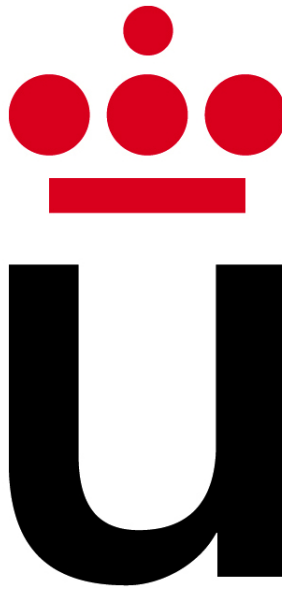


CASO PRÁCTICO II
Simulación y Metaheurísticas
GENERACIÓN DE VARIABLES ALEATORIAS

José Ignacio Escribano



MÓSTOLES, 1 DE MAYO DE 2016

Índice de tablas

1.	Comparativa de distintos parámetros muestrales de tres distribuciones normales con distinto número de muestras	2
2.	Comparativa de distintos parámetros muestrales de tres distribuciones normales generadas con el generador estándar de R y usando el método de 12 uniformes	2

Índice de figuras

1.	Histograma de una distribución normal con 10, 100 y 10000 muestras . . .	1
2.	Comparativa de los histogramas de la distribución normal generados usando distintos métodos	3
3.	10000 muestras de una distribución exponencial	3
4.	10000 muestras de una distribución normal bivalente	5
5.	Función de densidad de 10000 muestras de una distribución normal bivalente	5

Índice

1. Introducción	1
2. Resolución del caso práctico	1
2.1. Cuestión 1	1
2.2. Cuestión 2	2
2.3. Cuestión 3	3
2.4. Cuestión 4	4
3. Conclusiones	6
4. Código R utilizado	7

1. Introducción

En este caso práctico utilizaremos distintos métodos para generar variables aleatorias; tanto unidimensionales como multidimensionales. En el primer caso generaremos la distribución exponencial y la normal. En el caso multidimensional, generaremos la normal.

2. Resolución del caso práctico

A continuación resolveremos cada una de las cuestiones planteadas.

2.1. Cuestión 1

Usando el comando `rnorm` generaremos 10, 100 y 10000 observaciones de una distribución normal de parámetros $(\mu = 3, \sigma^2 = 6^2)$.

La Figura 1 muestra los histogramas con cada número de muestras.

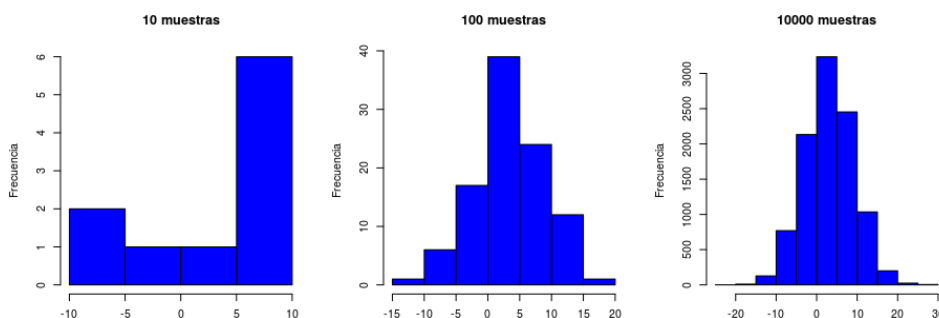


Figura 1: Histograma de una distribución normal con 10, 100 y 10000 muestras

Se puede observar como, a medida que se aumenta el número de muestras de la distribución, el histograma toma la forma característica de una distribución normal: simétrica y con forma de campana.

Usaremos distintos estadísticos muestrales para comprobar el hecho anterior. Estos estadísticos serán la media, varianza, primer y tercer cuartil, la desviación típica, la moda, el kurtosis y la asimetría.

La Tabla 1 muestra una comparativa de los distintos estadísticos según el número de muestras de la distribución normal.

Se puede observar que, según aumentamos el número de muestras que tomamos, los resultados de los estadísticos se acercan más a los resultados analíticos. Hay que tener en

Tabla 1: Comparativa de distintos parámetros muestrales de tres distribuciones normales con distinto número de muestras

Distribución	Media	Varianza	1Q	3Q	Desviación típica	Moda	Kurtosis	Asimetría	Mediana
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$ con 10 muestras	2.788	35.593	-2.168	6.826	5.966	2.788	1.820	-0.767	5.735
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$ con 100 muestras	3.537	31.452	0.148	7.501	5.608	3.537	3.319	-0.129	3.573
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$ con 10000 muestras	3.089	36.470	-0.954	6.039	7.172	3.088	3.045	0.021	3.0340
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$	3	36	-1.046	7.046	6	3	3	0	3

cuenta que para tener un resultado más fiable habría que repetir la simulación varias veces y calcular la media de esas simulaciones.

2.2. Cuestión 2

Simulamos 10000 observaciones de una distribución normal $\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$ con el método de suma de 12 uniformes. Recordemos que este método genera una distribución normal $\mathcal{N}(0, 1)$. Si queremos obtener una distribución normal $\mathcal{N}(\mu, \sigma^2)$ haremos

$$X = \mu + \sigma Y$$

donde $X \sim \mathcal{N}(\mu, \sigma^2)$ y $Y \sim \mathcal{N}(0, 1)$.

Además, lo comparamos con 10000 observaciones del generador normal de R y con los resultados de la Cuestión 2.1. Los resultados de todos ellos se puede ver en la Tabla 2 y la Figura 2 muestra una comparación entre los histogramas.

Tabla 2: Comparativa de distintos parámetros muestrales de tres distribuciones normales generadas con el generador estándar de R y usando el método de 12 uniformes

Distribución	Media	Varianza	1Q	3Q	Desviación típica	Moda	Kurtosis	Asimetría	Mediana
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$ de la Cuestión 2.1	3.089	36.470	-0.954	6.039	7.172	3.088	3.045	0.021	3.0340
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$ con 10000 muestras	2.993	37.691	-1.169	7.154	6.139	2.993	2.970	-0.051	3.096
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$ con 12 uniformes	3.037	36.614	-1.101	7.131	6.139	3.037	2.899	0.008	3.096
$\mathcal{N}(\mu = 3, \sigma^2 = 6^2)$	3	36	-1.046	7.046	6	3	3	0	3

Se puede observar que todos los métodos obtienen estadísticos descriptivos muy parecidos a su valor analítico. Esto hace que estemos ante buenos métodos de generación de distribuciones aleatorias, en este caso, una distribución normal.

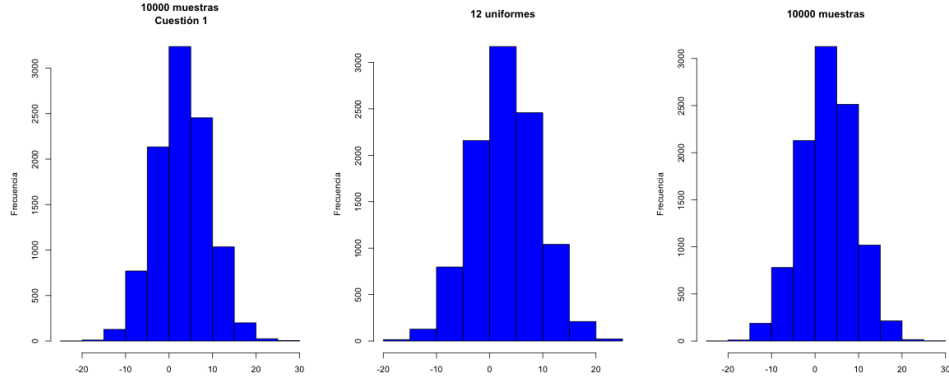


Figura 2: Comparativa de los histogramas de la distribución normal generados usando distintos métodos

2.3. Cuestión 3

Usaremos el método de inversión para obtener 10000 muestras de una distribución $\mathcal{Exp}(\lambda = 9)$. El histograma de las muestras generadas se muestra en la Figura 3.

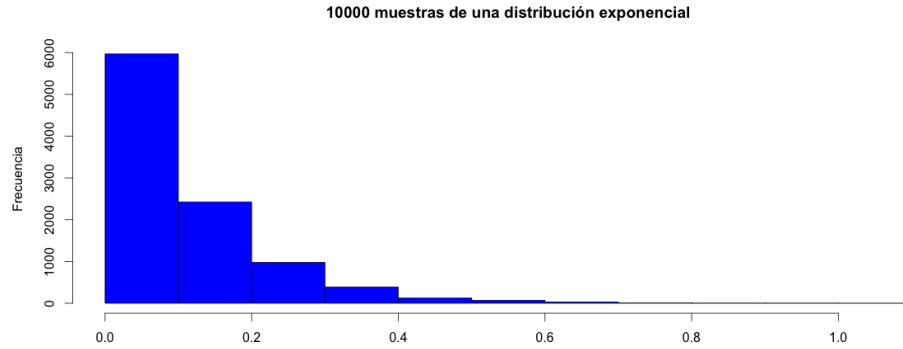


Figura 3: 10000 muestras de una distribución exponencial

Calculamos la media y el coeficiente de variación de la muestra. Recordemos que el coeficiente de variación de la muestra x con media μ y desviación típica σ viene dado por:

$$CV = \frac{\sigma}{|\mu|}$$

La media de la muestra es de 0.109 y el coeficiente de variación es de 0.998.

Recordemos que la media de una distribución exponencial de parámetro λ es $\frac{1}{\lambda}$. En nuestro caso, $\lambda = 9$ y, $\frac{1}{9} = 0.111 \approx 0.109$.

Por otra parte, se tiene que la varianza de una distribución exponencial de parámetro λ es $\frac{1}{\lambda^2}$, y por tanto, la desviación típica es $\frac{1}{\lambda}$. Por tanto, el coeficiente de variación de una distribución exponencial es 1.

Todo lo anterior hace que estemos ante una muy buena aproximación de la distribución exponencial.

2.4. Cuestión 4

A continuación generaremos 10000 muestras de una distribución normal bivalente dado por el vector de medias $\mu = (\mu_x = 3, \mu_y = 2)$, y matriz de varianzas y covarianzas Σ , donde Σ viene dada por

$$\begin{pmatrix} \sigma_x^2 = 7^2 & \sigma_{xy}^2 = 4^2 \\ \sigma_{xy}^2 = 4^2 & \sigma_y^2 = 5^2 \end{pmatrix}$$

En primer lugar, calculamos la matriz de Cholesky de Σ , obteniéndose,

$$L = \begin{pmatrix} 2.645751 & 0.000000 \\ 1.511858 & 1.647509 \end{pmatrix}$$

Se puede comprobar que, efectivamente, $\Sigma = L \cdot L^T$.

La muestra generada tiene las siguientes medias:

$$\mu_x = 2.970041, \quad \mu_y = 1.982205$$

Las varianzas son:

$$\sigma_x = 6.911890, \quad \sigma_y = 5.011215$$

Y la covarianza es:

$$\sigma_{xy} = 3.97125$$

De acuerdo a estos resultados, estamos ante a una muy buena aproximación de una distribución normal bivalente.

En la Figura 4 se muestra un scatterplot de las 10000 muestras generadas.

Recordemos que la función de densidad de una distribución normal multivariante viene dada por

$$f(\mathbf{x} = (x_1, \dots, x_n) | \mu, \Sigma) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \right)$$

La Figura 5 muestra la función de densidad de la distribución generada.

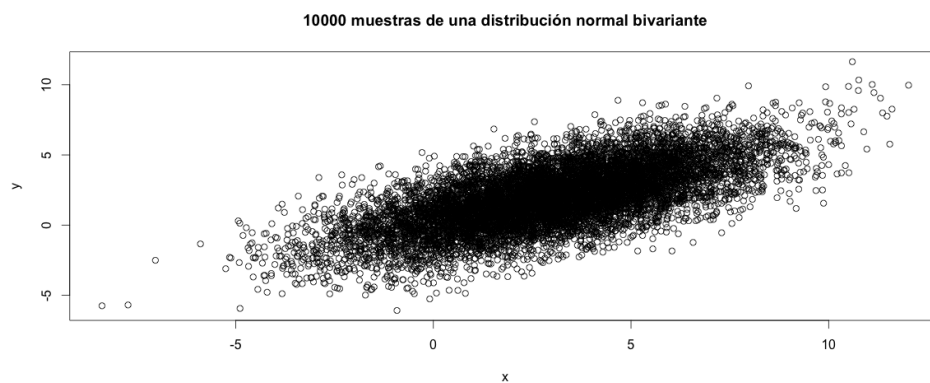


Figura 4: 10000 muestras de una distribución normal bivalente

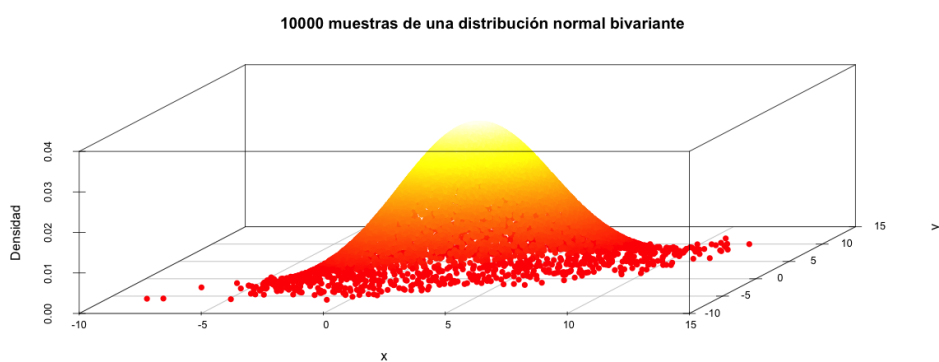


Figura 5: Función de densidad de 10000 muestras de una distribución normal bivalente

3. Conclusiones

En este caso práctico hemos aprendido a generar distintos tipos de variables aleatorias tanto unidimensionales como multidimensionales, aplicando además, varios métodos de generación: método de las 12 uniformes para la generación de una variable aleatoria normal, el método de inversión para generar variables aleatorias cuya función de distribución es creciente (como en el caso de la distribución exponencial), entre otros.

Además, gracias a la gran cantidad de funciones disponibles y la gran capacidad de cálculo de R ha permitido ahorrar mucho tiempo al implementar los distintos métodos de generación de las distintas variables aleatorias.

4. Código R utilizado

A continuación se muestra el código utilizado para la realización de este caso práctico.

```
#####  
# Cuestión 1  
#####  
  
# Número de muestras  
num_samples = c(10, 100, 10000)  
  
# Parámetros de la distribución normal  
mu = 3  
sigma = 6  
  
# Guardamos las muestras de cada distribución generada  
samples_10 = rnorm(num_samples[1], mu, sigma)  
samples_100 = rnorm(num_samples[2], mu, sigma)  
samples_10000 = rnorm(num_samples[3], mu, sigma)  
  
# Construimos los histogramas  
par(mfrow = c(1,3))  
hist(samples_10, main = "10 muestras",  
      xlab = "", ylab = "Frecuencia",  
      col = "blue")  
hist(samples_100, main = "100 muestras",  
      xlab = "", ylab = "Frecuencia",  
      col = "blue")  
hist(samples_10000, main = "10000 muestras",  
      xlab = "", ylab = "Frecuencia",  
      col = "blue")  
  
# Restauramos la vista de los plots  
par(mfrow=c(1,1))  
  
# Calculamos los estadísticos más relevantes  
  
# Mínimo y máximo, primer y tercer cuartil, media y mediana  
summary(samples_10)  
summary(samples_100)  
summary(samples_10000)  
  
# Calculamos la desviación típica
```

```

sd(samples_10)
sd(samples_100)
sd(samples_10000)

# Calculamos la moda
# Si no está disponible el paquete, lo instalamos
if(!require("modeest")) { install.packages("modeest") }

# Cargamos el paquete
require("modeest")

mlv(samples_10, method = "mfv")
mlv(samples_100, method = "mfv")
mlv(samples_10000, method = "mfv")

# Kurtosis y asimetría (es necesario el paquete `moments`)

# Si no está disponible, lo instalamos
if(!require("moments")) { install.packages("moments") }

# Cargamos el paquete
require("moments")

# Calculamos el kurtosis
kurtosis(samples_10)
kurtosis(samples_100)
kurtosis(samples_10000)

# Calculamos la asimetría
skewness(samples_10)
skewness(samples_100)
skewness(samples_10000)

#####
# Cuestión 2
#####

# Usamos el método de suma de 12 uniformes para generar
# 10000 muestras de una normal (0,1)
x = array()
n = 10000
for(i in 1:n){
  unif_samples = runif(12)

```

```

    x[i] = sum(unif_samples) - 6
}

# Desplazamos la distribución generada para que tenga media 3
# y desviación típica 6
x = sigma*x + mu

# Calculamos algunos descriptivos estadísticos
summary(x)
mean(x)
var(x)
sd(x)
kurtosis(x)
skewness(x)
mlv(x, method = "mfv")

# Generamos 10000 muestras de con la función `rnorm`
x1 = rnorm(n, mu, sigma)

# Calculamos algunos descriptivos estadísticos
summary(x1)
mean(x1)
var(x1)
sd(x1)
kurtosis(x1)
skewness(x1)
mlv(x1, method = "mfv")

# Comparamos los 3 histogramas
par(mfrow = c(1,3))
hist(samples_10000, main = "10000 muestras\n Cuestión 1",
      xlab = "", ylab = "Frecuencia", col = "blue")
hist(x, main = "12 uniformes",
      xlab = "", ylab = "Frecuencia", col = "blue")
hist(x1, main = "10000 muestras",
      xlab = "", ylab = "Frecuencia", col = "blue")

# Restauramos la vista de los plots
par(mfrow=c(1,1))

#####
# Cuestión 3

```

```
#####

# Generamos 10000 muestras de una distribución exponencial
# de parámetro lambda = 9

lambda = 9

exp_samples = array()
for(i in 1:n){
  u = runif(1)
  exp_samples[i] = -1/lambda * log(u)
}

# Dibujamos el histograma de la muestra generada
hist(exp_samples, main="10000 muestras de una distribución exponencial",
      xlab="", ylab = "Frecuencia", col = "blue")

# Calculamos el coeficiente de variación
cv = sd(exp_samples)/abs(mean(exp_samples))

# Calculamos la media
mean(exp_samples)

#####
# Cuestión 4
#####

# Usando rnorm generamos una distribución normal bivalente

# Parámetros de la función bivalente
Mu = c(3, 2)

Sigma = matrix(c(7, 4, 4, 5), ncol = 2, nrow = 2)

# Calculamos la matriz de Cholesky de la matriz de varianzas
# y covarianzas
L = t(chol(Sigma))

# Generamos 10000 muestras de la distribución
multi_dist = matrix(nrow = 2, ncol = n)
for(i in 1:n){
  Z = rnorm(length(Mu))
  X = Mu + L%*%Z
}
```

```

    multi_dist[1,i] = X[1]
    multi_dist[2,i] = X[2]
}

# Calculamos las medias
apply(multi_dist, 1, mean)

# Calculamos las varianzas
apply(multi_dist, 1, var)

# Calculamos las covarianzas
x = multi_dist[1,]
y = multi_dist[2,]

cov(x,y)

# Representamos el scatterplot

# Si no está el paquete `scatterplot3d`, lo instalamos
if(!require("scatterplot3d")) { install.packages("scatterplot3d") }

# Si no está el paquete `mvtnorm`, lo instalamos
if(!require("mvtnorm")) { install.packages("mvtnorm") }

# Cargamos el paquete `scatterplot3d`
require("scatterplot3d")

# Cargamos el paquete `mvtnorm`
require("mvtnorm")

# Función de densidad de una distribución normal mutivariante
cdf = function(x){
  const = 1/(2*pi*det(Sigma)^(1/2))
  return (const*exp(-1/2*t(x-Mu)%*%solve(Sigma)%*%(x-Mu)))
}

# Aplicamos la función de densidad con
# cada punto muestreado
z = apply(t(multi_dist), 1, cdf)

# Colores para el degradado de la función de densidad
colors = heat.colors(101)

```

```

zcolor = colors[(z - min(z))/diff(range(z))*100 + 1]

# Función de densidad usando la función `cdf` definida anteriormente
scatterplot3d(x, y, z, color = zcolor,
  angle = 55, scale.y = 0.7, pch = 16,
  main = "10000 muestras de una distribución normal bivalente",
  zlab = "Densidad")

# Función de densidad usando la función `dmvnorm`
scatterplot3d(x, y, dmvnorm(t(multi_dist), mean = Mu, sigma = Sigma),
  color = zcolor,
  angle = 55, scale.y = 0.7, pch = 16,
  main = "10000 muestras de una distribución normal bivalente",
  zlab = "Densidad")

# Scatterlot
plot(x,y, main = "10000 muestras de una distribución normal bivalente")

```