Santos Bravo

Una función random poco aleatoria

SANTOS BRAVO YUSTE Y
HÉCTOR SÁNCHEZ-PAJARES

Héctor Sánchez

El impacto del ordenador en todos los ámbitos de la actividad humana, y en la Ciencia en particular, es sin duda enorme. El ordenador permite tanto resolver problemas científicos anteriormente inabordables como corroborar teorías o soluciones "clásicas".

En Física, y especialmente en Física Estadística, una cualidad muy apreciada de los ordenadores es su capacidad de generar números (pseudo)aleatorios muy rápidamente. Esto los hace idóneos para la simulación de procesos estocásticos.

Veamos un ejemplo sencillo. Sea una partícula (caminante aleatorio) inicialmente en $(x,y)=(0,0)$ que, cada intervalo de tiempo τ, da un salto al azar de forma equiprobable bien hacia la derecha, la izquierda, arriba o abajo. ¿Cuál es la probabilidad $P(x,y,t)$ de encontrar la partícula en la posición (x,y) en el instante t?

Podemos responder a esta cuestión *simulando* $M \gg 1$ veces los t/τ saltos de la partícula: $P(x,y,t)$ vendrá dada aproximadamente por $m(x,y,t)/M$, siendo $m(x,y,t)$ el número de veces (del total de M "experiencias") en el que la partícula está en la posición (x,y) en el instante t. Esta simulación se puede llevar a cabo arrojando un par de monedas bien equilibradas para decidir hacia dónde salta el caminante (por ejemplo, si sale cara y cara, hacia arriba, si sale cara y cruz, a la derecha, etc.) pero esto es muy penoso si M y/o t/τ son grandes. Dado que decidir dar un salto en alguna de las cuatro direcciones de un modo equiprobable es una tarea simple que se repite una y otra vez, es natural acudir a un ordenador para simular este proceso estocástico. Para ello podríamos hacer que el ordenador tome números sucesivos de una tabla de números realmente aleatorios [1] distribuidos uniformemente en el intervalo [0,1) y asignar paso a la izquierda cada vez que el número de la tabla esté comprendido entre [0,1/4), hacia arriba si está entre [1/4,1/2), etc.

Por supuesto, este procedimiento es poco eficiente si se necesita, como es habitual en las aplicaciones científicas, una serie muy larga de números aleatorios. Lo ideal sería que el ordenador fuera capaz de *generarlos*. Pero esto es descabellado: es imposible que una máquina completamente determinista pueda generar números aleatorios. No obstante, sí existen algoritmos deterministas (conocidos como generadores de números aleatorios o GNA) que producen series de números que, *a efectos prácticos*, son indistinguibles de series de números generados de un modo *realmente* aleatorio (por ejemplo, obtenidos mediante el lanzamiento de monedas u otros procedimientos [2]). A los números generados mediante un GNA se los denomina, en ocasiones, números pseudoaleatorios. Aparte de su buena "aleatoriedad", un buen GNA debe [3] (i) tener un periodo largo, es decir, que el tamaño de la serie de números aleatorios más larga que pueda generarse sea próxima al rango de números que el ordenador pueda manejar de forma exacta, y (ii) debe ser rápido y "portátil", es decir, no debería ser dependiente en exceso del tipo de máquina o compilador en el que se va a correr el programa.

Existen muchos métodos para construir GNAs (como describe Raúl Toral en el artículo precedente de esta sección [4]), aunque el procedimiento más popular (probablemente por ser muy rápido) para generar una secuencia X_n de números (pseudo)aleatorios es el método congruencial lineal: X_n se obtiene a partir del anterior X_{n-1} mediante la relación $X_n=(a\,X_{n-1}+c)$ MOD m, siendo a el multiplicador, c el incremento y m el módulo. La operación x MOD m es simplemente el resto de la división de x entre m, es decir, x MOD $m = x-m\,[x/m]$ siendo $[y]$ la parte entera de y. Los valores de a, c y m deben ser *cuidadosamente* escogidos para que el GNA tenga todas las cualidades antes mencionadas.

Esto no es una tarea sencilla como da testimonio la extensa lista, compilada por Park y Miller [3] (una lista más reciente puede encontrarse en [5]) de malos o simplemente mediocres GNAs que se han usado profusamente durante años.

En este artículo vamos a mostrar, creemos que de un modo a la vez convincente y espectacular, que el GNA dado por la función RND del lenguaje MS-Basic (QBasic, QuickBASIC, GW-BASIC, Basic Interpreter para MS-DOS y Apple Macintosh, Visual Basic, ...) de Microsoft debe incluirse en esta lista de generadores poco afortunados. (Este generador es posiblemente uno de los más populares pues, hasta hace poco tiempo, el lenguaje QBasic venía incluido con el sistema operativo DOS y Windows de Microsoft.) La función RND es un generador congruencial lineal [6] con a = =1140671485, c = 12820163 y $m = 2^{24}$, periodo 2^{24} y cuya semilla inicial por defecto es X_0=327680.

En la figura 1 se representan los lugares de una red cuadrada bidimensional que han sido visitados por $N=2^9=512$ caminantes aleatorios que inicialmente partieron del punto central de la figura cuando la función RND de MS-Basic se utiliza para escoger de forma equiprobable una de las cuatro direcciones posibles en cada uno de los pasos de los N caminantes (el programa de esta simulación se da en el apéndice). Es evidente que hay algo fundamentalmente erróneo en esta simulación pues algunos caminantes siguen trayectorias que no son aleatorias en absoluto. Este comportamiento anómalo se da cuando N es una potencia de dos; en la figura 2 se ve que el territorio explorado por N=511 caminantes sí exhibe un aspecto razonable.

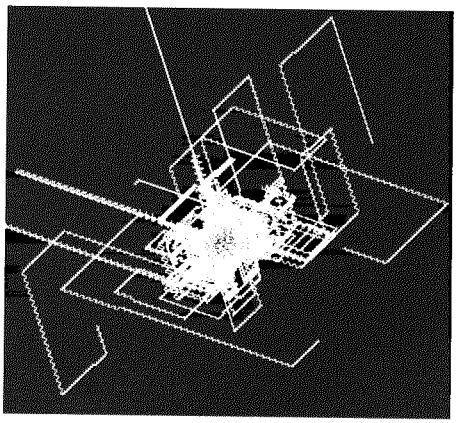

Figura 1. Territorio explorado (en blanco) por $N=2^9=512$ caminantes aleatorios (puntos) tras 1600 pasos cuando se usa el GNA dado por la función RND de MS-Basic. Los caminantes partieron incialmente de un mismo lugar.

Podría pensarse que la culpa de esta anomalía se encuentra en la simplicidad de los generadores congruenciales lineales, pero esto no es cierto. Por ejemplo, los parámetros $a=7^5=16807$, $c=0$ y $m=2^{31}-1$ conducen a un GNA, conocido como "estándar mínimo" [3], satisfactorio para la mayoría de las aplicaciones [7] y que se ha revelado como muy bueno tras muchos años de uso y tras superar numerosos test que métodos más complicados supuestamente mejores (por ejemplo, el R250 descrito en el artículo precedente [4]) no han pasado [3, 8].

En contra de lo que uno pudiera suponer a la vista de la figura 1, el GNA de MS-Basic *no es obviamente malo*. De hecho, supera con buenas notas muchos de los test téoricos estándar que se emplean en el análisis de los GNAs. Uno de los más importantes es el "test espectral" [9], el cual se basa en lo siguiente: cuando los sucesivos n números proporcionados por un GNA lineal congruencial (esto también sucede para otros tipos de generadores) se toman como las n coordenadas de un punto en un espacio n dimensional, entonces los puntos así formados no tienden a rellenar completamente el espacio sino que se disponen sobre hiperplanos paralelos de dimensión $(n-1)$. De entre todas las familias de hiperplanos paralelos, se escoge aquella en la que los hiperplanos están más separados y se llama d_n a esta distancia. Para cada dimensión n existe una distancia $d^*_n(m)$ mínima posible (ideal) que depende del módulo m del generador. Los puntos estarán tanto más uniformemente distribuidos sobre el espacio n dimensional, y por tanto el GNA será en principio tanto mejor, cuanto más próximo esté a uno el parámetro $S_n=d^*_n(m)/d_n$. En el cuadro 1 se dan los resultados de este test cuando se aplica a varios generadores. El GNA de MS-Basic es el que obtiene mejores calificaciones superando claramente al GNA *estándar mínimo*. Este resultado muestra que los test teóricos son útiles pero no pueden reemplazar a los test empíricos (test en los que se comparan los resultados de las simulaciones con resultados teóricos bien establecidos).

La dificultad a la hora de elegir un GNA reside en que sus defectos (no hay GNA perfecto) pueden pasar desapercibidos (o, con mayor precisión, no ser relevantes) en muchos test teóricos y empíricos y, por desgracia, ponerse de manifiesto con la mayor crueldad justamente en la simulación que más nos importa: la nuestra. Esto es justamente lo que ha sucedido en la simulación del territorio explorado con el GNA de MS-Basic: para 511 caminantes la simulación es razonable, para 512 es absurda. Lo malo es que estas situaciones son imprevisibles y, quizás, más frecuentes de lo que uno desearía. Por ejemplo, según Stauffer [7], cuando en la simulación de la magnetización de un modelo de Ising pentadimensional se usa el GNA *estándar mínimo* aparecen "errores dramáticos" si el lado de la red es una potencia de dos. Por otro lado, los GNAs más sofisticados que en principio habrían de ser más fiables son también, comparativamente, muy lentos [8].

	S_2	S_3	S_4	S_5	S_6	S_7	S_8
MS-Basic	0.790	0.782	0.418	0.666	0.661	0.632	0.685
Estándar mín.	0.338	0.441	0.575	0.736	0.645	0.571	0.610
Derive	0.097	0.555	0.548	0.321	0.643	0.521	0.673
TurboPascal	0.748	0.395	0.599	0.600	0.760	0.345	0.488

Cuadro 1. Resultados del test espectral normalizado para varios GNAs. Además del GNA de MS-Basic y el *estándar mínimo*, hemos incluido el GNA congruencial lineal usado en el programa de cálculo simbólico *Derive* ($a=2^{32}$, $m=3141592653$, $c=1$) y en *TurboPascal 7.0* ($a=2^{32}$, $m=134775813$, $c=1$) [5].

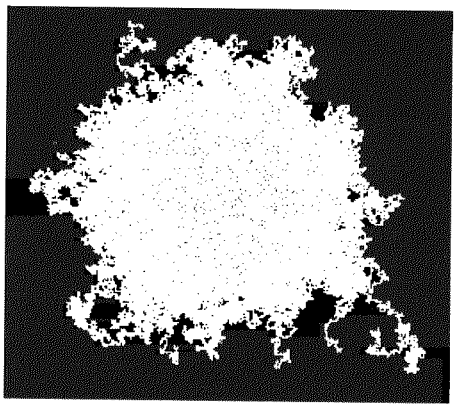

Figura 2. Territorio explorado (en blanco) por $N=511$ caminantes aleatorios (puntos), inicialmente situados en un mismo lugar, cuando se usa el GNA dado por la función RND de MS-Basic.

En resumen, la generación de buenos números aleatorios mediante ordenador no es una tarea fácil por lo que el GNA a emplear en nuestros cálculos debe escogerse con cuidado. Lo más sano es no fiarse del generador que se esté usando y comprobar que los resultados no cambian si se emplea un GNA completamente distinto (cuanto más diferente, mejor).

Apéndice: El programa de simulación

El programa MS-Basic que genera la figura 1 es:

```
SCREEN 12
cx% = 320: cy% = 240
numca% = 512
DIM x%(numca%), y%(numca%)
tmax% = 3000
CLS
FOR n% = 1 TO numca%
  x%(n%) = 0: y%(n%) = 0
NEXT n%
FOR t% = 1 TO tmax%
  FOR n% = 1 TO numca%
  PSET (x%(n%) + cx%, y%(n%) + cy%)
  dir% = INT(4! * RND)
  SELECT CASE dir%
       CASE 0: y%(n%) = y%(n%) + 1
       CASE 1: y%(n%) = y%(n%) - 1
       CASE 2: x%(n%) = x%(n%) + 1
       CASE 3: x%(n%) = x%(n%) - 1
  END SELECT
  PSET (x%(n%) + cx%, y%(n%) + cy%), 4
  NEXT n%
NEXT t%
```

La variable numca\% es el número N de caminantes aleatorios empleados en la simulación. Cambiando el valor de numca\% a 511 se obtiene la figura 2.

Bibliografía

[1] G. MARSAGLIA ha producido un CD-ROM con una tabla de 4.8×10^9 bits "inobjetablemente" aleatorios, generados mediante la combinación de los resultados de varios de los mejores GNAs y los producidos por diversas fuentes de ruido (entre ellas, música *rap* digitalizada). Véase http://stat.fsu.edu/\~geo/diehard.html.

[2] En la página web http://lavarand.sgi.com puede verse un modo singular (y espectacular) de generar números aleatorios mediante lámparas "Lava Lite" (lámparas decorativas, de moda en los años 70, en las que masas globulares de colores ascienden y descienden dentro de un fluido).

[3] S. K. PARK Y K. W. MILLER, *Random number generators: good ones are hard to find*, Communications of the ACM **31** (10), 1192-1201 (1988).

[4] R. TORAL, *Revista Española de Física*, artículo precedente.

[5] http://crypto.mat.sbg.ac.at/results/karl/server/server.html.

[6] http://support.microsoft.com/support/kb/articles/ Q231/8/47.ASP.

[7] W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING, AND B.P. FLANNERY, *Numerical Recipes in FORTRAN. The Art of Scientific Computing*. Cambridge University Press, segunda edición, 1992. Puede encontrarse una versión electrónica en http://www.nr.com/.

[8] D. STAUFFER, *Random number generation*. En: K. H. Hoffmann y M. Schreiber (editores), *Computational Physics*, Springer, Berlin (1996).

[9] D.E. KNUTH. *The Art of Computer Programming, volume 2: Seminumerical Algorithms*. Addison-Wesley, Reading, segunda edición, 1981. Véase también la referencia [5]. En http://random. mat.sbg.ac.at/results/karl/spectraltest/index.html se proporcionan programas en C y *Ma-thematica* que llevan a cabo el test espectral.

Santos Bravo Yuste y Héctor Sánchez-Pajares
están en Departamento de Física, Universidad de Extremadura. Badajoz

LIBROS Y PUBLICACIONES RECIBIDOS

- **Manual de Física.** F. Kurt Kneubühl. Herber. Barcelona, 2001. 588 pp.
- **Problemas resueltos de Química cuántica y espectroscopia molecular**. J. M. Pérez Martínez, A. L. Esteban Elum, Mª Paz Galache Payá. Publicaciones Universidad de Alicante. Alicante, 2001. 198 pp.
- **Proyecto Penélope**. El papel de la Historia de la Ciencia en la Enseñanza Secundaria. Editor: M. Hernández González. La Orotava, 2002. 241 pp.
- **Materiales de Historia de la Ciencia 1. Del Flogisto al Oxígeno.** J. Cartwright. Fundación Canaria Orotava de Historia de la Ciencia. La Orotava, 2000. 53 pp.
- **Materiales de Historia de la Ciencia 2. Prólogo a la traducción de la Historia Natural del Conde de Buffon.** J. Clavijo y Fajardo. Estudio preliminar J.L. Prieto. Fundación Canaria Orotava de Historia de la Ciencia. La Orotava, 2001. 90 pp.
- **Revista Mexicana de Física.** Publicación Bimestral de la Sociedad Mexicana de Física. Vol. 48, nº 1.
- **Physics Today.** Febreuary 2002. Vol. 55, nº 2.
- **Physics Word.** March 2002. Vol. 15, nº 3.
- **Avances en Física Médica.** 2001. Sociedad Española de Física Médica. Málaga, 2001. 232 pp.
- **Catálago Iberoamericano de programas y recursos humanos en Física.** 2001-2002. UNAM, CONACyT, CINVESTAV. México, 2002. 410 pp.
- **Progress of Theoretical Physics.** Vol. 106, nº 6, December 2001. The Physical Society of Japan.
- **Canadian Journal of Physics.** Vol. 79, nº 11/12. November/December 2001. National Research Council of Canada.
- **La textura del món. Les partícules elementals: dels quarks a la web.** J. Velasco. Premi Europeu de Divulgació Científica Estudi General 2000. Bromera, Publicacions de la Universitat de València. Valencia, 2000. 285 pp.
- **Revista de la Academia de Ciencias Exactas, Físicas, Químicas y Naturales de Zaragoza**. Vol. 56, Serie 2ª. 2001.
- **IUPAP, News Bulletin.** Spring 2002 Edition.
- **Informes Técnicos Ciemat.** Resultados de la I Campaña de Evaluación de la Trazabilidad de Activímetros de los Servicios de Medi-cina Nuclear en la Comu-nidad Autónoma de Madrid. E. García-Toraño Martínez, J. M. Los Arcos Merino y M. Roteta Ibarra. Nº 987, Febrero, 2002.
- **Informes Técnicos Ciemat.** The Normal-incidence Vacuum-ultraviolet Spectrometer for the TJ-II and First Experimental Results. K. J. McCarthy, B. Zurro y A. Baciero. Nº 988, Marzo, 2002.
- **Participación del Ciemat en la 27 Reunión Anual de la Sociedad Nuclear Española.** Editorial Ciemat. Madrid, 2002. 150 pp.
- **Complex Systems: Chaos and Beyond. A constructive approach with applications in life science.** K. Kaneko and I. Tsuda. Springer. Alemania, 2001. 273 pp.
- **Critical Phenomena in Natural Sciences.** Chaos, fractals, selforganization and disorder: concepts and tools. D. Sornette. Springer. Alemania, 2000. 434 pp.
- **Gaussian Self-Affinity and Fractacts.** B. B. Mandelbrot. Springer. EEUU, 2002. 654 pp.
- **Random Heterogeneous Materials.** Microstructure and macroscopic properties. S. Torquato. Springer. EEUU, 2002. 701 pp.
- **Revista Brasileira de Ensino de Física.** Sociedad Brasileña de Física. Vol. 23, nº 4, diciembre, 2001.
- **Ciencia, innovación y futuro.** F. Castro Díaz-Balart. Edición Grijalbo. Barcelona, 2002. 617 pp.
- **Cuba. Amanecer del tercer milenio**. Ciencia, Socie-dad y Tecnología. Coordina-dor-Editor, F. Castro Díaz-Balart. Editorial Debate. Madrid, 2002. 413 pp.