Simulación I

Prof. José Niño Mora

Investigación Operativa, Grado en Estadística y Empresa, 2011/12

Universidad Carlos III de Madrid
www.uc3m.es

Esquema

- Modelos de simulación y el método de Montecarlo
- Ejemplo: estimación de un área
- Ejemplo: estimación de integrales
- Ejemplo: el problema del quiosco
- Generación de números pseudo-aleatorios
- Generación de números pseudo-uniformes

Modelos de simulación

- Los analistas de sistemas que aparecen en aplicaciones de, p. ej. fabricación, telecomunicaciones, aeronáutica, logística, finanzas, etc., a menudo han de investigar el efecto de cambios en el diseño o la política de operación del sistema

- Resulta demasiado costoso investigar tales efectos haciendo los cambios en el sistema real

- Más sencillo y flexible: se construye un **modelo de ordenador** del sistema, y se investiga el efecto de los cambios sobre tal modelo

Modelos de simulación (cont.)

- A menudo, el **modelo es probabilístico**: incluye cantidades modelizadas como **variables aleatorias (v.a.)** con distribución conocida (p.ej: tiempo de vida de una componente, tiempo de servicio, etc.)

- Tenemos entonces un **modelo de simulación**, que **simula** el funcionamiento del sistema real de interés

- Estos modelos suelen ser muy difíciles de analizar

- En lugar de analizarlos de forma exacta, **la simulación aplica métodos estadísticos para estimar el rendimiento medio del sistema**

Estimación de funciones de v.a.

- En modelos que aparecen en aplicaciones, a menudo se quiere **estimar** la **media de una función** $\boxed{g}$ **de un vector** $\boxed{\mathbf{X} = (X_1, X_2, \ldots)}$ **de variables aleatorias (v.a.) con distribución conocida**:

$$\boxed{\widehat{g} \approx \mathsf{E}\left[g(\mathbf{X})\right]}$$

- Por ejemplo, podemos querer estimar el beneficio medio de una política, que es función de la demanda aleatoria

- A menudo es muy difícil calcular de forma exacta la media $\boxed{\mathsf{E}\left[g(\mathbf{X})\right]}$ buscada

- En tales casos se busca estimar la media con un **estimador** $\boxed{\widehat{g}}$, que sea más sencillo de calcular

El método de Montecarlo

- ¿Qué estimador $\boxed{\widehat{g}}$ de $\boxed{\mathrm{E}\left[g(\mathbf{X})\right]}$ se utiliza?

- El estimador natural es la **media muestral**

- El **método de Montecarlo** para estimar $\boxed{\mathrm{E}\left[g(\mathbf{X})\right]}$:

1. Generamos una **muestra i.i.d** de $\boxed{\mathbf{X}}$: $\boxed{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n}$

2. Estimamos la **media poblacional** $\boxed{\mathrm{E}\left[g(\mathbf{X})\right]}$ por la **media muestral**:

$$\boxed{\widehat{g} \triangleq \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i)}$$

El método de Montecarlo (cont.)

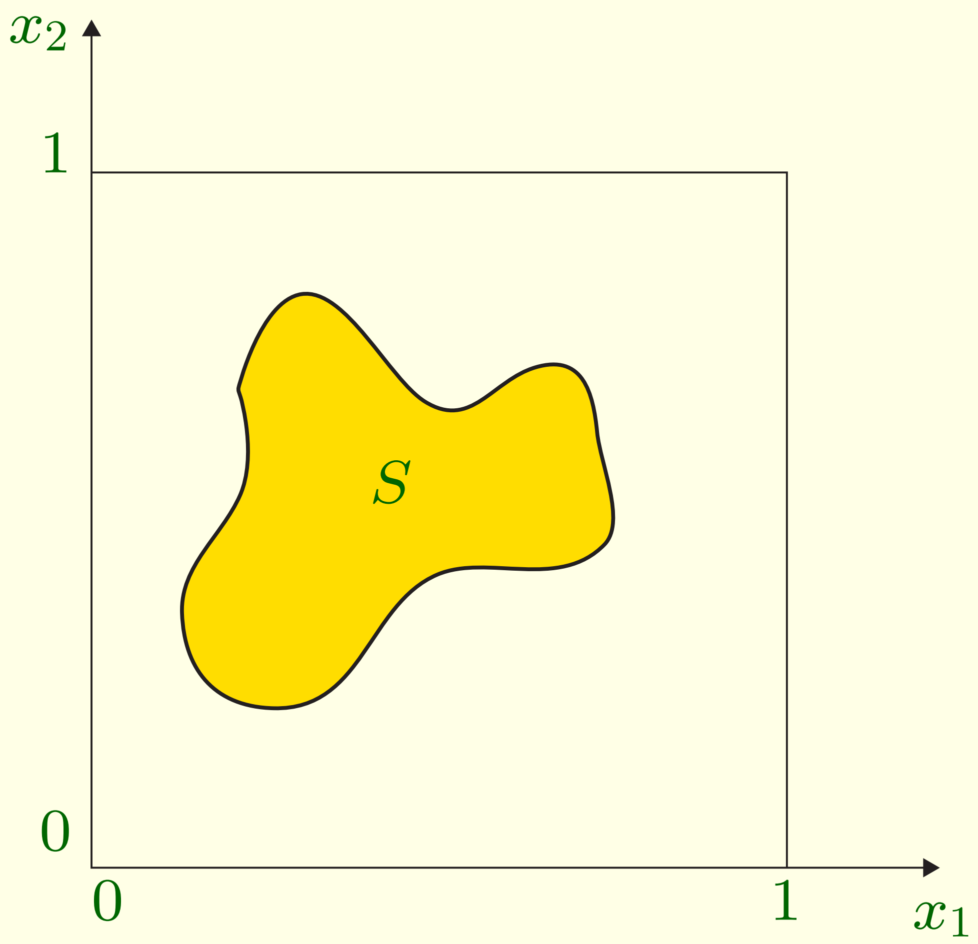
- Cuestiones que abordaremos para aplicar el método de Montecarlo:

1. ¿Cómo generamos una muestra i.i.d. de un vector de v.a. $\boxed{\mathbf{X}}$ con una distribución conocida dada?

2. ¿Cómo de grande hemos de elegir el **tamaño de la muestra** $\boxed{n}$ para que el **error de estimación** sea menor que un cierto valor con un nivel de confianza dado?

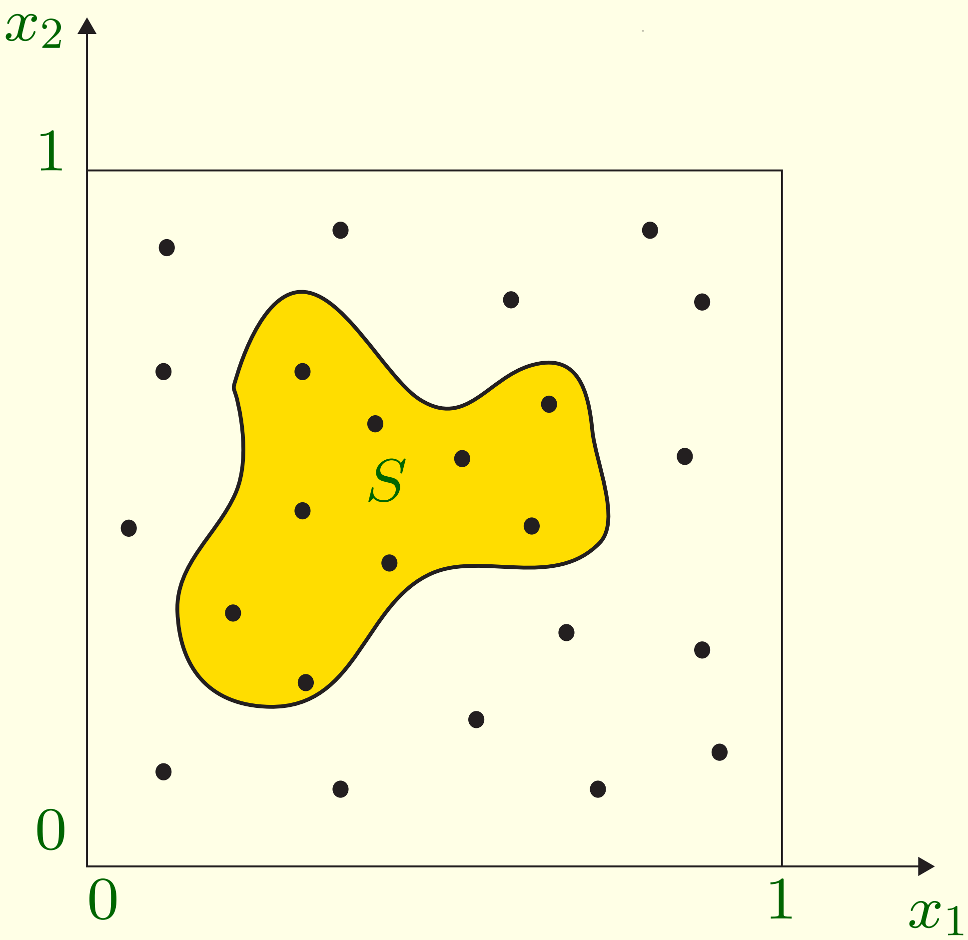
Ej: Estimación de un área

- ¿Cómo podemos estimar el área $\boxed{a}$ de la región $\boxed{S}$?



Ej: Estimación de un área

- Intuición: lanzamos $\boxed{n}$ dardos al cuadrado $\boxed{[0,1]^2}$ que contiene $\boxed{S}$; la proporción de dardos que caen en $\boxed{S}$ nos da una estimación $\boxed{\widehat{g}}$ de su área $\boxed{a}$ (ej: $\boxed{\widehat{g} = 9/26}$)



Ej: Estimación de un área

- Formalización: Generamos una muestra i.i.d de tamaño $\boxed{n}$ de un par de v.a. independientes $\boxed{\mathbf{X} = (X_1, X_2)}$, con $\boxed{X_1, X_2 \sim U[0,1]}$ (**distribución uniforme** en $\boxed{[0,1]}$): $\boxed{\mathbf{X}_1, \ldots, \mathbf{X}_n}$, con $\boxed{\mathbf{X}_i = (X_{1i}, X_{2i})}$
- $\boxed{X_i}$: coordenada $\boxed{x_i}$ en el lanzamiento $\boxed{i}$-ésimo
- Definimos la función $\boxed{g(x_1, x_2)}$ de interés como

$$\boxed{g(x_1,x_2) \triangleq 1_{\{(x_1,x_2)\in S\}} = \begin{cases} 1 \text{ si } (x_1,x_2) \in S \\ 0 \text{ en otro caso} \end{cases}}$$

- El área $\boxed{a}$ de $\boxed{S}$ es:

$$\boxed{a = \mathsf{E}\left[g(X_1, X_2)\right]}$$

Ej: Estimación de un área

- El área $\boxed{a}$ de $\boxed{S}$ es:

$$\boxed{a = \mathsf{E}\left[g(X_1, X_2)\right]}$$

- Estimación del área por el método de Montecarlo:

$$\boxed{\widehat{g} = \frac{1}{n}\sum_{i=1}^{n} g(X_{1i}, X_{2i}) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{(X_{1i}, X_{2i}) \in S\}}}$$

Ejemplo: estimación de integrales

- Supongamos que queremos estimar el valor de la **integral definida**

$$\boxed{\int_0^1 g(x)\,dx}$$

- Podemos aplicar el **Método de Montecarlo**

- Observamos que, tomando una v.a. $\boxed{X \sim U[0,1]}$, podemos representar la integral como una esperanza:

$$\boxed{\int_0^1 g(x)\,dx = \mathsf{E}\left[g(X)\right]}$$

Ejemplo: estimación de integrales

- Estimación de la integral por el **Método de Montecarlo**:

1. Generar una muestra i.i.d. $\boxed{X_1, X_2, \ldots, X_n \sim U[0,1]}$

2. Estimar la integral por la **media muestral**

$$\boxed{\widehat{g} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)}$$

Otro ejemplo: El problema del quiosco

- El dueño de un quiosco tiene que decidir cuántos períodicos debe comprar cada día a su proveedor

- Cada períodico le cuesta $\boxed{c\,€}$, y lo vende al público por un precio de $\boxed{p\,€}$, con $\boxed{p > c}$

- Los períodicos no vendidos los devuelve al proveedor, obteniendo un reembolso por unidad de $\boxed{r\,€}$, con $\boxed{r < c}$

- A partir de la experiencia, se conoce la **distribución de la demanda diaria** $\boxed{X}$ de periódicos:

$$\boxed{\mathsf{P}\{X = 100 + 50k\} = p_k, \quad k = 0, 1, \ldots, K\}}$$

- Problema: ¿Cuál es el número óptimo $\boxed{a^*}$ de períodicos que debe comprar cada día el quiosquero a su proveedor para maximizar su beneficio esperado?

Ej: El problema del quiosco

- Si el quiosquero compra $\boxed{a}$ períodicos, y la demanda es de $\boxed{X}$ períodicos, su beneficio es (nota: $\boxed{x^+ = \max(0, x)}$):

$$\boxed{g_a(X) \triangleq p \min(X, a) + r(a - X)^+ - ca}$$

- Así, formulamos el problema del quiosco como:

$$\boxed{\max_{a \geq 0} \mathsf{E}\left[g_a(X)\right]}$$

Ej: El problema del quiosco

- Enfoque de resolución aproximada por el método de Montecarlo:

1. Generamos una muestra i.i.d de tamaño $\boxed{n}$ de la demanda diaria: $\boxed{X_1, \ldots, X_n}$
2. Calculamos el beneficio esperado para algunos valores de $\boxed{a \geq 0}$:

$$\boxed{\widehat{g}_a \triangleq \frac{1}{n} \sum_{i=1}^{n} g_a(X_i)}$$

3. Resolvemos el problema aproximado

$$\boxed{\max_{a \geq 0} \widehat{g}_a,}$$

obteniendo una solución aproximada $\boxed{\widehat{a}^*}$

Generación de números pseudo-aleatorios

- ¿Podemos generar **números aleatorios** por ordenador, para aplicar el método de Montecarlo?

- **No podemos**: un ordenador es una **máquina determinista**

- Sin embargo, sí podemos generar **números pseudo-aleatorios**: se construyen de forma determinista, pero parecen aleatorios y se pueden utilizar como tales

La distribución uniforme

- La **distribución de probabilidad** más importante en simulación es $\boxed{X \sim U[0,1]}$ (uniforme en $\boxed{[0,1]}$)

- La **función de densidad** de una v.a. $\boxed{X \sim U[0,1]}$ es:

$$\boxed{f(x) = \begin{cases} 1 & \text{si } x \in [0,1] \\ 0 & \text{en otro caso} \end{cases}}$$

- La **función de distribución** de $\boxed{X \sim U[0,1]}$ es:

$$\boxed{F(x) = \int_{-\infty}^{x} f(x)\,dx = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \in [0,1] \\ 1 & \text{si } x > 1 \end{cases}}$$

La distribución uniforme

- La media de una v.a. $\boxed{U \sim U[0,1]}$ es

$$\boxed{\mathsf{E}\left[U\right] = \int_0^1 x f(x)\, dx = \int_0^1 x\, dx = \frac{1}{2}}$$

y su varianza es

$$\boxed{\mathrm{Var}\left[U\right] = \mathsf{E}\left[U^2\right] - \left\{\mathsf{E}\left[U\right]\right\}^2 = \int_0^1 x^2\, dx - \frac{1}{4} = \frac{1}{12}}$$

Generación de números pseudo-uniformes

- Para generar números pseudo-uniformes se emplea el **Método de Congruencias Lineales (MCL)**:

1. Elegir valores enteros positivos para los parámetros $\boxed{a, c, m}$

2. Elegir un valor inicial, o **semilla**, $\boxed{V_0 \in \{0, \ldots, m-1\}}$

3. Generar la sucesión $\boxed{V_0, V_1, V_2, V_3, \ldots}$ mediante:

$$\boxed{V_{i+1} = (aV_i + c) \mod m, \quad i = 0, 1, 2, \ldots}$$

   es decir, $\boxed{V_{i+1} \in \{0, \ldots, m-1\}}$ es el resto, o residuo, de dividir $\boxed{aV_i + c}$ entre $\boxed{m}$

4. Calcular la sucesión $\boxed{U_0, U_1, U_2, U_3, \ldots}$ por $\boxed{U_i = V_i / m}$

Generación de números pseudo-uniformes

- La sucesión $\boxed{U_0, U_1, U_2, U_3, \ldots}$ generada por el MCL satisface: $\boxed{U_i \in \{0, 1/m, \ldots, (m-1)/m\}}$

- Además, la sucesión generada se repite (tiene un **período finito**): existe algún $\boxed{n \geq 1}$ tal que $\boxed{U_n = U_0}$

- A pesar de ello, tomando $\boxed{m}$ muy grande, y eligiendo los parámetros $\boxed{a, c}$ de forma apropiada, la sucesión generada $\boxed{U_0, U_1, U_2, U_3, \ldots}$ se puede utilizar **como si fuese** $\boxed{U[0,1]}$

- Una buena elección utilizada en la práctica:

$$\boxed{a = 7^5, m = 2^{31} - 1, c = 0}$$

Generación de números pseudo-uniformes

- En Excel, se obtiene una sucesión de números pseudoaleatorios $\boxed{U_0, U_1, \ldots, \sim U[0,1]}$ mediante la función

$$\boxed{\text{ALEATORIO()}}$$

- Esta función no tiene argumento; su valor cambia mediante la tecla para **recalcular**: F9