

Una breve introducción al método de Monte Carlo

por

Elena Castilla y Pedro J. Chocano

RESUMEN. Hacemos un repaso al origen y evolución del método de Monte Carlo, explicando alguna de las técnicas más importantes para la generación de números y variables aleatorias. También repasamos algunos tests para comprobar si nuestros datos son aleatorios o bien provienen de una distribución dada. Por último, comentamos algunas aplicaciones básicas del método de Monte Carlo, como la metodología bootstrap o la aproximación de integrales definidas y soluciones de ecuaciones diferenciales.

El método de Monte Carlo, también conocido como simulación de Monte Carlo o simplemente como simulación estocástica [34, 35], fue definido por Halton (1970, [19]) como «un método para calcular un parámetro de una hipotética distribución, usando una secuencia de números aleatorios para construir una muestra de esta población y, a partir de ella, calcular su estimación estadística». Aunque puede variar, el método de Monte Carlo suele tener la siguiente estructura:

- i. Definir un espacio muestral de posibles datos o *inputs*.
- ii. Generar datos en ese dominio mediante una distribución de probabilidad determinada.
- iii. Realizar un cálculo determinístico a partir de esos datos generados.
- iv. Aplicar estos cálculos al resultado final.

Un ejemplo clásico de divulgación del método de Monte Carlo es el de la estimación de la superficie de un lago. Supongamos que tenemos un lago cuya superficie desconocemos, pero este lago está dentro de un terreno cuya área sí es conocida, por ejemplo, un cuadrado o rectángulo. Se dispone de una máquina que dispara al azar n bolas de goma en esta área (ver la figura 1).

Tras hacer una tanda de disparos, contamos el número de bolas que han caído al lago, k (que también se puede contar como $n - m$, donde m es el número de bolas que han caído fuera del lago). Entonces, se puede establecer la siguiente relación:

$$\frac{\text{área(lago)}}{\text{área(terreno)}} \approx \frac{k}{n},$$

$$\text{área(lago)} \approx \frac{k}{n} \times \text{área(terreno)}.$$

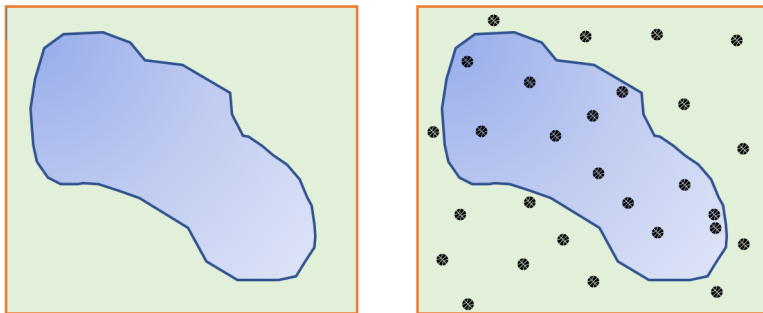


Figura 1: Estimación de la superficie de un lago.

Este mismo método puede utilizarse para estimar el valor de π . Si tomamos un cuadrado de perímetro $4l$ su circunferencia inscrita tendrá radio $l/2$. Si generamos n puntos aleatorios en el cuadrado (ver la figura 2), se tiene la siguiente aproximación:

$$\frac{\text{área(círculo)}}{\text{área(cuadrado)}} = \frac{\pi(l/2)^2}{l^2} = \frac{\pi}{4} \approx \frac{k}{n},$$

donde k es el número de puntos que caen dentro de la circunferencia. Así, una estimación de π puede realizarse mediante

$$\pi \approx \frac{4k}{n}.$$

Naturalmente, la estimación mejorará cuanto mayor sea la cantidad de números

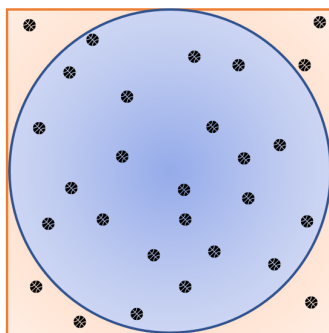


Figura 2: Estimación de π .

generados, y dependerá de que estos se generen de manera totalmente aleatoria y uniformemente en el cuadrado. Es por tanto necesario desarrollar una técnica adecuada para la generación de estos números. A lo largo de este artículo hacemos un repaso al origen y evolución del método de Monte Carlo, explicando alguna de las

técnicas más importantes para la generación de números y variables aleatorias. También proporcionaremos tests para la comprobación de aleatoriedad o de que nuestros datos provienen de una distribución dada e introduciremos el *bootstrap* como mayor representante de las llamadas *técnicas de remuestreo*. Por último, comentaremos algunas aplicaciones básicas del método de Monte Carlo, como la aproximación de integrales definidas o de soluciones de ecuaciones diferenciales.

1. ORÍGENES DEL MÉTODO DE MONTE CARLO

En 1777, Georges Louis Leclerc, conde de Buffon, resolvió un problema planteado por él mismo años antes y que ahora conocemos como el problema de la aguja de Buffon. Esta solución fue publicada en el Suplemento IV de la enciclopedia *Histoire Naturelle* [7] escrita entre otros por el propio Buffon durante 50 años, y en la que se intentó plasmar todo el saber de la época concerniente a la ciencia (figura 3). Años después, Laplace [23] corrigió un error en la solución, por lo que algunos autores también se refieren a este problema como el de la aguja de Buffon-Laplace.

El problema consiste en lanzar una aguja de longitud l sobre un papel en el que se han trazado rectas paralelas distanciadas entre sí una cantidad t , con $l \leq t$. Buffon se preguntaba por la probabilidad de que la aguja cruzara alguna de las líneas. Se puede demostrar que dicha probabilidad es $\frac{2l}{t\pi}$. Por tanto, podemos hacer una estimación del número π repitiendo el experimento un número suficientemente alto de veces. Así, si n es el número total de intentos y k el número de veces que la aguja ha cruzado alguna línea, tenemos

$$\pi \approx \frac{2nl}{tk}.$$

En 1901 el matemático italiano Lazzarine [24] realizó el experimento de Buffon utilizando una aguja de $l = 2.5$ cm y líneas de $t = 3$ cm de separación. De 3408 agujas, 1808 cruzaron la línea, dando una aproximación de

$$\pi \approx \frac{2 \times 3408 \times 2.5}{3 \times 1808} = \frac{355}{133} = 3.1415929 \dots$$

Dado que $\pi = 3.1415926 \dots$, la aproximación de Lazzarine dio una precisión de seis decimales! Una interesante discusión sobre este experimento puede encontrarse en [2]. Hoy en día hay muchas aplicaciones en internet que permiten simular el experimento de la aguja de Buffon de una manera bastante ilustrativa, por ejemplo, <https://mste.illinois.edu/activity/buffon/>.

Muchos autores consideran el experimento de la aguja de Buffon como el origen del método de Monte Carlo. Sin embargo, esta terminología no fue usada hasta 200 años después por el matemático Nicholas Metropolis, a mediados de siglo XX. Según el propio Metropolis, dos fueron los motivos del desarrollo de estas técnicas: (1) la aparición de las primeras computadoras, como por ejemplo la ENIAC [15], y (2) el trabajo del propio Metropolis con Stanislaw Ulam y John Von Neumann para resolver el problema de difusión de neutrones en material fisionable usando el muestreo estadístico, ya que este problema era analíticamente intratable.

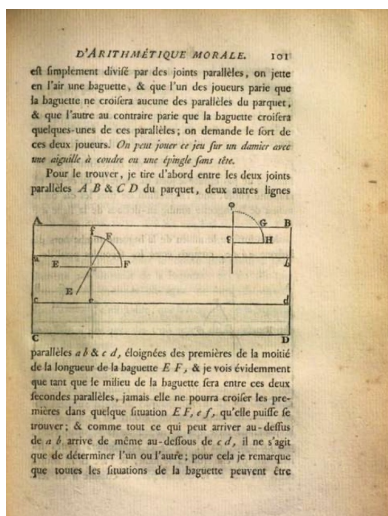


Figura 3: El problema de la aguja de Buffon.

Cuentan que Ulam, mientras hacía solitarios con una baraja de cartas, se planteó cómo calcular la probabilidad de ganar una partida. Como los cálculos combinatorios podían resultar demasiado tediosos, a Ulam se le ocurrió si, en lugar de hacer estos cálculos, podría emular una gran cantidad de partidas y contar los casos favorables. Ulam pensó rápidamente en cómo poder aplicar su idea a cuestiones físicas como el problema de la difusión de neutrones, entre otros. Poco después, en 1946, le comenta sus ideas a Metropolis y Von Neumann y comienzan a realizar cálculos. Esta técnica fue aplicada en el conocido proyecto Manhattan para la resolución de muchos de los problemas complejos que surgieron durante la creación de la bomba atómica. Fue en 1949 cuando Metropolis y Ulam usaron por primera vez el término «Monte Carlo» en un texto científico [30].

2. GENERACIÓN DE NÚMEROS (PSEUDO)ALEATORIOS

La primera tabla conocida de números aleatorios fue publicada en 1927 por Leonard Henry Caleb Tippet [38], un estudiante de Karl Pearson. La tabla de números aleatorios de Tippet consiste en 41 600 dígitos agrupados en 10 400 números de 4 cifras, y fue presentada en un total de 26 páginas, cada una de ellas con 32 columnas y 50 filas. Estos números fueron cogidos de forma «aleatoria» del registro del censo, provocando la primera lista de números aleatorios con periodo 10 000. Otra tabla muy conocida de números aleatorios fue dada por Fisher y Yates en 1938. Esta tabla constaba de 10 000 cifras tomadas «aleatoriamente» de las tablas de logaritmos [13].

Una de las tablas más usadas fue la generada por la Corporación RAND, una organización que ofrecía servicios de investigación a las Fuerzas Armadas de Estados Unidos, en el año 1955, con el objetivo de ser usada en aquellos experimentos en los

que se aplicaba el incipiente método de Monte Carlo. La tabla RAND, también conocida como la tabla de un millón de dígitos por su tamaño (ver figura 4), fue presentada en 400 páginas, cada una conteniendo 50 líneas de 50 dígitos, agrupados de cinco en cinco (oscilando entre el 00000 y el 99999), así como en forma de tarjetas perforadas. Los números fueron generados de manera física con una rueda de ruleta conectada a una computadora, y luego fueron cuidadosamente filtrados antes de ser pasados a la tabla [6].

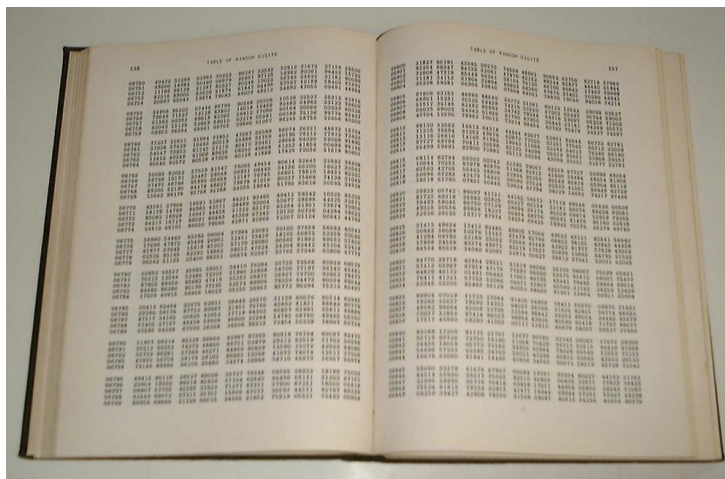


Figura 4: Un millón de dígitos aleatorios.

En los años posteriores, la aparición de ordenadores de alta velocidad promovió el desarrollo de métodos de generación de números pseudoaleatorios, evitando así tener que leerlos de tablas de tamaños inabarcables. Una secuencia de números se considera *pseudoaleatoria* si es aparentemente aleatoria, aunque se haya producido bajo un proceso totalmente determinístico y reproducible. En concreto, se parte de un número original (semilla) a partir del cual, y por un método determinístico, se produce una secuencia de números aleatorios que siguen una distribución uniforme en el intervalo $[0, 1]$.

2.1. MÉTODO DE VON NEUMANN

Uno de los primeros métodos para generar números pseudoaleatorios es el *método de los cuadrados medios* propuesto por Von Neumann en 1946 y publicado en 1951 [31]. Como Von Neumann comenta en su artículo, si cogiéramos las diez cifras centrales de un número de diez cifras elevado al cuadrado e iteráramos, se irían gradualmente mezclando los dígitos produciendo cierta pseudoaleatoriedad. De manera general, el método de los cuadrados medios de Von Neumann consiste en los siguientes pasos:

- i. Elegir un número par de n cifras, x_0 (semilla).

- ii. Elevar x_0 al cuadrado, x_0^2 .
- iii. Tomar los n dígitos centrales de x_0^2 , obteniendo x_1 .
- iv. Volver al paso ii y repetir el proceso.
- v. Para producir números en el intervalo $[0, 1]$, dividir la secuencia entre 10^n :

$$u_i = \frac{x_i}{10^n}.$$

Véase un esquema visual del método en la figura 5. La secuencia obtenida a partir de un cierto punto en adelante será nuestra secuencia de números pseudo-aleatorios. Este método tiene sus desventajas; por ejemplo, los números generados pueden repetirse cíclicamente después de un periodo (la longitud del ciclo) corto.

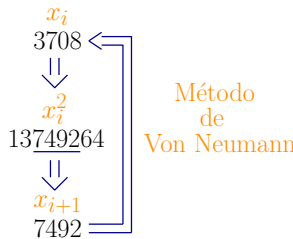


Figura 5: Método de Von Neumann.

2.2. MÉTODOS CONGRUENCIALES

Los generadores congruenciales lineales fueron introducidos por Lehmer en 1951 [25]. Decimos que dos enteros a y b son congruentes (módulo m) si $a - b$ es múltiplo de m . Un *método congruencial* consiste en:

- i. Elegir un número x_0 inicial (semilla).
- ii. Para $i \geq 1$ obtener de manera recursiva

$$x_i = ax_{i-1} + b \text{ módulo } m.$$

- ii. Para producir números en el intervalo $[0, 1]$, dividir la secuencia entre m :

$$u_i = \frac{x_i}{m}.$$

Si $b = 0$, se dice que es un generador multiplicativo y, en caso contrario, se dice que es mixto. La pseudoaleatoriedad de los métodos congruenciales se refleja en la siguiente proposición, donde se observa de forma clara tanto el algoritmo determinista que es empleado como que condiciones iniciales iguales producen mismos resultados.

PROPOSICIÓN 1. *Si se aplica un método congruencial, se verifica*

$$x_i = a^i x_0 + b \frac{a^i - 1}{a - 1} \text{ módulo } m.$$

Un gran inconveniente de este método es que los valores u_i solo pueden tomar valores i/m , $i = 0, \dots, m$. Además, es inevitable que se produzcan ciclos, ya que el periodo nunca puede exceder al módulo. Así, se tiene que tomar un valor suficientemente grande de m para que el conjunto de posibles valores de la sucesión se asemeje a la de una variable continua uniforme en el intervalo $[0, 1]$ y que el ciclo no sea demasiado corto. En concreto, nos podemos preguntar cómo elegir los parámetros para que el ciclo sea completo, es decir, para que se atraviesen todos los posibles estados (números) antes de volver a la semilla original. Hull y Dobell establecieron en 1962 [21] el siguiente teorema:

TEOREMA 2. *Un generador congruencial tiene periodo completo si y solo si se cumplen las siguientes condiciones:*

1. m y b son primos entre sí.
2. Si q es un número primo que divide a m , entonces q divide a $(a - 1)$.
3. Si 4 divide a m , entonces 4 divide a $(a - 1)$.

Nótese que un generador de ciclo completo lo es con independencia de la semilla x_0 escogida. Sin embargo, los generadores de ciclo no completo tendrán diferente longitud dependiendo de la semilla escogida.

En la década de los 70, fue muy difundido el generador multiplicativo RANDU, implementado en IBM, con $a = 2^{16} + 3$, $b = 0$, $m = 2^{31}$ y x_0 un número impar. Sin embargo, este generador es considerado como uno de los peores generadores de números aleatorios creados [14]. Como Marsaglia destacó ya en 1968 [28], un problema asociado con los métodos congruenciales es su naturaleza «reticular». La secuencia de valores generados se sitúa en un número normalmente bajo de hiperplanos paralelos, que dependerá de la elección de los parámetros (véase la figura 6, donde el eje de abscisas representa el número de iteraciones y el eje de ordenadas el valor asociado a la iteración). En particular, para el generador RANDU, se puede demostrar que el número de hiperplanos es 15, en un cubo de dimensión 3. Este generador fue sistemáticamente desapareciendo, aunque algunos compiladores de FORTRAN lo mantuvieron hasta finales del siglo XX.

2.3. CÓMO COMPROBAR LA ALEATORIEDAD (TEST DE WALD-WOLFOWITZ)

Este test, también conocido como test de las rachas, debe su nombre a los estadísticos Abraham Wald y Jacob Wolfowitz ([39], 1940) que lo presentaron originalmente como un test no paramétrico para comprobar si dos muestras vienen de la misma población. En su versión más conocida, es un test para comprobar la aleatoriedad de una secuencia de datos. Para aplicarlo, necesitamos los siguientes supuestos:

1. Los datos deben considerarse en el orden de ocurrencia.
2. Los datos deben reducirse a dos posibles valores.

Definimos una «racha» como una serie de valores consecutivos iguales. Por ejemplo, en la secuencia de 20 elementos

+ + + - - + - + - + + - - + + - + - - -

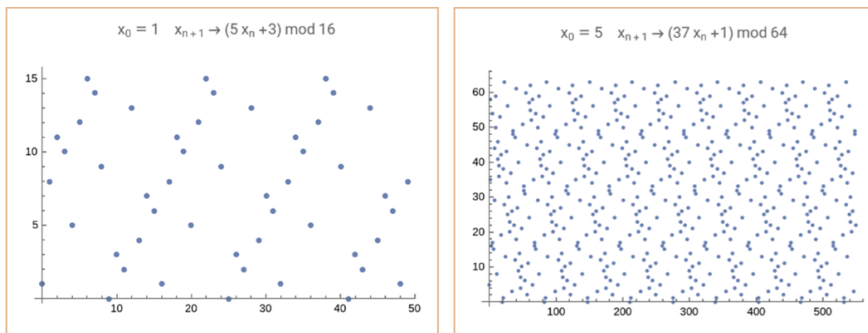


Figura 6: Generadores congruenciales. Imágenes generadas con [4].

tendremos 12 rachas, 6 correspondientes al símbolo «+» y 6 correspondientes al símbolo «-». Se puede demostrar que, bajo la hipótesis nula de aleatoriedad de la muestra, el número total de rachas se comporta asintóticamente como una normal de media μ y varianza σ^2 , con

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1, \quad \sigma^2 = \frac{(\mu - 1)(\mu - 2)}{n_1 + n_2 - 1},$$

donde n_1 y n_2 son el número de elementos asociados a cada una de las dos categorías.

Así, dada una secuencia de valores en el orden original y categorizada en dos clases, se calcula el estadístico

$$Z = \frac{N - \mu}{\sigma},$$

donde N es el número total de rachas observado. Dado un nivel de significación α , la hipótesis de aleatoriedad se rechaza si $|Z| \geq Z_{\alpha/2}$, donde $Z_{\alpha/2}$ es el cuantil $1 - \alpha/2$ de la distribución normal estándar.

EJEMPLO 3. *En una clase se anota el género de 20 alumnos a la hora de entregar el examen. Se desea conocer con un nivel de significancia del 5% si la posición de alumnos y alumnas es aleatoria. En la muestra aparecen 10 alumnos y 10 alumnas, y el número de rachas es igual a 12. La colocación de los alumnos es como sigue:*

♂♂♂♂♀♀♂♂♂♂♀♀♂♂♂♂♀♀♂♂♂♂♂♂

En este caso, $|Z| \approx 0.46 < Z_{0.05/2} = 1.96$, por lo que no rechazamos la hipótesis de aleatoriedad. En otra clase se ha obtenido la siguiente sucesión:

♀♀♀♀♀♀♀♀♀♀♂♂♂♂♂♂♀♀♀♀♀♀

En este caso tenemos 5 alumnos y 15 alumnas, $N = 3$ y $|Z| \approx 3.43 > 1.96$, por lo que rechazamos la hipótesis de aleatoriedad.

El test de Wald-Wolfowitz puede utilizarse para comprobar la aleatoriedad de una secuencia de números aleatorios tomando los números en el mismo orden de la secuencia y marcando con «+» si el número es mayor que la media (o mediana o cualquier otro estadístico a considerar) y con «-» si es menor que el valor elegido.

EJEMPLO 4. *Dada la sucesión*

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| 0.50 | 0.51 | 0.89 | 0.98 | 0.74 | 0.91 | 0.51 | 0.62 | 0.34 | 0.24 |
| 0.19 | 0.72 | 0.83 | 0.08 | 0.54 | 0.04 | 0.01 | 0.36 | 0.16 | 0.28 |
| 0.18 | 0.01 | 0.95 | 0.69 | 0.22 | 0.49 | 0.23 | 0.32 | 0.82 | 0.60 |
| 0.31 | 0.42 | 0.73 | 0.04 | 0.83 | 0.45 | 0.13 | 0.57 | 0.63 | 0.44 |

la sucesión de signos asociados (considerando la media como punto crítico) es

+ + + + + + + - - - + + - + - - - -
 - - + + - + - - + + - - + - + - - + + -

Entonces $N = 18$ y $n_1 = n_2 = 20$. En este caso $|Z| \approx 0.96 < 1.96$, por lo que a un nivel de significación del 5% no rechazamos la hipótesis de aleatoriedad.

En la siguiente sección veremos algunas técnicas para la generación de variables aleatorias (continuas) a partir de números aleatorios que, como hemos visto, generalmente se considera que pertenecen a una variable aleatoria uniforme de intervalo $[0, 1]$.

3. GENERACIÓN DE VARIABLES ALEATORIAS CONTINUAS

3.1. MÉTODO DE LA TRANSFORMADA INVERSA

El método de la transformada inversa, probablemente el método más «sencillo» para la generación de variables aleatorias, se basa en el siguiente resultado:

PROPOSICIÓN 5. *Supongamos una variable aleatoria X que tiene función de distribución continua $F_X(x)$ estrictamente creciente siempre que $0 < F_X(x) < 1$. Sea U una variable aleatoria que sigue una distribución uniforme en el intervalo $(0, 1)$, $U \sim U(0, 1)$. Entonces, la variable aleatoria $F_X^{-1}(U)$ tiene como función de distribución a F_X .*

El método de la transformada inversa para generar valores de una variable aleatoria X descrita por una función de distribución $F_X(x)$ consiste en los siguientes pasos:

- i. Se genera un número aleatorio u de una variable aleatoria $U(0, 1)$.
- ii. Se calcula $x = F_X^{-1}(u)$.

EJEMPLO 6. *Vamos a generar valores de una variable aleatoria exponencial mediante el método de la transformada inversa. Si $X \sim \text{Exp}(\lambda)$, $F_X(x) = 1 - \exp(-\lambda x)$. Si hacemos $u = F_X(x)$, entonces*

$$u = 1 - e^{-\lambda x},$$

$$x = -\frac{\log(1 - u)}{\lambda}.$$

Nótese que $U \sim U(0,1)$ implica que $1 - U \sim U(0,1)$, por lo que el método de la transformada inversa se simplificaría en: dado un número aleatorio u de una variable aleatoria $U(0,1)$, calcular

$$x = -\frac{\log(u)}{\lambda}.$$

EJEMPLO 7. Vamos a generar valores de una variable aleatoria Gamma mediante el método de la transformada inversa. En este caso, si $X \sim \text{Gamma}(n, \lambda)$ y

$$F_X(x) = \int_0^x \frac{\lambda^n}{(n-1)!} \exp(-\lambda y) y^{n-1} dy,$$

no es posible aplicar el método de la transformada inversa de forma directa. Sin embargo, podemos utilizar la siguiente propiedad: Sean X_1, \dots, X_n variables aleatorias independientes tales que $X_i \sim \text{Exp}(\lambda)$ para todo i ; entonces

$$\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda).$$

Por tanto, se puede usar el siguiente algoritmo:

- i. Generar u_1, \dots, u_n independientes de una variable aleatoria $U(0,1)$.
- ii. Calcular $x = -\sum_{i=1}^n \frac{\log(u_i)}{\lambda}$.

Aunque este método es aplicable a cualquier función continua, a veces no es posible encontrar una función explícita para $F_X^{-1}(u)$ o, en caso de poderse, puede suceder que su evaluación requiera demasiado tiempo de computación.

3.2. MÉTODO DE RECHAZO

Este método fue propuesto por Von Neumann (ver la figura 7) en su ya citado artículo *Various techniques used in connection with random digits* (1951, [31]). Supongamos que queremos generar valores de una variable aleatoria X con función de densidad $f(x)$. Supongamos también que sabemos muestrear una variable aleatoria Y con función de densidad $g(x)$ tal que $\sup_x f(x)/g(x) \leq c$, con $c > 0$. El método del rechazo (también llamado de *aceptación y rechazo*) consiste en los siguientes pasos:

- i. Generar u de una variable aleatoria $U(0,1)$ e y de la variable aleatoria Y .
- ii. Si $u \leq \frac{f(y)}{cg(y)}$, hacer $x = y$.
- iii. En caso contrario, volver al paso i.

EJEMPLO 8. Supongamos que queremos generar valores de la variable aleatoria $X \sim \text{Beta}(\alpha, \beta)$, cuya función de densidad es

$$f(x) = B(\alpha, \beta) x^{\alpha-1} (1-x)^{\beta-1},$$



Figura 7: John Von Neumann (1903–1957), uno de los padres de la simulación estadística.

donde $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$. Supongamos el caso particular $\alpha = \beta$. Esto es: $f(x) = B(\alpha, \alpha)(x(1-x))^{\alpha-1}$. Cojamos g de una uniforme $U(0, 1)$. Se puede demostrar fácilmente que

$$f(x)/g(x) = f(x) \leq B(\alpha, \alpha)(1/4)^\alpha.$$

Así, aplicando el método del rechazo:

- i. Generar u_1 y u_2 independientes de una variable aleatoria $U(0, 1)$.
- ii. Si $u_2 \leq 4^\alpha (u_1(1-u_1))^\alpha$, hacer $x = u_1$.
- iii. En caso contrario, volver al paso i.

Nótese que, en este caso, no es necesario calcular $B(\alpha, \alpha)$.

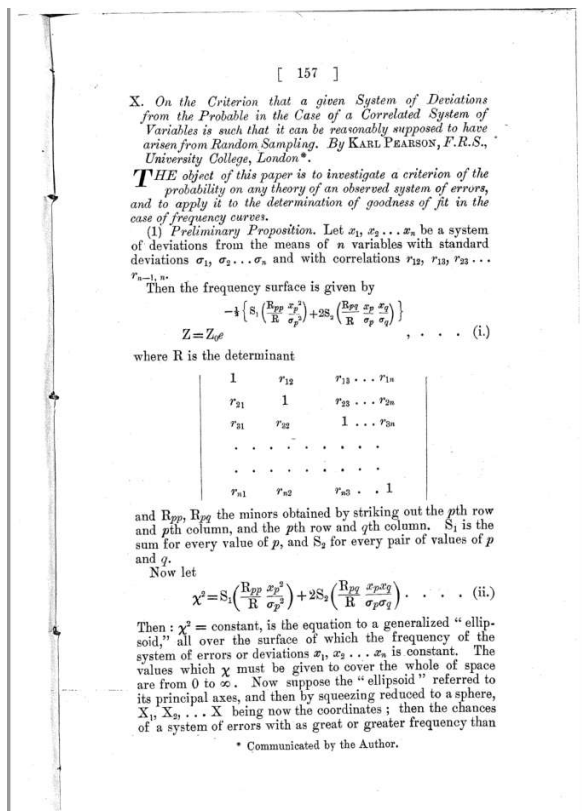
El método de aceptación y rechazo es un método alternativo al de inversión cuando se dispone de una función de densidad sencilla de acotar.

3.3. CÓMO COMPROBAR QUE LOS DATOS PERTENECEN A UNA DISTRIBUCIÓN

3.3.1. TEST DE LA χ^2 DE PEARSON

Este test fue introducido por Karl Pearson en 1900 ([32], véase la figura 8). En este artículo Pearson critica las prácticas habituales en esa época, en las que se tendía a asumir que los datos seguían una distribución normal. En particular, cita el trabajo de George Biddell Airy en su libro *Theory of Errors of Observation* [1]. La prueba de la chi-cuadrado permite contrastar si nuestros datos provienen de una distribución en concreto, digamos $F_0(x)$, que consideramos aquí totalmente especificada.

Supongamos que tenemos n observaciones en una muestra aleatoria clasificados en $k > 1$ clases disjuntas. Para cada $j \in \{1, \dots, k\}$, denotemos como O_j al número de observaciones de cada clase y como e_j al número de observaciones esperadas (suponiendo que la variable aleatoria X sigue la distribución F_0), calculada como

Figura 8: Pearson introduce el test de la χ^2 .

$e_j = np_j$, con p_j la probabilidad de caer en la j -ésima clase. El estadístico de Pearson viene dado por

$$\chi^2 = \sum_{j=1}^k \frac{(O_j - e_j)^2}{e_j},$$

que sabemos que asintóticamente se comporta como una χ^2 con $k - 1$ grados de libertad. Así, bajo un nivel de significación α , si $\chi^2 \geq \chi_{k-1, \alpha}^2$ se rechaza la hipótesis de que nuestra muestra provenga de $F_0(x)$.

Mann y Wald presentaron en 1942 [27] una técnica para decidir el número óptimo de intervalos para aplicar el test de la χ^2 . La idea era elegir los intervalos de tal manera que la probabilidad de rechazar $F_0(x)$ como verdadera distribución, siendo $F_1(x)$ la distribución verdadera, nunca fuera menor que 0.5.

3.3.2. EL TEST DE KOLMOGOROV-SMIRNOV

Supongamos que tenemos n observaciones independientes e idénticamente distribuidas x_1, \dots, x_n . El *estadístico de Kolmogorov-Smirnov* (en ocasiones abreviado como K-S) para una función de distribución dada viene dado por

$$D = \sup_x |F_n(x) - F_0(x)|,$$

donde $F_n(x)$ es la función de distribución empírica, definida como

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(x_i),$$

donde $I_{(-\infty, x]}(x_i)$ es la función indicatriz, igual a 1 si $x_i \leq x$ y a 0 en caso contrario. Dado un nivel de significación α , la hipótesis de que los datos provengan de la distribución $F_0(x)$ se rechaza si

$$D \geq D_{n, \alpha},$$

siendo D_n la distribución de Kolmogorov-Smirnov.

Este test fue propuesto independientemente por los matemáticos rusos Andréi Kolmogorov [22] y Nikolai Smirnov [36] en 1933 y 1939, respectivamente. Mientras que la distribución asintótica del estadístico D fue derivada en ambos trabajos, una primera tabla con los valores de la correspondiente distribución fue presentada por Smirnov en [37]. Destacamos los trabajos de Feller [12] y Doob [10], que simplificaron y unificaron la teoría de ambos. Uno de los primeros autores en hacer una comparación entre el test de la χ^2 de Pearson y el de Kolmogorov-Smirnov fue F. J. Massey en 1951 [29], artículo que ya presentaba un pequeño estudio de Monte Carlo. Algunas de las conclusiones obtenidas por Massey fueron las siguientes:

1. En general no podemos saber la potencia del test de la χ^2 (Mann y Wald [27], por ejemplo, consideraban solo el caso 0.5), mientras que es sencillo calcular una cota inferior de la potencia de D .
2. A diferencia del test de Kolmogorov-Smirnov, el test de la χ^2 depende de agrupar las observaciones, con lo que podemos perder información.
3. El test de Kolmogorov-Smirnov no se puede aplicar directamente a distribuciones discretas, mientras que el de la χ^2 sí.

En el caso en que la hipótesis nula corresponda a toda una familia paramétrica en la que el parámetro no está especificado y hay que estimarlo, el contraste de bondad de ajuste de la χ^2 podría llevarse a cabo mediante la modificación de los grados de libertad de este. Por otra parte, es posible llevar a cabo el contraste de Kolmogorov-Smirnov para una hipótesis nula que corresponda a una familia paramétrica en la que el parámetro no esté especificado y haya que estimarlo mediante técnicas de computación intensiva como el bootstrap, el cual desarrollamos más adelante.

4. LA LEY DE LOS GRANDES NÚMEROS

La inferencia estadística es el conjunto de métodos que permiten inducir, a través de una muestra estadística, el comportamiento de una determinada población. A mayor cantidad de datos, mayor será la precisión de nuestras técnicas. Esta afirmación, que hasta ahora hemos asumido como «natural» o «intuitiva», puede justificarse de manera teórica gracias al siguiente resultado:

TEOREMA 9 (Ley fuerte de los grandes números). *Sea F una distribución de probabilidad en \mathbb{R}^d , A un conjunto medible en \mathbb{R}^d y F_n la distribución empírica asociada a una muestra de tamaño n de F . Entonces,*

$$F_n(A) \xrightarrow[n \rightarrow \infty]{c.s.} F(A),$$

donde *c.s.* denota la convergencia casi segura.

El matemático italiano Gerolamo Cardano (1501–1576) en su manual para juegos de azar *Liber de Ludo Aleae* editado en 1663 [9] (aunque escrito mucho antes), ya destacó, sin pruebas, que «cuantas más partidas se celebran de un juego de azar mejor predice la probabilidad matemática el resultado». Sin embargo, tuvieron que pasar cien años hasta que Jacob Bernoulli (1654–1705) en el *Ars Conjectandi* (1713, [3]) publicara la primera formulación de la ley de los grandes números para una variable aleatoria binaria. Este resultado, al que Bernoulli se refirió como su «teorema dorado» y al que otros muchos conocen como «teorema de Bernoulli», fue denominado por primera vez como «ley de los grandes números» por Poisson en 1837 [33]. Hay versiones variadas de la ley de los grandes números probadas por matemáticos como Khinchin, Kolmogorov o Chebychev, entre otros.

5. TÉCNICAS DE REMUESTREO

Es por tanto conveniente trabajar con grandes cantidades de datos. Sin embargo, esto no es siempre posible, o puede resultar demasiado costoso. Además, cuando tomamos una muestra solo tenemos una estimación única del parámetro de población, con poca idea de la variabilidad o incertidumbre en la estimación. Las técnicas de remuestreo permiten mejorar la precisión de nuestras estimaciones a partir de las muestras originales y sin necesidad de tomar otras. Esencialmente, consisten en tomar muestras de forma repetida del conjunto completo de muestras que tenemos [8].

A continuación detallamos la *metodología bootstrap*, cuya esencia es el principio del *plug-in*, que sustituye la distribución de probabilidad de la población original, que es desconocida, por la distribución empírica asociada a la muestra. En caso de que la cantidad que depende de la distribución empírica no se pueda calcular, se aplica una técnica de remuestreo.

5.1. METODOLOGÍA BOOTSTRAP

Supongamos que tenemos una muestra de tamaño n . La metodología bootstrap se basa en tomar, de esta muestra original, K muestras al azar con reemplazamiento.

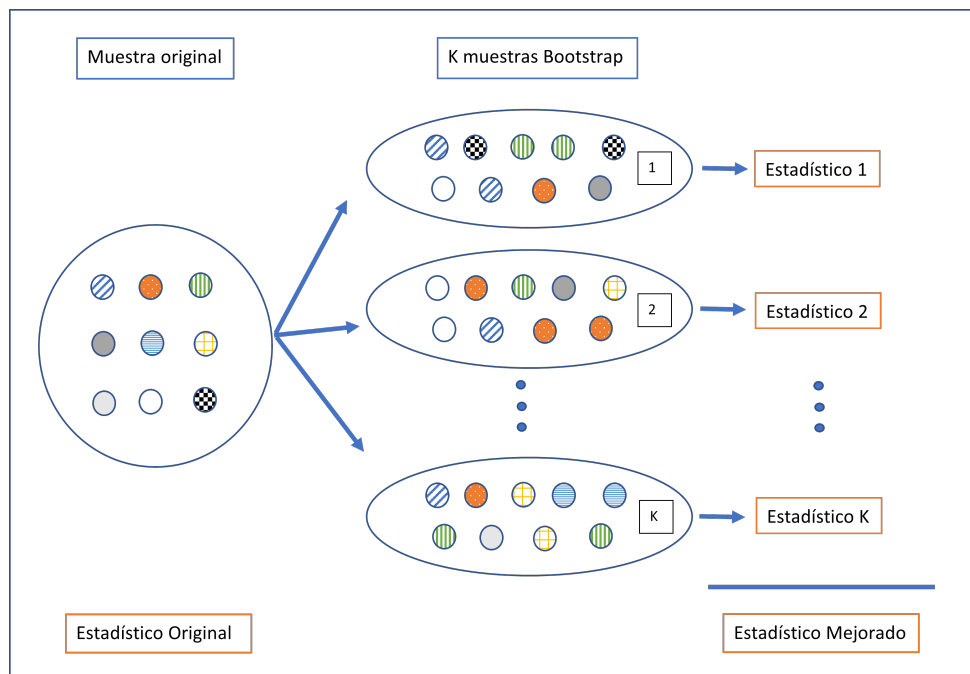


Figura 9: Esquema de la metodología bootstrap.

Con cada muestra se calcula el coeficiente estadístico buscado, y se espera que la media de todos ellos sea similar al coeficiente de la distribución real que generó el conjunto de entrenamiento original (véase la figura 9).

El término bootstrap deriva de la frase «to pull oneself up by one's bootstrap», tomada del libro de 1781 *Las Aventuras del Barón de Munchausen*, de Rudolph E. Raspe, en un momento en el que el barón se había caído en el fondo de un profundo lago y se le ocurre escapar tirando de los cordones de sus propias botas. Se trata de una técnica computacional intensiva, debido a la fuerza del método de Monte Carlo, que es necesario salvo en raras ocasiones. Sin embargo, su implementación no suele ser complicada y, gracias al desarrollo de los ordenadores en la actualidad, no debe causar demasiados problemas computacionales. Esta técnica fue originalmente presentada por el estadístico americano Bradley Efron en 1979 [11], y estudiada y desarrollada en muchas formas por diversos autores, entre los que destacamos a Peter Hall [16, 17, 18]. La metodología bootstrap permite resolver muchos problemas estadísticos difícilmente tratables desde el punto de vista analítico. En particular, si tenemos una muestra grande, se puede usar para mejorar los intervalos de confianza de estadísticos de localización como la media muestral, la mediana o los percentiles muestrales. También es la base de una de las técnicas más conocidas en *machine learning*, el *random forest*, introducido por Leo Breiman en 2001 [5], asimismo autor de los árboles de clasificación y regresión, entre otras técnicas.

6. ALGUNAS APLICACIONES DEL MÉTODO DE MONTE CARLO

6.1. INTEGRACIÓN POR EL MÉTODO DE MONTE CARLO

Supongamos que queremos calcular la integral definida

$$\int_a^b f(x) dx,$$

con $f : [a, b] \rightarrow \mathbb{R}$ integrable en el intervalo (a, b) pero difícil de integrar de forma analítica. Sea $U^* \sim U(a, b)$; se puede ver que

$$\int_a^b f(x) dx = (b - a)E[f(U^*)] \approx \frac{(b - a)}{n} \sum_{i=1}^n f(u_i^*),$$

donde u_1^*, \dots, u_n^* son observaciones independientes de una uniforme en el intervalo (a, b) .

EJEMPLO 10. *Es bien sabido que*

$$\int_0^1 \sqrt{1 - x^2} dx = \frac{\pi}{4}.$$

Vamos a aplicar el método de Monte Carlo para obtener una aproximación a dicho valor. En la figura 10 se pueden observar las distintas aproximaciones que se obtienen al valor de la integral variando los valores de n entre 1 y 10 000.

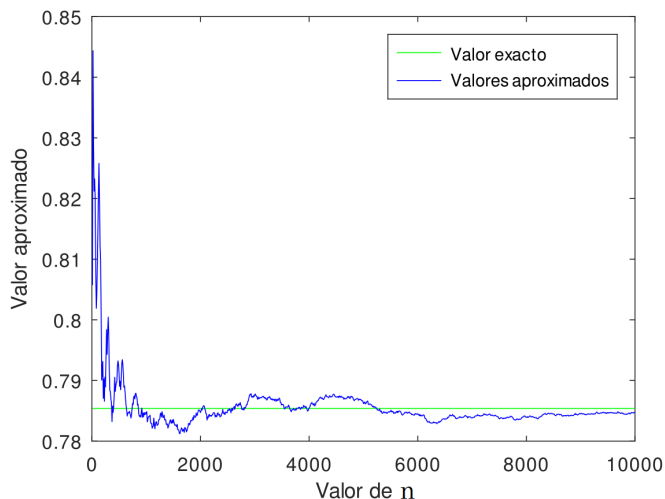


Figura 10: Representación gráfica de las aproximaciones y el valor de $\frac{\pi}{4}$ (ejemplo 10).

Aunque el método de Monte Carlo, frente a otros métodos numéricos, no es muy eficiente para el cálculo de integrales en una variable, cobra importancia en el caso de integrales múltiples. Supongamos que queremos integrar una función de M variables

$$\int_0^1 \cdots \int_0^1 f(x_1, \dots, x_M) dx_1 \cdots dx_M.$$

Utilizamos que, dadas U_1, \dots, U_M uniformes independientes en $(0, 1)$ y dadas $\{u_j^{(i)}\}$, $j \in \{1, \dots, M\}$, $i \in \{1, \dots, n\}$ observaciones de estas uniformes, se tiene

$$\int_0^1 \cdots \int_0^1 f(x_1, \dots, x_M) dx_1 \cdots dx_M = E[f(U_1, \dots, U_M)] \approx \frac{1}{n} \sum_{i=1}^n f(u_1^{(i)}, \dots, u_M^{(i)}).$$

6.2. APROXIMACIÓN DE SOLUCIONES DE ECUACIONES DIFERENCIALES

Una ecuación diferencial de primer orden es una expresión como sigue:

$$\frac{dy}{dx} = f(x, y(x)),$$

donde $x \in [a, b] \subset \mathbb{R}$ y f puede ser una función no lineal y «complicada». Añadiendo a nuestra ecuación una condición inicial $y(x_0 = a) = y_0$ obtenemos un problema de valor inicial. Asumiendo que f tiene ciertas propiedades (es continua en su dominio y verifica la condición de Lipschitz para y) se puede demostrar que el problema de valor inicial planteado tiene existencia y unicidad de solución. A este resultado se le conoce como teorema de Picard-Lindelöf y debe su nombre al francés Charles Émile Picard y al finlandés Ernst Leonard Lindelöf, quien generalizó la teoría de Picard tras su muerte [26]. Nótese que, de forma general, un problema de valor inicial siempre se puede convertir en una ecuación integral

$$y(x) = y(x_0) + \int_{x_0}^x f(\tau, y(\tau)) d\tau.$$

Si $a = x_0 < x_1 < \cdots < x_N = b$ es una discretización del intervalo $[a, b]$, tenemos la siguiente expresión:

$$y(x) = y(x_0) + \int_{x_0}^x f(\tau, y(\tau)) d\tau = y(x_0) + \sum_{i=1}^N \int_{x_{i-1}}^{x_i} f(\tau, y(\tau)) d\tau.$$

Vamos a restringir nuestra atención a ecuaciones diferenciales de primer orden, lineales, independientes de y ; esto es, $\frac{dy}{dx} = f(x)$ con $f(x_0) = y_0$. Usando la aplicación vista en la anterior sección para aproximar integrales, obtenemos

$$y(x) = y(x_0) + \sum_{i=1}^N (x_i - x_{i-1}) E[f(U_i)] \approx y(x_0) + \sum_{i=1}^N \left(\frac{x_i - x_{i-1}}{K} \sum_{j=1}^K f(u_j^{(i)}) \right),$$

donde $U_i \sim U(x_{i-1}, x_i)$ y $u_j^{(i)}$ son observaciones independientes de U_i con $i \in \{1, \dots, N\}$ y $j \in \{1, \dots, K\}$, siendo K el número de muestras que tomamos en cada intervalo. Este método se puede convertir en un proceso iterativo pues a partir de la anterior fórmula de aproximación se obtiene

$$y(x_i) = y(x_{i-1}) + \frac{x_i - x_{i-1}}{K} \sum_{j=1}^K f(u_j^{(i)}).$$

Una de las ventajas de este método es su sencilla implementación en lenguajes de programación. Además, se obtiene una aproximación a la solución de una ecuación diferencial usando solamente muestras de una uniforme y evaluándolas en f .

El mismo proceso puede adaptarse para aproximar soluciones en ecuaciones diferenciales de primer orden más generales. Por ejemplo, si consideramos $\frac{dy}{dx} = f(x, y(x))$ con condición inicial $y(x_0) = y_0$, si $a = x_0 < x_1 < \dots < x_N = b$ es una discretización del intervalo $[a, b]$, entonces

$$\begin{aligned} y(x) &= y(x_0) + \sum_{i=1}^N (x_i - x_{i-1}) E[f(U_i, y(U_i))] \\ &\approx y(x_0) + \sum_{i=1}^N \left(\frac{x_i - x_{i-1}}{K} \sum_{j=1}^K f(u_j^{(i)}, y(u_j^{(i)})) \right), \end{aligned}$$

y se puede obtener el proceso iterativo

$$y(x_i) = y_{i-1} + \frac{x_i - x_{i-1}}{K} \sum_{j=1}^K f(u_j^{(i)}, y_{i-1}),$$

donde $y_i \approx y(x_i)$. Nótese que en la fórmula del proceso iterativo se está reemplazando el valor $y(u_k^{(i)})$ por y_{i-1} , esto se debe a cuestiones computacionales. En [40] se pueden encontrar más detalles.

EJEMPLO 11. *Consideremos el siguiente problema de valor inicial:*

$$\begin{cases} y'(x) = y(x) + x, & x \in [0, 1], \\ y(0) = 1. \end{cases}$$

La solución exacta a este problema es $y(x) = 2e^x - x - 1$. Podemos obtener una aproximación a esa solución usando el método desarrollado, tomando la discretización de $[0, 1]$ dada por $x_i = \frac{i}{10}$ con $i = 0, \dots, 10$ y $K = 50$. En la figura 11 pueden verse tanto la representación de la solución analítica como la de su aproximación.

6.3. REDUCCIÓN DE VARIANZA MEDIANTE VARIABLES ANTITÉTICAS

Esta técnica fue introducida por Hammersley y Morton en 1956 [20]. Supongamos que queremos aproximar $E[X]$ con $\text{Var}[X] = \sigma^2$. Sean $Y \sim Z \sim X$; dada una muestra

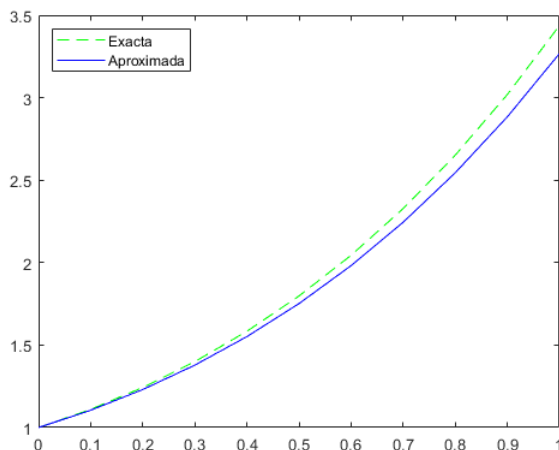


Figura 11: Representación de la solución analítica de la ecuación diferencial y su aproximación mediante el método de Monte Carlo (ejemplo 11).

aleatoria simple de tamaño n se tiene que

$$\begin{aligned} \text{Var} \left[\frac{\bar{Y} + \bar{Z}}{2} \right] &= \frac{1}{4} (\text{Var} [\bar{Y}] + \text{Var} [\bar{Z}] + 2 \text{Cov}[\bar{Y}, \bar{Z}]) \\ &= \frac{\sigma^2}{2n} + \frac{1}{2n} \text{Cov}[Y, Z] \\ &= \frac{\sigma^2}{2n} (1 + \text{Cov}[Y, Z]). \end{aligned}$$

Con este método se obtiene la misma varianza que si hubiéramos empleado $2n$ muestras de la variable original, pero con una reducción de $-100(1 + \rho(Y, Z)) \%$. Nótese que, para variables uniformes, si tomamos $Y \sim U(0, 1)$ y $Z = 1 - Y$, entonces Z también será uniforme en el intervalo $(0, 1)$ y $\text{Cov}(Y, Z) \leq 0$, por lo que el método reduce la varianza original.

EJEMPLO 12. *Supongamos que queremos aproximar*

$$\int_0^1 f(x) dx.$$

Podemos emplear el método de las variables antitéticas,

$$\int_0^1 f(x) dx = E \left[\frac{f(U) + f(1 - U)}{2} \right] \approx \frac{1}{2n} \sum_{i=1}^n (f(u_i) + f(1 - u_i)),$$

donde u_i , $i \in \{1, \dots, n\}$ son observaciones de una uniforme $U \sim U(0, 1)$. Este método, además, solo necesita de la generación de n observaciones.

AGRADECIMIENTOS. Los autores quieren expresar su agradecimiento al revisor anónimo por su profunda revisión del artículo y por sus comentarios, que han ayudado sin duda a mejorarlo.

REFERENCIAS

- [1] G. B. AIRY, *On the algebraical and numerical theory of errors of observations and the combination of observations*, Macmillan, London, 1861.
- [2] L. BADGER, Lazzarini's lucky approximation of π , *Math. Mag.* **67** (1994), no. 2, 83–91.
- [3] J. BERNOULLI, *Ars conjectandi: opus posthumum: accedit Tractatus de seriebus infinitis; et Epistola gallice scripta de ludo pilae reticularis*, Impensis Thurnisiorum, Fratrum, 1713.
- [4] J. BOLTE, Linear Congruential Generators, 2011, <http://demonstrations.wolfram.com/LinearCongruentialGenerators/>
- [5] L. BREIMAN, Random forests, *Machine Learning* **45** (2001), no. 1, 5–32.
- [6] G. W. BROWN, *History of RAND's Random Digits: Summary*, RAND Corporation, Santa Monica, California, 1949.
- [7] G. BUFFON, *Essai d'arithmétique morale*, Histoire naturelle, générale et particulière, Supplément, tome quatrième, 46–148, Imprimerie Royale, Paris, 1777.
- [8] R. CAO Y R. FERNÁNDEZ, *Técnicas de Remuestreo*, 2022, https://rubenfcasal.github.io/book_remuestreo/book_remuestreo.pdf.
- [9] G. CARDANO, *Liber de ludo Aleae*, Opera Omnia, ed. facsímil, F. Frommann Verlag, republicación, Stuttgart, 1966.
- [10] J. L. DOOB, Heuristic approach to the Kolmogorov-Smirnov theorems, *Ann. Math. Statistics* **20** (1949), 393–403.
- [11] B. EFRON Bootstrap methods: another look at the jackknife, *Ann. Statist.* **7** (1979), no. 1, 1–26.
- [12] W. FELLER, On the Kolmogorov-Smirnov limit theorems for empirical distributions, *Ann. Math. Statistics* **19** (1948), 177–189.
- [13] R. A. FISHER Y F. YATES, *Statistical tables for biological, agricultural and medical research*, Oliver and Boyd, London, 1938.
- [14] G. S. FISHMAN Y L. R. MOORE, A statistical evaluation of multiplicative congruential random number generators with modulus $2^{31} - 1$, *J. Amer. Statist. Assoc.* **77** (1982), no. 377, 129–136.
- [15] H. H. GOLDSTINE Y A. GOLDSTINE, The electronic numerical integrator and computer (ENIAC), *Math. Tables Aids Comput.* **2** (1946), no. 15, 97–110.
- [16] P. HALL, On the bootstrap and confidence intervals, *Ann. Statist.* **14** (1986), 1431–1452.
- [17] P. HALL, On efficient bootstrap simulation, *Biometrika* **76** (1989), no. 3, 613–617.

- [18] P. HALL, Methodology and Theory for the Bootstrap, *Handbook of Econometrics* **4** (1994), 2341–2381.
- [19] J. H. HALTON, A retrospective and prospective survey of the Monte Carlo method, *SIAM Rev.* **12** (1970), no. 1, 1–63.
- [20] J. HAMMERSLEY Y K. MORTON, A new Monte Carlo technique: antithetic variates, *Proc. Cambridge Philos. Soc.* **52** (1956), 449–475.
- [21] T. E. HULL Y A. R. DOBELL, Random number generators, *SIAM Rev.* **4** (1962), no. 3, 230–254.
- [22] A. KOLMOGOROV, Sulla determinazione empirica di una legge di distribuzione, *Giorn. Inst. Ital. Attuari* **4** (1933), 83–91.
- [23] P. S. LAPLACE, *Théorie analytique des probabilités*, V. Courcier, Paris, 1812.
- [24] M. LAZZARINI, Un'applicazione del calcolo della probabilità alla ricerca sperimentale di un valore approssimato di π , *Periodico di Matematica per l'Insegnamento Secondario* **4** (1901), 140–143.
- [25] D. H. LEHMER, Mathematical Methods in Large-scale Computing Units, *Proceedings of the Second Symposium on Large Scale Digital Computing Machinery* (1949), 141–146, Harvard University Press, Cambridge, Mass, 1951.
- [26] E. LINDELÖF, Sur l'application de la méthode des approximations successives aux équations différentielles ordinaires du premier ordre, *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **116** (1894), 454–457.
- [27] H. B. MANN Y A. WALD, On the choice of the number of class intervals in the application of the chi-square test, *Ann. Math. Statistics* **13** (1942), 306–317.
- [28] G. MARSAGLIA, Random numbers fall mainly in the planes, *Proc. Nat. Acad. Sci. U.S.A.* **61** (1968), 25–28.
- [29] F. J. MASSEY JR, The Kolmogorov-Smirnov test for goodness of fit, *J. Amer. Statist. Assoc.* **46** (1951), no. 253, 68–78.
- [30] N. METROPOLIS Y S. ULAM, The Monte Carlo method, *J. Amer. Statist. Assoc.* **44** (1949), no. 247, 335–341.
- [31] J. VON NEUMANN, Various techniques used in connection with random digits, *J. Res. Nat. Bur. Stand. Appl. Math. Series* **12** (1951), 36–38.
- [32] K. PEARSON, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine Series 5* **50** (1900), no. 302, 157–175.
- [33] S. D. POISSON, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile: précédées des règles générales du calcul des probabilités*, Bachelier, Paris, 1837.
- [34] B. D. RIPLEY, *Stochastic simulation*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, 1987.
- [35] S. S. SAWIŁOWSKY, You think you've got trivials?, *J. Mod. Appl. Stat. Methods* **2** (2003), no. 1, 218–225.

- [36] N. SMIRNOV, Sur les écarts de la courbe de distribution empirique, *Rec. Math. N.S. (Mat. Sbornik)* **6** (1939), 3–26.
- [37] N. SMIRNOV, Table for estimating the goodness of fit of empirical distributions, *Ann. Math. Statistics* **19** (1948), 279–281.
- [38] L. H. C. TIPPET, *Random sampling numbers*, Tracts for Computers **15**, Cambridge University Press, London, 1927.
- [39] A. WALD Y J. WOLFOWITZ, On a test whether two samples are from the same population, *Ann. Math. Statistics* **11** (1940), no. 2, 147–162.
- [40] W. ZHONG Y T. ZHOU, Solving initial value problem of ordinary differential equations by Monte Carlo method, *International Conference on Multimedia Technology* (Hangzhou, 2011), 2577–2579, IEEE, 2011.

ELENA CASTILLA, DPTO. DE MATEMÁTICA APLICADA, CIENCIA E INGENIERÍA DE LOS MATERIALES Y TECNOLOGÍA ELECTRÓNICA, ESCET, UNIVERSIDAD REY JUAN CARLOS, MÓSTOLES (MADRID)

Correo electrónico: elena.castilla@urjc.es

Página web: <https://sites.google.com/view/elenacastillagonzlez/>

PEDRO J. CHOCANO, DPTO. DE MATEMÁTICA APLICADA, CIENCIA E INGENIERÍA DE LOS MATERIALES Y TECNOLOGÍA ELECTRÓNICA, ESCET, UNIVERSIDAD REY JUAN CARLOS, MÓSTOLES (MADRID)

Correo electrónico: pedro.chocano@urjc.es

Página web: <https://sites.google.com/view/pedrojchocanofeito/>