

CAPÍTULO 4

NÚMEROS ALEATORIOS E INTEGRACIÓN MONTE-CARLO

Bruno Julia Diaz, última revisión Oct 2015

Al tener a nuestra disposición ordenadores de gran potencia y con capacidad para realizar ingentes cantidades de operaciones por segundo, se han extendido una gran cantidad de métodos numéricos que se basan en la generación de variables aleatorias que exploren sus respectivos dominios de definición permitiendo desde realizar integrales numéricas hasta buscar soluciones a ecuaciones diferenciales o simular la solución de problemas complejos.

Los métodos más conocidos son los denominados métodos de Montecarlo, denominados así por la ciudad de los casinos, y entre los primeros algoritmos relevantes tenemos el de Metropolis¹, que se diseñó para poder generar números aleatorios con una determinada distribución de probabilidad.

4.1 Variables aleatorias

La esencia de estos métodos son las variables aleatorias. En los textos de estadística y probabilidad se definen las variables aleatorias como aquellas cuyo valor esta sujeto a variaciones debidas al azar. Normalmente se definen, en el sentido matemático, mediante, por un lado, el conjunto de valores que pueden tomar (denominado soporte) y, por otro, mediante una función probabilidad de tomar dichos valores. En física el comportamiento aleatorio de las variables puede ser debido a un conjunto de factores:

- La naturaleza intrínsecamente probabilística del fenómeno, como ocurre en física cuántica, por ejemplo, el instante en que un núcleo radioactivo se desintegra, el valor de la posición de

¹Denominado así por Nicholas Metropolis, uno de los científicos del laboratorio nacional de Los Alamos en la época del proyecto Manhattan, que fue uno de los primeros grandes proyectos en requerir cálculo numérico intensivo.

un electrón en un momento determinado, etc.

- Ser el resultado de un número de factores externos incontrolado, por ejemplo, cara o cruz de una moneda, resultado de un lanzar un dado, posición de caída de las gotas de lluvia, volumen del ruido de fondo, etc.

En particular en cualquier medida experimental en la que haya un número suficientemente alto de factores externos que no controlemos el resultado de cada medida será aleatorio, requiriendo múltiples medidas para poder determinar el observable con precisión.

Es importante observar que repitiendo un determinado experimento un gran número de veces obtendremos una serie de resultados que normalmente estarán razonablemente concentrados alrededor de un valor medio. Tendremos después medidas más o menos exóticas, muy alejadas de la media, cuya significación estadística es esencialmente nula (no contribuyen al promedio) pero que puede que sean de extraordinaria relevancia. Un ejemplo interesante es el del Prof. Rutherford que bombardeó con partículas alfa una lámina de oro obteniendo una mancha de partículas alfa que parecían no encontrar nada en su camino ². La materia estaba realmente vacía, pero, aparte de la mayoría, observó varios eventos realmente extraordinarios, uno de cada mil, en los que la partícula alfa rebotaba como si chocase contra una pared. En un análisis estadístico habitual estos puntos serían irrelevantes, sin embargo, desde el punto de vista físico eran los más importantes al demostrar la presencia de algo muy pesado en el interior de los átomos: el núcleo atómico.

También es bueno al comenzar a hablar de variables aleatorias o estocásticas, pensar en la dificultad intrínseca que tenemos los humanos para entender la aleatoriedad. Esto se pone por ejemplo de manifiesto si pedimos a un amigo que nos escriba una cadena de cincuenta valores cara-cruz al azar, como si fueran resultado de lanzar una moneda al aire. La cadena escrita casi nunca pasa criterios mínimos de aleatoriedad, como son por ejemplo que en 50 tiradas tiene que haber con alta probabilidad alguna cadena de 4 caras o cruces seguidas.

Las variables aleatorias pueden ser discretas, como el resultado de la suma de dos dados, o continuas, como el valor de la posición, la energía cinética, etc. Ambos tipos se distinguen pues según sea la estructura del soporte, conjunto de valores que puede tomar, de la variable aleatoria.

Si el soporte es un conjunto numerable de valores, x_k , con $k = 1, \dots$, tendremos una variable discreta. Se definen en este caso un conjunto numerable de probabilidades, P_k , tales que,

$$P_k \equiv \lim_{N \rightarrow \infty} \frac{\text{Numero de veces que se obtiene el resultado } x_k}{\text{Numero total de medidas}} \quad (4.1)$$

En el caso de variables continuas se define la densidad de probabilidad, o función de probabilidad de la variable aleatoria x , $p(x)$,

$$P\{x|x_1 < x < x_1 + dx\} \equiv \lim_{N \rightarrow \infty} \frac{\text{Numero de veces que se obtiene el resultado en } [x_1, x_1 + dx]}{\text{Numero total de medidas}} \quad (4.2)$$

Con lo que,

$$P\{x|x_1 < x < x_1 + dx\} \equiv p(x_1)dx. \quad (4.3)$$

Como antes, $p(x) > 0$ y $\int_S p(x)dx = 1$.

4.1.1 Funciones de variables aleatorias

Dada x variable aleatoria con densidad de probabilidad $p(x)$ tendremos que $y = f(x)$ también será una variable aleatoria. La función de probabilidad de la variable y , $g(y)$, se puede obtener igualando,

$$p(x)dx = g(y)dy \quad (4.4)$$

²Más detalles, <https://www.aip.org/history/exhibits/rutherford/sections/alpha-particles-atom.html>

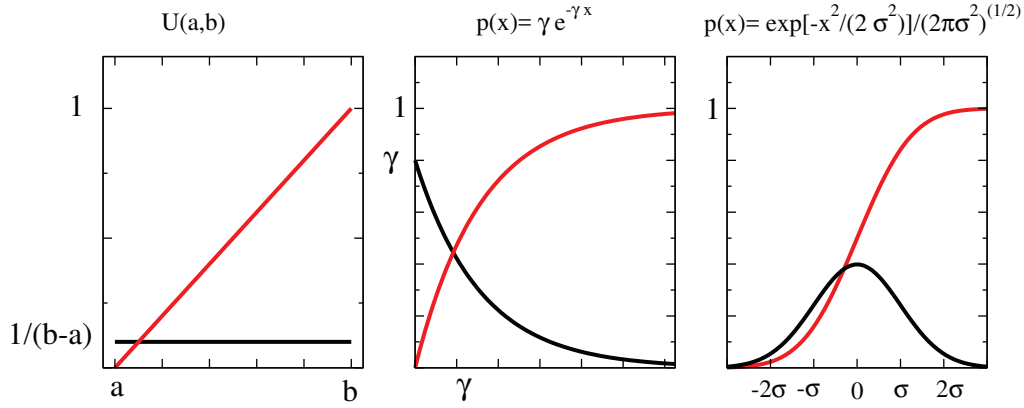


Figura 4.1: Ejemplos de distribuciones de probabilidad: uniforme (izquierda), exponencial (centro) y normal (derecha), junto con sus probabilidades acumuladas.

siempre que la función sea biunívoca (si no fuese así, habría que sumar varias contribuciones). En este caso, tendremos,

$$g(y) = p(x) \left| \frac{dx}{dy} \right| \quad (4.5)$$

que se puede escribir como,

$$g(y) = p(f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right|. \quad (4.6)$$

4.1.2 Probabilidad acumulada

Dada una variable aleatoria x con función densidad de probabilidad $p(x)$, se define la función de probabilidad acumulada como,

$$\mathcal{P}(x) \equiv \int_{x_{\min}}^x p(x') dx' \quad (4.7)$$

de donde,

$$\frac{d\mathcal{P}(x)}{dx} = p(x). \quad (4.8)$$

Nótese que $\mathcal{P}(x)$ es una función no decreciente, debido a que $p(x) \geq 0$ i que varia entre $\mathcal{P}(x_{\min}) = 0$ y $\mathcal{P}(x_{\max}) = 1$. Además, se tiene que la probabilidad de obtener un resultado entre dos valores dados,

$$P(x|a < x < b) = \int_a^b p(x) dx = \mathcal{P}(b) - \mathcal{P}(a). \quad (4.9)$$

Las variables discretas pueden tratarse utilizando estas fórmulas sin más que definir las mediante deltas de Dirac,

$$p(x) = \sum_k p_k \delta(x - x_k) \quad (4.10)$$

que nos lleva a la función probabilidad acumulada correspondiente,

$$\mathcal{P}(x) = \int_{-\infty}^x p(x') dx' = \sum_{x_k < x} p_k. \quad (4.11)$$

4.1.3 Valores esperados y momentos de la distribución

Dada una variable aleatoria x con densidad de probabilidad $p(x)$ se define el promedio de la función $g(x)$ como,

$$E(g(x)) = \langle g(x) \rangle = \int_S p(x) g(x) dx \quad (4.12)$$

donde S es el soporte de la variable x . En el caso de $g(x) = x$, tendremos el promedio o valor esperado de la propia variable x ,

$$E(x) = \langle x \rangle = \int_S x p(x) dx. \quad (4.13)$$

Para variables discretas tenemos las definiciones equivalentes,

$$E(g(x)) = \langle g(x) \rangle = \sum_k p_k g(x_k). \quad (4.14)$$

Los valores esperados satisfacen varias propiedades sencillas,

$$\begin{aligned} \langle f(x) + a \rangle &= \langle f(x) \rangle + a \quad \forall a \in R \\ \langle a f(x) \rangle &= a \langle f(x) \rangle \quad \forall a \in R \\ \langle f(x) + g(x) \rangle &= \langle f(x) \rangle + \langle g(x) \rangle. \end{aligned} \quad (4.15)$$

Se definen finalmente todos los momentos de la distribución de probabilidad,

$$m_n \equiv \langle x^n \rangle = \int_S x^n p(x) dx. \quad (4.16)$$

Para caracterizar mínimamente la distribución de probabilidad se calcula su varianza, $\text{Var}(x)$, que viene dada por,

$$\text{Var}(x) = \langle (x - \langle x \rangle)^2 \rangle = \int_S (x - \langle x \rangle)^2 p(x) dx = m_2 - m_1^2. \quad (4.17)$$

De la propia definición se observa que la varianza da una medida de cuan alejados están los diferentes valores posibles de la variable x de su valor medio, $\langle x \rangle$.

Así, por ejemplo, la varianza sólo puede ser cero si el único valor posible de la variable es el valor medio. La desviación estándar no es más que la raíz cuadrada de la varianza, $\sigma_x = \sqrt{\text{Var}(x)}$ y mide la anchura de la distribución. Nótese que la variable adimensional normalizada, $x^* = \frac{(x - \langle x \rangle)}{\sigma_x}$ tiene media igual a 0 y desviación estándar igual a 1.

Si antes hemos definido los momentos de la distribución, se pueden definir los llamados momentos centrales de la distribución,

$$\mu_n = \langle (x - \langle x \rangle)^n \rangle. \quad (4.18)$$

4.2 Ejemplos y obtención de distribuciones

4.2.1 Ejemplos de distribuciones discretas

Distribución de Bernoulli

Consideramos una variable aleatoria i , cuyo soporte es $S = \{0, 1\}$. Esto es, sólo puede tomar dos valores, con probabilidad de $q = 1 - p$ y p , respectivamente. Sus valores esperados pueden calcularse,

$$m_n = \langle i^n \rangle = \sum_{i=0}^1 p_i i^n = 0^n \times (1 - p) + 1^n \times p = p. \quad (4.19)$$

La varianza resulta, $\text{Var}(i) = \langle (i - \langle i \rangle)^2 \rangle = m_2 - m_1^2 = p(1 - p) = pq$. Normalmente al suceso con valor 1 se le denomina éxito, mientras que al de 0 se le denomina fracaso.

Ejemplo 4.1: Un ejemplo típico sería el resultado de lanzar una moneda al aire, donde en este caso tendríamos $p = 1 - p$ de donde, $p = 1/2$. Existen múltiples ejemplos, uno sería el resultado de preguntar si un individuo mide más de 150cm, con si/no correspondientes a 0 y 1. En este caso, $p \neq 1/2$. Otro sería el éxito/fracaso de un tratamiento médico, y un largo etc.

Ejemplo 4.2: Considera la variable ¿Es el resultado de lanzar un dado el valor 5? La variable tiene como soporte $S = \{0, 1\}$ (0=no, 1=si). El valor de $p = 1/6$.

Distribución binomial

Consideremos una variable aleatoria que sea el resultado de obtener k éxitos (valor 1) en n experimentos de Bernoulli idénticos e independientes. O, lo que es lo mismo, la probabilidad de que la suma de dichos n experimentos de Bernoulli sea k . El soporte de la variable k es $\{0, 1, \dots, n\}$. En este caso, la distribución de la variable k viene dada por,

$$B(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k}. \quad (4.20)$$

Comprobemos la normalización,

$$\sum_{k=0}^n B(k; n, p) = (p + 1 - p)^n = 1. \quad (4.21)$$

El valor medio, varianza y desviación estándar resultan,

$$\begin{aligned} \langle k \rangle &= \sum_{k=0}^n k B(k; n, p) = np \\ \langle k^2 \rangle &= \sum_{k=0}^n k^2 B(k; n, p) = np(1 + (n-1)p) \\ \text{Var}(k) &= np(1 - p). \end{aligned} \quad (4.22)$$

Ejemplo 4.3: La distribución binomial aparece en múltiples contextos y tiene aplicación siempre que contemos el número de éxitos de un conjunto de eventos independientes en los que la probabilidad de éxito/fracaso es la misma.

Distribución de Poisson

La distribución de Poisson se puede obtener como un límite de la distribución Binomial. Tomando el límite $n \rightarrow \infty$, que hace que el soporte de k sea ahora $\{0, 1, \dots\}$, pero manteniendo el valor

medio $\lambda \equiv np$ constante. Tomando el límite en la Eq. (4.20), tenemos,

$$\begin{aligned}
 P(k; \lambda) &= \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \lim_{n \rightarrow \infty} \frac{n!}{(n-k)! k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &= \lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} \frac{\lambda^k}{k!} \left[\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda} \right]^{-\lambda} \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &= e^{-\lambda} \frac{\lambda^k}{k!}.
 \end{aligned} \tag{4.23}$$

Los momentos bajos son,

$$\begin{aligned}
 \langle k \rangle &= \lambda \\
 \langle k^2 \rangle &= \lambda(\lambda + 1) \\
 \text{Var}(k) &= \lambda
 \end{aligned} \tag{4.24}$$

Describe procesos que involucran muchos procesos elementales con probabilidad constante y que no dependen de la historia pasada,

- Desexcitaciones por segundo de una muestra radioactiva
- Número de gotas de lluvia en una región acotada
- Número de emails recibidos por día
- La suma de dos variables de Poisson con λ_1 y λ_2 sigue una distribución de Poisson con $\lambda = \lambda_1 + \lambda_2$.

4.2.2 Ejemplos de distribuciones continuas

Normalmente el ordenador proporciona números pseudo-aleatorios distribuidos uniformemente en el intervalo $(0, 1)$.

4.2.3 Método del cambio de variable

Así un problema habitual es generar números aleatorios con una distribución distinta, esto es, dada una variable $y \in (a, b)$ con distribución $g(y)$ buscar el cambio de variables $y = f(x)$ tal que la variable x esté distribuida uniformemente y la y resulte distribuida según $g(y)$. Como vimos en la sección 4.1.1, tenemos,

$$g(y) = p(f^{-1}(y)) \left| \frac{df^{-1}(y)}{dy} \right|. \tag{4.25}$$

Ahora fijemos la variable x como distribuida uniformemente en $(0, 1)$, esto es $p(x) = 1$, con lo que $f(0) = a$ y $f(1) = b$, de donde,

$$g(y) = \left| \frac{df^{-1}(y)}{dy} \right| \quad \text{de donde } \mathcal{P}(y) = \int_a^y g(y') dy' = f^{-1}(y) = x. \tag{4.26}$$

Esta última expresión nos proporciona un procedimiento: dada una función de distribución, $g(y)$, calculamos su distribución acumulada, $\mathcal{P}(y)$, esta última resulta ser la inversa del cambio de variable que necesitamos hacer.

Distribución uniforme

Para el caso de una **variable continua** $y \in [a, b]$, la distribución uniforme $U(a, b)$ se escribe, $p(y) = 1/(b - a)$. Su distribución acumulada es,

$$\mathcal{P}(y) = \int_a^y p(y') dy' = \frac{y - a}{b - a}. \quad (4.27)$$

El valor medio de la variable y es, $\langle y \rangle = \int_a^b p(y) dy = (a + b)/2$ y su varianza $\text{Var}(y) = \langle y^2 \rangle - \langle y \rangle^2 = (a - b)^2/12$ de donde la desviación estándar es $\sigma_y = (b - a)/\sqrt{12}$.

En este caso, si partimos de una variable x originalmente distribuida uniformemente en el intervalo $(0, 1)$, necesitamos invertir la función acumulativa (4.27),

$$x = \frac{y - a}{b - a} \Rightarrow y = (b - a)x + a \quad \text{con } x \in U(0, 1). \quad (4.28)$$

En el caso de **variable discreta** i consideramos que puede tomar los valores $\{1, 2, \dots, n\}$ con equiprobabilidad $p_i = 1/n$. La probabilidad está normalizada, $\sum_i p_i = 1$. El valor medio, varianza y desviación estándar resultan,

$$\begin{aligned} \langle i \rangle &= \sum_{i=1}^n p_i i = \frac{1}{n} \frac{n + 1}{2} = \frac{n + 1}{2} \\ \langle i^2 \rangle &= \frac{(n + 1)(2n + 1)}{6} \\ \text{Var}(i) = \langle i^2 \rangle - \langle i \rangle^2 &= \frac{n^2 - 1}{12} \\ \sigma_i &= \sqrt{\frac{n^2 - 1}{12}}. \end{aligned} \quad (4.29)$$

Distribución exponencial

Esta distribución aparece de modo natural en física cuántica al estudiar el decaimiento de núcleos radioactivos. Para una variable $y \in (0, \infty)$ se escribe,

$$p(y) = \gamma e^{-\gamma y} \quad \text{con } \mathcal{P}(y) = \int_0^y p(y') dy' = 1 - e^{-\gamma y}. \quad (4.30)$$

Invirtiendo la distribución acumulada tenemos,

$$y = -\frac{1}{\gamma} \log(x) \quad \text{si } x \in U(0, 1). \quad (4.31)$$

Distribución gaussiana o normal

La distribución de Gauss, con soporte todos los reales, valor medio μ y varianza σ^2 , viene dada por la expresión,

$$N(x, \mu, \sigma) \equiv \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}. \quad (4.32)$$

Es fácil comprobar que está correctamente normalizada y que verifica que el valor esperado y varianza son los que corresponden. Su distribución acumulada es la denominada función error,

$$\text{Err}(x) = \int_{-\infty}^x N(x', \mu, \sigma) dx' \quad (4.33)$$

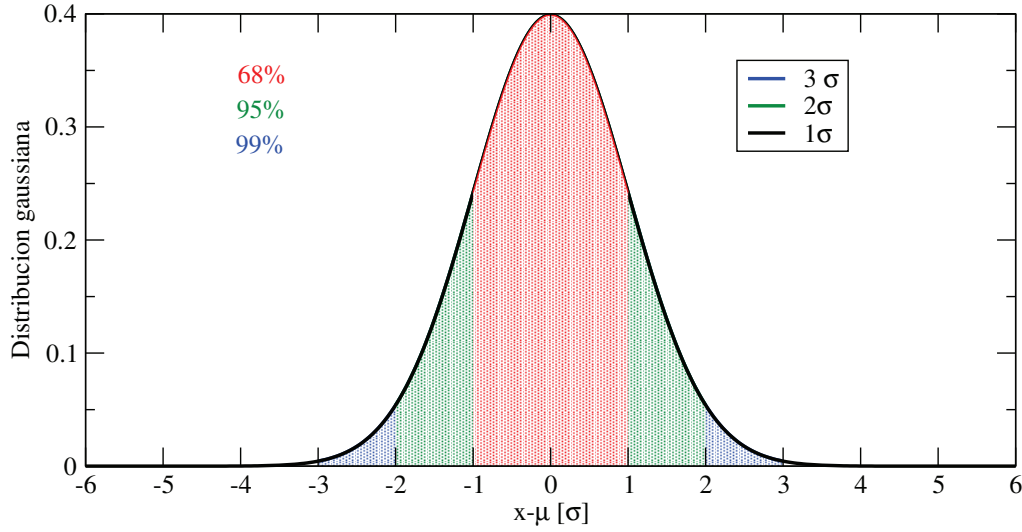


Figura 4.2: Ilustración de la probabilidad de que la variable gaussiana tome un valor comprendido entre $(\mu - k\sigma, \mu + k\sigma)$ con $k = 1, 2$ y 3 .

y no tiene una expresión analítica sencilla.

La distribución de Gauss adquiere una importancia capital al aparecer en el teorema del límite central, ver sección , como función a la que se aproxima el promedio de un número infinito de variables aleatorias independientes.

La distribución gaussiana o normal juega un papel extraordinariamente relevante gracias al teorema del límite central que veremos más adelante. La pregunta habitual es, dada una variable aleatoria que sigue una distribución gaussiana, ¿Cual es la probabilidad de que la variable tome un valor comprendido entre $(\mu - \sigma, \mu + \sigma)$? ¿Y entre $(\mu - 2\sigma, \mu + 2\sigma)$, o $(\mu - 3\sigma, \mu + 3\sigma)$? Veamos,

$$\int_{\mu-k\sigma}^{\mu+k\sigma} dx \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} = \text{Erf}(k/\sqrt{2}) \quad (4.34)$$

Y toma los valores, 0.682689, 0.9545, 0.9973, 0.999937 y 0.9999994, para $k = 1, 2, 3, 4$ y 5 , respectivamente. Normalmente se da el intervalo, $(\mu - \sigma, \mu + \sigma)$ como intervalo de confianza, con un 68% de probabilidad de que la variable tome un valor comprendido en dicho rango. Si queremos una estimación más conservadora del error podemos usar por ejemplo $(\mu - 3\sigma, \mu + 3\sigma)$, que contiene un 99.73% de la probabilidad.

4.3 Múltiples variables

Podemos generalizar las definiciones y conceptos al caso de tener un conjunto de variables aleatorias, x_1, x_2, \dots, x_n , cuya función densidad de probabilidad $p(x_1, x_2, \dots, x_n)$ cuantifica el número de eventos que cumplen, que las variables x_i toman un valor entre $[x_i, x_i + dx_i]$, simultáneamente. Como antes la densidad de probabilidad ha de ser definida positiva y estar normalizada,

$$\int_S dx_1 dx_2 \dots dx_n p(x_1, x_2, \dots, x_n) = 1. \quad (4.35)$$

Análoga a la definición para el caso de una variable tenemos el valor esperado de una función de las n variables,

$$\langle g(x_1, x_2, \dots, x_n) \rangle = \int_S dx_1 dx_2 \dots dx_n g(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n). \quad (4.36)$$

Podemos definir los promedios de cada una de las variables,

$$\langle x_i \rangle = \int_S dx_1 dx_2 \dots dx_n x_i p(x_1, x_2, \dots, x_n). \quad (4.37)$$

Ahora tenemos toda una serie de conceptos que estudian la correlación entre las distintas variables. El primero de ellos es la matriz de covarianza,

$$\text{Cov}(x_i, x_j) = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = \text{Cov}(x_j, x_i) \quad (4.38)$$

Los elementos diagonales de la matriz de covarianza son las covariancias o variancias de las variables,

$$\text{Cov}(x_i, x_i) = \langle x_i^2 \rangle - \langle x_i \rangle^2 \equiv \sigma_i^2. \quad (4.39)$$

A partir de las covarianzas se definen los coeficientes de correlación entre dos variables,

$$\rho_{i,j} = \rho(x_i, x_j) \equiv \frac{\text{Cov}(x_i, x_j)}{\sigma_i \sigma_j} = \frac{\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle}{\sqrt{\langle x_i^2 \rangle - \langle x_i \rangle^2} \sqrt{\langle x_j^2 \rangle - \langle x_j \rangle^2}}. \quad (4.40)$$

A partir de $p(x_1, x_2, \dots, x_n)$ podemos definir la probabilidad marginal de la variable x_i , del siguiente modo,

$$p_i(x_i) = \int_S dx_1 dx_2 dx_{i-1} dx_{i+1} \dots dx_n p(x_1, x_2, \dots, x_n). \quad (4.41)$$

Que es la probabilidad de medir el valor x_i , sin medir, o independientemente del valor de, el resto de variables.

Un concepto fundamental es el de variables independientes, se dice que las variables x_1, x_2, \dots, x_n son independientes si

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n g_i(x_i) \quad (4.42)$$

donde cada densidad de probabilidad $g_i(x_i)$ está normalizada.

En el caso de variables independientes es fácil comprobar que,

$$\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle \quad (4.43)$$

por lo que todas las covarianzas no diagonales son cero,

$$\rho(x_i, x_j) = 0. \quad (4.44)$$

Hay que tener en cuenta que hemos demostrado que si las variables son independientes las covarianzas no diagonales son cero, el recíproco, sin embargo, no es cierto.

4.3.1 Método de Box-Müller

Siendo la distribución de Gauss particularmente importante necesitamos tener métodos para generar números distribuidos con dicha distribución. El método más sencillo es el de Box-Müller. La idea es la siguiente, consideramos dos variables aleatorias x_1 y x_2 que obedecen cada una de ellas una distribución gaussiana de media $\mu = 0$ y desviación estándar $\sigma = 1$, el soporte de ambas es la recta real. Tomemos cada punto (x_1, x_2) como puntos de un plano, en este caso, la probabilidad del punto (x_1, x_2) será, $N(x_1, 0, 1)N(x_2, 0, 1)dx_1dx_2$. Consideremos un cambio de variables, $x_1 = r \cos \phi$ y $x_2 = r \sin \phi$. Con $r \in [0, \infty)$ y $\phi \in [0, 2\pi)$, en estas variables tenemos,

$$P(x_1, x_2)dx_1dx_2 = \frac{1}{2\pi}e^{-\frac{x_1^2+x_2^2}{2}}dx_1dx_2 = \frac{1}{2\pi}e^{-\frac{r^2}{2}}rdrd\phi. \quad (4.45)$$

Que podemos leer como la densidad de probabilidad de dos variables independientes, r y ϕ , con $p(r)dr = e^{-r^2/2}rdr$ y $g(\phi)d\phi = 1/(2\pi)d\phi$, que cumplen,

$$\begin{aligned} \int_0^\infty r e^{-r^2/2} dr &= 1 \\ \int_0^{2\pi} \frac{1}{2\pi} d\phi &= 1. \end{aligned} \quad (4.46)$$

Lo interesante es que es sencillo obtener números distribuidos según ambas distribuciones mediante cambio de variable a partir de variables distribuidas según $U(0, 1)$. Dadas $\xi_1 \in U(0, 1)$ y $\xi_2 \in U(0, 1)$ tenemos,

$$r = \sqrt{-2\log(\xi_1)} \quad \phi = 2\pi\xi_2. \quad (4.47)$$

Donde hemos utilizado que la distribución acumulada de $p(r)$ es,

$$P(r) = \int_0^r p(r')dr' = 1 - e^{-r^2/2} = \xi_1 \Rightarrow r = \sqrt{-2\log(1 - \xi_1)} \quad (4.48)$$

Que normalmente se escribe, $r = \sqrt{-2\log(\xi_1)}$ ya que con $\xi_1 \in U(0, 1)$, $-\log(1 - \xi_1)$ y $-\log(\xi_1)$ producen los mismos números aleatorios.

Este procedimiento permite obtener pares de números distribuidos según la distribución gaussiana $N(x, 0, 1)$, recordemos el cambio $y = \mu + x\sigma$, nos permite obtener números y distribuidos según la gaussiana general $N(y, \mu, \sigma)$.

4.4 Método para estimar la densidad de probabilidad de una variable aleatoria

Supongamos que nos proporcionan un conjunto de N números $\{x_1, \dots, x_N\}$, que se supone están generados independientemente siguiendo una densidad de probabilidad $\rho(x)$. El problema es a partir de los números intentar estimar o reconstruir la densidad original.

Un método habitual es de construir un histograma. El procedimiento se resume del modo siguiente,

- 1) Determinamos los valores máximo y mínimo de $\{x_k\}$, x_M y x_m , respectivamente.
- 2) Construimos una partición dentro del intervalo $[x_m, x_M]$, $z_1 = x_m < z_2 < \dots < z_{NB+1} = x_M$, NB es el número de subintervalos de la partición, que se corresponde con el número de barras del histograma. La anchura del intervalo k -ésimo es $\omega_k = z_{k+1} - z_k$.

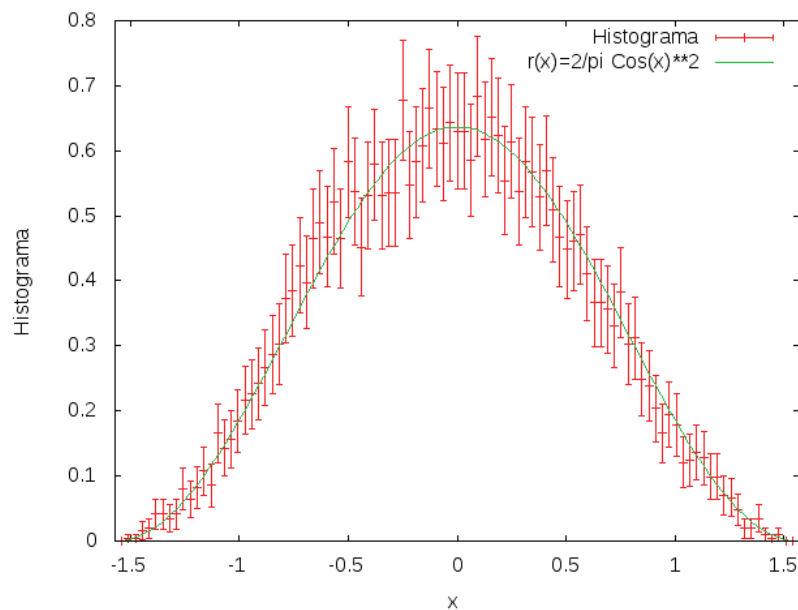


Figura 4.3: Ejemplo de comparación entre la distribución teórica, $\rho(x) = (2/\pi) \cos^2(x)$ y el estimador construido con un histograma de 100 barras y 10000 números generados mediante el método de aceptación y rechazo. Las barras de error se corresponden a $2\sigma_k$, donde N_k ha sido aproximado por una distribución Binomial como se explica en el texto.

- 3) Contamos el número de valores x_j dentro de cada subintervalo k -ésimo, N_k .
- 4) Un estimador de la distribución de probabilidad será,

$$p_N(x) = \sum_{k=1}^{NB} p_k W(x; z_k, z_{k+1}) \quad \text{con } p_k = \frac{N_k}{N\omega_k} \quad (4.49)$$

y W una función rectángulo, definida como $W(x; z_k, z_{k+1}) = 1$ si $z_k < x \leq z_{k+1}$ y cero si x está fuera del subintervalo k -ésimo.

La primera propiedad que debe cumplir el estimador es,

$$\int_S dx p_N(x) = 1. \quad (4.50)$$

Veamos,

$$\int_S dx p_N(x) = \int_S dx \sum_{k=1}^{NB} p_k W(x; z_k, z_{k+1}) = \sum_{k=1}^{NB} p_k \omega_k = \sum_{k=1}^{NB} \frac{N_k}{N\omega_k} \omega_k = \frac{\sum_k N_k}{N} = 1 \quad (4.51)$$

y es siempre un buen test de que todo va razonablemente bien.

El valor de p_k es una aproximación al valor de la distribución original en el intervalo k -ésimo. Curiosamente se puede obtener de forma sencilla una estimación del error en este número. El quid es estudiar la distribución del número de cuentas en un subintervalo, N_k . El que un número aleatorio de la secuencia original caiga o no en el subintervalo k -ésimo es una dicotomía y por

tanto corresponde a un problema de Bernouilli. En el caso de que los números sean realmente aleatorios cada número será independiente del anterior y por tanto la distribución de N_k será una distribución binomial. El problema ahora es el siguiente, ¿Cuál es el valor de p , la probabilidad de éxito? Este lo podemos aproximar razonablemente por N_k/N , y esta aproximación será tanto mejor cuanto mayor sea N para un número fijo de NB.

Así pues el número de cuentas en cada subintervalo sigue una distribución binomial,

$$B(N_k, N, N_k/N). \quad (4.52)$$

Esto nos permite estudiar por ejemplo las propiedades de la variable p_k , altura de la barra, en particular su varianza vendrá dada por,

$$\text{Var}(p_k) = \frac{1}{\omega_k^2 N^2} \text{Var}(N_k) = \frac{1}{\omega_k^2 N^2} N \frac{N_k}{N} \left(1 - \frac{N_k}{N}\right) \quad (4.53)$$

y por tanto la desviación standard será,

$$\sigma = \sqrt{\text{Var}(p_k)} = \frac{1}{\omega_k \sqrt{N}} \sqrt{\frac{N_k}{N} \left(1 - \frac{N_k}{N}\right)}. \quad (4.54)$$

Esta última expresión muestra que para disminuir el error de cada valor del histograma y mejorar nuestro conocimiento de la distribución $\rho(x)$ necesitaremos aumentar el número N , la mejora es lenta va como $1/\sqrt{N}$.

Ejemplo 4.4: Construye un histograma, estimando el error de cada barra como $2\sigma_k$, con 100 barras y 1000 números generados siguiendo la distribución de probabilidad $\rho(x) = (2/\pi) \cos^2(x)$ con $x \in [-\pi/2, \pi/2]$.

4.5 Teorema del límite central

4.5.1 Suma de variables aleatorias

Supongamos que tenemos una suma de dos variables aleatorias, $y = x_1 + x_2$. En este caso, tenemos

$$\text{Cov}(x_1, x_2) = \langle (x_1 - \langle x_1 \rangle)(x_2 - \langle x_2 \rangle) \rangle = \langle x_1 x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle. \quad (4.55)$$

Veamos la varianza de y ,

$$\text{Var}(y) = \langle (x_1 + x_2)^2 \rangle - \langle x_1 + x_2 \rangle^2 = \text{Var}(x_1) + \text{Var}(x_2) + 2\text{Cov}(x_1, x_2). \quad (4.56)$$

Si las dos variables son independientes tenemos que la covarianza mutua es cero por lo que la varianza de la suma es la suma de las varianzas. En general para variables independientes se tiene,

$$\text{Var}\left(\sum_i a_i x_i\right) = \sum_i a_i^2 \text{Var}(x_i). \quad (4.57)$$

En particular para el caso de dos variables independientes la densidad de probabilidad de la variable suma $y = x_1 + x_2$ se puede escribir como,

$$p(y) = \int_{S_1} dx_1 \int_{S_2} dx_2 p_1(x_1) p_2(x_2) \delta(y - x_1 - x_2) = \int_{S_1} dx_1 p_1(x_1) p_2(y - x_1) \quad (4.58)$$

que es la convolución de ambas densidades de probabilidad.

4.5.2 Teorema del límite central

Consideremos una variable aleatoria x distribuida ³ según la función de distribución $f(x)$, y construimos una segunda variable aleatoria z del siguiente modo,

$$z = \frac{x_1 + x_2 + \dots + x_N}{N}. \quad (4.59)$$

Donde x_1, \dots, x_N son N resultados de la variable x , esto es, hemos promediado un grupo de N resultados de la variable aleatoria original.

¿Cuál será la función de distribución de la nueva variable z ? La respuesta a esta pregunta, para cualquier distribución $f(x)$, la proporciona el teorema del límite central. Veamos,

Demostración. La función de probabilidad de la variable z se puede escribir como,

$$g(z) = \int dx_1 dx_2 \dots dx_N f(x_1) f(x_2) \dots f(x_N) \delta \left(\frac{x_1 + x_2 + \dots + x_N}{N} - z \right) \quad (4.60)$$

donde hemos tenido en cuenta que las N variables (o los N resultados) x_1, x_2 , etc, son independientes. Esta integral la podemos atacar con una representación integral de la delta de Dirac,

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixp} dp \quad (4.61)$$

de donde tenemos,

$$\begin{aligned} g(z) &= \int dx_1 dx_2 \dots dx_N f(x_1) f(x_2) \dots f(x_N) \frac{1}{2\pi} \int_{-\infty}^{\infty} dp e^{ip \left(\frac{x_1 + x_2 + \dots + x_N}{N} - z \right)} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp e^{-ip(z-\mu)} \left[\int dx e^{\frac{ip(x-\mu)}{N}} f(x) \right]^N. \end{aligned} \quad (4.62)$$

Donde μ es el valor medio de la distribución $f(x)$. Consideremos el desarrollo de Taylor del integrando,

$$\begin{aligned} \int dx e^{\frac{ip(x-\mu)}{N}} f(x) &\simeq \int dx f(x) \left[1 + ip \frac{(x-\mu)}{N} - \frac{p^2}{2N^2} (x-\mu)^2 + \dots \right] \\ &= 1 - \frac{p^2}{2N^2} \sigma_x^2 + \dots \end{aligned} \quad (4.63)$$

Ahora tenemos,

$$\begin{aligned} g(z) &\simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} dp e^{-ip(z-\mu)} \left[1 - \frac{p^2}{2N^2} \sigma_x^2 + \dots \right]^N \\ &\simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} dp e^{-ip(z-\mu)} e^{-\frac{p^2}{2N} \sigma_x^2} = \frac{1}{\sqrt{2\pi \sigma_x^2 / N}} e^{-\frac{(\mu-z)^2}{2\sigma_x^2 / N}}. \end{aligned} \quad (4.64)$$

Que es una distribución gaussiana para la variable y de media, $\langle y \rangle = \mu$ y desviación estándar, $\sigma_y = \sigma_x / \sqrt{N}$. c.q.d.

Este teorema proporciona un resultado contundente. Sea cual sea la distribución de probabilidad de la variable aleatoria x , si hacemos promedios de un número suficientemente grande de números aleatorios que sigan dicha distribución de probabilidad, el promedio estará distribuido según una distribución gaussiana centrada en el valor medio de la distribución original y con una desviación estándar $\sigma = \sigma_x / \sqrt{N}$.

³Seguimos la discusión en el libro de Guardiola, et al.

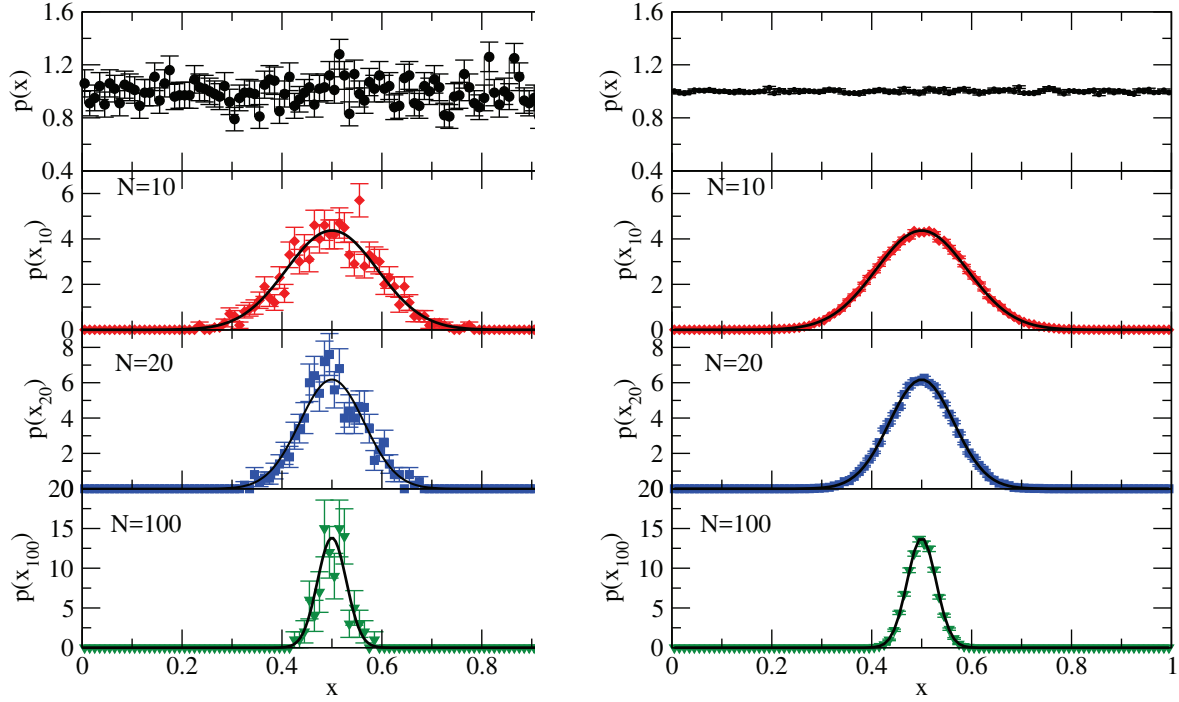


Figura 4.4: Ilustración del teorema del límite central para la distribución $U(0,1)$. Arriba, histograma de 100 cajas de la distribución $U(0,1)$, los paneles de la izquierda están calculados con 10000 puntos, y los de la derecha con 1000000. Debajo, histogramas de 100 cajas para las variables x_{10} , x_{20} y x_{100} , donde x_k son promedios de k valores. Comparados con la predicción gaussiana del límite central, curvas solidas negras. Los errores estadísticos de cada barra, aproximación binomial, crecen al promediar sobre más valores porque el número total de valores va bajando. Esto es, en los paneles de la izquierda, si promediamos de 10 en 10, tenemos 1000 valores, si promediamos de 20 en 20 tendremos 500, y de 100 en 100 tan sólo 100 de los 10000 números iniciales. En los paneles de la derecha, calculados con 10^6 puntos se observa como la predicción del teorema es muy precisa.

Ejemplo 4.5: Veamos un ejemplo sencillo, consideremos una variable aleatoria $x \in U(0,1)$. Genere-mos 10000 números, construyamos después otra variable z que sea el resultado de promediar los x de 10 en 10, tendremos 1000 números z . Construyamos un histograma con los números $\{z_k\}$. En la figura 4.4 mostramos histogramas calculados promediando en grupos de 10, 20 y 100.

Generalización del teorema del límite central

El teorema anterior, probado para el caso de la variable $y = (1/N) \sum_k x_k$ puede extenderse a valores esperados de cualquier función de una variable aleatoria $\langle g(x) \rangle = \int dx g(x) \rho(x)$,

$$y_N = \frac{1}{N} \sum_{k=1}^N g(x_k). \quad (4.65)$$

Se puede demostrar que si N es suficientemente grande, la variable y_N sigue una distribución Gaussiana cuyo valor medio es $\langle g(x) \rangle$ y cuya desviación estándar va como,

$$\sigma_{y_N} = \sigma_g / \sqrt{N} \quad (4.66)$$

donde $\sigma_g = \sqrt{\langle g^2 \rangle - (\langle g \rangle)^2}$. Y estos promedios están calculados sobre la densidad de probabilidad ρ .

4.6 Integrales definidas

Una de las aplicaciones más relevantes de los números aleatorios es el cálculo de integrales definidas. La mayoría de los métodos pertenecen a los denominados métodos de Montecarlo. Antes de llegar a ellos, veremos primero un algoritmo para producir números distribuidos según una distribución arbitraria y una aplicación del mismo método a evaluar integrales definidas de funciones sin nodos, como son áreas de figuras, volúmenes, etc.

4.6.1 Método de aceptación y rechazo

En la primera parte del tema hemos visto como generar números aleatorios con algunas distribuciones de probabilidad. En algunos casos podía invertirse la distribución acumulada y obtener así un cambio de variable adecuado para pasar de $U(0, 1)$ a la distribución deseada. Para la distribución gaussiana vimos el método de Box-Müller. Existe un método general que permite obtener números distribuidos según una distribución arbitraria que se conoce como el método de *aceptación y rechazo*.

Los requisitos que han de cumplir la variable aleatoria y y su densidad de probabilidad son que la primera esté definida en un dominio finito y la segunda ha de estar acotada, $\rho(y) < M$ y $y \in [a, b]$. El método, realizado en la práctica 5 es como sigue,

A1) Obtenemos dos números al azar a partir de distribuciones uniformes, $x \in U(a, b)$ y $p \in U(0, M)$. Estos se pueden generar a partir de $x_1, x_2 \in U(0, 1)$ con el cambio de variable, $x = (b - a)x_1 + a$ y $p = Mx_2$.

A2) Si $\rho(x) \geq p$ aceptamos el valor de x , $y = x$, en caso contrario volvemos a A1.

Este método tiene una determinada tasa de aceptación, el cociente entre números x producidos en la distribución $U(a, b)$ y los aceptados y . La tasa de aceptación aumenta si la cota superior es más baja. El paso A2) es fundamental ya que es el que implica que los números x cuya $\rho(x)$ sea mayor serán los más probables.

La demostración es sencilla, veamos, la probabilidad de los números y se puede escribir,

$$\bar{\rho}(y) = \int_a^b dx \frac{1}{b-a} \int_0^M dp \frac{1}{M} \delta(x-y) \Theta(\rho(x) - p) \quad (4.67)$$

que resulta

$$\bar{\rho}(y) = \rho(y) \frac{1}{M(b-a)}. \quad (4.68)$$

Donde la distribución $\bar{\rho}(y)$ debemos normalizarla ya que hemos ido perdiendo norma, $M(b-a) > 1$ al rechazar algunos de los números.

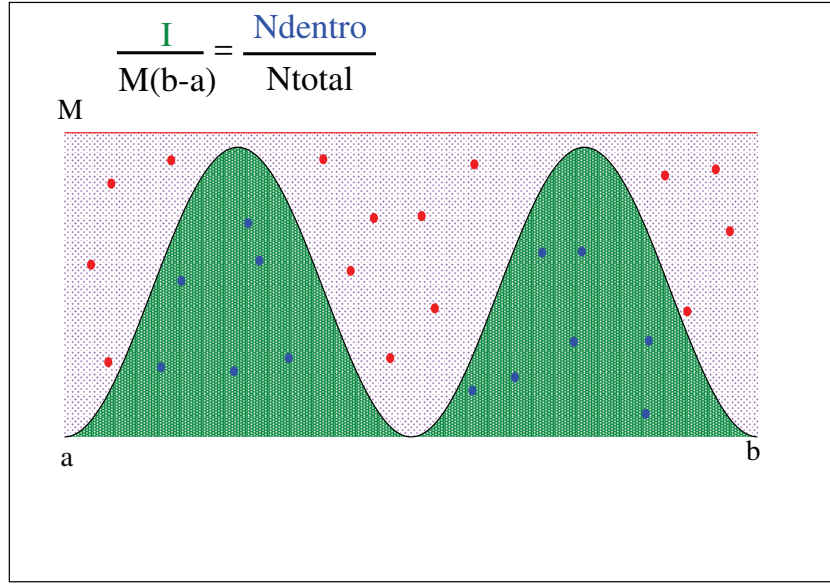


Figura 4.5: Método de aceptación y rechazo utilizado para evaluar integrales definidas de funciones $0 \leq f(x) < M$.

4.6.2 Método de acierto/fallo para evaluar integrales definidas

El método de aceptación y rechazo junto con el teorema del límite central nos proporciona el primer método para realizar integrales de funciones definidas positivas y acotadas, $0 \leq f(x) < M$. En este caso podemos escribir,

$$I = \int_a^b f(x) dx = \int_a^b dx \int_0^M dp \theta(f(x) - p). \quad (4.69)$$

Donde la primera integral se puede leer como $(b-a)$ veces el promedio de la función $f(x)$ con la variable x distribuida según $U(a, b)$. La segunda integral recuerda directamente al método de aceptación y rechazo. Utilicemos el teorema del límite central para evaluarla,

$$I = M(b-a) \int_a^b \frac{dx}{b-a} \int_0^M \frac{dp}{M} \theta(f(x) - p) \simeq \frac{M(b-a)}{N} \sum_{k=1}^N \theta(f(x_k) - p_k). \quad (4.70)$$

$$\frac{I}{M(b-a)} \simeq \frac{\sum_{k=1}^N \theta(f(x_k) - p_k)}{N}$$

Donde estamos viendo la última integral como el valor esperado de la función $\theta(f(x) - p)$ calculado para una distribución $(x, p) \in U(0, 1) \times U(a, b)$, y que por tanto, si aplicamos el teorema del límite central podemos estimar a partir de un promedio de N suficientemente grande. La última expresión, tiene una lectura muy sencilla, consiste en generar números $x \in U(a, b)$ y $p \in U(0, M)$ y contar los que caen por debajo de la función, N_{dentro} . El cociente entre el área bajo nuestra curva, I y el área del rectángulo $M(b-a)$ es aproximadamente igual al cociente entre puntos bajo la curva y todos los lanzados. En la figura 4.5 mostramos como funciona este método. El error del método puede estimarse teniendo en cuenta que el problema individual es de Bernouilli, ¿Cae el punto bajo la curva? En este caso, podemos hacer una estimación del error utilizando la distribución binomial

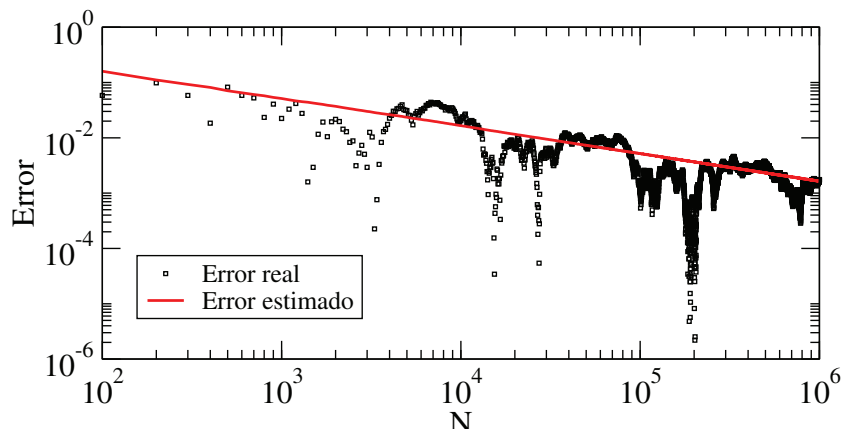


Figura 4.6: Convergencia del método de acierto fallo para el cálculo del valor de π utilizando parejas de puntos generados en $U(-1,1)$ y contando cuantas de ellas caen dentro del círculo de radio unidad. Comparamos el error real cometido con el error estimado según la ecuación (4.71).

para la variable N_{dentro} . El área viene dada por, $I = M(b-a)N_{\text{dentro}}/N$, como hicimos en la Ec. (4.54),

$$\sigma_I = \frac{M(b-a)}{N} \sqrt{N \frac{N_{\text{dentro}}}{N} \left(1 - \frac{N_{\text{dentro}}}{N}\right)} = \frac{M(b-a)}{\sqrt{N}} \sqrt{\frac{N_{\text{dentro}}}{N} \left(1 - \frac{N_{\text{dentro}}}{N}\right)}. \quad (4.71)$$

Esta estimación del error, como siempre con la estimación binomial, muestra que el error desciende como $1/\sqrt{N}$ ya que a medida que N crece, el cociente N_{dentro}/N tiende a un valor determinado. Es bueno también notar que cuanto menor sea la cota utilizada M mejor será la estimación de nuestra integral.

Ejemplo 4.6: Ejemplos típicos de uso de este método son cálculos de áreas de superficies o volúmenes de objetos en múltiples dimensiones, e.g. calcular π viendo cuando puntos caen dentro de un círculo circunscrito dentro de un cuadrado, ver figura 4.6, calcular el volumen de una hipersfera en 10 dimensiones, etc.

4.6.3 Método de Montecarlo

Otra serie de métodos potentes para evaluar integrales definidas, ahora sin la restricción necesaria en el apartado anterior, son los llamados métodos de Montecarlo. De nuevo se basan profundamente en el teorema del límite central y en su generalización, sección 4.5.2.

Veamos primero el caso unidimensional, para el que realmente no representan ninguna mejora respecto a los métodos vistos anteriormente, que permite entender la base del método.

Consideremos la integral más general,

$$I = \int_a^b f(x) dx \quad (4.72)$$

y hagamos un cambio de variable para situarnos en el intervalo $(0,1)$,

$$I = \int_0^1 h(t) dt \quad (4.73)$$