

## 1.2 Números aleatorios y Monte-Carlo

En el corazón de los Métodos de Monte-Carlo encontramos un generador de números aleatorios, es decir un procedimiento que produce un flujo infinito de variables aleatorias, que generalmente se encuentran en el intervalo (0, 1); los cuales son independientes y están uniformemente distribuidos de acuerdo a una distribución de probabilidad. La mayoría de los lenguajes de programación hoy en día contienen un generador de números aleatorios por defecto al cual simplemente debemos ingresarle un valor inicial, generalmente llamado **seed** o semilla, y que luego en cada invocación nos va a devolver un secuencia uniforme de variables aleatorias independientes en el intervalo (0, 1).

El concepto de una secuencia infinita de variables aleatorias es una abstracción matemática que puede ser imposible de implementar en una computadora. En la práctica, lo mejor que se puede hacer es producir una secuencia de números pseudoaleatorios con propiedades estadísticas que son indistinguibles de las de una verdadera secuencia de variables aleatorias.

Entonces se requiere sortear variables aleatorias según una distribución de probabilidad. Por ejemplo, para hallar  $\int dx f(x)g(x)$  los valores de X deben distribuirse según  $f(x)$  y la integral es el valor medio de  $g(x)$  sobre el conjunto de X. En este procedimiento es fundamental muestrear X siguiendo  $f(x)$ . Veamos qué quiere decir exactamente muestrear. Consideremos un espacio  $\Omega$  formado por todos los posibles valores de una o varias variables aleatorias  $X = X_1, X_2, \dots$  y sea  $f(x)$  su pdf, que cumple

$$\int_{\Omega} dx f(x) = 1.$$

El muestreo es un algoritmo que *produce una secuencia de variables aleatorias* X tales que para cualquier  $\Omega' \subset \Omega$ ,

$$P\{X \in \Omega'\} = \int_{\Omega'} dx f(x) \leq 1.$$

En particular, si tenemos una sola variable aleatoria X que se distribuye según una pdf unidimensional definida sobre el intervalo [0, 1], lo anterior significa que

$$P\{X \in (a, b)\} = \int_a^b dx f(x), \quad 0 < a < b < 1,$$

o, más informalmente, si  $d(x) = b - a$ ,

$$P\{X \in dx\} = f(x)dx.$$

Podemos producir variables aleatorias  $X_1, X_2, \dots$  según una pdf cualquiera si disponemos de variables aleatorias  $\xi_1, \xi_2, \dots$  que se distribuyan uniformemente en el intervalo [0, 1]. Las variables aleatorias uniformemente distribuidas se imitan mediante un ordenador, pues siendo generadas por una rutina determinista, un generador de números pseudoaleatorios (prng), no son por tanto realmente aleatorias. Sin embargo, son satisfactorias siempre que cumplan ciertos criterios de calidad. Estudiaremos los números pseudoaleatorios (prn) en esta sección.

## 1. INTRODUCCION

Los modelos de simulación son útiles para describir algunos fenómenos de la realidad. Así la simulación trata fundamentalmente de construir modelos abstractos en una computadora de tal forma que describan la parte esencial del comportamiento de un sistema de interés, pudiéndose entonces diseñar y realizar experimentos con el modelo y posteriormente extraer conclusiones de sus resultados. Los modelos de simulación tienen su origen en los trabajos de John Von Neumann, pero solo es apreciable su poder cuando se introduce el computador moderno.

Para la implementación de un modelo de simulación se hace necesario entonces, tener en primera instancia una fuente de números aleatorios que representen las observaciones del fenómeno, unas transformaciones de estas observaciones para alimentar el modelo, un procesamiento de las salidas del modelo y un análisis de las respectivas estadísticas resultantes.

**Nuestro objetivo** es mostrar como los números aleatorios, además de ser la base fundamental de cualquier modelo de simulación, también permite, entre otras, hacer evaluaciones de integrales definidas, lo cual la hace una herramienta útil para el calculo.

## 2. INTEGRACIÓN

Sea  $f(x)$  una función, deseamos resolver entonces el problema de calcular la siguiente integral:

$$\theta = \int_0^1 f(x)dx$$

Para calcular el valor de la integral observemos que si  $U$  es una variable aleatoria distribuida uniformemente en el intervalo  $(0,1)$ , entonces es posible expresar a  $\theta$  como:

$$\theta = E(f(u))$$

Si  $U_1, \dots, U_p$  son Variables aleatorias independientes y uniformes, entonces  $g(U_1), \dots, g(U_p)$  son variables aleatorias independientes e idénticamente distribuidas con valor esperado  $\theta$ .

Utilizando la ley fuerte de los grandes números que enuncia que con probabilidad 1.

$$\sum_{i=1}^p \frac{f(U_i)}{p} \rightarrow E(f(u)) = \theta$$

Por consiguiente, es posible aproximar el valor de la integral, generando una gran cantidad de números aleatorios  $U_i$ , evaluar  $f(x)$  en cada  $U_i$ , sumar estas cantidades y dividir entre el numero de números aleatorios; obteniendo entonces una aproximación al verdadero valor de la integral cuando la cantidad de números aleatorios tiende al infinito.

## 3. LEY DE LOS GRANDES NUMEROS

Aunque el tema central de este documento es la integración a través del uso de números aleatorios, es conveniente hacer algunas claridades y justificaciones sobre el teorema que sustenta todo el método.

Primero que todo se debe mostrar un teorema de Concentración de distribuciones llamado “Desigualdad de Chebyshev” que va a ser la base de todo nuestro análisis.

Si  $X$  es una variable aleatoria con media  $\mu$  y varianza  $\sigma^2$ , entonces para cualquier valor de  $k > 0$  se tiene que

$$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

Esta desigualdad muestra que la cantidad de área de la distribución, situada fuera del intervalo

$\mu - k\sigma \leq x \leq \mu + k\sigma$  es a lo sumo igual a  $\frac{1}{k^2}$ , dando así una idea de la concentración de una variable aleatoria sin importar como se distribuya.

Ahora si se tiene una sucesión de variables aleatorias  $X_1, X_2, \dots, X_n$  Independientes con media  $\mu$  y desviación estándar  $\sigma^2$  se tiene para cada  $\varepsilon > 0$ , aplicando la desigualdad de Chebyshev [4]

$$P\left[\left|\frac{\sum x_i}{n} - \mu\right| \geq \frac{k\sigma}{\sqrt{n}}\right] \leq \frac{1}{k^2}$$

Si se reemplaza  $\frac{k\sigma}{\sqrt{n}}$  por  $\varepsilon$  se tiene

$$P\{|\bar{X} - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2} \quad (6)$$

Si  $n \rightarrow \infty$  infinito en (6), se convierte en

$$P\{|\bar{X} - \mu| \geq \varepsilon\} = 0 \quad (7)$$

De este modo, la probabilidad de que  $\bar{X}$  difiera de la media  $\mu$  en mas de  $\varepsilon$ , con epsilon pequeño, tiende a ser igual a cero a medida que aumente el tamaño de la muestra.

Este teorema llamado “Ley Débil de los Grandes Números” sería suficiente para nosotros si lo que interesara fuera la convergencia en probabilidad de

$$\frac{\sum f(U_i)}{n} \text{ a } E(f(u)) \quad (8)$$

En realidad, el interés radica en la convergencia casi segura del promedio de variables aleatorias

O en otras palabras el problema radica en garantizar la convergencia

$$\lim_{n \rightarrow \infty} \frac{\sum f(U_i)}{n} = E(f(u))$$

Afortunadamente el problema fue solucionado ya por los matemáticos a través de un teorema llamado “Ley Fuerte de los Grandes Números”, que enuncia que bajo algunas consideraciones muy generales el promedio de variables aleatorias de una muestra converge con probabilidad 1 a la esperanza matemática de la variable en cuestión.

Así es la “Ley Fuerte de los Grandes Números” el teorema que justifica el método de integración Monte Carlo como llaman algunos autores.