

SIMULACION POR COMPUTADOR: LO CASUAL A TRAVES DE LO DETERMINADO

OSCAR H. BUSTOS

1. Conceptos y objetivos

En el uso coloquial el término "simulación" tiene una connotación marcadamente peyorativa, siendo sinónimo de términos tales como "disfraz", "fingimiento", "aparentar", "disimular", etc. Todo lo contrario de lo que sugiere el uso actual de la simulación de procesos reales en las ciencias aplicadas y en la Estadística. En el sentido técnico, "simulación" consiste en "observar" un fenómeno de la vida real sin observarlo realmente; esto es, obteniendo realizaciones de un modelo matemático que pretende describir lo más exactamente posible lo esencial de dicho fenómeno. Si en ese modelo se usan variables aleatorias tenemos un "modelo estocástico".

La Física Clásica, por ejemplo, nos enseña que un modelo matemático adecuado para describir la velocidad (v) que lleva un móvil en función del tiempo (t) en que recorre cierto espacio (e) es:

$$v = e/t.$$

Un modelo de este tipo es llamado determinístico.

Un ejemplo simple de modelo estocástico lo tenemos en la siguiente representación de los posibles valores de una cierta variable Y , cuya medición sobre cada uno de los

integrantes de una población puede tomar diversos valores:

$$Y = \mu + \varepsilon,$$

donde μ es un cierto valor de esa magnitud física correspondiente a la población considerada globalmente (por ejemplo: "la media de Y en la población"), y ε representa un "error de observación", que da cuenta de las desviaciones de Y respecto de μ cuando se observa Y en cada individuo.

Otro término que se acostumbra usar como sinónimo de "simulación estocástica" es el de "método de Monte Carlo". Esta denominación se acuñó durante la Segunda Guerra Mundial para las simulaciones estocásticas de modelos de choques atómicos (procesos de ramificación). Actualmente se usan ambos términos como sinónimos, es lo que haremos, pero a veces se reserva el término "Monte Carlo" para referirse a "hacer algo inteligente con simulación".

Una simulación puede tener varios objetivos. Esto hace imposible dar una guía de validez general para conducir buenas aplicaciones de esa metodología. Frecuentemente se considera a la simulación un arte (Tocher (1963) escribió uno de los primeros libros sobre esta materia y le llamó: "The Art of Simulation"). Pese a los considerables esfuerzos por dotar a la simulación de un cuerpo teórico sistemático, continúa siendo al día de hoy un arte, cuyo resultado depende marcadamente de la habilidad y experiencia de quien lo aplica.

Sin lugar a dudas, la profunda transformación en la vida de toda sociedad que están provocando la difusión y mejoramiento técnico de los computadores, prometen hacer de la

simulación (bien hecha) una de las herramientas fundamentales en todas las investigaciones científicas que tienen como objetivo entender y modificar para lo que se considere mejor, este "maravilloso Universo Nuestro". Quien esté interesado en obtener una excelente, aunque super optimista, imagen de las posibilidades en este sentido en el campo de la biomedicina, podrá leer el libro de "divulgación científica" de Silverstein (1979).

Lo primero que necesitamos para realizar una simulación estocástica, es un mecanismo de generación de "aleatoriedad". Debido a que suponemos que haremos uso de un computador, nos restringiremos a tratar de aprender a usarlo como fuente de aleatoriedad lo más genuina posible, adecuada a las necesidades de cada experiencia. Frecuentemente, por desgracia, se supone ingenuamente que el fabricante de un computador o del compilador de un lenguaje de programación, ha tenido el cuidado de vendernos un buen generador de sucesiones numéricas. La mayoría de las veces esto no es así, sobre todo en los tiempos actuales, en que las fábricas de computadores (micros sobre todo) han surgido como la hierba después de la lluvia, con el afán de lucro más que el de brindar una herramienta confiable para investigaciones científicas. A este respecto, debemos tener presente lo que se dice en el Prólogo de Ripley (1987): *It might be disastrous to believe in your computer manufacturer!*.

A decir verdad, el computador no puede describir fielmente lo que se entiende por "fenómeno aleatorio". En efecto, toda "función aleatoria" implementada en un computador sólo es capaz de generar una sucesión numérica determinística; en ese sentido es que a tales funciones las

deberíamos llamar más correctamente "generadores de números pseudoaleatorios". Pero, para los fines prácticos, lo que se requiere es una sucesión numérica que imite, lo mejor posible, la propiedad del fenómeno aleatorio que consideramos esencial para los fines de una determinada investigación. Una de esas propiedades esenciales es que sea "prácticamente imprevisible". Esto es: que conociendo los términos presentes de la sucesión sea muy difícil, del punto de vista práctico, saber cuál será el próximo número de la sucesión... por supuesto que sin saber la definición formal de la función aleatoria que da origen a tal sucesión.

Por ejemplo, la sucesión:

13, 8, 1, 2, 11, 14, 7, 12, 13, 12, 17, 2, 11, 10, 3, 4, 17, 4, 5, 6, 15, 6, ...

ha sido generada por medio de una regla determinística bastante simple, pero... podría el lector, sin saber cuál es esa regla, determinar cuál será el próximo número de la sucesión?

Claro, todo depende de cuántos términos de la sucesión conozcamos. Por ejemplo, si la sucesión presente es:

1, 0, 0, 1, 1, 2, 1, 1,

es muy difícil saber si el próximo número será 0, 1 ó 2. Pero si esa sucesión presente es:

(*) 1, 0, 0, 1, 1, 2, 1, 1, 1, 0, 2, 0, 2, 1, 0, 0, 1, 1, 2, 1, 1, 1, 0, 2, 0, 2, 1, 0, 0,
con un poco de reflexión concluiríamos, correctamente, que el próximo número es 1.

Así es que, mientras más largas sean las sucesiones de números que no se repiten, más chance tendremos de que

gocen de esa característica de "imprevisibilidad" práctica de que hablamos.

Ahora, cómo definir una regla determinística que genere una sucesión con una "buena" propiedad de "imprevisibilidad" práctica?.

La regla determinística que genera la primera sucesión:

$$x_1=2, x_2=11, x_3=14, x_4=7, \dots, \text{es:}$$

$$x_i = (x_{i-1} + x_{i-2} + x_{i-3}) \text{ mod } (20), \quad i=1,2,3,\dots$$

con $x_0=1$, $x_{-1}=8$ y $x_{-2}=13$.

Se puede ver que tal sucesión se comienza a repetir después de 248 términos. Tal característica la torna inútil para la mayoría de los fines prácticos.

Otra sucesión basada en una regla similar es la (*), sólo que esta vez es:

$$x_i = (x_{i-1} + x_{i-2} + x_{i-3}) \text{ mod } (3), \quad i=1,2,3,\dots$$

con $x_0=0$, $x_{-1}=0$ y $x_{-2}=1$, que comienza a repetirse tan sólo después de 13 términos.

Como podemos apreciar, el saber si una cierta regla determinística genera una sucesión aleatoriamente genuina, es algo bastante complicado.

Toda simulación tiene un propósito: obtener "observaciones" de un modelo estocástico. Ahora, cómo obtener "buenas" muestras? Una simulación es un experimento

matemático. Del mismo modo que al realizar una experiencia de laboratorio, se debe ser cuidadoso en el diseño de la experiencia implicada en una simulación. Así, las técnicas para el diseño y análisis de simulaciones son herramientas importantes y constituyen una área rica bajo investigación.

Al planificar una simulación, primero que nada, debemos tener presente el objetivo perseguido al proponer un cierto modelo estocástico como representación (siempre aproximada) de un fenómeno real. El uso de simulación se justifica cuando el estudio analítico del modelo se hace muy difícil o imposible. También es adecuado cuando queremos estudiar lo que sucede al variar ciertas condiciones del modelo.

Por otra parte, al decidirnos a realizar un trabajo de simulación, dos factores debemos tener en cuenta muy especialmente: eficiencia y costo computacional. La mayoría de las veces obtener buenas performances en ambos criterios es un asunto conflictuante. Casi siempre se hace necesario adoptar una solución de compromiso.

Varias técnicas tomadas de la rama de la Estadística llamada "Diseños de Experimentos" pueden ayudarnos en la adecuada planificación de una experiencia Monte Carlo que contemple los varios factores aquí expuestos. Un conjunto de ellas, que ha recibido el nombre de "técnicas de reducción de varianza", ha mostrado ser de gran utilidad. Según Ripley (1987) las podemos agrupar en las siguientes categorías:

a) Las que atienden al proceso de "muestreo". A veces, es conveniente obtener muestras no directamente de la distribu-

ción de interés, sino de otra distribución relacionada con ella. Otras usan recursos propios de muestreo estratificado.

b) Uso de variables de control y de variables correlacionadas negativamente con las variables que se desean estudiar (en inglés, esto se conoce como técnicas de *Control and antithetic variates*).

c) Las que usan propiedades de variables condicionadas.

d) Las que usan varias veces la misma sucesión de números pseudoaleatorios, imitando así las técnicas de agrupamiento de unidades experimentales en bloques homogéneos. En la jerga propia de la simulación se las llama "técnicas basadas en números aleatorios comunes".

En lo que se refiere al uso de simulación en Inferencia Estadística podemos distinguir, entre otras, las siguientes aplicaciones:

1) Verificación de que una distribución teórica describe razonablemente bien los resultados aleatorios de una experiencia. Relacionados con esta aplicación están los llamados "tests Monte Carlo".

2) Verificación de coherencia entre cálculos analíticos respecto de parámetros de una distribución y sus valores obtenidos por simulación.

3) Para tener una idea de cuán bien se ajustan los resultados asintóticos con los correspondientes sobre muestras finitas.

4) Para comparar el comportamiento de diversos estimadores

de un mismo parámetro sobre muestras finitas.

5) Para calcular la varianza de un estimador realizando un trabajo de simulación basado en la distribución empírica de la muestra (estimadores *bootstrap*).

6) Construcción de intervalos de confianza ("intervalos de confianza Monte Carlo", según los llama Buckland (1984)).

Una vez más: un trabajo de simulación es un experimento matemático cuyo objetivo consiste en mostrar resultados difíciles o imposibles de ser obtenidos analíticamente. Estos últimos están enunciados por medio de teoremas cuyas demostraciones se ajustan a las leyes de la lógica. Los resultados de una simulación no gozan de tal propiedad. Por eso, cuando se presentan conclusiones basadas en resultados de trabajos de simulación, se debe suministrar toda la información de cómo se procedió para llegar a tales conclusiones. Una parte importante de tal informe debe contener los procedimientos de análisis efectuados sobre los resultados de la simulación. Algunas de las principales indicaciones sobre este asunto, fueron sugeridas por Hoaglin y Andrews (1975).

2. Una breve revisión histórica

Los primeros investigadores que hicieron uso de simulación empleaban procedimientos físicos para la obtención de sucesiones numéricas pseudoaleatorias. También es probable que nosotros mismos hayamos arrojado monedas y dados como ejercicios prácticos en nuestros primeros cursos sobre Probabilidad. Experimentos de simulación basados en tales experimentos tienen una larga historia y aún hoy son usados

en juegos y loterías: Bingo, dados, ruletas, etc..

Tippett (1927) produjo una tabla de 40000 dígitos tomados aleatoriamente de informes censales. Posteriormente surgieron tablas de números pseudoaleatorios construidas en base a fenómenos de ruido electrónico. Un ejemplo notable es la de RAND Corporation (1955).

Todos estos métodos físicos han gozado de gran aceptabilidad, quizás en base a experiencias o tests. Posiblemente también a que una sucesión numérica producida por una "buena" ruleta electrónica goce de esa propiedad de "imprevisibilidad práctica" de la que ya hablamos en la Sección anterior. En efecto, apenas una ligera modificación en las condiciones físicas iniciales de la ruleta, producirán sucesiones numéricas bastante diferentes.

Sin embargo, muchos de los mecanismos usados en esas experiencias, cuando se analizaron con mayor cuidado, mostraron fallas que conducían a sesgos y relaciones de dependencia inaceptables para los fines de simulaciones refinadas como las que se realizan en la actualidad. Por ejemplo, la máquina usada por la RAND tenía fallas mecánicas en el mecanismo de registro que comprometía severamente la "aleatoriedad teórica" sugerida por la Electrónica.

Uno de los primeros fines a los que se destinaron los computadores, fue al de producir sucesiones numéricas pseudoaleatorias. En efecto, cuando se descubrieron los defectos de ciertos mecanismos físicos usados anteriormente y cuando se constató la inviabilidad práctica de incorporar tablas como la de la RAND a la memoria o a los periféricos

del computador, se buscó producir tales sucesiones numéricas apelando a esquemas más formales.

Ante todo, hubo que ponerse de acuerdo sobre qué considerar como sucesión numérica "pseudoaleatoria". Surgió así el siguiente concepto que también nosotros adoptaremos:

Una sucesión de N números pseudoaleatorios (U_i) con distribución uniforme en $(0,1)$ es una sucesión determinística que tiene (aproximadamente) las mismas "propiedades estadísticas relevantes" que una muestra "verdadera" de tamaño N de la distribución $\mathcal{U}(0,1)$ (uniforme en $(0,1)$)

Ahora, cómo formalizar "propiedades estadísticas relevantes"? Para los fines prácticos, parece suficiente pedir que: "para todo k "razonable" (digamos $k \leq 6$) (U_i, \dots, U_{i+k-1}) tenga una distribución aproximadamente uniforme sobre el hipercubo $(0,1)^k$."

Los primeros "generadores de sucesiones numéricas pseudoaleatorias" que aparecieron, fueron los llamados "congruenciales". Sin entrar en detalles, digamos que son los basados en algoritmos como los que generan las sucesiones

13, 8, 1, 2, 11, 14, 7, 12, 13, 12, 17, 2, 11, 10, 3, 4, 17, 4, 5, 6, 15, 6, ...

y 1, 0, 0, 1, 1, 2, 1, 1, 1, 0, 2, 0, 2, 1, 0, 0, 1, 1, 2, 1, 1, 1, 0, 2, 0, 2, 1, 0, 0, vistas en la Sección anterior. Obviamente, como ya vimos, no cualquiera sirve. A veces es fácil rechazarlos sin hacer ninguna cuenta. Por ejemplo, el basado en el algoritmo:

$$U_i = (U_{i-1} + U_{i-2}) \text{ mod } (1)$$

hace que nunca se produzca $U_{i-1} < U_i < U_{i-2}$ lo que la hace "previsible" en cierta manera.

Quizás antes de estos generadores, Von Neumann sugirió uno basado en un "esquema no linear recursivo" que se denominó *middle square*. Veámoslo: supongamos que queremos producir una sucesión con cuatro dígitos decimales cada número. Consideremos el siguiente algoritmo:

- 1) Sea X_0 un número entero entre 1000 y 9999. Sea $i = 0$.
- 2) Calcular X_1^2 .
- 3) Tomar como X_{i+1} los cuatro dígitos "centrales" de X_i^2 .
- 4) Hacer $i = i + 1$ y volver a 2).

Por ejemplo, para $X_0 = 7466$, obtenemos:

$$X_1 = 7411, X_2 = 9229, X_3 = 1744, X_4 = 4153, X_5 = 2474, \dots$$

Parece razonable, ... pero qué pasa cuando $X_0 = 2100$? En tal caso obtenemos:

$$X_1 = 4100, X_2 = 8100, X_3 = 6100, X_4 = 2100, X_5 = 4100, \dots$$

lo que es evidentemente inaceptable.

Modernamente, la Criptografía (esa ciencia-arte de los James Bond 's) ha producido algunos de los mejores generadores. Tienen más que ver con sucesiones aleatorias de bits y el uso de *hardware* especial. En este campo, las preferencias están por los llamados "generadores *shift-registers*".

Es interesante notar que ligeras modificaciones en las implementaciones de un algoritmo de generador del tipo *shift-register* o congruencial, puede llevar a sucesiones numéricas con comportamientos bastante diferentes.

Finalmente, en la actualidad, varios compiladores de

lenguajes de alto nivel como BASIC y PASCAL, tienen "embutida" una sub rutina de generación de números pseudoaleatorios uniformemente distribuida en (0,1), la mayoría de las veces sin ninguna referencia al método de generación que se usó. Esto es más notable en los compiladores ofrecidos por los vendedores de "micros" y "minis". Respecto a esto, debemos estar advertidos de lo que se dice en Ripley (1987) pag. 45:

Microcomputer implementations work incorrectly surprisingly often from faulty compilers or side effects of operating systems.

En lo que se refiere a *mainframes* parece que las cosas son un poco mejor... aunque siempre es bueno dar una "mirada" a lo que realmente está produciendo un computador como sucesión numérica pseudoaleatoria.

3. Comentario final

El desarrollo exponencial de la Computación en el día de hoy y las posibilidades de estudio de fenómenos reales simulándolos por medio de un computador ha abierto un amplio campo de investigación no sólo para la Matemática Aplicada a la Computación sino para varias ramas consideradas clásicamente como de interés apenas para los matemáticos "puros", por ejemplo: Teoría de Números, Algebra, etc.. Es indudable que día a día la Matemática en general está llamada a ser una protagonista por excelencia en la formidable aventura del hombre como "punta de lanza" de la Evolución del Universo.

Referencias:

ANDREWS, D. F. , BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H., and TUKEY, J. W. (1972), "*Robust Estimates of Location*", Princeton University Press, Princeton.

BUCKLAND, S. T. (1984), "Monte Carlo Confidence Intervals", *Biometrics*, 40, 811 - 817.

HOAGLIN, D. C., and ANDREWS, D. F. (1975), "The Reporting of Computation Based Results in Statistics", *Amer. Statist.*, 29, 122 - 126.

RAND Corporation (1955), "*A Million Random Digits with 100,000 Normal Deviates*", Free Press, Glenco, IL, U.S.A.

RIPLEY, B. D. (1987), "*Stochastic Simulation*", Wiley, New York.

SILVERSTEIN, A. (1979), "*Conquest of Death*", Mac Millan Publishing Company, New York.

TIPPETT, L. H. C. (1927), "*Random Sampling Numbers*", Tracts for Computers XV, Cambridge University Press, London.

TOCHER, K. D. (1963), "*The Art of Simulation*", English University Press, London.

Oscar H. Bustos
IMPA
Estrada Dona Castorina 110
Jardim Botânico
22460 Rio de Janeiro (RJ)
BRASIL