

1. Stopping generating new simulation data

Write a program to generate standard normal random variables until you have generated n of them, where $n \geq 100$ is such that $S/\sqrt{n} < 0,01$, where S is the sample standard deviation of the n data values. Note that this is the “Method for Determining When to Stop Generating New Data”. Also, answer the following questions:

Debido a que se va a generar variables aleatorias **normales estándar** X_i , es decir, $X_i \sim \mathcal{N}(0, 1)$ y se debe parar de generar cuando $\frac{S}{\sqrt{n}} < 0,01$.

```
x <- numeric()
s <- 1
n <- 0

while(n < 100 || s / sqrt(n) >= 0.01) {
  x <- c(x, rnorm(1))
  n <- length(x)
  s <- sd(x)
}

cat("Cantidad de Xi generados", n, "\n")

## Cantidad de Xi generados 10045

cat("Desviación estándar", round(s, 5), "\n")

## Desviación estándar 1.00221

cat("Error estándar S/sqrt(n)", round(s/sqrt(n), 5), "\n")

## Error estándar S/sqrt(n) 0.01
```

En el código se declara una variable s inicializada en 1 para que luego sea nuevamente computada a la desviación estándar de x que es un vector que va a almacenar todas las variables normales aleatorias generadas. Según el enunciado tenemos una condición de que al menos deben haber 100 variables aleatorias generadas ($n \geq 100$) y que el error estándar sea menor a 0.01 ($S/\sqrt{n} < 0,01$). En el ciclo `while`, por lo tanto, tiene sentido que el ciclo continúe si alguna de las afirmaciones anteriores son falsas, por eso queda la condición de esa manera. Se hace uso de la función `rnorm(1)` para que genere una variable aleatoria normal con los valores por defecto (promedio 0 y desviación 1).

1.1. How many normals do you think will be generated? Give an analytic estimate.

1.1.1. Respuesta

Aunque ya se tengan los resultados de la simulación, se puede hacer un estimado analítico con la condición $S/\sqrt{n} < 0,01$ donde se puede despejar n para saber cuántas variables se necesitan para parar el criterio.

Se tiene en primer lugar la inecuación:

$$\frac{S}{\sqrt{n}} < 0,01$$

Se eleva ambas partes con menos 1:

$$\left(\frac{S}{\sqrt{n}}\right)^{-1} > 0,01^{-1}$$

$$\frac{\sqrt{n}}{S} > 100$$

Multiplicando ambas partes por S :

$$\cancel{S} \times \frac{\sqrt{n}}{\cancel{S}} > 100 \times S$$

$$\sqrt{n} > 100 \times S$$

Ahora se cancela la raíz elevando ambas partes al cuadrado:

$$(\sqrt{n})^2 > (100 \times S)^2$$

Quedando entonces:

$$n > 10000 \times S^2$$

Para hacer el ejercicio se utilizó en R la función `rnorm(1, mean = 0, sd = 1)`, de esta manera genera únicamente 1 valor con media 0 y desviación 1, por lo tanto, podemos hacer $S = 1$ para estimar cuántos n necesitamos para que se cumpla condición y deje de generar variables aleatorias, por lo tanto:

$$n > 10000 \times (1)^2$$

$$n > 10000$$

Este valor se acerca bastante al que se imprime en la simulación.

1.2. How many normals did you generate?

1.2.1. Respuesta

Se han generado $n= 10045$ normales.

1.3. What is the sample mean of all the normals generated?

1.3.1. Respuesta

Utilizando el comando `mean()` al vector `x` se obtiene 0,0062027.

1.4. What is the sample variance?

1.4.1. Respuesta

La varianza muestral se calcula como:

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

Esto se puede sacar con la función `var()` de R, por lo tanto, aplicando eso al vector `x` se obtiene 1,0044249.

1.5. Comment on the results of (1.3) and (1.4). Were they surprising?

1.5.1. Respuesta

El resultado de 1.3 es la media de `x` que es 0,0062027 y el de 1.4 es la varianza muestral de `x` dando 1,0044249. Estos valores muy sorprendentes no fueron... Se espera que al simular muchas veces una distribución normal, esta nos dé los valores de la media y la varianza 1,0044249