Tema 2. Generación de Números Aleatorios y Simulación

Modelos Estadísticos y Simulación 2024-2025

Antonio Elías
Area de Estadística e Investigación Operativa
Universidad de Málaga

Introducción a la simulación

Qué es la simulación

- **Definición de simulación** según Robert E. Shanon:

> "La simulación es el proceso de diseñar un modelo de un sistema real y llevar a cabo experiencias con él, con la finalidad de aprender el comportamiento del sistema o de evaluar el funcionamiento del mismo".

La simulación es una herramienta muy importante:

- evita el uso de experimentos reales
- ayuda a resolver problemas o aproximar las soluciones

Ejemplos: red eléctrica, túnel del viento, sistema cardiovascular, tsunamis etc.

Tipos de números aleatorios

Lo primero que necesitamos a la hora de simular un fenómeno aleatorio es disponer de algo que podamos considerar como debido al azar.

Números pseudoaleatorios

Simulación de números pseudoaleatorios

Se llama números pseudoaleatorios a una sucesión, u_i para $i = 1, \ldots, n$, determinística de números en el intervalo $(0, 1)$ que tiene las mismas propiedades estadísticas que una sucesión de números aleatorios.

En realidad no son realmente aleatorios pero lo parecen, es decir, pueden ser considerados aleatorios en un sentido estadístico.

Una sucesión de números pueden considerarse aleatorios si:

Generación por congruencias

Se han propuesto distintos métodos para la generacion de numeros pseudoaleatorios, generación por congruencias es de los más utilizados actualmente.

Para generar la sucesión aleatoria $x_1, x_2, \ldots x_n$ la generación por congruencias consiste en:

$$x_{i+1} \equiv a x_i + c \pmod{m}$$

Donde:

- x_0 es la semilla. Nos permite tener reproducibilidad y replicabilidad.
- a, c, m son enteros positivos

Distribución uniforme en (0,1)

En todas las distribuciones que simulemos será necesario obtener números aleatorios en el intervalo $(0,1)$.

La distribución que representa a este hecho es la distribución uniforme.

$$f(x) = \begin{cases} 1 & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

y su función de distribución acumulada viene dada por

$$F(x) = \begin{cases} 0 & \text{si } 0 < x, \\ x & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Generación de números pseudoaleatorios y distribuciones en R

Distribuciones de Probabilidad

En R, hay un estandar en la terminología para nombrar a las diferentes funciones relacionadas con una distribución concreta, llamémosla *XYZ*:

- `dXYZ()`: devuelve la densidad.
- `pXYZ()`: devuelve la función de distribución.
- `qXYZ()`: devuelve la función cuantíl.
- `rXYZ()`: genera números aleatorios de esa distribución.

Los argumentos específicos de estas funciones para cada distribución dependerán principalmente del número de parámetros que tenga la distribución. Las distribuciones más conocidas se encuentran en el paquete R base.

Funciones de utilidad

- `set.seed()`
- `rNGkind()`
- Función `sample()`: toma una muestra aleatoria de un vector.
- Función `replicate()`: “envuelve” a la función sapply y simplifica.
- `simulate()`

Funciones de utilidad

- Función `manipulate::manipulate()`: permite ejecutar un gráfico con diferentes parámetros.

```r
library(manipulate)
manipulate(
  hist(rnorm(n)),
  n = slider(5, 1000))
```

- Tiempos de ejecución: `proc.time()`

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

Métodos de generación de variables aleatorias

Simulación de la distribución de Bernoulli

Es un experimento aleatorio que sólo tiene dos posibles resultados comunmente denominados como Éxito y Fracaso.

La variable aleatoria tiene funcion de probabilidad $\mathbb{P}(\text{Exito}) = \mathbb{P}(X = 1) = p$ y $\mathbb{P}(\text{Fracaso}) = \mathbb{P}(X = 0) = 1 - p$.

Diremos que $X \sim Bernoulli(p)$.

Simulación de la distribución de Bernoulli

Simulación:

1. Generar un número aleatorio u de $U(0, 1)$
2. Si $u \leq p$, asignar $X = 1$, sino $X = 0$

Simulación de la distribución de Binomial

Realizamos n pruebas independientes de Bernoulli todas con probabilidad de éxito igual a p. Definimos la v.a. $Y = \{\text{número total de éxitos en esas pruebas}\}$. Esta v.a. se distribuye según una binomial de parámetros n y p. Diremos que $Y \sim Bi(n, p)$

La distribución de probabilidad de esta v.a. viene dada de la forma

$$\mathbb{P}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 1, 2, \ldots, n.$$

Si Y se distribuye como una binomial y $X \sim Ber(p)$ es su bernouilli asociada, podemos decir que

$$Y = X_1 + X_2 + \cdots + X_n$$

Simulación de la distribución de Binomial

El algoritmo para generar valores de una distribuci´on binomial se basa en esta descomposición.

1. Hacer $i = 1$
2. Generar U_i con densidad uniforme en $(0, 1)$
3. Si $U_i \leq p$, devolver $X_i = 1$; en otro caso, hacer $X_i = 0$
4. Hacer $i = i + 1$. Si $i < n$, pasar a 2.
5. Devolver $Y = X_1 + X_2 + \cdots + X_n$

Simulación de la Distribución Geométrica

Realizamos tantas pruebas independientes de Bernoulli todas con probabilidad de éxito igual a p, como sean necesarias hasta obtener el primer éxito.

Definimos la v.a.
$Y = \{\text{numero de fracasos antes del primer éxito}\}$.
Esta v.a. se distribuye según una distribución geométrica de parámetro p. Diremos que $Y \sim Ge(p)$.

La distribución de probabilidad viene dada de la forma

$$\mathbb{P}(Y = y) = (1-p)^y p \qquad y = 0, 1, 2, 3 \ldots$$

Simulación de la Distribución Geométrica

Simulación:

1. Hacer $i = 1$
2. Generar U_i con densidad uniforme en $(0, 1)$
3. Si $U_i > p$, devolver $X_i = 1$; en otro caso, hacer $X_i = 0$
4. Hacer $i = i + 1$. Si $X_i = 0$, pasar a 2
5. Devolver $Y = X1 + X2 + \cdots + Xi$

Simulación de la Binomial Negativa

Realizamos pruebas independientes de Bernoulli todas con probabilidad de éxito igual a p hasta obtener el n-ésimo éxito.

Definimos la v.a.
$Y = \{\text{número de fracasos hasta obtener el enésimo éxito}$
Esta v.a. se distribuye según una binomial negativa de parámetros n y p, $Y \sim BN(n,p)$.

La distribucion de probabilidad de esta v.a. viene dada de la forma

$$\mathbb{P}(Y=y)=\binom{n+y-1}{y}p^n(1-p)^y, \qquad y=0,1,2,\ldots$$

Si Y se distribuye como una binomial negativa y $X \sim Ge(p)$ es su geométrica asociada, podemos decir que

$$Y = X_1 + X_2 + \cdots + X_n$$