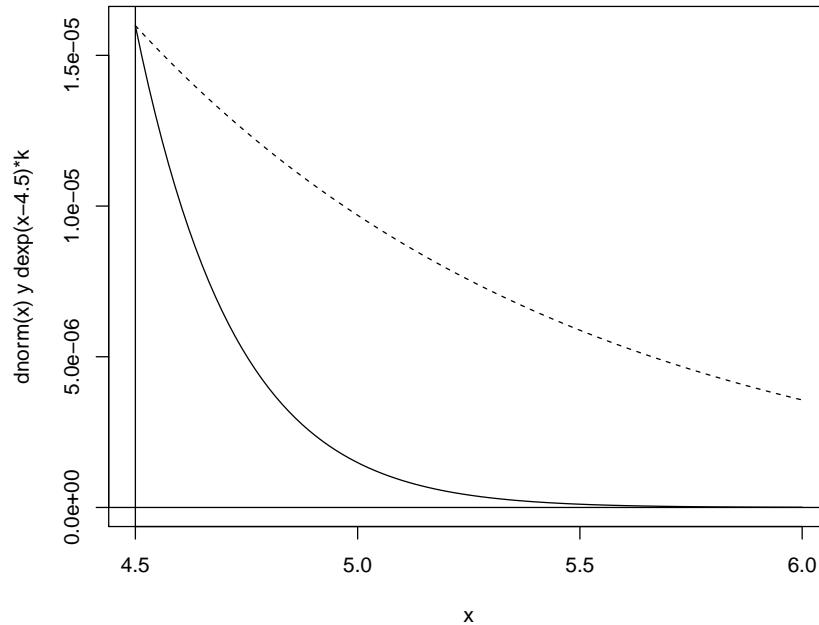
9.2.1 Ejercicio 3

Aproximar la integral del ejercicio anterior empleando muestreo por importancia considerando como densidad auxiliar una exponencial de parámetro $\lambda = 1$ truncada en t :

$$g(x) = \lambda e^{-\lambda(x-t)}, x > t,$$

(emplear `rexp(n)+t` y `dexp(y-t)`). Comparar $h(x)f(x)$ con $g(x)f(4.5)$ y representar gráficamente la aproximación en función de n .

```
curve(dnorm(x), 4.5, 6, ylab = "dnorm(x) y dexp(x-4.5)*k")
abline(v = 4.5)
abline(h = 0)
escala <- dnorm(4.5) # Reescalado para comparación...
curve(dexp(x - 4.5) * escala, add = TRUE, lty = 2)
```



Se generan los valores de la densidad auxiliar y se calculan los pesos:

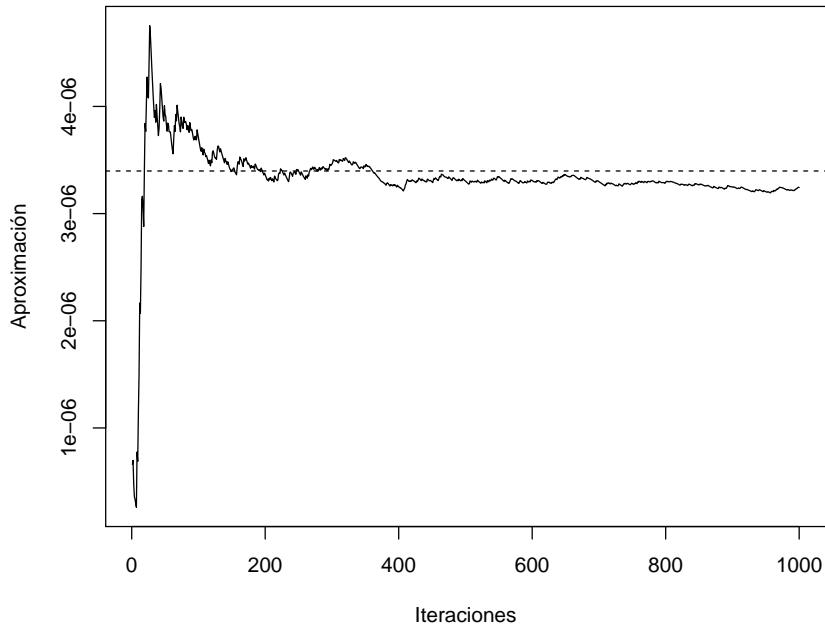
```
set.seed(54321)
nsim <- 10^3
y <- rexp(nsim) + 4.5      # Y ~ g
w <- dnorm(y)/dexp(y - 4.5)
```

La aproximación por simulación sería `mean(w * h(y))`:

```
# h(x) <- function(x) x > 4.5  # (1 si x > 4.5 => h(y) = 1)
mean(w) # w*h(y)
```

```
## [1] 3.243321e-06
pnorm(-4.5) # valor teórico
```

```
## [1] 3.397673e-06
plot(cumsum(w)/1:nsim, type = "l", ylab = "Aproximación", xlab = "Iteraciones")
abline(h = pnorm(-4.5), lty = 2)
```



El error estandar de la aproximación sería `sqrt(var(w * h(y))/nsim)`:

```
sqrt(var(w)/nsim) # sd(w*h(y))/sqrt(nsim)
```

```
## [1] 1.386216e-07
```

Empleando la aproximación tradicional:

```
est <- mean(rnorm(nsim) > 4.5)
est
```

```
## [1] 0
sqrt(est * (1 - est)/nsim)
```

```
## [1] 0
```

9.2.2 Ejercicio 4

Aproximar $P(2 < X < 6)$ siendo $X \sim \text{Cauchy}(0, 1)$ empleando muestreo por importancia y considerando como densidad auxiliar la normal estandar $Y \sim N(0, 1)$. Representar gráficamente la aproximación y estudiar los pesos $w(y_i)$.

Nota: En este caso van a aparecer problemas (la densidad auxiliar debería tener colas más pesadas que la densidad objetivo; sería adecuado si intercambiaramos las distribuciones objetivo y auxiliar, como en el ejercicio siguiente).

Se trata de aproximar `pcauchy(6) - pcauchy(2)`, i.e. $f(y) = \text{dcauchy}(y)$ y $h(y) = (y > 2) * (y < 6)$, empleando muestreo por importancia con $g(y) = \text{dnorm}(y)$.

```
nsim <- 10^5
set.seed(4321)
```

```
y <- rnorm(nsim)
w <- dcauchy(y)/dnorm(y) # w <- w/sum(w) si alguna es una cuasidensidad
```

La aproximación por simulación es `mean(w(y) * h(y))`:

```
mean(w * (y > 2) * (y < 6))
```

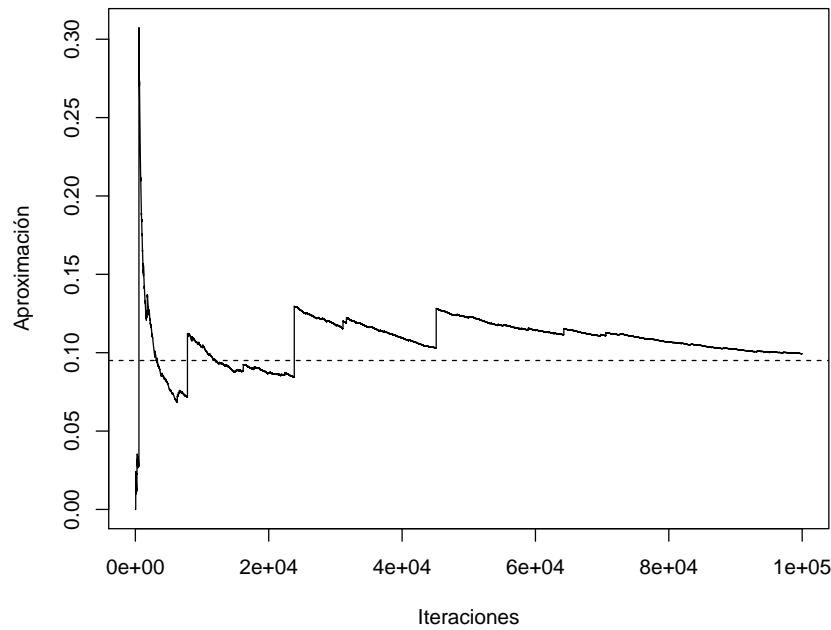
```
## [1] 0.09929348
```

```
pcauchy(6) - pcauchy(2) # Valor teórico
```

```
## [1] 0.09501516
```

Si se estudia la convergencia:

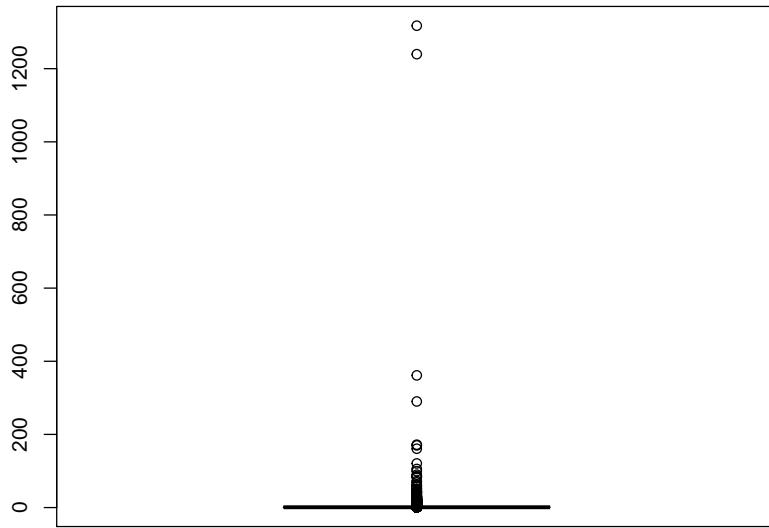
```
plot(cumsum(w * (y > 2) * (y < 6))/1:nsim, type = "l", ylab = "Aproximación", xlab = "Iteraciones")
abline(h = pcauchy(6) - pcauchy(2), lty = 2)
```



Lo que indica es una mala elección de la densidad auxiliar...

La distribución de los pesos debería ser homogénea. Por ejemplo, si los reescalamos para que su suma sea el número de valores generados, deberían tomar valores en torno a uno:

```
boxplot(nsim * w/sum(w))
```



9.2.3 Remuestreo (del muestreo) por importancia

Cuando f y/o g son cuasi-densidades, para evitar calcular constantes normalizadoras, se emplea como aproximación:

$$\theta \approx \frac{\sum_{i=1}^n w(y_i)h(y_i)}{\sum_{i=1}^n w(y_i)}.$$

Adicionalmente, puede verse que con un muestreo de $\{y_1, y_2, \dots, y_n\}$ ponderado por $w(y_i)$ (prob. $= w(y_i) / \sum_{i=1}^n w(y_i)$) se obtiene una simulación aproximada de f (*Sample importance resampling*, Rubin, 1987).

9.2.4 Ejercicio 5

Generar 1000 simulaciones de una distribución (aprox.) $N(0, 1)$ mediante remuestreo del muestreo por importancia de 10^5 valores de una $Cauchy(0, 1)$.

Se trata de simular una normal a partir de una Cauchy (Sampling Importance Resampling). En este caso $f(y) = dnorm(y)$ y $g(y) = dcauchy(y)$, al revés del ejercicio anterior...

```
# Densidad objetivo
# f <- dnorm # f <- function(x) ...

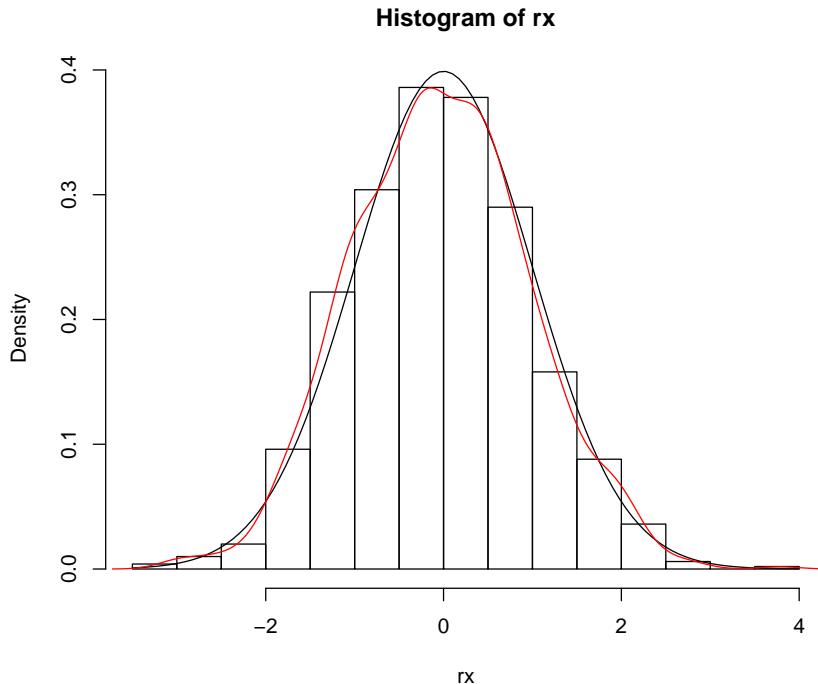
nsim <- 10^3
# El nº de simulaciones de la densidad auxiliar debe ser mucho mayor:
nsim2 <- 10^5
set.seed(4321)
y <- rcauchy(nsim2)
```

```
w <- dnorm(y)/dcauchy(y) # w <- w/sum(w) si alguna es una cuasidensidad

# Si se pidiera aproximar una integral
# h(y) = y si es la media # h <- function(y) y
# mean(w * h(y))
```

Sampling Importance Resampling:

```
rx <- sample(y, nsim, replace = TRUE, prob = w/sum(w))
hist(rx, freq = FALSE)
curve(dnorm, add = TRUE)
lines(density(rx), col = "red")
```



Nota: Si f o g fuesen cuasidensidades y se pidiese aproximar la integral, habría que reescalar los pesos: $w <- f(y)/g(y)$; $w <- w/sum(w)$, y la aproximación por simulación sería $\sum(w * h(y))$ y en el análisis de convergencia se emplearía $\text{cumsum}(w * h(y))$ (sin dividir por el número de simulaciones).

9.3 Optimización Monte Carlo

Supongamos que estamos interesados en la minimización de una función:

$$\min_{\mathbf{x} \in D} f(\mathbf{x}).$$

Hay una gran cantidad de algoritmos numéricos para resolver problemas de optimización no lineal multidimensional, por ejemplo los basados en el método de Newton-Raphson.

La idea original consiste en buscar los ceros de su primera derivada (o del gradiente) empleando una aproximación iterativa:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - [Hf(\mathbf{x}_i)]^{-1} \nabla f(\mathbf{x} =_i),$$

donde $Hf(\mathbf{x}_i)$ es el hessiano de la función (matriz de segundas derivadas) y $\nabla f(\mathbf{x}_i)$ el gradiente (vector de primeras derivadas). Estos métodos normalmente funcionan muy bien cuando la función objetivo no tiene mínimos locales (ideal f cuadrática). Los resultados obtenidos pueden ser muy malos en caso contrario (especialmente en el caso multidimensional) y dependen en gran medida del punto inicial. Un ejemplo donde es habitual que aparezcan este tipo de problemas es en la estimación por máxima verosimilitud (la función objetivo puede ser multimodal).

9.3.1 Ejercicio 6

La mixtura de distribuciones normales:

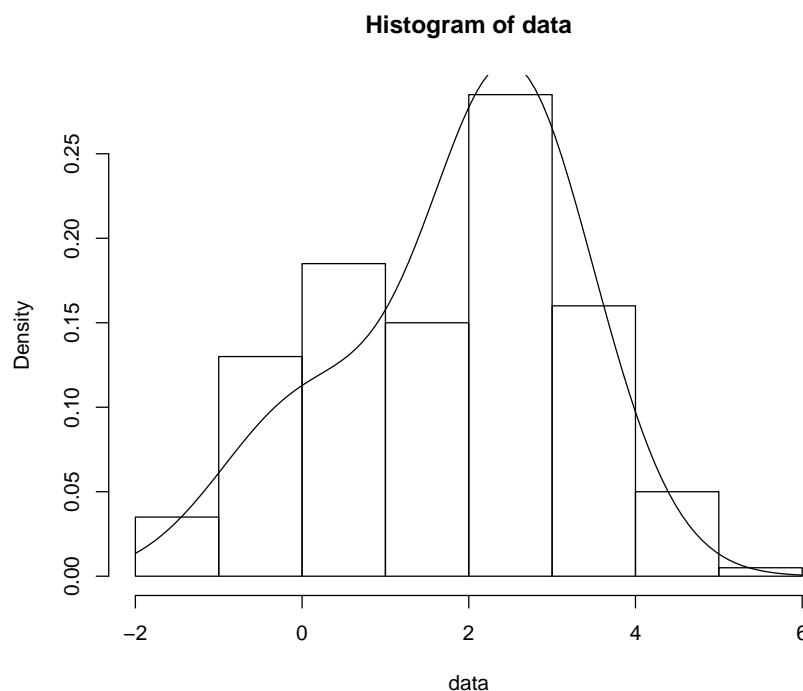
$$\frac{1}{4}N(\mu_1, 1) + \frac{3}{4}N(\mu_2, 1),$$

tiene una función de verosimilitud asociada bimodal. Generar una muestra de 200 valores de esta distribución con $\mu_1 = 0$ y $\mu_2 = 2.5$, construir la correspondiente función de verosimilitud y representarla graficamente. Obtener la estimación por máxima verosimilitud de los parámetros empleando la rutina `nlm`. Muestra (simulación mixtura dos normales)

```
nsim <- 200
mu1 <- 0
mu2 <- 2.5
sd1 <- sd2 <- 1

set.seed(12345)
p.sim <- rbinom(nsim, 1, 0.25)
data <- rnorm(nsim, mu1*p.sim + mu2*(1-p.sim), sd1*p.sim + sd2*(1-p.sim))

hist(data, freq = FALSE, breaks = 'FD')
curve(0.25 * dnorm(x, mu1, sd1) + 0.75 * dnorm(x, mu2, sd2), add = TRUE)
```



Logaritmo (negativo) de la función de verosimilitud (para la estimación de las medias)

```
like <- function(mu)
  -sum(log((0.25 * dnorm(data, mu[1], sd1) + 0.75 * dnorm(data, mu[2], sd2))))
  # NOTA: Pueden aparecer NA/Inf por log(0)
```

Si queremos capturar los valores en los que se evalúa esta función, podemos proceder de forma similar a como se describe en el capítulo Function operators del libro “Advanced R” de Hadley Wickham: Behavioural FOs leave the inputs and outputs of a function unchanged, but add some extra behaviour.

```
tee <- function(f) {
  function(...) {
    input <- if(nargs() == 1) c(...) else list(...)
    output <- f(...)
    # Hacer algo ...
    # ... con output e input
    return(output)
  }
}
```

En este caso queremos representar los puntos en los que el algoritmo de optimización evalúa la función objetivo (especialmente como evoluciona el valor óptimo)

```
tee.optim2d <- function(f) {
  best.f <- Inf      # Suponemos que se va a minimizar (opción por defecto)
  best.par <- c(NA, NA)
  function(...) {
    input <- c(...)
    output <- f(...)
    ## Hacer algo ...
    points(input[1], input[2], col = "lightgray")
    if(best.f > output) {
      lines(rbind(best.par, input), lwd = 2, col = "blue")
      best.f <-> output
      best.par <-> input
      # points(best.par[1], best.par[2], col = "blue", pch = 20)
      # cat('par = ', best.par, 'value = ', best.f, '\n')
    }
    ## ... con output e input
    return(output)
  }
}
```

Representar la superficie del logaritmo de la verosimilitud, los puntos iniciales y las iteraciones en la optimización numérica con `nlm`:

```
mmu1 <- mmu2 <- seq(-2, 5, length = 128)
lli <- outer(mmu1, mmu2, function(x,y) apply(cbind(x,y), 1, like))

par(mar = c(4, 4, 1, 1))
image(mmu1, mmu2, -lli, xlab = expression(mu[1]), ylab = expression(mu[2]))
contour(mmu1, mmu2, -lli, nlevels = 50, add = TRUE)

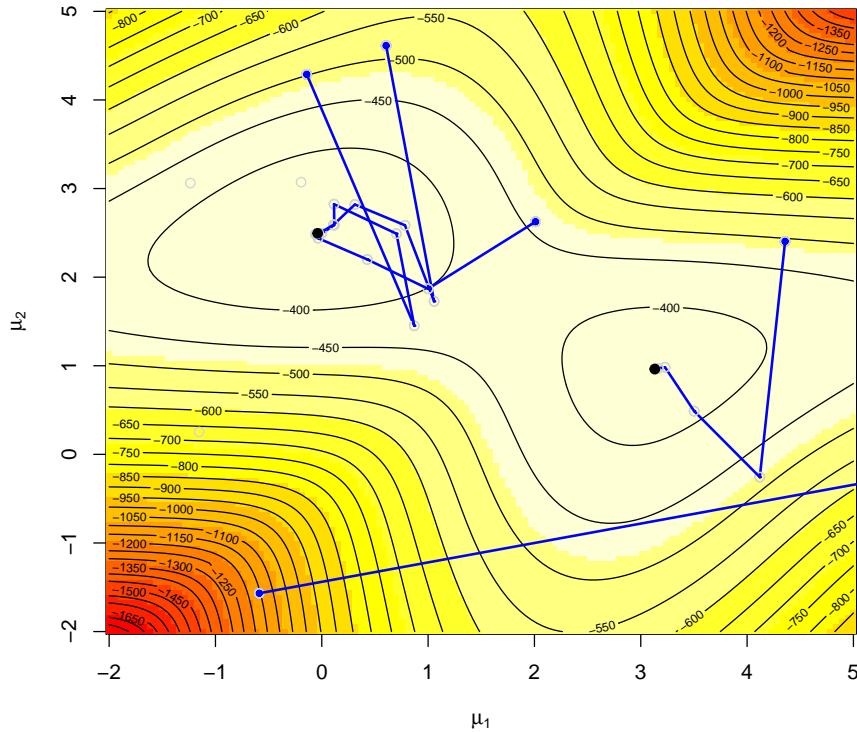
# Valores iniciales aleatorios
nstarts <- 5
set.seed(1)
starts <- matrix(runif(2*nstarts, -2, 5), nrow = nstarts)
```

```

points(starts, col = "blue", pch = 19)

# Minimización numérica con nlm
for (j in 1:nstarts){
  # Normalmente llamaríamos a nlm(like, start)
  res <- nlm(tee.optim2d(like), starts[j, ])
  points(res$estimate[1],res$estimate[2], pch = 19)
  cat('par = ', res$estimate, 'value = ', res$minimum, '\n')
}

```



```

## par = -0.03892511 2.494589 value = 361.5712
## par = -0.03892501 2.494589 value = 361.5712
## par = -0.03892507 2.494589 value = 361.5712
## par = 3.132201 0.9628536 value = 379.3739
## par = 20.51013 1.71201 value = 474.1414

```

9.3.2 Algoritmos de optimización Monte Carlo

Una alternativa sería tratar de generar valores aleatoriamente de forma que las regiones donde la función objetivo es menor tuviesen mayor probabilidad y menor probabilidad las regiones donde la función objetivo es mayor. Por ejemplo, se podría pensar en generar valores de acuerdo a una densidad (transformación Boltzman-Gibbs):

$$g(x) \propto \exp(-f(x)/T),$$

donde $T > 0$ es un parámetro (denominado temperatura) seleccionado de forma que se garantize la integrabilidad.

Entre los métodos de optimización Monte Carlo podríamos destacar:

- Métodos con gradiente aleatorio.

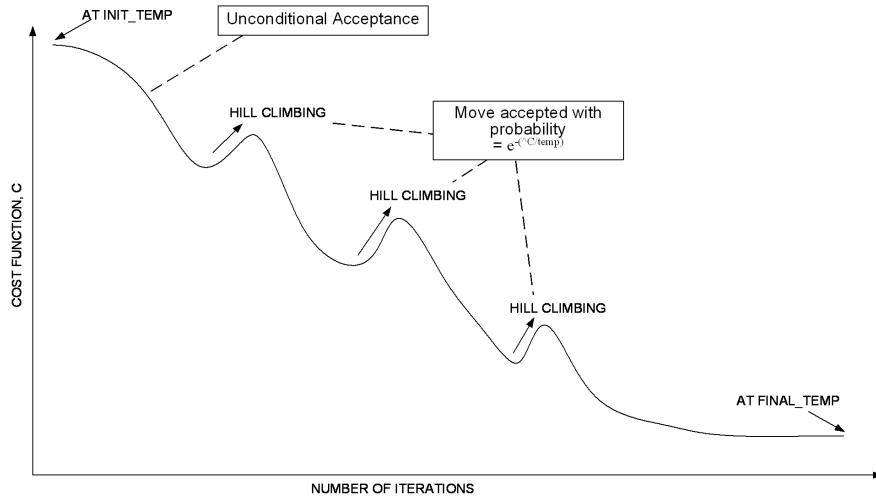
- Temple simulado.
- Algoritmos genéticos.
- Montecarlo EM.
- ...

9.4 Temple simulado

Método inspirado en el templado de un metal (se calienta el metal a alta temperatura y se va enfriando lentamente). En cada paso se reemplaza la aproximación actual por un valor aleatorio “cercano”, elegido con una probabilidad que depende de la mejora en la función objetivo y de un parámetro T (denominado temperatura) que disminuye gradualmente durante el proceso.

- Cuando la temperatura es grande los cambios son bastante probables en cualquier dirección.
- Al ir disminuyendo la temperatura los cambios tienden a ser siempre “cuesta abajo”.

Al tener una probabilidad no nula de aceptar una modificación “cuesta arriba” se trata de evitar quedar atrapado en un óptimo local.



9.4.1 Algoritmo:

```

temp <- TEMP.INIT
place <- INIT.PLACEMENT()
cost.place <- COST(place)
while(temp < TEMP.FINAL) {
  while LOOP.CRITERION() {
    place.new <- PERTURB(place, temp)
    cost.new <- COST(place.new)
    cost.inc <- cost.new - cost.place
    temp <- SCHEDULE(temp)
    if ((cost.inc < 0) || (runif(1) > exp(-(cost.inc/temp)))) break
  }
  place <- place.new
  cost.place <- cost.new
  # temp <- SCHEDULE(temp)
}

```

```

}
COST <- function(place, ...) {...}
SCHEDULE <- function(temp, ...) {...}
INIT.PLACEMENT <- function(...) {...}
LOOP.CRITERION <- function(...) {...}

```

Adaptado de Premchand Akella (ppt).

Este algoritmo se puede ver como una adaptación del método de Metropolis-Hastings que veremos más adelante (Tema 12 Introducción a los métodos de cadenas de Markov Monte Carlo).

9.4.2 Ejercicio 7

Repetir el ejercicio anterior empleando el algoritmo del temple simulado.

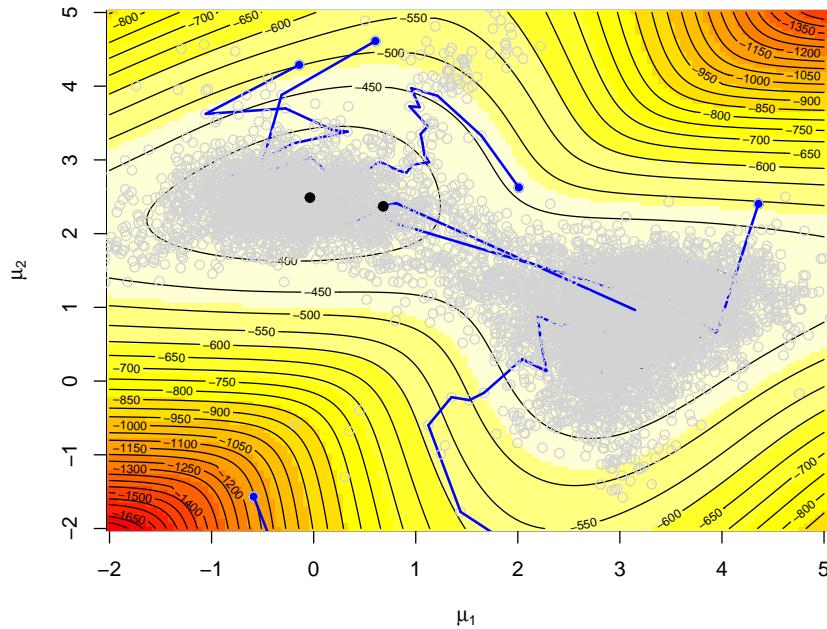
Minimización “SANN” con optim:

```

# Representar la superficie del logaritmo de la verosimilitud
image(mmu1, mmu2, -lli, xlab = expression(mu[1]), ylab = expression(mu[2]))
contour(mmu1, mmu2, -lli, nlevels = 50, add = TRUE)
points(starts, col = "blue", pch = 19)

set.seed(1)
for (j in 1:nstarts){
  # Normalmente llamaríamos a optim(start, like, method = "SANN")
  # Note that the "SANN" method depends critically on the settings of the control parameters.
  # For "SANN" maxit gives the total number of function evaluations: there is no other stopping criterio
  res <- optim(starts[j, ], tee.optim2d(like), method = "SANN", control = list(temp = 100, maxit = 2000))
  points(res$par[1],res$par[2], pch = 19)
  cat('par = ', res$par, 'value =', res$value, '\n')
}

```



```
## par =  0.0002023461 2.473437 value = 361.6372
## par = -0.182735 2.455585 value = 362.0255
## par = -0.0281341 2.484467 value = 361.5801
## par = -0.03642928 2.488626 value = 361.5732
## par =  0.6814165 2.370026 value = 374.839
```

Alternativa: función basada en el algoritmo empleado en el ejemplo 5.9 del libro: Robert y Casella, Introducing Monte Carlo Methods with R, Springer, 2010.

```
SA <- function(fun, pini, lower = -Inf, upper = Inf, tolerance = 10^(-4), factor = 1) {
  temp <- scale <- iter <- dif <- 1
  npar <- length(pini)
  the <- matrix(pini, ncol = npar)
  curfun <- hval <- fun(pini)
  while (dif > tolerance) {
    prop <- the[iter, ] + rnorm(npar) * scale[iter]
    # Se decide si se acepta la propuesta
    if (any(prop < lower) || any(prop > upper) ||
        (temp[iter] * log(runif(1)) > -fun(prop) + curfun)) prop <- the[iter, ]
    curfun <- fun(prop)
    hval <- c(hval, curfun)
    the <- rbind(the, prop)
    iter <- iter + 1
    temp <- c(temp, 1/log(iter + 1)) # Actualizar la temperatura
    # Se controla el nº de perturbaciones aceptadas
    ace <- length(unique(the[(iter/2):iter, 1]))
    if (ace == 1)
      # si es muy pequeño se disminuye la escala de la perturbación
      factor <- factor/10
    if (2 * ace > iter)
      # si es muy grande se aumenta
```

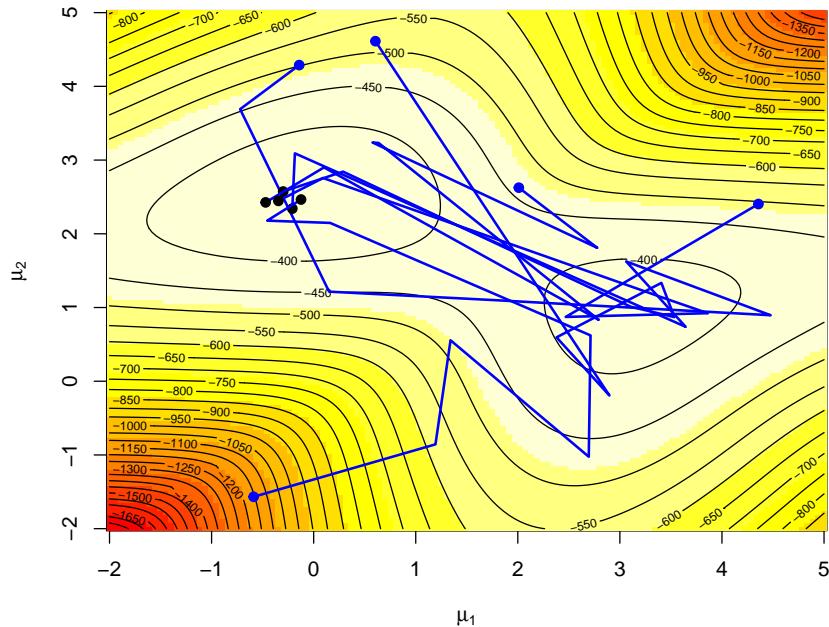
```

    factor <- factor * 10
    scale <- c(scale, max(2, factor * sqrt(temp[iter]))) # Actualizar la escala de la perturbación
    dif <- (iter < 100) + (ace < 2) + (max(hval) - max(hval[1:(iter/2)]))
}
list(theta = the, val = hval, ite = iter)
}

# Representar la superficie del logaritmo de la verosimilitud
image(mmu1, mmu2, -lli, xlab = expression(mu[1]), ylab = expression(mu[2]))
contour(mmu1, mmu2, -lli, nlevels = 50, add = TRUE)
points(starts, col = "blue", pch = 19)

set.seed(1)
for (j in 1:nstarts) {
  sar <- SA(like, starts[j, ])
  lines(sar$the[, 1], sar$the[, 2], lwd = 2, col = "blue")
  points(sar$the[sar$it, 1], sar$the[sar$it, 2], pch = 19)
}

```



9.5 Algoritmos genéticos

Los algoritmos genéticos tratan de encontrar la mejor solución (entre un conjunto de soluciones posibles) imitando los procesos de evolución biológica:

- **Población:** formada por n individuos \mathbf{x}_i codificados en **cromosomas**.
- $f(\mathbf{x}_i)$ ajuste/capacidad/**adaptación** del individuo \mathbf{x}_i .
- **Selección:** los individuos mejor adaptados tienen mayor probabilidad de ser **padres**.

- **Cruzamiento:** los cromosomas de dos padres se combinan para generar hijos.
- **Mutación:** modificación al azar del cromosoma de los hijos (variabilidad).
- **Elitismo:** el mejor individuo pasa a la siguiente generación.

Los paquetes de R `DEoptim` y `gafit` implementan algunos de estos tipos de algoritmos.

9.5.1 Ejercicio 8

Repetir el ejercicio anterior empleando la función `DEoptim`.

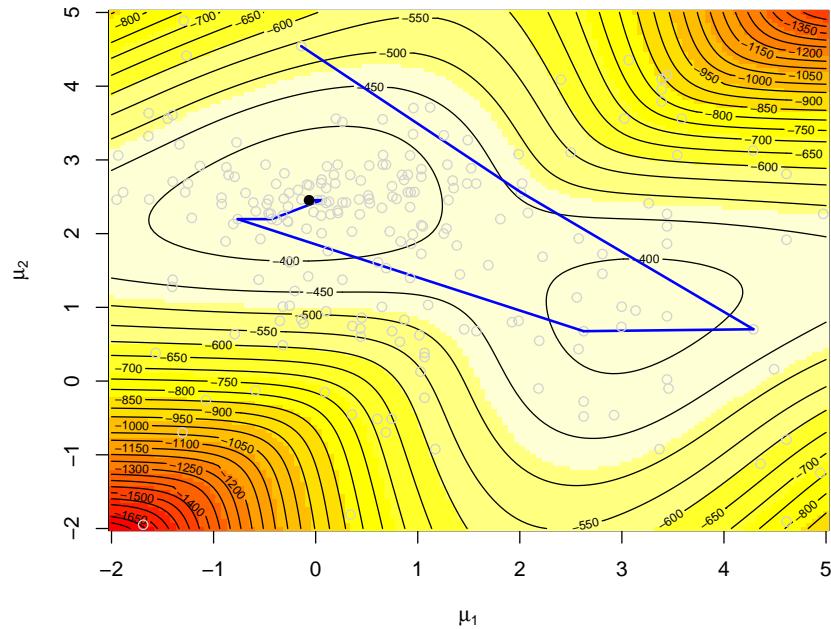
Optimización con algoritmo genético implementado en `DEoptim`:

```
require(DEoptim)

# Representar la superficie del logaritmo de la verosimilitud
image(mmu1, mmu2, -lli, xlab = expression(mu[1]), ylab = expression(mu[2]))
contour(mmu1, mmu2, -lli, nlevels = 50, add = TRUE)
# Estos algoritmos no requieren valores iniciales (los generan al azar en el rango)

lower <- c(-2, -2)
upper <- c(5, 5)
set.seed(1)
# DEoptim(like, lower, upper)
der <- DEoptim(tee.optim2d(like), lower, upper, DEoptim.control(itermax = 10))
```

```
## Iteration: 1 bestvalit: 373.132461 bestmemit: -0.764103 2.196961
## Iteration: 2 bestvalit: 367.580379 bestmemit: -0.430095 2.196961
## Iteration: 3 bestvalit: 367.580379 bestmemit: -0.430095 2.196961
## Iteration: 4 bestvalit: 367.580379 bestmemit: -0.430095 2.196961
## Iteration: 5 bestvalit: 361.906887 bestmemit: 0.058951 2.455186
## Iteration: 6 bestvalit: 361.906887 bestmemit: 0.058951 2.455186
## Iteration: 7 bestvalit: 361.906887 bestmemit: 0.058951 2.455186
## Iteration: 8 bestvalit: 361.657986 bestmemit: -0.064005 2.452184
## Iteration: 9 bestvalit: 361.657986 bestmemit: -0.064005 2.452184
## Iteration: 10 bestvalit: 361.657986 bestmemit: -0.064005 2.452184
# Por defecto fija el tamaño de la población a NP = 10*npar = 20
# Puede ser mejor dejar el valor por defecto itermax = 200
points(der$optim$bestmem[1], der$optim$bestmem[2], pch = 19)
```



Capítulo 10

Técnicas de reducción de la varianza

10.1 Reducción de la varianza

Éstas técnicas son aplicadas normalmente cuando se pretende ofrecer respuestas lo más precisas posibles (con menor costo computacional) y principalmente sobre cantidades medias.

- Supongamos que **estamos interesados en aproximar la media** de un estadístico mediante simulación y **no nos interesa aproximar su varianza**.

Existe un sinfín de técnicas encaminadas a reducir la varianza en un estudio de simulación (respecto a una aproximación estandar). Algunas de ellas son:

- Muestreo por importancia.
- Variables antitéticas.
- Muestreo estratificado.
- Variables de control.
- Números aleatorios comunes.
- Métodos de remuestreo.
- Condicionamiento.
- ...

10.2 Variables antitéticas

Supongamos que pretendemos aproximar

$$\theta = E(Z)$$

con $Var(Z) = \sigma^2$. Si generamos n pares $(X_1, Y_1), \dots, (X_n, Y_n)$ de $X \sim Y \sim Z$ con $Cov(X, Y) < 0$, el estimador combinado tiene menor varianza:

$$\begin{aligned} Var\left(\frac{\bar{X} + \bar{Y}}{2}\right) &= \frac{1}{4} (Var(\bar{X}) + Var(\bar{Y}) + 2Cov(\bar{X}, \bar{Y})) \\ &= \frac{\sigma^2}{2n} + \frac{1}{2n} Cov(X, Y) \\ &= \frac{\sigma^2}{2n} (1 + \rho(X, Y)), \end{aligned}$$

que es equivalente a una muestra unidimensional independiente con el mismo número de observaciones $2n$ (con una reducción del $-100\rho(X, Y)\%$).

10.2.1 Ejemplo: Integración Monte Carlo

Para aproximar:

$$I = \int_0^1 h(x) dx,$$

a partir de x_1, x_2, \dots, x_n i.i.d. $\mathcal{U}(0, 1)$. Podemos emplear:

$$\begin{aligned} I &= E\left(\frac{h(U) + h(1-U)}{2}\right) \\ &\approx \frac{1}{2n} \sum_{i=1}^n (h(x_i) + h(1-x_i)). \end{aligned}$$

10.2.2 Generación de variables antitéticas

Cuando se utiliza el método de inversión resulta sencillo obtener pares de variables con correlación negativa:

- $U \sim \mathcal{U}(0, 1)$ para simular X .
- $1 - U$ para simular la variable antitética Y .

En el caso general, si $X = h(U_1, \dots, U_d)$ y h es monótona puede verse (e.g. Ross, 1997) que $Y = h(1 - U_1, \dots, 1 - U_d)$ está negativamente correlada con X .

Si $X \sim \mathcal{N}(\mu, \sigma)$ puede tomarse como variable antitética

$$Y = 2\mu - X$$

En general esto es válido para cualquier variable simétrica respecto a un parámetro μ . (e.g. $X \sim \mathcal{U}(a, b)$ e $Y = a + b - X$).

Ejercicio 10.1. Variables antitéticas en integración Monte Carlo

Crear una función que implemente la técnica de variables antitéticas para aproximar integrales del tipo:

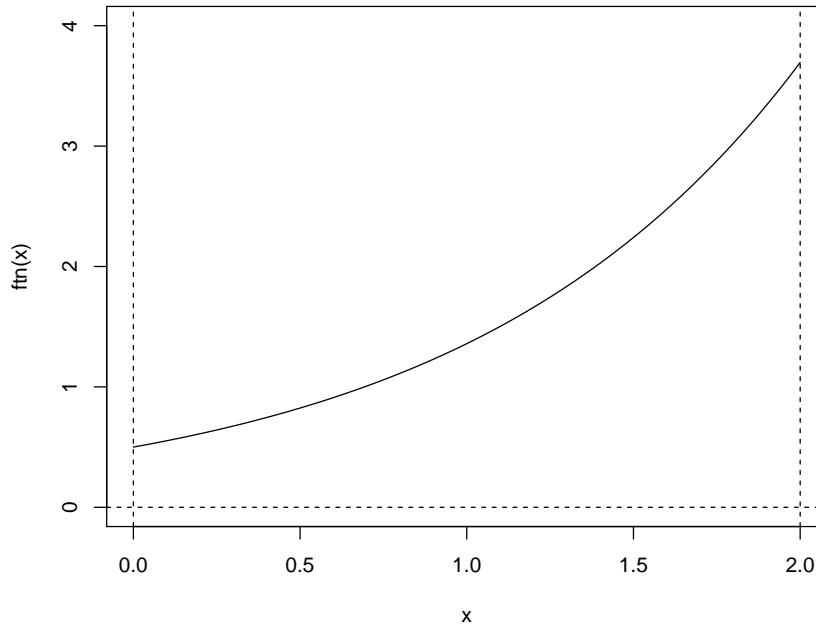
$$I = \int_a^b h(x) dx.$$

Emplearla para aproximar:

$$E\left(e^{\mathcal{U}(0,2)}\right) = \int_0^2 \frac{1}{2} e^x dx \approx 3.194,$$

y representar gráficamente la aproximación en función de n . Función objetivo:

```
a <- 0; b <- 2
ftn <- function(x) return(exp(x)/(b-a))
curve(ftn, a, b, ylim=c(0,4))
abline(h=0,lty=2)
abline(v=c(a,b),lty=2)
```



Se trata de calcular la media de $e^{U(0,2)}$:

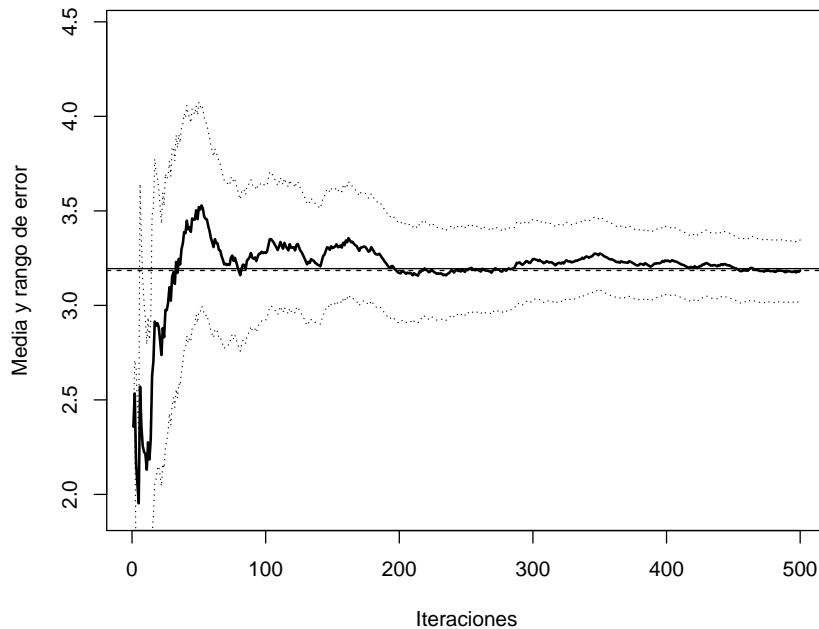
```
teor <- (exp(b)-exp(a))/(b-a)
teor
```

```
## [1] 3.194528
```

Para la aproximación por integración Monte Carlo podemos emplear la función del capítulo anterior:

```
mc.integral <- function(ftn, a, b, n, plot=TRUE) {
  fx <- sapply(runif(n, a, b), ftn)*(b-a)
  if (plot) {
    estint <- cumsum(fx)/(1:n)
    esterr <- sqrt(cumsum((fx-estint)^2))/(1:n)
    plot(estint, ylab="Media y rango de error", type="l", lwd= 2,
         ylim=mean(fx)+2*c(-esterr[1],esterr[1])), xlab="Iteraciones")
    abline(h = estint[n], lty=2)
    lines(estint+2*esterr, lty = 3)
    lines(estint-2*esterr, lty = 3)
    return(list(valor=estint[n], error=2*esterr[n]))
  } else return(list(valor=mean(fx), error=2*sd(fx)/sqrt(n)))
}

set.seed(54321)
res <- mc.integral(ftn, a, b, 500)
abline(h = teor)
```



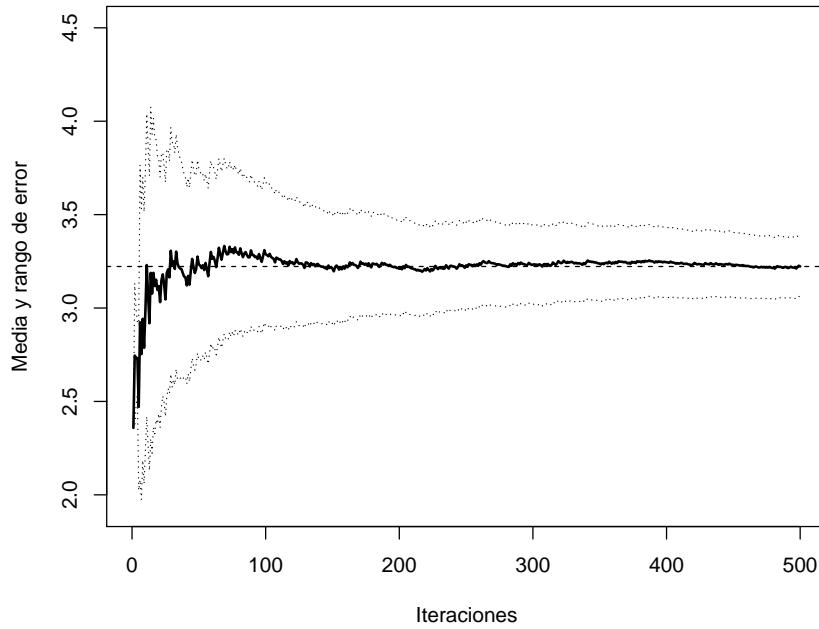
res

```
## $valor
## [1] 3.184612
##
## $error
## [1] 0.1619886
```

Para la integración Monte Carlo con variables antitéticas podríamos considerar:

```
mc.integrala <- function(ftn, a, b, n, plot=TRUE,...) {
  # n es el nº de evaluaciones de la función objetivo (para facilitar comparaciones, solo se genera la primera)
  x <- runif(n%/%2, a, b)
  # La siguiente linea solo para representar alternando
  x <- as.numeric(matrix(c(x,a+b-x),nrow=2,byrow=TRUE))
  # bastaría con emplear p.e. c(x,a+b-x)
  fx <- sapply(x, ftn)*(b-a)
  if (plot) {
    estint <- cumsum(fx)/(1:n)
    esterr <- sqrt(cumsum((fx-estint)^2))/(1:n)
    plot(estint, ylab="Media y rango de error",type="l", lwd = 2,
         ylim=mean(fx)+2*c(-esterr[1],esterr[1]),xlab="Iteraciones",...)
    abline(h = estint[n], lty=2)
    lines(estint+2*esterr, lty = 3)
    lines(estint-2*esterr, lty = 3)
    return(list(valor=estint[n],error=2*esterr[n]))
  } else return(list(valor=mean(fx),error=2*sd(fx)/sqrt(n)))
}

set.seed(54321)
res <- mc.integrala(ftn, a, b, 500)
```



```
res
```

```
## $valor
## [1] 3.222366
##
## $error
## [1] 0.1641059
```

Pero aunque aparentemente converge antes, parece no haber una mejora en la precisión de la aproximación. Si calculamos el porcentaje (estimado) de reducción del error:

```
100*(0.1619886-0.1641059)/0.1619886
```

```
## [1] -1.307067
```

El problema es que en este caso se está estimando mal la varianza (asumiendo independencia). Hay que tener cuidado con las técnicas de reducción de la varianza si uno de los objetivos de la simulación es precisamente estimar la variabilidad. En este caso, una versión de la función anterior para integración Monte Carlo con variables antitéticas, con aproximación del error bajo dependencia podría ser:

```
mc.integrala2 <- function(ftn, a, b, n, plot = TRUE,...) {
  # n es el nº de evaluaciones de la función objetivo (para facilitar comparaciones, solo se genera la primera)
  x <- runif(n%/%2, a, b)
  # La siguiente línea solo para representar alternando
  x <- matrix(c(x,a+b-x),nrow=2,byrow=TRUE)
  # bastaría con emplear p.e. c(x,a+b-x)
  fx <- apply(x, 1, ftn)*(b-a)
  corr <- cor(fx[,1], fx[,2])
  fx <- as.numeric(fx)
  return(list(valor=mean(fx), error=2*sd(fx)/sqrt(n)*sqrt(1+corr)))
}

set.seed(54321)
```

```
res <- mc.integrala2(ftn, a, b, 500)
res

## $valor
## [1] 3.222366
##
## $error
## [1] 0.05700069
```

Porcentaje estimado de reducción del error:

```
100*(0.1619886-0.05700069)/0.1619886
```

```
## [1] 64.81191
```

En este caso puede verse que la reducción teórica de la varianza es del 96.7%

10.3 Estratificación

Si se divide la población en estratos y se genera en cada uno un número de observaciones proporcional a su tamaño (a la probabilidad de cada uno) nos aseguramos de que se cubre el dominio de interés y se puede acelerar la convergencia.

- Por ejemplo, para generar una muestra de tamaño n de una $\mathcal{U}(0, 1)$, se pueden generar $l = \frac{n}{k}$ observaciones ($1 \leq k \leq n$) de la forma:

$$U_{j_1}, \dots, U_{j_k} \sim \mathcal{U}\left(\frac{(j-1)}{k}, \frac{j}{k}\right) \text{ para } j = 1, \dots, k.$$

Si en el número de observaciones se tiene en cuenta la variabilidad en el estrato se puede obtener una reducción significativa de la varianza.

Ejemplo 10.1. Muestreo estratificado de una exponencial (libro Ricardo)

Supóngase el siguiente problema (absolutamente artificial pero ilustrativo para comprender esta técnica). Dada una muestra de tamaño 10 de una población con distribución:

$$X \sim \exp(1),$$

se desea aproximar la media poblacional (es sobradamente conocido que es 1) a partir de 10 simulaciones. Supongamos que para evitar que, por puro azar, exista alguna zona en la que la exponencial toma valores, no representada en la muestra simulada de 10 datos, se consideran tres estratos. Por ejemplo, el del 40% de valores menores, el siguiente 50% de valores (intermedios) y el 10% de valores más grandes para esta distribución.

El algoritmo de inversión (optimizado) para simular una $\exp(1)$ es:

1. Generar $U \sim U(0, 1)$.
2. Hacer $X = -\ln U$.

Dado que, en principio, simulando diez valores $U_1, U_2, \dots, U_{10} \sim U(0, 1)$, no hay nada que nos garantice que las proporciones de los estratos son las deseadas (aunque sí lo serán en media). Una forma de garantizar el que obtengamos **4, 5 y 1** valores, respectivamente, en cada uno de los tres estratos, consiste en simular:

- 4 valores de $U[0.6, 1]$ para el primer estrato,
- 5 valores de $U[0.1, 0.6)$ para el segundo y
- uno de $U[0, 0.1)$ para el tercero.

Otra forma de proceder consistiría en rechazar valores de U que caigan en uno de esos tres intervalos cuando el cupo de ese estrato esté ya lleno (lo cual no sería computacionalmente eficiente).

El algoritmo con la estratificación propuesta sería como sigue:

1. Para $i = 1, 2, \dots, 10$:
2. Generar U_i :
 - 2a. Generar $U \sim U(0, 1)$.
 - 2b. Si $i \leq 4$ hacer $U_i = 0.4 \cdot U + 0.6$.
 - 2c. Si $4 < i \leq 9$ hacer $U_i = 0.5 \cdot U + 0.1$.
 - 2d. Si $i = 10$ hacer $U_i = 0.1 \cdot U$.
3. Devolver $X_i = -\ln U_i$.

No es difícil probar que:

- $\text{Var}(X_i) = 0.0214644$ si $i = 1, 2, 3, 4$,
- $\text{Var}(X_i) = 0.229504$ si $i = 5, 6, 7, 8, 9$ y
- $\text{Var}(X_{10}) = 1$.

Como consecuencia:

$$\text{Var}(\bar{X}) = \frac{1}{10^2} \sum_{i=1}^{10} \text{Var}(X_i) = 0.022338$$

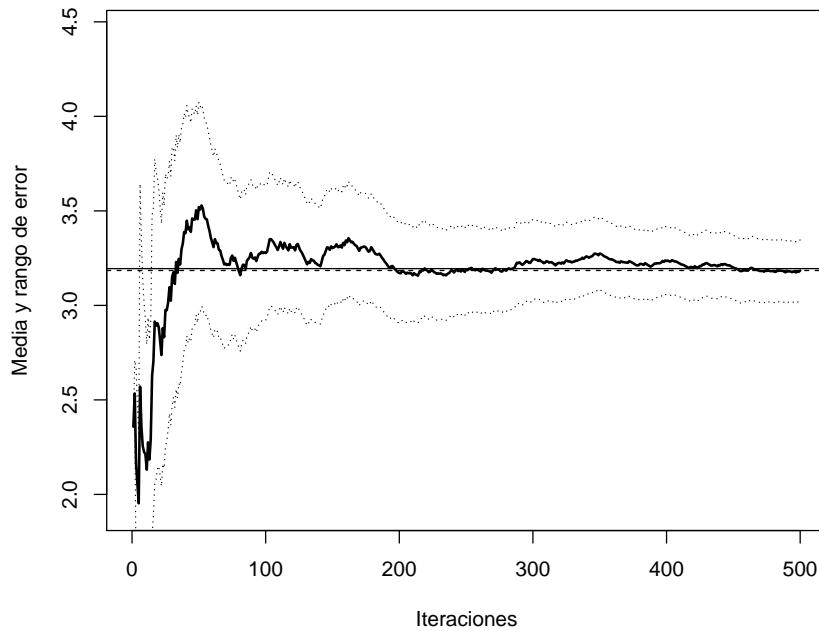
que es bastante menor que 1 (la varianza en el caso de muestreo aleatorio simple no estratificado).

Ejercicio 10.2. Integración Monte Carlo con estratificación

Aproximar la integral anterior empleando la técnica de estratificación, considerando k subintervalos regularmente espaciados en el intervalo $[0, 2]$. ¿Cómo varía la reducción en la varianza dependiendo del valor de k ?

```
mc.integrale <- function(ftn, a, b, n, k) {
  # Integración Monte Carlo con estratificación
  l <- n%/%k
  int <- seq(a, b, len=k+1)
  x <- runif(l*k, rep(int[-(k+1)], each=1), rep(int[-1], each=1))
  # l uniformes en cada uno de los intervalos [(j-1)/k , j/k]
  fx <- sapply(x, ftn)*(b-a)
  return(list(valor=mean(fx), error=2*sd(fx)/sqrt(n)))  # error mal calculado
}

set.seed(54321)
res <- mc.integrale(ftn, a, b, 500)
abline(h = teor)
```



```
res
```

```
## $valor
## [1] 3.184612
##
## $error
## [1] 0.1619886

set.seed(54321)
mc.integrale(ftn, a, b, 500, 50)

## $valor
## [1] 3.193338
##
## $error
## [1] 0.1597952

set.seed(54321)
mc.integrale(ftn, a, b, 500, 100)

## $valor
## [1] 3.193927
##
## $error
## [1] 0.1599089
```

De esta forma no se tiene en cuenta la variabilidad en el estrato. El tamaño de las submuestras debería incrementarse hacia el extremo superior.

Ejercicio 10.3.

Repetir el ejemplo anterior considerando intervalos regularmente espaciados en escala exponencial.

10.4 Variables de control

En este caso se trata de sacar partido tanto a una covarianza positiva como negativa. La idea básica es emplear una variable Y , con media conocida μ_Y , para controlar la variable X (con media desconocida), de forma que ambas variables estén “suficientemente” correlacionadas. La versión “controlada” de X será:

$$X^* = X + \alpha(Y - \mu_Y)$$

con $E(X^*) = E(X) = \theta$. Puede verse que $Var(X^*) = Var(X) + \alpha^2 Var(Y) + 2\alpha Cov(X, Y)$ es mínima para:

$$\alpha^* = -\frac{Cov(X, Y)}{Var(Y)},$$

con $Var(X^*) = Var(X)(1 - \rho^2(X, Y))$ (lo que supone una reducción del $100\rho^2(X, Y)\%$).

En la práctica normalmente α^* no es conocida. Para estimarlo se puede realizar ajuste lineal de X sobre Y (a partir de los datos simulados X_i e Y_i , $1 \leq i \leq n$):

- Si $\hat{x} = \hat{\beta}_0 + \hat{\beta}_1 y$ es la recta ajustada, con $\hat{\beta}_1 = \frac{S_{XY}}{S_Y^2}$ y $\hat{\beta}_0 = \bar{X} - \hat{\beta}_1 \bar{Y}$, la estimación sería:

$$\hat{\alpha}^* = -\hat{\beta}_1$$

- Adicionalmente, para aproximar θ :

$$\begin{aligned}\hat{\theta} &= \bar{X}^* = \bar{X} - \hat{\beta}_1(\bar{Y} - \mu_Y) \\ &= \hat{\beta}_0 + \hat{\beta}_1 \mu_Y\end{aligned}$$

- Si $\mu_Y = 0 \Rightarrow \hat{\theta} = \bar{X}^* = \hat{\beta}_0$.

Ejercicio 10.4. Integración Monte Carlo con variables de control

Aproximar la integral anterior empleando la variable $U \sim \mathcal{U}(0, 2)$ para controlar la variable e^U .

Se trata de calcular la media de $\exp(\mathcal{U}(a, b))$:

```
a <- 0; b <- 2
teor <- (exp(b)-exp(a))/(b-a)
teor
```

```
## [1] 3.194528
```

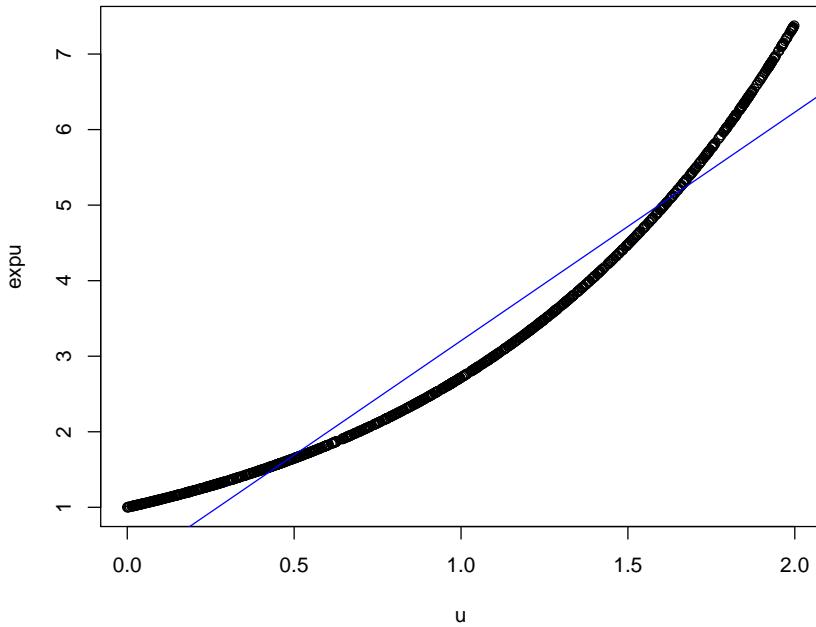
Aproximación clásica por simulación:

```
set.seed(54321)
nsim <- 1000
u <- runif(nsim, a, b)
expu <- exp(u)
mean(expu)
```

```
## [1] 3.182118
```

Con variable control:

```
plot(u, expu)
reg <- lm(expu ~ u)$coef
abline(reg, col='blue')
```



```
# summary(lm(expu ~ u)) # R-squared: 0.9392
reg[1]+reg[2] # Coincidirá con la solución mean(expuc)
```

```
## (Intercept)
##      3.204933
```

Lo siguiente ya no sería necesario:

```
expuc <- expu - reg[2]*(u-1)
mean(expuc)
```

```
## [1] 3.204933
```

Estimación del porcentaje de reducción en la varianza:

```
100*(var(expu)-var(expuc))/var(expu)
```

```
## [1] 93.91555
```

10.5 Números aleatorios comunes

Se trataría de una técnica básica del diseño de experimentos: realizar comparaciones homogéneas (bloquear). Por ejemplo cuando se diseña un experimento para la comparación de la media de dos variables, se pueden emplear las denominadas muestras apareadas, en lugar de muestras independientes.

Supóngamos que estamos interesados en las diferencias entre dos estrategias (e.g. dos estimadores):

$$E(X) - E(Y) = E(X - Y).$$

Para ello se generan dos secuencias X_1, X_2, \dots, X_n , e Y_1, Y_2, \dots, Y_n y se calcula:

$$\bar{X} - \bar{Y} = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)$$

- Si las secuencias se generan de modo independiente:

$$Var(\bar{X} - \bar{Y}) = \frac{1}{n} (Var(X) + Var(Y))$$

- Si se generan las secuencias empleando la **misma semilla**, los datos son dependientes:

$$Cov(X_i, Y_i) > 0$$

y tendríamos que:

$$\begin{aligned} Var(\bar{X} - \bar{Y}) &= \frac{1}{n^2} \sum_{i=1}^N Var(X_i - Y_i) = \frac{1}{n} Var(X_i - Y_i) \\ &= \frac{1}{n} (Var(X_i) + Var(Y_i) - 2Cov(X_i, Y_i)) \\ &\leq \frac{1}{n} (Var(X_i) + Var(Y_i)) \end{aligned}$$

En el capítulo de aplicaciones de la simulación se empleó esta técnica para comparar distribuciones de estimadores...

10.6 Ejercicios fin de práctica

Ejercicio 10.5.

Aproximar mediante integración Monte Carlo (clásica) la media de una distribución exponencial de parámetro 1/2:

$$I = \int_0^\infty \frac{x}{2} e^{-\frac{x}{2}} dx$$

y representar gráficamente la aproximación en función de n . Comparar los resultados con los obtenidos empleando variables antitéticas, ¿se produce una reducción en la varianza?

Nota: Puede ser recomendable emplear el método de inversión para generar las muestras (antitéticas) de la exponencial.

MC clásico:

```
nsim <- 1000
lambda <- 0.5
set.seed(1)
x <- - log(runif(nsim)) / lambda
# Aprox por MC da media
mean(x) # valor teor 1/lambda = 2

## [1] 1.97439
# Aprox da precisión
var(x)

## [1] 3.669456
```

MC con variables antitéticas:

```
# xa <-
# mean(xa) # Aprox por MC da media (valor teor 1/lambda = 2)
# var(xa) # Aprox da precisión suponiendo independencia
# corr <- cor(x1,x2)
# var(xa)*(1 + corr) # Estimación varianza suponiendo dependencia
```

Estimación del porcentaje de reducción en la varianza

```
# 100*(var(x) - var(xa))/var(x)
```

Referencias

Bibliografía básica

- Cao, R. (2002). *Introducción a la simulación y a la teoría de colas*. NetBiblo.
- Gentle, J.E. (2003). *Random number generation and Monte Carlo methods*. Springer-Verlag.
- Jones, O. et al. (2009). *Introduction to Scientific Programming and Simulation Using R*. CRC.
- Ripley, B.D. (1987). *Stochastic Simulation*. John Wiley & Sons.
- Robert, C.P. y G. Casella (2010). *Introducing Monte Carlo Methods with R*. Springer.
- Ross, S.M. (1999). *Simulación*. Prentice Hall.
- Suess, E.A. y Trumbo, B.E. (2010). *Introduction to probability simulation and Gibbs sampling with R*. Springer.

Bibliografía complementaria

- Azarang, M. R. y García Dunna, E. (1996). *Simulación y análisis de modelos estocásticos*. McGraw-Hill.
- Bratley, P., Fox, B. L. y Schrage L. E. (1990). *A guide to simulation*. Springer-Verlag.
- Devroye, L. (1986). *Non-uniform random variate generation*. Springer-Verlag.
- Evans, M. y Swartz, T. (2000). *Approximating integrals via Monte Carlo and deterministic methods*. Oxford University Press.
- Gentle, J.E. (1998). *Random number generation and Monte Carlo methods*. Springer-Verlag.
- Hörmann, W. et al. (2004). *Automatic Nonuniform Random Variate Generation*. Springer.
- Law, A.M. y Kelton, W.D. (1991). *Simulation, modeling and analysis*. McGraw-Hill.
- Moeschlin, O., Grycko, E., Pohl, C. y Steinert, F. (1998). *Experimental stochastics*. Springer-Verlag.
- Nelson, R. (1995). *Probability, stochastic processes, and queueing theory: the mathematics of computer performance modelling*. Springer-Verlag.
- Pardo, L. y Valdés, T. (1987). *Simulación. Aplicaciones prácticas a la empresa*. Díaz de Santos.
- Robert, C.P. y G. Casella (2004). *Monte Carlo statistical methods*. Springer.

Apéndice A

Enlaces

Recursos para el aprendizaje de R

A continuación se muestran algunos recursos que pueden ser útiles para el aprendizaje de R y la obtención de ayuda (basados en el post <https://rubenfcasal.github.io/post/ayuda-y-recursos-para-el-aprendizaje-de-r>, que puede estar más actualizado).

Ayuda online:

- Ayuda en línea sobre funciones o paquetes: RDocumentation
- Buscador RSeek
- StackOverflow

Cursos: algunos cursos gratuitos:

- Coursera:
 - Introducción a Data Science: Programación Estadística con R
 - Mastering Software Development in R
- DataCamp:
 - Introducción a R
- Stanford online:
 - Statistical Learning
- Curso UCA: Introducción a R, R-commander y shiny
- Udacity: Data Analysis with R
- Swirl Courses: se pueden hacer cursos desde el propio R con el paquete swirl.

Para información sobre cursos en castellano se puede recurrir a la web de R-Hispano en el apartado formación. Algunos de los cursos que aparecen en entradas antiguas son gratuitos. Ver: Cursos MOOC relacionados con R.

Libros

- *Iniciación:*
 - 2011 - The Art of R Programming. A Tour of Statistical Software Design, (No Starch Press)
 - R for Data Science (online, O'Reilly)

- Hands-On Programming with R: Write Your Own Functions and Simulations, by Garrett Grolemund (O'Reilly)
- **Avanzados:**
 - 2008 - Software for Data Analysis: Programming with R - Chambers (Springer)
 - Advanced R by Hadley Wickham (online: 1^a ed, 2^a ed, Chapman & Hall)
 - R packages by Hadley Wickham (online, O'Reilly)
- **Bookdown:** el paquete `bookdown` de R permite escribir libros empleando R Markdown y compartirlos. En <https://bookdown.org> está disponible una selección de libros escritos con este paquete (un listado más completo está disponible aquí). Algunos libros en este formato en castellano son:
 - Técnicas de Remuestreo (disponible en el repositorio de GitHub [rubencasal/book_remuestreo](https://github.com/rubencasal/book_remuestreo)).
 - Introducción al Análisis de Datos con R (disponible en el repositorio de GitHub [rubencasal/intror](https://github.com/rubencasal/intror)).
 - Prácticas de Tecnologías de Gestión y Manipulación de Datos (disponible en el repositorio de GitHub [gltaboadal/tgdbook](https://github.com/gltaboadal/tgdbook)).
 - Escritura de libros con bookdown (disponible en el repositorio de GitHub [rubencasal/bookdown_intro](https://github.com/rubencasal/bookdown_intro)).
 - R para profesionales de los datos: una introducción.
 - Estadística Básica Edulcorada.

Material online: en la web se puede encontrar mucho material adicional, por ejemplo:

- CRAN: Other R documentation
- RStudio:
 - Online learning, Webinars
 - tidyverse: dplyr, tibble, tidyr, stringr, readr.
 - CheatSheets: RMarkdown, Shiny, dplyr, tidyr, stringr.
- Blogs en inglés:
 - <https://www.r-bloggers.com/>
 - <https://www.littlemissdata.com/blog/rstudioconf2019>
 - RStudio: <https://blog.rstudio.com>
 - Microsoft Revolutions: <https://blog.revolutionanalytics.com>
- Blogs en castellano:
 - <https://www.datanalytics.com>
 - <http://oscarperpinan.github.io/R>
 - <http://rubencasal.github.io>
- Listas de correo:
 - Listas de distribución de r-project.org: <https://stat.ethz.ch/mailman/listinfo>
 - Búsqueda en R-help: <http://r.789695.n4.nabble.com/R-help-f789696.html>
 - Búsqueda en R-help-es: <https://r-help-es.r-project.narkive.com>
 - <https://grokbase.com/g/r/r-help-es>

- Archivos de R-help-es: <https://stat.ethz.ch/pipermail/r-help-es>

Apéndice B

Integración numérica

B.1 Introducción

En muchos casos nos puede interesar la aproximación de una integral definida. En estadística, además del caso de Inferencia Bayesiana (que trataremos más adelante empleando Integración Montecarlo), nos puede interesar por ejemplo aproximar mediante simulación el error cuadrático integrado medio (MISE) de un estimador. En el caso de una densidad univariante sería de la forma:

$$MISE \left\{ \hat{f} \right\} = E \int (\hat{f}(x) - f(x))^2 dx$$

Cuando el numero de dimensiones es pequeño, nos puede interesar emplear un método numérico para aproximar este tipo de integrales.

B.2 Integración numérica unidimensional

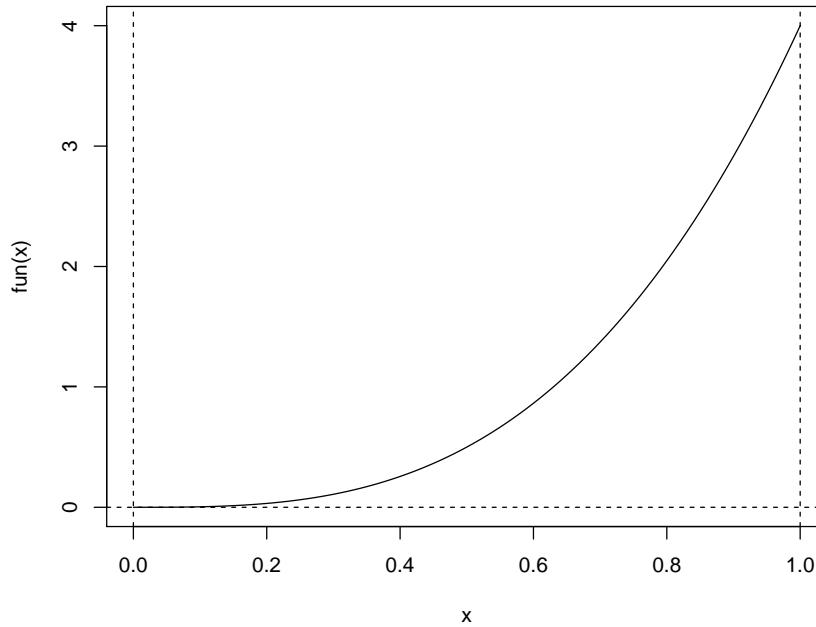
Supongamos que nos interesa aproximar una integral de la forma:

$$I = \int_a^b h(x) dx$$

Consideraremos como ejemplo:

$$\int_0^1 4x^3 dx = 1$$

```
fun <- function(x) return(4 * x^3)
curve(fun, 0, 1)
abline(h = 0, lty = 2)
abline(v = c(0, 1), lty = 2)
```



B.2.1 Método del trapezoide

La regla de los trapecios es una forma de aproximar la integral utilizando n trapecios. Si se consideran n subintervalos en $[a, b]$ de longitud $h = \frac{b-a}{n}$ (i.e. $n+1$ puntos regularmente espaciados cubriendo el dominio), y se aproxima linealmente la función en cada subintervalo, se obtiene que:

$$\int_a^b f(x) dx \approx \frac{h}{2} [f(a) + 2f(a+h) + 2f(a+2h) + \dots + f(b)]$$

```
trapezoid.vec <- function(f.vec, h = 0.01) {
  # Integración numérica unidimensional entre a y b
  # utilizando el método del trapezoide
  # (se approxima f linealmente en cada intervalo)
  n <- length(f.vec)
  return(h*(f.vec[1]/2 + sum(f.vec[2:(n-1)]) + f.vec[n]/2))
}

trapezoid <- function(fun, a = 0, b = 1, n = 100) {
  # Integración numérica de fun (función unidimensional) entre a y b
  # utilizando el método del trapezoide con n subdivisiones
  # (se approxima f linealmente en cada intervalo)
  # Se asume a < b y n entero positivo
  h <- (b-a)/n
  x.vec <- seq(a, b, by = h)
  f.vec <- sapply(x.vec, fun)
  return(trapezoid.vec(f.vec, h))
}

trapezoid(fun, 0, 1, 20)
```

```
## [1] 1.0025
```

El error en esta aproximación se corresponde con:

$$\frac{(b-a)^3}{12n^2} f''(\xi),$$

para algún $a \leq \xi \leq b$ (dependiendo del signo de la segunda derivada, i.e. de si la función es cóncava o convexa, el error será negativo ó positivo). El error máximo absoluto es $\frac{(b-a)^3}{12n^2} \max_{a \leq \xi \leq b} |f''(\xi)|$. En el caso general multidimensional sería $O(n^{-\frac{2}{d}})$.

B.2.2 Regla de Simpson

Se divide el intervalo n subintervalos de longitud $h = \frac{b-a}{n}$ (con n par), considerando $n+1$ puntos regularmente espaciados $x_i = a + ih$, para $i = 0, 1, \dots, n$. Aproximando de forma cuadrática la función en cada subintervalo $[x_{j-1}, x_{j+1}]$ (considerando 3 puntos), se obtiene que:

$$\int_a^b f(x) dx \approx \frac{h}{3} \left[f(x_0) + 2 \sum_{j=1}^{(n/2)-1} f(x_{2j}) + 4 \sum_{j=1}^{n/2} f(x_{2j-1}) + f(x_n) \right],$$

```
simpson <- function(fun, a, b, n = 100) {
  # Integración numérica de fnt entre a y b
  # utilizando la regla de Simpson con n subdivisiones
  # (se aproxima fun de forma cuadrática en cada par de intervalos)
  # fnt es una función de una sola variable
  # Se asume a < b y n entero positivo par
  n <- max(c(2*(n %% 2), 4))
  h <- (b-a)/n
  x.vec1 <- seq(a+h, b-h, by = 2*h)
  x.vec2 <- seq(a+2*h, b-2*h, by = 2*h)
  f.vec1 <- sapply(x.vec1, fun)
  f.vec2 <- sapply(x.vec2, fun)
  return(h/3*(fun(a) + fun(b) + 4*sum(f.vec1) + 2*sum(f.vec2)))
  # Una cota del error en valor absoluto es:
  # h^4*(b-a)*max(c(f.vec1, fvec.2))^4/180.
}
simpson(fun, 0, 1, 20)
```

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

El máximo error (en el caso unidimensional) viene dado por la expresión:

$$\frac{(b-a)^5}{180n^4} \max_{a \leq \xi \leq b} |f^{(4)}(\xi)|.$$

En el caso general multidimensional sería $O(n^{-\frac{4}{d}})$.

B.2.3 Cuadratura adaptativa

En lugar de evaluar la función en una rejilla regular (muestrear por igual el dominio), puede interesar ir añadiendo puntos sólo en los lugares “necesarios”.

```

quadrature <- function(fun, a, b, tol=1e-8) {
  # numerical integration using adaptive quadrature

  simpson2 <- function(fun, a, b) {
    # numerical integral using Simpson's rule
    # assume a < b and n = 2
    return((b-a)/6 * (fun(a) + 4*fun((a+b)/2) + fun(b)))
  }

  quadrature_internal <- function(S.old, fun, a, m, b, tol, level) {
    level.max <- 100
    if (level > level.max) {
      cat ("recursion limit reached: singularity likely\n")
      return (NULL)
    }
    S.left <- simpson2(fun, a, m)
    S.right <- simpson2(fun, m, b)
    S.new <- S.left + S.right
    if (abs(S.new-S.old) > tol) {
      S.left <- quadrature_internal(S.left, fun,
                                      a, (a+m)/2, m, tol/2, level+1)
      S.right <- quadrature_internal(S.right, fun,
                                      m, (m+b)/2, b, tol/2, level+1)
      S.new <- S.left + S.right
    }
    return(S.new)
  }

  level = 1
  S.old <- (b-a) * (fun(a) + fun(b))/2
  S.new <- quadrature_internal(S.old, fun,
                                a, (a+b)/2, b, tol, level+1)
  return(S.new)
}

quadrature(fun, 0, 1)

## [1] 1

```

Fuente r-blogger Guangchuang Yu?

B.2.4 Comandos de R

```

integrate(fun, 0, 1)    # Permite límites infinitos

## 1 with absolute error < 1.1e-14
## Cuidado: fun debe ser vectorial...

require(MASS)
area(fun, 0, 1)

## [1] 1

```

B.3 Integración numérica bidimensional

Supongamos que nos interesa aproximar una integral de la forma:

$$I = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y) dy dx$$

Consideraremos como ejemplo:

$$\int_{-1}^1 \int_{-1}^1 (x^2 - y^2) dx dy = 0$$

```
f2d <- function(x,y) x^2 - y^2
```

Es habitual (especialmente en simulación) que la función se evalúe en una rejilla:

```
ax = -1
ay = -1
bx = 1
by = 1
nx = 21
ny = 21
x <- seq(ax, bx, length = nx)
y <- seq(ay, by, length = ny)
z <- outer(x, y, f2d)

hx <- x[2]-x[1]
hy <- y[2]-y[1]
```

B.3.1 Representación gráfica

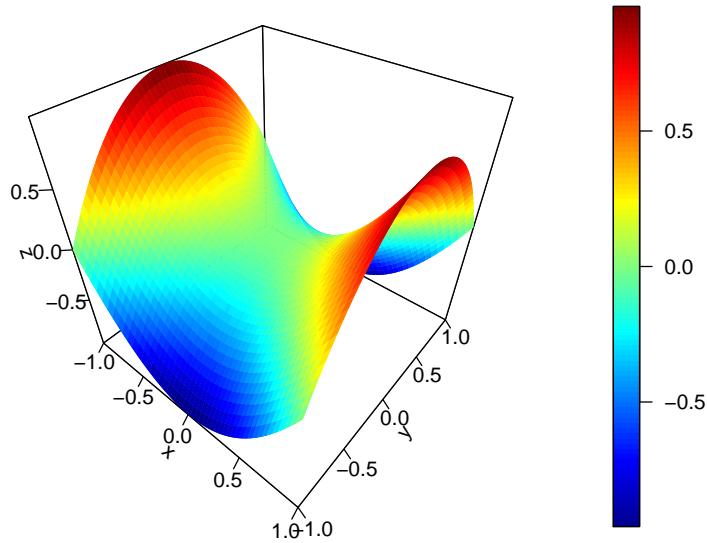
Puede ser de utilidad las herramientas de los paquetes `plot3D` y `plot3Drgl` (también se pueden utilizar las funciones `spersp`, `simage`, `spoints` y `splot` del paquete `npsp`).

```
if(!require(plot3D)) stop('Required package `plot3D` not installed.')

# persp3D(z = z, x = x, y = y)

persp3D.f2d <- function(f2d, ax=-1, bx=1, ay=-1, by=1, nx=21, ny=21, ...) {
  x <- seq(ax, bx, length = nx)
  y <- seq(ay, by, length = ny)
  hx <- x[2]-x[1]
  hy <- y[2]-y[1]
  z <- outer(x, y, f2d)
  persp3D(x, y, z, ...)
}

persp3D.f2d(f2d, -1, 1, -1, 1, 50, 50, 1, ticktype = "detailed")
```



B.3.2 Método del trapezoide

Error $O(n^{-\frac{2}{3}})$.

```

trapezoid.mat <- function(z, hx, hy) {
# Integración numérica bidimensional
# utilizando el método del trapezoide (se aproxima f linealmente)
  f.vec <- apply(z, 1, function(x) trapezoid.vec(x, hx))
  return(trapezoid.vec(f.vec, hy))
}

# trapezoid.mat(z, hx, hy)

trapezoid.f2d <- function(f2d, ax=-1, bx=1, ay=-1, by=1, nx=21, ny=21) {
  x <- seq(ax, bx, length = nx)
  y <- seq(ay, by, length = ny)
  hx <- x[2]-x[1]
  hy <- y[2]-y[1]
  z <- outer(x, y, f2d)
  trapezoid.mat(z, hx, hy)
}

trapezoid.f2d(f2d, -1, 1, -1, 1, 101, 101)

## [1] -8.881784e-18

```

B.3.3 Comandos de R

Suponiendo que la función es vectorial, podemos emplear:

```
integrate( function(y) {  
  sapply(y, function(y) {  
    integrate(function(x) f2d(x,y), ax, bx)$value }) },  
  ay, by)  
  
## -2.775558e-17 with absolute error < 1.1e-14
```

Si la función no es vectorial y solo admite parámetros escalares:

```
integrate(function(y) {  
  sapply(y, function(y) {  
    integrate(function(x) {  
      sapply(x, function(x) f2d(x,y)) }, ax, bx)$value }) },  
  ay, by)
```

Fuente tolstoy.newcastle.edu.au.

Alternativamente se podría emplear la función `adaptIntegrate` del paquete `cubature`.