Tema 5

RESOLUCION NUMERICA DE ECUACIONES NO LINEALES Y DIAGONALIZACION DE MATRICES

Este es el último tema dedicado al Análisis Numérico y se ocupará, fundamentalmente, de la resolución numérica de *ecuaciones no lineales* (raíces reales) y de un problema estrechamente ligado al anterior: la *diagonalización* de matrices.

Basándonos en el *teorema de Rolle* introduciremos el procedimiento para separar las raíces reales de una ecuación dentro de un intervalo, es decir cómo acotar los intervalos en los que se sitúan las raíces. Esto tiene gran importancia a la hora de aplicar los métodos de cálculo. El primer método que se presenta es el de *bisección*. En ciertas condiciones, el *teorema de Cantor* sobre intervalos encajados garantiza la posibilidad de calcular por este método la raíz, a la que puede asignarse una cota sencilla del error cometido. Como generalización de este método está la *regula falsi*, que converge con mayor rapidez. Como cota del error introduciremos el valor absoluto de la diferencia entre dos aproximaciones sucesivas. Atendiendo a este criterio, general para todo este tipo de cálculos, podremos llevar la convergencia hasta donde deseemos. Después consideraremos los métodos *iterativo* y de *Newton*, y el análisis de las *raíces múltiples*, para concluir con un ejemplo de aplicación extraído de la radiación del cuerpo negro.

La segunda parte del tema trata de los *sistemas de ecuaciones* y está orientado, principalmente, a la diagonalización de matrices cuadradas. Así tras una somera revisión de dos métodos para resolver estos sistemas (*Newton-Raphson*, *Gauss*), nos dirigiremos rápidamente al cálculo de los *valores propios* de una matriz. Este problema posee una importancia capital en todas las aplicaciones prácticas de la Química Cuántica. Nos limitaremos al caso común de diagonalizar matrices *reales y simétricas*, y

expondremos los métodos del *polinomio característico* y de *Jacobi*. Como el tamaño de las matrices, en las aplicaciones comunes, suele ser muy grande (80 × 80 o más) el segundo método resulta más eficaz que el primero. Como Ejemplo de Aplicación probaremos ambos en el estudio del radical ciclopropenilo por el método de *Hückel*.

1. ECUACIONES NO LINEALES

1.1. Separación de raíces en un intervalo

Comenzaremos enunciando el teorema de Rolle para una función real de variable real y continua, $y(x)$, definida en un intervalo (a, b) $\equiv a < x < b$ (Apostol, 1972):

Si $y(a) \cdot y(b) < 0$, entonces existe un punto c interior al intervalo, $a < c < b$, tal que $f(c) = 0$. Si además $y'(x)$ existe y tiene signo constante en (a, b), el punto c anterior (raíz) es único.

Apoyándonos en este teorema es fácil separar los intervalos en los que habrá raíces reales para $y(x) = 0$ en (a, b). Si $y(a) \cdot y(b) < 0$, existe al menos una raíz. Deberá determinarse si es única, lo que puede hacerse particionando sucesivamente (a, b) y aplicando el mismo criterio a los subintervalos (a_n, b_n) resultantes. Por otra parte, si la primera derivada $y'(x)$ es continua y las raíces de $y'(x) = 0$ son fáciles de calcular, la separación para $y(x) = 0$ puede realizarse ordenadamente determinando el signo de $y(x)$ en los ceros de la derivada y en los extremos del intervalo. Nada mejor para ilustrar la manera de proceder que con un ejemplo.

Ejercicio de aplicación

Sea la función $y(x) = x^3 - 3x^2 - 6x + 8$, de la que deseamos acotar los intervalos en los que se encuentran las raíces. Procediendo por tanteos, fáciles de programar en una calculadora de escritorio, se construye la tabla siguiente:

x	$-\infty$	-3	0	3	6	∞
signo $[y(x)]$	$-$	$-$	$+$	$-$	$+$	$+$

de la que deducimos que $y(x)$ tiene tres raíces reales comprendidas en los intervalos $(-3,0)$, $(0,3)$ y $(3,6)$.

Si derivamos $y(x)$ tenemos: $y'(x) = 3x^2 - 6x - 6$, que es continua y con ceros en $x = 1 \pm 3^{1/2}$. Utilizando esta información tenemos:

x	$-\infty$	$1-\sqrt{3}$	$1+\sqrt{3}$	$+\infty$
signo $[y(x)]$	$-$	$+$	$-$	$+$

que nos indica la existencia de tres raíces dentro de los intervalos $(-\infty, 1-\sqrt{3})$, $(1-\sqrt{3}, 1+\sqrt{3})$, $(1+\sqrt{3}, +\infty)$.

Reuniendo la información de ambas tablas acotamos la situación de las raíces en: $(-3,1-\sqrt{3})$, $(1-\sqrt{3}, 1+\sqrt{3})$, $(3,6)$.

La operación de separación es fundamental para lo que seguirá a continuación, pero antes de ir a ello conviene hacer algunas advertencias. La primera es que la información que dan estas tablas es sólo sobre raíces reales. Recuérdese que si la función es polinómica tendrá tantas raíces (reales y/o complejas) como grado posea (Ejercicio 1). La segunda es que la multiplicidad de la raíz no se ve reflejada en las tablas, por lo que deberá procederse con cautela al identificarlas. Por último, hay casos en los que la ecuación a resolver contiene infinitas soluciones, lo que sucede cuando hay involucradas funciones trigonométricas, y entre todas esas soluciones suele ser normal estar interesado sólo en algunas. Esto obligará a efectuar el correspondiente análisis.

Un criterio general para el error de una raíz aproximada es el siguiente. Sea $y(x)$ diferenciable en $a \leqslant x \leqslant b$, α una raíz de $y(x) = 0$ en ese intervalo y x_n una aproximación a α. Por el teorema del Valor Medio, existe un valor c, $a \leqslant c \leqslant b$, entre α y x_n, tal que (Arteaga, 1975):

$$y(\alpha) - y(x_n) = (\alpha - x_n)y'(c) \qquad [1]$$

Si para todo x de $a \leqslant x \leqslant b$ se cumple $|y'(x)| \geqslant M > 0$, como $y(\alpha) = 0$, escribimos:

$$|y(x_n)| = |\alpha - x_n| \cdot |y'(c)| \geqslant |\alpha - x_n| M \qquad [2]$$

de donde:

$$|\alpha - x_n| \leqslant \frac{|y(x_n)|}{M} \qquad [3]$$

expresión que sirve para evaluar el error de la aproximación x_n. Más adelante daremos otra medida del error muy útil en el cálculo.

1.2. Método de bisección

Sea una función real de variable real $y(x)$, continua en $a \leqslant x \leqslant b$, y tal que $y(a) \cdot y(b) < 0$. La búsqueda de una raíz para la ecuación $y(x) = 0$ en el intervalo puede hacerse siguiendo el siguiente proceso:

i) Dividir $a \leqslant x \leqslant b$ en dos partes iguales: $a \leqslant x \leqslant (a + b)/2$ y $(a + b)/2 \leqslant x \leqslant b$. Si $y((a + b)/2) = 0$, entonces la raíz es $\alpha = (a + b)/2$. Si $y((a + b)/2) \neq 0$, se continúa:

ii) Eligiendo el subintervalo en el que $y(x)$ cambia de signo. Sea éste, por ejemplo, el primero que renombramos:

$$a_1 \leqslant x \leqslant b_1 \equiv a \leqslant x \leqslant (a + b)/2 \qquad [4]$$

iii) Con $a_1 \leqslant x \leqslant b_1$ recomenzamos el proceso en i), y así sucesivamente.

Procediendo de esta manera, o se obtiene la raíz exacta α o se obtiene una sucesión de intervalos encajados:

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n] \supset \ldots \qquad [5]$$

tales que $y(a_n) \cdot y(b_n) \leqslant 0$ $(n = 1, 2, ...)$, teniéndose:

$$b_n - a_n = \frac{1}{2^n}(b - a) \qquad [6]$$

La sucesión $\{a_n\}$ es no decreciente y está acotada superiormente, en tanto que la $\{b_n\}$ es no creciente y acotada inferiormente. En consecuencia (Apostol, 1972):

$$\alpha = \lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n \qquad [7]$$

Este valor α es la raíz buscada, pues al ser $y(x)$ continua:

$$\lim_{n \to \infty} [y(a_n) \cdot y(b_n)] = [y(\alpha)]^2 \leqslant 0 \qquad [8]$$

de donde $y(\alpha) = 0$ (Ejercicio 2a). El error cometido con este método es evidentemente:

$$0 \leqslant |\alpha - x_n| \leqslant \frac{1}{2^n}(b - a) \qquad [9]$$

1.3. Regula falsi

Este método recibe también los nombres de método de las partes proporcionales, o método de la *falsa posición*. Dada una función real de variable real $y(x)$, continua en $a \leqslant x \leqslant b$ y tal que $y(a) \cdot y(b) < 0$, para calcular una raíz de $y(x) = 0$, se divide $a \leqslant x \leqslant b$ en dos partes que están en la relación:

$$\frac{-y(a)}{y(b)} \qquad [10]$$

A partir del extremo $x_0 = a$ se calcula un valor aproximado a la raíz:

$$x_1 = a - \frac{y(a)}{y(b) - y(a)} \cdot (b - a) \qquad [11]$$

y se repite este procedimiento en aquél de los intervalos resultantes, $a \leqslant x \leqslant x_1$, $x_1 \leqslant x \leqslant b$, en el que $y(x)$ cambie de signo. Se obtendría con ello una aproximación mejorada x_2, y así sucesivamente. Geométricamente este método equivale a reemplazar $y(x)$ por la *cuerda* que conecta $(a, y(a))$ con $(b, y(b))$ (Fig. 1).

En las condiciones dadas el proceso es convergente. Supongamos que la raíz ha sido separada y que $y''(x)$ tiene signo constante en $a \leqslant x \leqslant b$. Tomemos $y''(x) > 0$ en ese intervalo (función cóncava; sería análogo para $y''(0)$. Puede darse, bien que $y(a) > 0$, o bien que $y(a) < 0$. Fijémonos en $y(a) > 0$, pues sería análogo para el otro caso.

Si $y(a) > 0$ se elige $x = a$ como punto fijo y hacemos:

$$x_0 = b$$

$$x_{n+1} = x_n - \frac{y(x_n)}{y(x_n) - y(a)}(x_n - a) \qquad [12]$$

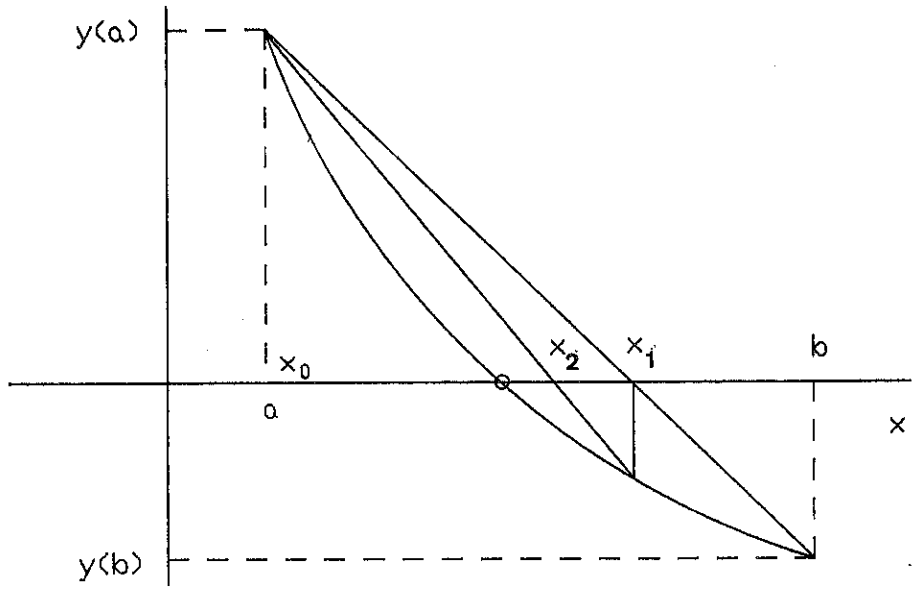


Figura 1.—Representación geométrica de la regula falsi.

La sucesión $\{x_n\}$ así generada está acotada inferiormente y es decreciente:

$$a < \alpha < \ldots < x_{n+1} < x_n < \ldots < x_0 = b \qquad [13]$$

por lo que tendrá límite:

$$\lim_{n \to \infty} x_n = \alpha \qquad [14]$$

que es justamente la raíz buscada, ya que tomando $\lim (n \to \infty)$ en [12] se llega a:

$$\alpha = \alpha - \frac{y(\alpha)}{y(\alpha) - y(a)}(\alpha - a) \qquad [15]$$

de donde evidentemente $y(\alpha) = 0$.

Podemos estimar el error a partir de la fórmula general [3], o bien dar otra cota superior para este error utilizando el valor absoluto de la diferencia entre dos aproximaciones sucesivas (Arteaga, 1975):

$$|\alpha - x_{n+1}| \leqslant |x_{n+1} - x_n| \qquad [16]$$

expresión que resulta de gran utilidad en los cálculos numéricos. Se exige y' continua y con signo constante en $a \leqslant x \leqslant b$.

1.4. Método de iteración

Vamos a tratar ahora con un método de resolución de ecuaciones no lineales que recibe el nombre de método de *iteración*, si bien hay que notar que por su esencia todos los métodos que tratamos son iterativos. Sea $y(x) = 0$ con $y(x)$ continua. Cuando $y(x) = 0$ puede reescribirse en la forma:

$$y(x) = 0 \quad \Leftrightarrow \quad x = g(x) \qquad [17]$$

donde $g(x)$ es continua, podemos intentar encontrar la raíz construyendo a partir de un valor inicial x_0 la sucesión:

$$x_n = g(x_{n-1}) \quad ; \quad n = 1, 2, \ldots \qquad [18]$$

Si esta sucesión x_n es convergente, entonces su límite α es una solución de [17]:

$$\alpha = \lim_{n \to \infty} x_n = \lim_{n \to \infty} g(x_n) = g\left[\lim_{n \to \infty} x_n\right] = g(\alpha) \qquad [19]$$

El valor α puede calcularse con tanta aproximación como se quiera sin más que aumentar el número de iteraciones (Ejercicio 2b y epígrafe 1.7).

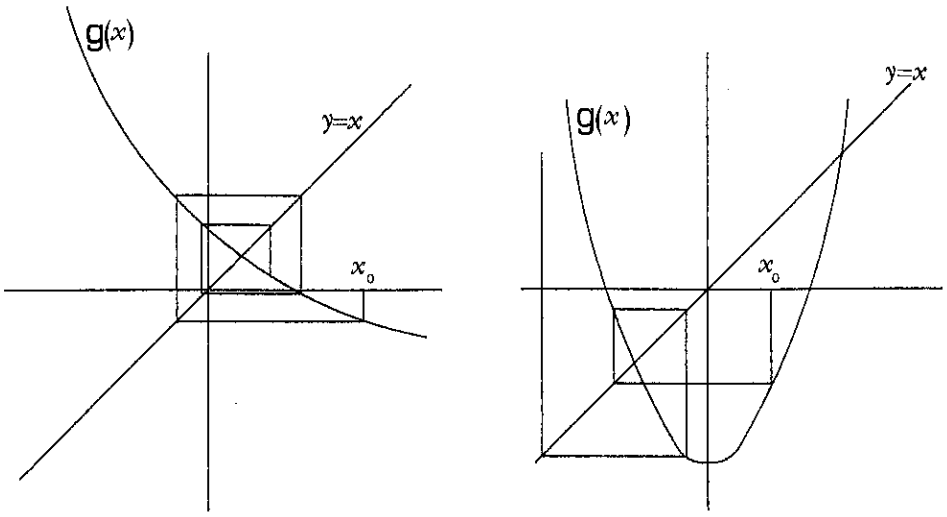
Cabe preguntarse por la convergencia de la sucesión [18]. Una condición suficiente de convergencia es (Scheid, 1972):

Si $|g'(x)| \leqslant K < 1$ para todo x de un intervalo centrado en α $(\alpha - a, \alpha + a)$ y con x_0 en ese intervalo, entonces:

$$\alpha = \lim_{n \to \infty} x_n \qquad [20]$$

El error absoluto del método lo estimaremos a través de la relación [16].

Es muy instructivo visualizar geométricamente el proceso contenido en [18]. En definitiva, se sustituye $y(x) = 0$ por un sistema de ecuaciones: $y = x$; $y = g(x)$, y el método busca la intersección de ambas funciones. Una buena elección del x_0 inicial puede ser crucial, tanto para una rápida convergencia a la raíz deseada, como para que ésta, en su caso, sea alcanzable (Fig. 2).



a) Convergencia *b*) Divergencia

Figura 2.—Método de iteración.

1.5. Método de Newton

Sea la ecuación $y(x) = 0$ de la que sabemos posee una raíz α en el intervalo $a \leqslant x \leqslant b$. Si las dos primeras derivadas $y'(x)$, $y''(x)$ son continuas en el intervalo señalado y mantienen su signo constante, entonces poniendo:

$$\alpha = x_n + h_n \qquad [21]$$

donde x_n es una aproximación a α, el desarrollo de Taylor en torno a $x = x_n$ de $y(x)$ queda (a primer orden):

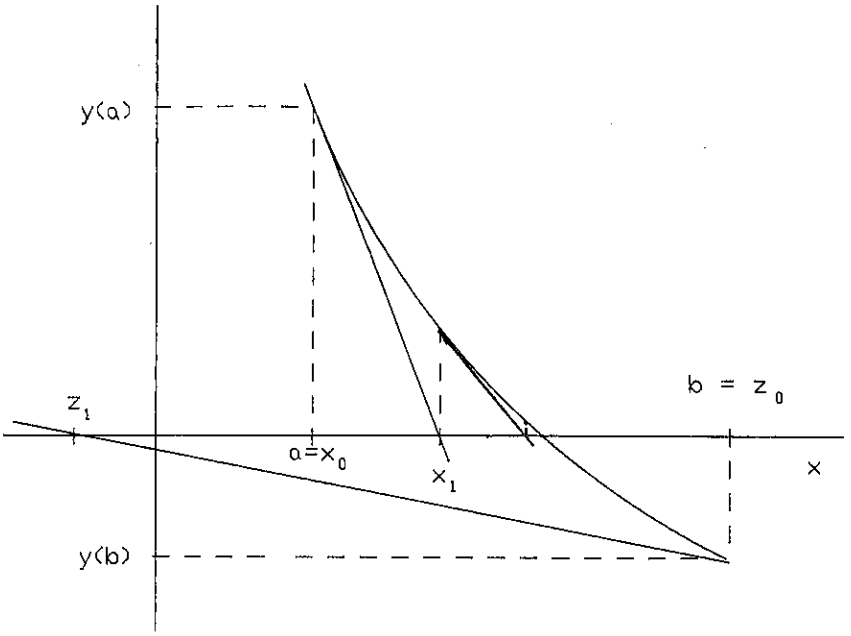
$$y(\alpha) = y(x_n + h_n) \simeq y(x_n) + h_n y'(x_n) \qquad [22]$$

Dado que α es la raíz, se tiene:

$$h_n = -\frac{y(x_n)}{y'(x_n)}$$

y por tanto se construye la sucesión (Ejercicio 2c):

$$x_{n+1} = x_n - \frac{y(x_n)}{y'(x_n)} \quad ; \quad n = 0, 1, 2, \ldots \qquad [23]$$



Convergencia con $x_0 = a$
Divergencia con $z_0 = b$

Figura 3.—Método de Newton.

Geométricamente, este método reemplaza el arco de secante de la *regula falsi* por la *tangente*. Consideremos una función cóncava en $a \leqslant x \leqslant b$ $(y''(x) > 0)$ (Fig. 3). Tomando como aproximación inicial $x_0 = a$, tal que $y(a) \cdot y''(a) > 0$, tracemos por $(a, y(a))$ la tangente a $y(x)$ y determinemos el corte x_1 de esta recta con el eje x. A continuación repetimos la operación trazando la tangente que pasa por $(x_1, y(x_1))$, determinamos x_2, y así sucesivamente. En general la ecuación de la tangente en un punto $(x_n, y(x_n))$ es:

$$y - y(x_n) = y'(x_n)(x - x_n)$$

que para $y = 0$ nos da la relación [23]. Es muy importante darse cuen-

ta de que la elección del valor x_0 es crucial, pues de haberse tomado $x_0 = z_0 = b$, el método o divergería o nos daría raíces fuera del intervalo.

Condiciones de validez del método se dan a continuación sin demostrar los teoremas correspondientes (Arteaga, 1975). La primera condición se expresa en el enunciado:

Si $y(a) \cdot y(b) < 0$, $y'(x)$ e $y''(x)$ son no nulas y continuas con signo constante en $a \leqslant x \leqslant b$, la raíz única de $y(x) = 0$ en ese intervalo puede calcularse por el método de Newton, con una aproximación tan grande como se desee, partiendo de un x_0 inicial que cumpla:

$$y(x_0) \cdot y''(x_0) > 0 \qquad [24]$$

La segunda condición se resume en los siguientes requisitos que debe cumplir en su totalidad $y(x)$:

i) $y(x)$ debe estar definida y ser continua en $-\infty < x < +\infty$

ii) $y(a) \cdot y(b) < 0$

iii) $y'(x) \neq 0$ en $a \leqslant x \leqslant b$

iv) $y''(x)$ existe en todas partes y tiene signo constante en $a \leqslant x \leqslant b$

entonces puede tomarse como x_0 cualquier punto x del intervalo, en particular $x_0 = a$ ó $x_0 = b$.

El error de la aproximación $(n + 1)$-ésima lo evaluaremos comúnmente con [16]. La convergencia del método es *cuadrática*, como se deduce de aplicar el desarrollo en serie de Taylor hasta segundo orden a $y(x)$ y utilizar el algoritmo [23]. Se llega así sin dificultad a la relación entre los errores de pasos consecutivos:

$$|\alpha - x_n| \simeq \left|\frac{y''(\alpha)}{2y'(\alpha)}\right| \cdot |\alpha - x_{n-1}|^2 \qquad [25]$$

Esta relación indica que el error en el paso n es proporcional al cuadrado del error previo $(n - 1)$, con lo que el número de cifras decimales correctas de la raíz *se dobla*, aproximadamente, con cada paso adicional. Esta convergencia rápida a medida que avanza el método indica que conviene utilizar buenas aproximaciones iniciales. Muchas veces se emplean algoritmos diferentes del de Newton para iniciar el cálculo y obtener buenos x_0, los cuales son velozmente refinados con la aplicación

final del método que nos ocupa. Todo ello está muy relacionado con el problema de la convergencia hacia raíces no esperadas cuando se elige un x_0 inicial próximo a regiones de tangente horizontal ($y'(x_0) \to 0$), lo que incide negativamente en el algoritmo [23] (Ejercicio 3).

1.6. Raíces múltiples

Hasta aquí hemos asumido que la raíz de $y(x) = 0$ contenida en $a \leqslant x \leqslant b$ era simple. El problema que planteamos ahora es cómo determinar las raíces múltiples de $y(x) = 0$. Por ejemplo, $(x - 2)^3 = 0$ tiene en $x = 2$ una raíz de orden tres. En este caso los métodos anteriores no son aplicables.

Una raíz múltiple de $y(x) = 0$ se presenta en $x = \alpha^*$ con multiplicidad m cuando:

$$\begin{cases} y(\alpha^*) = y'(\alpha^*) = y''(\alpha^*) = \ldots = y^{(m-1)}(\alpha^*) = 0 \\ y^{(m)}(\alpha^*) \neq 0 \end{cases} \qquad [26]$$

lo que puede expresarse diciendo que $y(x)$ es de la forma:

$$y(x) = (x - \alpha^*)^m \cdot f(x) \qquad [27]$$

con $f(x) \neq 0$ en las inmediaciones de $x = \alpha^*$. Veamos varios caminos para tratar de resolver este problema (Ralston, 1970; Rice, 1983).

a) Si construimos la función auxiliar:

$$g(x) = \frac{y(x)}{y'(x)} = \frac{(x - \alpha^*) f(x)}{m f(x) + (x - \alpha^*) f'(x)} \qquad [28]$$

es obvio que sólo poseerá raíces simples. Un sencillo cálculo demuestra que:

$$g'(\alpha^*) = 1/m \qquad [29]$$

y por tanto podemos cambiar el problema $y(x) = 0$ por el de $g(x) = 0$, cuando exista posibilidad de toparnos con una raíz múltiple. Aplicando a $g(x)$ los métodos ya estudiados y completando con [29] el problema queda resuelto.

b) *Métodos de Newton y de la* regula falsi *modificados*

Para construir el algoritmo modificado de Newton se hace la consideración:

$$y(x) \simeq \text{Constante} \cdot (x - \alpha^*)^m \qquad [30]$$

de modo que:

$$\frac{y(x)}{y'(x)} = \frac{x - \alpha^*}{m} \Rightarrow \alpha^* = x - m\frac{y(x)}{y'(x)} \qquad [31]$$

A partir de [31] se puede proponer el esquema iterativo:

$$x_{n+1} = x_n - m\frac{y(x_n)}{y'(x_n)} \qquad [32]$$

que constituye el método de Newton modificado. Resulta interesante visualizar esta ecuación [32] como la correspondiente a un algoritmo de Newton para raíces simples con pendiente efectiva $y'(x_n)/m$ (Ejercicio 4).

De manera análoga se obtiene el algoritmo de la *regula falsi* modificada:

$$x_{n+1} = x_n - m\frac{y(x_n)(x_n - x_{n-1})}{y(x_n) - y(x_{n-1})} \qquad [33]$$

La siguiente cuestión a analizar es, naturalmente, la identificación de las situaciones en las que una raíz múltiple puede estar involucrada. En general, esto ocurre cuando se da una convergencia anómala:

i) $y(x_n)$ se hace muy pequeño mientras que $|x_n - x_{n-1}|$ se mantiene relativamente grande.

ii) El error $|x_n - x_{n-1}|$ tiende a cero mucho más lentamente que lo esperado por los criterios teóricos de convergencia.

Otras veces son argumentos de tipo físico (*degeneración, simetría,* etc.) los que nos apuntan no sólo esta posibilidad, sino también la multiplicidad que cabe esperar.

Si nos decidimos a utilizar los métodos del apartado *b*), es necesario disponer de una estimación de m. El primer paso es utilizar una función auxiliar que simule el comportamiento aproximado de $y(x)$:

$$y(x) \simeq a(x - \alpha^*)^m = f_a(x) \qquad [34]$$

donde m debe fijarse *a priori*. Utilizando $f_a(x)$ calculamos una estimación α_1^* de la raíz α^*. Supongamos que la estimación resulta ser muy buena, pero que ha sido obtenida tras un número muy elevado de iteraciones. Esto indica que posiblemente el valor de m seleccionado es incorrecto. Tomaremos entonces dos valores x_1 y x_2, apartados de α_1^*, y evaluaremos:

$$y(x_1) \simeq a(x_1 - \alpha^*)^m$$
$$y(x_2) \simeq a(x_2 - \alpha^*)^m$$

que conducen a:

$$m \sim \frac{\log [y(x_1)/y(x_2)]}{\log [(x_1 - \alpha^*)/(x_2 - \alpha^*)]} \qquad [35]$$

Haciendo $\alpha^* \simeq \alpha_1^*$ se puede calcular un valor mejorado para m. Con este valor repetiríamos las operaciones previas para obtener una mejor aproximación α_2^* a la raíz, y así sucesivamente hasta alcanzar la precisión deseada.

1.7. Ejemplo de aplicación: Posición del máximo de la radiación del cuerpo negro

La ley de Planck para la densidad de energía, por unidad de volumen y de frecuencia, de la radiación del cuerpo negro es:

$$E(\nu) = \frac{8\pi h}{c^3} \cdot \frac{\nu^3}{\exp [h\nu/kT] - 1}$$

Nos proponemos localizar la posición del máximo de esta función a varias temperaturas ($T = 1\,500\,K$, $3\,000$ K, $6\,000$ K).

La ecuación a resolver numéricamente es:

$$\frac{h\nu}{kT} = 3(1 - \exp[-h\nu/kT])$$

y contiene en sí misma la posición ν de todos los máximos a cualquier temperatura. Por comodidad haremos:

$$x = h\nu/kT$$

quedándonos:

$$y = 3e^{-x} + x - 3 = 0$$

cuya derivada es $y' = -3e^{-x} + 1$, que posee un único cero en $x = \ln 3$. Tomando este punto encontramos:

x	$-\infty$	$\ln 3$	$+\infty$
signo y	$+$	$-$	$+$

de donde concluimos la existencia de dos raíces reales, una en $(-\infty, \ln 3)$ y otra en $(\ln 3, +\infty)$. Una de las raíces, por simple inspección, se ve que es $x = 0$ y será excluida del estudio subsiguiente por la propia naturaleza del problema ($v > 0 \Rightarrow x > 0$). Para calcular la solución aplicaremos dos métodos, el de iteración y el de Newton.

a) *Iteración*

Escribiremos la ecuación a resolver en la forma:

$$x = 3(1 - e^{-x})$$

Tomaremos primeramente $x_0 = 1{,}1$ ($> \ln 3$) para localizar la raíz en $(\ln 3, +\infty)$. Aplicando el algoritmo:

$$x_n = 3(1 - e^{-x_{n-1}})$$

obtenemos la tabla:

$$x_1 = 2{,}001386749$$
$$x_3 = 2{,}775963098$$
$$x_5 = 2{,}819949757$$
$$x_7 = 2{,}821391836$$
$$x_9 = 2{,}821437856$$
$$x_{11} = 2{,}821439324$$
$$x_{13} = 2{,}821439371$$
$$x_{15} = 2{,}821439372$$

cuyo último valor nos da la solución buscada $\alpha = 2{,}821439372$ con nueve cifras decimales.

b) *Newton*

El algoritmo a emplear ahora es:

$$x_n = x_{n-1} - \frac{3e^{-x_{n-1}} + x_{n-1} - 3}{-3e^{-x_{n-1}} + 1}$$

y deberemos tomar un x_0 tal que $y_0 \cdot y_0'' > 0$. Sea pues $x_0 = 4$. Los resultados se resumen en la breve tabla:

$$x_1 = 2{,}883716761$$
$$x_2 = 2{,}821838575$$
$$x_3 = 2{,}821439389$$
$$x_4 = 2{,}821439372$$

siendo de apreciar la rápida convergencia (cuadrática) de este método. El valor de la raíz calculada es el mismo que en *a*).

Consecuentemente, la frecuencia de los máximos de radiación estará localizada a una T dada en:

$$\nu_M = \frac{kT}{h} x$$

siendo k la constante de Boltzmann y h la de Planck. Es claro que al crecer la temperatura el máximo se desplaza a la derecha del espectro. Para las temperaturas indicadas tenemos ($k = 1{,}381 \times 10^{-23}$ jul/K, $h = 6{,}626 \times 10^{-34}$ jul. s):

T(K)	1 500	3 000	6 000
$\nu_M(s^{-1})$	$8{,}8207 \times 10^{13}$	$17{,}6414 \times 10^{13}$	$35{,}2829 \times 10^{13}$

2. SISTEMAS DE ECUACIONES

2.1. Método de Newton-Raphson

Consideremos el sistema general de dos ecuaciones con dos incógnitas:

$$\begin{cases} f(x, y) = 0 \\ g(x, y) = 0 \end{cases} \qquad [36]$$

del que deseamos calcular una solución (α, β) tomando como aproximación inicial (x_0, y_0). Siguiendo la idea del método de Newton (1.5), desarrollemos en serie de Taylor f y g en torno al punto inicial y quedémonos con la parte lineal:

$$\begin{cases} f(x, y) \simeq f(x_0, y_0) + (x - x_0) f_x(x_0, y_0) + (y - y_0) f_y(x_0, y_0) \\ g(x, y) \simeq g(x_0, y_0) + (x - x_0) g_x(x_0, y_0) + (y - y_0) g_y(x_0, y_0) \end{cases} \qquad [37]$$

La mejora x_1, y_1 a la aproximación inicial puede calcularse haciendo en [37] $f(x, y) = g(x, y) = 0$ con $x = x_1$ e $y = y_1$, y resolviendo el sistema lineal resultante. El proceso se repite tomando (x_1, y_1) como nuevo punto inicial, y así sucesivamente hasta alcanzar la precisión requerida para la solución (α, β).

En forma compacta el algoritmo se escribe:

$$\begin{cases} x_n = x_{n-1} + h_{n-1} \\ y_n = y_{n-1} + k_{n-1} \end{cases} \quad n = 1, 2, \ldots \qquad [38]$$

con h_{n-1} y k_{n-1} dadas por:

$$\begin{cases} f_x(x_{n-1}, y_{n-1}) \cdot h_{n-1} + f_y(x_{n-1}, y_{n-1}) \cdot k_{n-1} = -f(x_{n-1}, y_{n-1}) \\ g_x(x_{n-1}, y_{n-1}) \cdot h_{n-1} + g_y(x_{n-1}, y_{n-1}) \cdot k_{n-1} = -g(x_{n-1}, y_{n-1}) \end{cases} \qquad [39]$$

Su aplicación requiere, evidentemente, que las derivadas parciales f_x y g_x existan en los puntos (x_{n-1}, y_{n-1}). Como en el caso de ecuaciones no lineales este método da una convergencia *cuadrática* a la solución deseada, en tanto que la aproximación inicial sea suficientemente buena.

2.2. Método de Gauss para sistemas lineales

Sea el sistema lineal compatible de n ecuaciones con n incógnitas x_i:

$$\left\{\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = a_{1,n+1} \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = a_{2,n+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = a_{n,n+1} \end{array}\right. \qquad [40]$$

que abreviamos por:

$$Ax = A^c \qquad [41]$$

y que necesariamente es no singular: $\det(A) = |A| \neq 0$.

La resolución de este sistema por el método de Gauss consiste en reducirlo a forma *triangular superior*, procediendo recurrentemente a partir de ahí. Veamos los pasos que se dan.

i) Sea $a_{11} \neq 0$. Se resta de cada ecuación la primera multiplicada por el cociente que resulta de dividir el coeficiente de x_1 en dicha ecuación por a_{11}, y se obtiene:

$$\left.\begin{array}{r} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = a_{1,n+1} \\ a_{22}^{(1)}x_2 + \ldots + a_{2n}^{(1)}x_n = a_{2,n+1}^{(1)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n2}^{(1)}x_2 + \ldots + a_{nn}^{(1)}x_n = a_{n,n+1}^{(1)} \end{array}\right\} \quad a_{ij}^{(1)} = a_{ij} - \frac{a_{i1}}{a_{11}} \cdot a_{1j}$$

ii) Sea ahora $a_{22}^{(1)} \neq 0$. Análogamente a lo anterior se llega con las $n - 1$ ecuaciones inferiores a:

$$\left.\begin{array}{r} a_{11}x_1 + a_{12}x_2 + \ldots\ldots\ldots + a_{1n}x_n = a_{1,n+1} \\ a_{22}^{(1)}x_2 + \ldots\ldots\ldots + a_{2n}^{(1)}x_n = a_{2,n+1}^{(1)} \\ a_{33}^{(2)}x_3 + \ldots + a_{3n}^{(2)}x_n = a_{3,n+1}^{(2)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n3}^{(2)}x_3 + \ldots + a_{nn}^{(2)}x_n = a_{n,n+1}^{(2)} \end{array}\right\}$$

iii) Procediendo de la misma forma acabaríamos con el sistema triangularizado:

$$\left.\begin{array}{rl} a_{11}x_1 + a_{12}x_2 + \dots\dots\dots\dots + a_{1n}x_n & = a_{1,n+1} \\ \bar{a}_{22}x_2 + \dots\dots\dots\dots + \bar{a}_{2n}x_n & = \bar{a}_{2,n+1} \\ \dots\dots\dots\dots\dots\dots\dots\dots & \\ \bar{a}_{n-1,n-1}x_{n-1} + \bar{a}_{n-1,n}x_n & = \bar{a}_{n-1,n+1} \\ \bar{a}_{n,n}x_n & = \bar{a}_{n,n+1} \end{array}\right\} \qquad [42]$$

cuya resolución es directa.

Es importante resaltar la circunstancia de que la única fuente de error en este método directo es la de los redondeos, que dependen fuertemente del número de operaciones N que se realicen. Esta es una de las razones por las que el método de Cramér ($N \sim n!$) no se usa en la práctica. En el algoritmo de Gauss ($N \sim n^3/3$) aparece la división por un número (*pivote*) a_{11}, $a_{22}^{(1)}$, etc., que en el caso de que fuera próximo a cero contribuiría a disparar el error de redondeo. Para evitar esta circunstancia, deberá seleccionarse un elemento *pivote que difiera de cero todo lo posible.* Naturalmente esta selección conviene hacerla desde el principio, tomando siempre como pivote del paso j el elemento $a_{jn}^{(j-1)}$ que sea el mayor en valor absoluto de los coeficientes de todas las $n - j + 1$ ecuaciones finales a reducir en tal paso. Para técnicas más depuradas en este campo véase (Ralston, 1970).

3. PROBLEMAS DE VALORES PROPIOS Y DIAGONALIZACION DE MATRICES

Un gran número de problemas conducen a la resolución de un sistema lineal de ecuaciones de la forma:

$$\left\{\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = \lambda x_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = \lambda x_n \end{array}\right. \qquad [43]$$

o más compactamente:

$$A\bar{x} = \lambda\bar{x} \qquad [44]$$

siendo A la matriz $(n \times n)$ de los coeficientes a_{ij} y $\bar{x}$ la matriz columna $(n \times 1)$ que contiene a las incógnitas x_i. Es inmediato ver que éste es un sistema *homogéneo* y que tendrá soluciones distintas de la trivial ($x_i = 0$, $i = 1, 2, ..., n$) para algunos valores específicos del parámetro λ. Estos valores, en número de n, se pueden obtener notando que [44] puede escribirse:

$$(A - \lambda I)\bar{x} = 0 \qquad [45]$$

donde I es la matriz unidad. El determinante de la matriz cuadrada en [45] debe ser nulo

$$|A - \lambda I| = 0 \quad \text{(determinante secular)} \qquad [46]$$

o lo que es lo mismo:

$$P(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & a_{23} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0 \qquad [47]$$

3.1. Polinomio característico

El desarrollo del determinante [47] conduce al llamado *polinomio característico* $P(\lambda)$ asociado al sistema. Sus raíces son los valores λ buscados que reciben el nombre de *autovalores* o *valores propios* de A. Estos valores pueden ser tanto reales como complejos y, una vez calculados, deberemos evaluar los *vectores propios* (o *autovectores*) $\bar{x}$, *por la derecha* en el caso [45], asociados a cada uno de ellos. Incidentalmente señalaremos que los coeficientes de $P(\lambda)$ pueden calcularse a partir de los elementos a_{ij} sin necesidad de evaluar el determinante [47] (Ralston, 1970), y más adelante daremos algunas relaciones útiles. Las matrices A a tratar aquí tendrán n valores propios y n vectores propios *linealmente independientes*, siendo así *semejantes* a una matriz diagonal $(n \times n)$ $\Lambda = (\lambda_i \delta_{ij})$.

Sea un autovalor λ_i supuesto simple (*no degenerado*). Su sustitución en [43] nos lleva a un sistema lineal del que se puede suprimir una ecuación. Trabajando con las $n - 1$ primeras ecuaciones y asignando un valor

arbitrario a, por ejemplo, x_n se despejan las otras incógnitas $x_1, x_2, ..., x_{n-1}$. El resultado sería el vector columna:

$$\bar{x}^{(i)} = \begin{pmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \dots \\ x_n^{(i)} \end{pmatrix} \qquad [48]$$

que suele darse en forma normalizada exigiendo que:

$$[x_1^{(i)}]^2 + [x_2^{(i)}]^2 + \dots + [x_n^{(i)}]^2 = 1 \qquad [49]$$

Si la *degeneración* de λ_i fuese dos ($\lambda_i = \lambda_{i1} = \lambda_{i2}$), entonces sobrarían dos ecuaciones en [43] a la hora de calcular el primer vector $\bar{x}^{(i1)}$. El proceso sería análogo, pero ahora dando valores arbitrarios a dos incógnitas x_{n-1}, x_n. Para fijar el vector $\bar{x}^{(i1)}$ en forma normalizada se impone de nuevo la condición [49]. Resta por calcular el segundo vector propio $\bar{x}^{(i2)}$ asociado al autovalor λ_{i2}, y que debe ser *linealmente independiente* del anterior. Sobran como antes dos ecuaciones, pero ahora sólo fijamos el valor de una incógnita x_n e imponemos dos condiciones suplementarias:

i) ortogonalidad de $\bar{x}^{(i2)}$ a $\bar{x}^{(i1)}$:

$$x_1^{(i1)} \cdot x_1^{(i2)} + x_2^{(i1)} \cdot x_2^{(i2)} + \dots + x_n^{(i1)} \cdot x_n^{(i2)} = 0 \qquad [50]$$

ii) normalización de $\bar{x}^{(i2)}$:

$$[x_1^{(i2)}]^2 + [x_2^{(i2)}]^2 + \dots + [x_n^{(i2)}]^2 = 1 \qquad [51]$$

Si la degeneración g del autovalor λ_i fuese mayor ($g > 2$) se procedería análogamente generalizando el esquema previo. Notemos que habría que imponer $g(g-1)/2$ condiciones de ortogonalidad. Todas estas operaciones habrían de realizarse para los n autovalores λ.

Este método nos puede ayudar a diagonalizar matrices *simétricas o no simétricas*, pero resulta de naturaleza ardua. Así, aunque formalmente nos resuelve el problema, no cabe ser considerado como el más eficaz cuando el orden n es grande (en muchas aplicaciones $n > 80$).

3.2. Método de Jacobi

En las aplicaciones habituales los problemas de diagonalización suelen involucrar matrices *reales y simétricas*. Vamos pues a fijarnos en este caso y a resolverlo por el método de *Jacobi*, menos engorroso que el del polinomio característico y que nos resolverá de una vez el cálculo de valores y de vectores propios. Para diagonalizar matrices no simétricas por técnicas similares (transformaciones matriciales) a la de Jacobi remitimos al lector a referencias especializadas (Ralston, 1970; Rice, 1983).

La base del método de Jacobi se encuentra en el hecho de que toda matriz A real y simétrica ($a_{ij} = a_{ji}$) es diagonalizable mediante una *transformación de semejanza* con una *matriz ortogonal* O (Sokolnikoff, 1971). Esto significa que:

$$O^{-1} \cdot A \cdot O = \Lambda \quad \text{(diagonal)} \qquad [52]$$

o más explícitamente:

$$O^{-1} \cdot \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \cdot O = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix} \qquad [53]$$

en donde O es una matriz ortogonal que se caracteriza por cumplir:

$$O^{-1} = O^T \Leftrightarrow O_{ij}^{-1} = O_{ji} \quad (T = \text{transpuesta})$$

El teorema clave aquí establece que para matrices A reales y simétricas todos sus valores propios λ_i son *reales* y sus vectores propios son *ortogonales*. Evidentemente, los valores propios λ_i coinciden con las raíces del polinomio característico $P(\lambda)$ asociado a A.

El método de Jacobi construye la matriz O como un producto infinito de matrices de rotación O_n. A partir de la matriz original A se van eliminando en cada paso n una pareja de elementos $a_{ij}, a_{ji}\,(j \neq i)$ de fuera de la diagonal principal. En cada paso conviene, para cálculo manual, tomar para ser eliminado aquel elemento a_{ij} de mayor valor absoluto. Elegido éste se forma la matriz de rotación (ortogonal):

$$O_1 = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccccccc} & & & & i & & & j & & \\ & & & & | & & & | & & \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cos\phi & \cdots & -\operatorname{sen}\phi & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \operatorname{sen}\phi & \cdots & \cos\phi & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 1 \end{pmatrix} \end{array} \begin{array}{l} \\ \\ \\ -i \\ \\ -j \\ \\ \\ \end{array} \qquad [54]$$

El producto de matrices $O_1^{-1}AO_1 = B$ da para las cuatro posiciones que nos interesan (ii, ij, ji, jj):

$$\begin{aligned} b_{ii} &= a_{ii}\cos^2\phi + 2a_{ij}\operatorname{sen}\phi\cdot\cos\phi + a_{jj}\operatorname{sen}^2\phi \\ b_{ij} &= b_{ji} = (a_{jj} - a_{ii})\operatorname{sen}\phi\cdot\cos\phi + a_{ij}(\cos^2\phi - \operatorname{sen}^2\phi) \\ b_{jj} &= a_{ii}\operatorname{sen}^2\phi - 2a_{ij}\operatorname{sen}\phi\cdot\cos\phi + a_{jj}\cos^2\phi \end{aligned} \qquad [55]$$

Como pretendemos hacer $b_{ij} = b_{ji} = 0$, nos bastará con tomar un ángulo de rotación ϕ tal que:

$$tg\,2\phi = \frac{-a_{ij}}{(a_{jj} - a_{ii})/2} = \frac{\beta}{\gamma} \quad ; \quad \left.\begin{matrix}\beta = -a_{ij} \\ \xi^2 = \beta^2 + \gamma^2\end{matrix}\right\} \Rightarrow \cos\phi = \left(\frac{\xi + |\gamma|}{2\xi}\right)^{1/2} \qquad [56]$$

La operación se repite con la matriz B eliminando una nueva pareja de elementos vía $O_2^{-1}BO_2 = C$, y así sucesivamente.

Hay que notar, sin embargo, que cada etapa del algoritmo introduce correcciones no nulas en las posiciones que ya eran cero. A pesar de ello está garantizada la convergencia del método descrito, ya que se puede demostrar que la sucesión de matrices:

$$O_1^{-1}AO_1,\ O_2^{-1}O_1^{-1}AO_1O_2, \ldots, O_n^{-1}O_{n-1}^{-1}\ldots O_1^{-1}AO_1\ldots O_{n-1}O_n, \ldots \qquad [57]$$

converge hacia una matriz diagonal $O^{-1}AO = \Lambda$. Conviene tener presente que el algoritmo [56], con $|\phi| \leqslant \pi/4$, es la elección más estable numérica-

mente hablando (Ralston, 1970; Press y col., 1988). Los valores propios de A quedan en la diagonal principal de Λ, en tanto que los vectores propios son las columnas de la matriz O:

$$O = O_1 \cdot O_2 \cdot O_3 \cdot \ldots \cdot O_{n-1} \cdot O_n \ldots \qquad [58]$$

Es muy importante señalar que cada autovalor λ_i lleva asociado su autovector $\bar{x}_i = (x_{1i}, x_{2i}, \ldots, x_{ni})$ que está justamente situado en la columna i.

En la práctica se usan varios criterios para decidir cuándo se encuentra diagonalizada la matriz inicial. Uno de ellos consiste en exigir que todos los elementos no diagonales de la matriz resultante en el paso n sean menores en valor absoluto que una cierta cota de tolerancia (10^{-6}, etc.). Otros criterios usan comparaciones entre los elementos diagonales y los no diagonales (de la misma fila) para decidir la convergencia (Press y col., 1988).

Mencionaremos ahora un par de propiedades interesantes que pueden resultar de ayuda a la hora de comprobar si una diagonalización no es correcta. La primera de ellas es que la *traza* de la matriz A debe conservarse en Λ:

$$\mathrm{tr}\,(A) = \sum_{i=1}^{n} a_{ii} = \mathrm{tr}\,(\Lambda) = \sum_{i=1}^{n} \lambda_i \qquad [59]$$

es decir, la suma de los elementos de la diagonal principal se mantiene constante. La segunda es que el determinante de A también se conserva:

$$\det(A) = \det(\Lambda) = \prod_{i=1}^{n} \lambda_i = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n \qquad [60]$$

Se dice entonces que la traza y el determinante son cantidades *invariantes* bajo la diagonalización. La comprobación de estas propiedades es un ejercicio trivial que se deja a cargo del lector. Notemos que las condiciones [59]-[60] no nos dicen, caso de cumplirse, que la diagonalización efectuada sea correcta. No obstante, suelen ser suficientes para detectar errores de procedimiento.

Naturalmente, los cálculos de diagonalización salvo cuando la matriz es pequeña (4×4) no se realizan manualmente. Sin embargo, cuando determinadas características de *simetría* asociadas al problema permiten aplicar técnicas de *Teoría de Grupos* (Hernanz, 1991), el determinante secular puede quedar apropiadamente factorizado. Esto quiere decir que

el problema inicial de dimensión $(n \times n)$ se reduce a una serie de subproblemas independientes de órdenes tratables. Así una matriz A $(n \times n)$ puede sufrir una transformación de semejanza y convertirse en una matriz $\bar{A}$ diagonal por cajas:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \xrightarrow[\text{Grupos}]{\text{Teoría de}} \bar{A} =$$

$$= \begin{pmatrix} \boxed{\begin{matrix} \bar{a}_{11} & \bar{a}_{12} \\ \bar{a}_{21} & \bar{a}_{22} \end{matrix}} & 0 & \cdots & 0 & 0 \\ 0 & \boxed{\bar{a}_{33}} & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & & \\ 0 & 0 & & \boxed{\begin{matrix} \bar{a}_{n-1,n-1} & \bar{a}_{n-1,n} \\ \bar{a}_{n,n-1} & \bar{a}_{nn} \end{matrix}} & \\ 0 & 0 & 0 & \cdots & \end{pmatrix} \qquad [61]$$

La diagonalización de A es equivalente a diagonalizar las cajas de $\bar{A}$ por separado.

A pesar de todo, no siempre la reducción anterior nos deja tamaños manejables para el cálculo manual. En estos casos, es imprescindible utilizar un ordenador y diseñar un programa de diagonalización. ¿Cómo averiguar que tal programa diagonaliza correctamente, más allá de las comprobaciones [59]-[60]? Se utiliza entonces el denominado test de «la matriz de unos» $A(a_{ij} = 1;\ i = 1, 2, ..., n;\ j = 1, 2, ..., n)$. El método funcionará correctamente con respecto al cálculo de los autovalores cuando éstos sean todos nulos salvo uno de ellos que coincidirá con el orden de la matriz original. Si además los autovectores forman un conjunto *ortonormal*, ortogonales dos a dos y normalizados, entonces la diagonalización completa es correcta. Para captar el funcionamiento del método de Jacobi y aclarar la naturaleza de este test conviene realizar el Ejercicio 5.

Señalaremos, para terminar, que en los cálculos no manuales la opción más eficaz consiste en aniquilar ordenadamente los elementos no diagonales (*método cíclico*), en contra de lo indicado anteriormente para cálculo manual (selección del a_{ij} con mayor valor absoluto). Explícitamente, aniquilando por filas se eliminarían y por este orden: a_{12}, a_{13}..., a_{1n}, a_{23}..., etc.

3.3. Ejemplo de aplicación: Cálculo de orbitales moleculares de Hückel para el radical ciclopropenilo

La geometría del ciclopropenilo es triangular (Fig. 4) y hay tres electrones π, por lo que habrá una deslocalización electrónica por encima y debajo del plano del radical. Para estudiar este sistema π por el método de Hückel elegimos como base, para obtener los orbitales moleculares, a los tres orbitales atómicos $2p_z$ de los tres carbonos. La función de prueba a introducir en el cálculo variacional (con ε = energía de los orbitales moleculares):

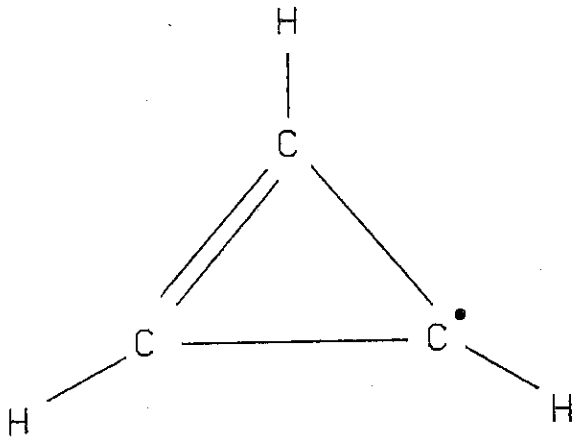


Figura 4.—Radical ciclopropenilo.

$$\begin{cases} \varepsilon = \langle\psi|H|\psi\rangle/\langle\psi|\psi\rangle = \int \psi^* H\psi \, d\tau / \int \psi^*\psi \, d\tau \\ \dfrac{\partial\varepsilon}{\partial c_i} = 0 \quad ; \quad i = 1, 2, 3 \end{cases} \qquad [62]$$

será la combinación lineal:

$$\psi = c_1(2p_z)_1 + c_2(2p_z)_2 + c_3(2p_z)_3 \qquad [63]$$

El sistema de ecuaciones seculares al que se llega es:

$$\sum_{i=1}^{3} c_i(H_{ji} - \varepsilon S_{ji}) = 0 \quad ; \quad j = 1, 2, 3 \qquad [64]$$

Las simplificaciones Hückel son: S_{ij} (solapamientos) $= \delta_{ij}$, donde δ es la función de Kroenecker; H_{ij} es:

i) la integral de Coulomb α cuando $i = j$

ii) la integral de resonancia $\beta < 0$ cuando los átomos i, j son contiguos

iii) nula cuando i, j no son contiguos.

Estamos frente a un problema de valores propios, ya que para que [64] tenga solución no trivial debe suceder:

$$|H_{ij} - \varepsilon\delta_{ij}| = 0 \quad ; \quad i,j = 1,2,3 \qquad [65]$$

Dividiendo todos los elementos del determinante por β y haciendo:

$$x = (\alpha - \varepsilon)/\beta \qquad [66]$$

obtenemos:

$$\begin{vmatrix} x & 1 & 1 \\ 1 & x & 1 \\ 1 & 1 & x \end{vmatrix} = 0 \qquad [67]$$

Resolver esta ecuación es equivalente a diagonalizar la matriz simétrica (H_{ij}) que descomponemos en la forma:

$$\begin{pmatrix} \alpha & \beta & \beta \\ \beta & \alpha & \beta \\ \beta & \beta & \alpha \end{pmatrix} = \beta \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad [68]$$

quedando reducido el problema a diagonalizar la matriz que multiplica a β. Resolveremos el problema del ciclopropenilo vía los dos métodos: polinomio característico y Jacobi.

a) Polinomio característico

La ecuación secular para el ciclopropenilo es:

$$P(x) = x^3 - 3x + 2 = 0 \qquad [69]$$

que puede resolverse por simple inspección ($x_1 = -2$; $x_2 = x_3 = 1$), pero vamos a ilustrar el modo general de ataque desde el punto de vista numérico. No es difícil acotar dos raíces de la ecuación en los intervalos $(-3, -1)$ y $(-1, 2)$. Utilizaremos el método de Newton y calcularemos primero la raíz en $(-3, -1)$.

a.1) Raíz en $(-3, -1)$

Sea $x_0 = -3$, de modo que se satisfaga [24]. La aplicación del algoritmo:

$$x_{n+1} = x_n - \frac{x_n^3 - 3x_n + 2}{3x_n^2 - 3} \qquad [70]$$

conduce a la sucesión convergente:

$$x_1 = -2{,}333333333$$
$$x_2 = -2{,}055555556$$
$$x_3 = -2{,}001949318$$
$$x_4 = -2{,}000002528$$
$$x_5 = -2{,}000000000$$

en la que se observa la esperada rápida convergencia al valor $x_1 = -2$. Estos resultados se han obtenido con una calculadora de escritorio convencional programable que trabaja con 13 dígitos significativos. Por ello en cinco pasos la precisión es ya de nueve cifras decimales. Tomamos como raíz el valor mencionado (que coincide con el exacto) despreciando cifras decimales superiores.

a.2) Raíz en $(-1, 2)$

El valor inicial debe ser aquí $x_0 = 2$ y por aplicación reiterada de [70] se obtiene la sucesión siguiente:

$$x_1 = 1{,}555555556$$
$$x_3 = 1{,}155390199$$
$$x_5 = 1{,}040288435$$
$$x_7 = 1{,}010172323$$

$$x_9 = 1{,}002549528$$
$$x_{11} = 1{,}000637788$$
$$x_{13} = 1{,}000159473$$
$$x_{15} = 1{,}000039872$$
$$x_{17} = 1{,}000009980$$
$$x_{19} = 1{,}000002506$$
$$x_{21} = 1{,}000000800$$
$$x_{23} = 1{,}000000592$$

cuya convergencia es impresionantemente lenta comparada con la de la raíz anterior.

Si bien desde el punto de vista práctico podemos asegurar que esta raíz será aproximadamente la unidad, el comportamiento mencionado y la existencia de una tercera raíz real para [69] sugieren una multiplicidad dos para la raíz que nos ocupa. Modificando el algoritmo de Newton con $m = 2$ obtenemos que [32] es:

$$x_{n+1} = x_n - 2\frac{x_n^3 - 3x_n + 2}{3x_n^2 - 3} \quad ; \quad n = 1, 2, 3, \ldots \qquad [71]$$

o equivalentemente:

$$x_{n+1} = x_n - 2\frac{(x_n + 2)(x_n - 1)}{3(x_n + 1)} \qquad [71']$$

y utilizando la entrada $x_0 = 2$ llegamos a la sucesión convergente (trabajando con doce decimales en [71]):

$$x_1 = 1{,}111111111$$
$$x_2 = 1{,}001949318$$
$$x_3 = 1{,}000000633$$
$$x_4 = 1{,}000000633$$

siendo más que notable la velocidad del método modificado. Esto confirma la característica de ser de orden dos esta raíz. Aunque podríamos aplicar otros métodos como verificación adicional y mejora del resultado numérico, vamos a detenernos aquí. La raíz que estábamos buscando es $x_2 = x_3 = 1$ con una precisión de seis cifras decimales. De continuar iterando [71] ($n = 5, 6, \ldots$) no se mejoraría la calidad del resultado numérico para alcanzar el valor exacto que sabemos es la unidad. El cálculo [71] con mayor número de cifras decimales (16) conduciría al valor

$x = 1 + 0{,}677 \times 10^{-13}$ en cuatro iteraciones, manteniéndose estable a partir de ahí. Si en lugar de [71] se hubiera trabajado con [71'], el resultado exacto $x = 1$ (16 decimales exactos) se alcanzaría en cinco iteraciones. El lector debe ser capaz de justificar estas circunstancias.

Calculadas las raíces de [69], $x_1 = -2$, $x_2 = x_3 = 1$, los valores propios ε_i son:

$$\varepsilon_i = \alpha - \beta x_i = \begin{cases} \varepsilon_1 = \alpha + 2\beta \\ \varepsilon_2 = \varepsilon_3 = \alpha - \beta \end{cases} \qquad [72]$$

Quedan por determinar los vectores propios (orbitales moleculares π) ψ_i asociados a cada uno de estos autovalores. Para ello, escribamos [65] utilizando la relación entre ε y x de arriba:

$$\begin{cases} c_1 x + c_2 + c_3 = 0 \\ c_1 + c_2 x + c_3 = 0 \\ c_1 + c_2 + c_3 x = 0 \end{cases} \qquad [73]$$

Para el valor $x = -2$ tenemos un orbital *enlazante*, y el sistema se reduce a:

$$\left.\begin{array}{r} -2c_1 + c_2 + c_3 = 0 \\ c_1 - 2c_2 + c_3 = 0 \\ \hline c_1^2 + c_2^2 + c_3^2 = 1 \end{array}\right\} \qquad [74]$$

que una vez resuelto lleva a:

$$c_1 = c_2 = c_3 = 1/\sqrt{3} \qquad [75]$$

lo que cabría esperarse por simetría, siendo entonces:

$$\psi_1(\varepsilon_1 = \alpha + 2\beta) = \frac{1}{\sqrt{3}}[(2p_z)_1 + (2p_z)_2 + (2p_z)_3] \qquad [76]$$

Para el nivel degenerado $x = 1$ (orbitales *antienlazantes*) hay que proceder en dos etapas. Primero calcular un autovector a partir del sistema reducido a una ecuación:

$$c_1 + c_2 + c_3 = 0 \qquad [77]$$

en donde tenemos absoluta libertad para fijar dos incógnitas $c_2 = c_3 = 1$,

lo que implica $c_1 = -2$. Con la normalización el orbital molecular queda:

$$\psi_2(\varepsilon_2 = \alpha - \beta) = \frac{1}{\sqrt{6}}[-2(2p_z)_1 + (2p_z)_2 + (2p_z)_3] \qquad [78]$$

El cálculo del tercer vector propio se reduce a eliminar, como antes, dos ecuaciones y exigir que la nueva solución sea ortogonal a [78]:

$$\left.\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ -2c_1 + c_2 + c_3 &= 0 \end{aligned}\right\} \qquad [79]$$

Haciendo $c_3 = 1$, resulta $c_1 = 0$, $c_2 = -1$, y normalizando:

$$\psi_3(\varepsilon_3 = \alpha - \beta) = \frac{1}{\sqrt{2}}[-(2p_z)_2 + (2p_z)_3] \qquad [80]$$

Es fácil comprobar que las dos soluciones ψ_2 y ψ_3 son ortogonales a ψ_1. Por ejemplo:

$$\langle \psi_3 | \psi_1 \rangle = \int \psi_3^* \cdot \psi_1 \cdot d\tau =$$

$$= \frac{1}{\sqrt{6}} \cdot \int [-(2p_z)_2 + (2p_z)_3]^* \cdot [(2p_z)_1 + (2p_z)_2 + (2p_z)_3]\, d\tau =$$

$$= \frac{1}{\sqrt{6}} \left[-\int (2p_z)_2^2\, d\tau + \int (2p_z)_3^2\, d\tau \right] = 0$$

pues los productos cruzados son nulos ($S_{ij} = \delta_{ij}$).

Se podrían haber obtenido otras soluciones ψ'_2, ψ'_3, distintas de las anteriores sin más que tomar otras elecciones para c_2 y c_3. La solución así calculada sería equivalente a la anterior y no alteraría para nada los parámetros moleculares que se calcularan con estas nuevas funciones con respecto a las antiguas. La operación que acabamos de describir es, en definitiva, una *rotación* de la base $\{\psi_2, \psi_3\} \rightarrow \{\psi'_2, \psi'_3\}$, por lo que el espacio que describen ambos conjuntos es el mismo (Fig. 5).

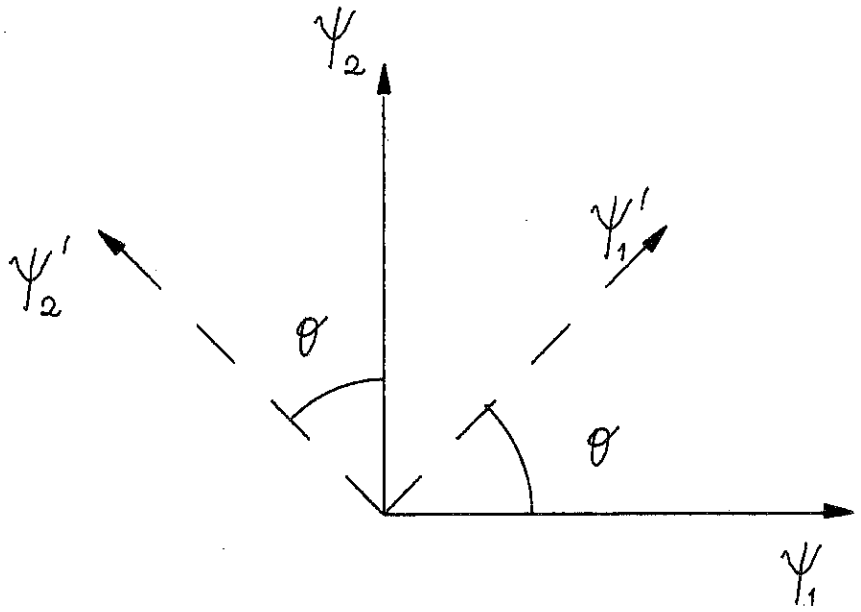


Figura 5.—Bases ortogonales equivalentes por rotación.

b) Jacobi

La ecuación [68] la reescribimos como:

$$H = \begin{pmatrix} \alpha & \beta & \beta \\ \beta & \alpha & \beta \\ \beta & \beta & \alpha \end{pmatrix} = \beta \cdot A + \alpha \cdot I \qquad [81]$$

Si deseamos diagonalizar H es obvio que:

$$E = O^{-1}HO = \beta \cdot (O^{-1}AO) + \alpha \cdot I \qquad [82]$$

pues I es ya diagonal. El problema se reduce a diagonalizar:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad [83]$$

Utilizando el método de Jacobi eliminemos la pareja $a_{12}, a_{21} = 1$:

$$\operatorname{tg} 2\phi = \frac{2a_{12}}{a_{11} - a_{22}} = \frac{2 \cdot 1}{0} \Rightarrow \phi = 45°$$

Así, se tiene la transformación:

$$B = O_1^{-1} A O_1 =$$

$$= \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & 0 & \sqrt{2} \\ 0 & -1 & 0 \\ \sqrt{2} & 0 & 0 \end{pmatrix}$$

A continuación eliminemos $b_{13} = b_{31} = \sqrt{2}$:

$$\operatorname{tg} 2\phi = \frac{2b_{13}}{b_{11} - b_{33}} = \frac{2\sqrt{2}}{1 - 0} = 2\sqrt{2}$$

Es inmediato obtener:

$$\operatorname{sen} \phi = 1/\sqrt{3} \quad ; \quad \cos \phi = \sqrt{2/3}$$

y con ellos la segunda transformación:

$$C = O_2^{-1} B O_2 =$$

$$= \begin{pmatrix} \sqrt{2/3} & 0 & \sqrt{1/3} \\ 0 & 1 & 0 \\ -\sqrt{1/3} & 0 & \sqrt{2/3} \end{pmatrix} \begin{pmatrix} 1 & 0 & \sqrt{2} \\ 0 & -1 & 0 \\ \sqrt{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2/3} & 0 & -\sqrt{1/3} \\ 0 & 1 & 0 \\ \sqrt{1/3} & 0 & \sqrt{2/3} \end{pmatrix} =$$

$$= \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Los valores propios de H a la vista de la relación [82] son:

$$\varepsilon_1 = \alpha + 2\beta \quad ; \quad \varepsilon_2 = \varepsilon_3 = \alpha - \beta \qquad [84]$$

y efectuando el producto $O = O_1 O_2$ se obtiene:

$$O = O_1 \cdot O_2 = \begin{pmatrix} 1/\sqrt{3} & \sqrt{2}/2 & -1/\sqrt{6} \\ 1/\sqrt{3} & -\sqrt{2}/2 & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & \sqrt{2/3} \end{pmatrix}$$

matriz cuyas columnas i son los orbitales moleculares ψ_i correspondientes a los ε_i:

$$\psi_1 = \frac{1}{\sqrt{3}}[(2p_z)_1 + (2p_z)_2 + (2p_z)_3] \qquad [85]$$

$$\psi_2 = \frac{1}{\sqrt{2}}[(2p_z)_1 - (2p_z)_2] \qquad [86]$$

$$\psi_3 = \frac{1}{\sqrt{6}}[-(2p_z)_1 - (2p_z)_2 + 2(2p_z)_3] \qquad [87]$$

Los resultados obtenidos concuerdan con los del polinomio característico, ya que los tres orbitales $2p_z$ son equivalentes, y un cambio de signo en una función de onda no afecta a sus propiedades probabilísticas.

El lector interesado en este tipo de cálculos puede encontrar provechosa la referencia (Dmitriev, 1981) en donde se presenta una atractiva discusión topológica del problema de las moléculas con enlaces π.

Bibliografía

1. APOSTOL, T., *Análisis Matemático,* Reverté, Barcelona, 1972, capítulos 5 y 12.

2. ARTEAGA, L., *Cálculo Numérico I,* Departamento de Apuntes de la Facultad de Ciencias Matemáticas, Universidad Complutense, Madrid, 1975, capítulo 4.

3. DMITRIEV, I. S., *Molecules without chemical bonds,* Mir, Moscú, 1981.

4. GALINDO, A., y PASCUAL, P., *Mecánica Cuántica,* Alhambra, Madrid, 1978, capítulo 4.

5. HERNANZ, A., *Métodos Teóricos de la Química Física* (vol. 2), Universidad Nacional de Educación a Distancia, Madrid, 1991.

6. PRESS, W. H.; FLANNERY, B. P.; TEUKOLSKI, S. A., y VETTERLING, W. T., *Numerical Recipes,* Cambridge University Press, Cambridge, 1988, capítulo 11.

7. RALSTON, A., *Introducción al Análisis Numérico,* Limusa, México, 1970, capítulos 8, 9 y 10.

8. RICE, J. R., *Numerical Methods, Software and Analysis,* McGraw-Hill, Nueva York, 1983, capítulos 6 y 8.

9. SCHEID, F., *Análisis Numérico,* McGraw-Hill, México, 1972, capítulo 25.

10. SOKOLNIKOFF, I. S., *Análisis Tensorial,* Index-Prial, Madrid, 1971, capítulo 1.

EJERCICIOS DE AUTOCOMPROBACION

1. Separar las raíces reales de $y(x) = x^4/2 - 2x - 7$.

2. Para la ecuación del ejercicio 1 calcular sus raíces reales aplicando los métodos:

   a) bisección

   b) iteración

   c) Newton

3. Una partícula cuántica monodimensional se encuentra atrapada en una caja de potencial con paredes finitas:

$$V(x) = \begin{cases} 0 & -\infty < x < -a \\ V_0 & -a < x < a \\ 0 & a < x < +\infty \end{cases} \quad ; \quad V_0 < 0$$

   Al resolver la ecuación de Schrödinger se pueden clasificar sus soluciones en pares e impares dada la simetría del potencial $V(x)$. Para las soluciones de paridad positiva sus valores propios asociados E (energías) se calculan a partir de la ecuación (Galindo y Pascual, 1978):

$$\cos \alpha a = \pm \alpha / \sqrt{|V_0|} \quad ; \quad \operatorname{tg} \alpha a > 0$$

en tanto que las de paridad negativa de:

$$\operatorname{sen} \alpha a = \pm \alpha / \sqrt{|V_0|} \quad ; \quad \operatorname{cotg} \alpha a < 0$$

En estas relaciones $\alpha = (E - V_0)^{1/2} > 0$. Obtener numéricamente los autovalores de los cuatro estados más bajos cuando $(a|V_0|^{1/2} = 9)$. (¡Cuidado con el cuadrante en el que deben estar las soluciones!)

4. Calcular los momentos principales de inercia $I^{(1)}$, $I^{(2)}$, $I^{(3)}$, para las moléculas CO_2, H_2O y CH_4 resolviendo la ecuación determinantal (secular) de tercer grado en I que involucra al *tensor de inercia* molecular:

$$P(I) = \begin{vmatrix} I_{xx} - I & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} - I & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} - I \end{vmatrix} = 0$$

en donde:

$$I_{xx} = \sum_i m_i(y_i^2 + z_i^2) \quad ; \quad I_{yy} = \sum_i m_i(x_i^2 + z_i^2) \quad ;$$

$$I_{zz} = \sum_i m_i(x_i^2 + y_i^2) \quad ; \quad I_{xy} = I_{yx} = \sum_i m_i x_i y_i \quad ;$$

$$I_{xz} = I_{zx} = \sum_i m_i x_i z_i \quad ; \quad I_{yz} = I_{zy} = \sum_i m_i z_i y_i$$

siendo m_i, x_i, y_i, z_i, la masa y las coordenadas de cada átomo i.

Datos: $l(C{-}O) = 1{,}162\ \text{Å}$; $O\hat{C}O = 180^\circ$; $l(O{-}H) = 0{,}960\ \text{Å}$;
$H\hat{O}H = 104{,}45^\circ$; $l(C{-}H) = 1{,}093\ \text{Å}$; $H\hat{C}H = 109{,}5^\circ$;
$m_H = 1{,}00797\ uma$; $m_c = 12{,}01115\ uma$;
$m_O = 15{,}9994\ uma$.

5. Diagonalizar por el método de Jacobi la matriz de tercer orden:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

(test de la «matriz de unos»).

SOLUCIONES A LOS EJERCICIOS DE AUTOCOMPROBACION

1. Calculando la primera derivada de la función $x^4/2 - 2x - 7$ encontramos:

$$y' = 2x^3 - 2$$

que posee un cero con multiplicidad tres en $x = 1$. Es claro entonces que el signo de y va como:

x	$-\infty$	1	$+\infty$
signo y	+	−	+

y habrá al menos dos raíces reales una en $(-\infty, 1)$ y otra en $(1, +\infty)$. No hacemos referencia a su posible multiplicidad, aunque un sencillo gráfico nos indica que son simples.

2. *a*) Tras alguna búsqueda localizamos las raíces reales en los intervalos $(-2, -1)$ y $(2, 3)$. Aplicando bisecciones sucesivas en cada caso obtenemos:

— Intervalo $(-2, -1)$:

$$(-\overset{+}{2}, -\overset{-}{1}) \to (-\overset{+}{2}, -\overset{-}{1{,}5}) \to (-\overset{+}{1{,}75}, -\overset{-}{1{,}5}) \to (-\overset{+}{1{,}75}, -\overset{-}{1{,}625}) \to$$

$$(-1,6\overset{+}{8}75,\ -1,\overset{-}{6}25) \rightarrow (-1,6\overset{+}{5}25,\ -1,\overset{-}{6}25) \rightarrow$$

$$(-1,65\overset{+}{6}25,\ -1,640\overset{-}{6}25) \rightarrow (-1,65\overset{+}{6}25,\ -1,648\overset{-}{4}375) \rightarrow \ldots$$

siendo la raíz: $-1,6\ldots$.

— Intervalo $(2, 3)$

$$(\overset{-}{2},\ \overset{+}{3}) \rightarrow (\overset{-}{2},\ 2,\overset{+}{5}) \rightarrow (\overset{-}{2},\ 2,\overset{+}{2}5) \rightarrow (2,\overset{-}{1}25,\ 2,\overset{+}{2}5) \rightarrow$$

$$(2,\overset{-}{1}25,\ 2,1\overset{+}{8}75) \rightarrow (2,1\overset{-}{5}625,\ 2,1\overset{+}{8}75) \rightarrow$$

$$(2,17\overset{-}{1}875,\ 2,1\overset{+}{8}75) \rightarrow \ldots$$

La raíz es $2,1\ldots$.

Como se ve, la convergencia es extremadamente lenta, pues a siete pasos sólo se garantiza una cifra decimal exacta.

b) La descomposición trivial de $y(x) = 0$ para aplicar el método de iteración es:

$$x_n = \frac{x_{n-1}^4}{4} - \frac{7}{2} = g(x_{n-1})$$

La condición suficiente de convergencia no se cumple en este caso, pues $|g'(x)| = |x^3| < 1$ solamente se da para $|x| < 1$, y sabemos por *a*) que las raíces no caen en $-1 < x < 1$. Por otra parte, el método basado en esta $g(x)$ es esencialmente divergente como se comprueba con un simple gráfico o con unos pocos cálculos. Incluso introduciendo un $x_0 = x_{\text{correcto}}$ hasta nueve cifras decimales (lo calcularemos enseguida) ¡conseguimos una rápida divergencia! Esto se debe a que los redondeos por encima de esas cifras actúan como fuente desestabilizadora.

Sin embargo, podemos proponer otra descomposición:

$$x_n = \pm\sqrt[4]{4x_{n-1} + 14} = g(x_{n-1})$$

a partir de la cual es inmediato concluir la convergencia con unos x_0 cercanos a las raíces (por la acotación de la derivada).

Así, tomando $x_0 = -1$ y la rama negativa:

$$x_n = -\sqrt[4]{4x_{n-1} + 14}$$

se calcula la sucesión convergente

$$\begin{aligned} x_1 &= -1{,}788279410 \quad & x_{12} &= -1{,}649453248 \\ x_2 &= -1{,}619965129 \quad & x_{13} &= -1{,}649453259 \\ x_3 &= -1{,}655985287 \quad & x_{14} &= -1{,}649453256 \\ x_4 &= -1{,}647995772 \quad & x_{15} &= -1{,}649453257 \\ &\ldots\ldots\ldots\ldots\ldots\ldots & & \end{aligned}$$

que converge a la raíz lentamente en forma de espiral.

Del mismo modo con $x_0 = 2$ y la rama positiva:

$$x_n = +\sqrt[4]{4x_{n-1} + 14}$$

tenemos

$$\begin{aligned} x_1 &= 2{,}165736771 \quad & x_5 &= 2{,}183586179 \\ x_2 &= 2{,}181871132 \quad & x_6 &= 2{,}183587554 \\ x_3 &= 2{,}183422808 \quad & x_7 &= 2{,}183587686 \\ x_4 &= 2{,}183571862 \quad & x_8 &= 2{,}183587698 \quad & x_9 &= 2{,}183587700 \end{aligned}$$

con una convergencia más rápida en forma monótona creciente.

Existen métodos para acelerar la convergencia y remitimos al lector a las referencias especializadas ya citadas. Es importante reparar en que según sea la descomposición de $y(x) = 0$ las soluciones pueden ser alcanzadas o no, por lo que al aplicar este método suele ser común tener que ensayar varias posibilidades.

c) Planteemos el método de Newton como:

$$x_n = x_{n-1} - \frac{\frac{x_{n-1}^4}{2} - 2x_{n-1} - 7}{2x_{n-1}^3 - 2} \quad ; \quad n = 1, 2, \ldots$$

Tomemos $x_0 = 3\,(y_0 \cdot y_0' > 0)$ para calcular la raíz en $(1, +\infty)$. En cinco iteraciones encontramos una solución con nueve decimales exactos:

$$x_1 = 2{,}471153846$$
$$x_2 = 2{,}233296676$$
$$x_3 = 2{,}185384128$$
$$x_4 = 2{,}183590148$$
$$x_5 = 2{,}183587700$$

Análogamente, con $x_0 = -2\,(y_0 \cdot y_0' > 0)$ para $(-\infty, 1)$ encontramos en cuatro iteraciones con nueve decimales correctos la raíz:

$$x_1 = -1{,}722222222$$
$$x_2 = -1{,}653202753$$
$$x_3 = -1{,}649463685$$
$$x_4 = -1{,}649453257$$

Comparando con *a*) y *b*) observamos una veloz convergencia aquí. Del comportamiento de estas soluciones deducimos que la ecuación posee dos raíces reales y dos complejas.

3. Las cuatro ecuaciones a resolver son $(\alpha a = x)$:

$$\cos x = \pm \frac{x}{9} \quad ; \quad \operatorname{tg} x > 0$$

$$\operatorname{sen} x = \pm \frac{x}{9} \quad ; \quad \operatorname{cotg} x < 0$$

Calculados los valores x solución, los cuatro más bajos, las energías de los estados ligados pedidos serán:

$$E = (x^2 - 81)/a^2$$

El algoritmo de Newton deberá aplicarse a los cuatro casos que explicitamos a continuación:

$$a)\quad x_n = x_{n-1} - \frac{9\cos x_{n-1} - x_{n-1}}{-9\,\text{sen}\, x_{n-1} - 1} \quad (y = 9\cos x - x = 0)$$

$$b)\quad x_n = x_{n-1} - \frac{9\cos x_{n-1} + x_{n-1}}{-9\,\text{sen}\, x_{n-1} + 1} \quad (y = 9\cos x + x = 0)$$

$$c)\quad x_n = x_{n-1} - \frac{9\,\text{sen}\, x_{n-1} - x_{n-1}}{9\cos x_{n-1} - 1} \quad (y = 9\,\text{sen}\, x - x = 0)$$

$$d)\quad x_n = x_{n-1} - \frac{9\,\text{sen}\, x_{n-1} + x_{n-1}}{9\cos x_{n-1} + 1} \quad (y = 9\,\text{sen}\, x + x = 0)$$

Con ayuda de unas sencillas representaciones gráficas podemos dar primeras estimaciones fiables a x_0 para cada uno de ellos. Estos valores iniciales se acogen a las condiciones i)-iv) del epígrafe 1.5. Se obtienen así:

a) $x_0 = \pi/4$ (radianes), con el intervalo $(\pi/4, \pi/2)$

$x_1 = 1{,}542947374$

$x_2 = 1{,}413668338$

$x_3 = 1{,}413129533$

$x_4 = 1{,}413129513$

$$x_{a1} = 1{,}413129513 \quad (\text{tg}\, x_{a1} > 0)$$

$$x_0 = 5\pi/2 \Rightarrow x_{a2} = 6{,}96842222 \quad (\text{tg}\, x_{a2} > 0)$$

b) $x_0 = 3\pi/2$ (radianes), con el intervalo $(\pi, 2\pi)$

$x_1 = 4{,}241150082$

$x_2 = 4{,}223938133$

$x_3 = 4{,}223869731$

$x_4 = 4{,}223869730$

$$x_{b1} = 4{,}223869730 \quad (\text{tg}\, x_{b1} > 0)$$

c) $x_0 = \pi$ (radianes), con el intervalo $(\pi/2, \pi)$

$x_1 = 2{,}825435750$

$x_2 = 2{,}822589849$

$x_3 = 2{,}822588658$

$$x_{c1} = 2{,}822588658 \quad (\text{tg}\, x_{c1} < 0)$$

$$x_0 = 5\pi/2 \Rightarrow x_{c2} = 8{,}261820644 \quad (\text{tg}\, x_{c2} < 0)$$

d) $x_0 = 2\pi$ (radianes), con el intervalo $(3\pi/2, 2\pi)$

$x_1 = 5{,}654866776$

$x_2 = 5{,}610814999$

$x_3 = 5{,}610163879$

$x_4 = 5{,}610163731$

$$x_{d1} = 5{,}610163731 \quad (\text{tg}\, x_{d1} < 0)$$

Las cuatro raíces más bajas son $x_{a1}(+)$, $x_{c1}(-)$, $x_{b1}(+)$, $x_{d1}(-)$, correspondientes alternativamente a estados de paridad positiva y negativa. Las energías resultan:

x	1,413129513	2,822588658	4,22386973
Ea^2	−79,00306498	−73,03299327	−63,1589245

5,610163731

−49,52606291

estando la profundidad del pozo en $V_0a^2 = -81$.

4. Hay que determinar primeramente las coordenadas atómicas. Hasta seis decimales éstas son

Molécula	Atomo	x	y	z
CO_2	C	0	0	0
	O	0	0	1,162
	O	0	0	−1,162
H_2O	O	0	0	0
	H	0	0	0,96
	H	0	0,929631	−0,239554
CH_4	C	0	0	0
	H	0,631044	0,631044	0,631044
	H	−0,631044	−0,631044	0,631044
	H	0,631044	−0,631044	−0,631044
	H	−0,631044	0,631044	−0,631044

Los tensores de inercia resultantes son (de nuevo 6 decimales):

$$CO_2: \begin{pmatrix} 43{,}206188 & 0 & 0 \\ 0 & 43{,}206188 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$H_2O: \begin{pmatrix} 1{,}857890 & 0 & 0 \\ 0 & 0{,}986789 & 0{,}224472 \\ 0 & 0{,}224472 & 0{,}871102 \end{pmatrix}$$

$$CH_4: \begin{pmatrix} 3{,}211123 & 0 & 0 \\ 0 & 3{,}211123 & 0 \\ 0 & 0 & 3{,}211123 \end{pmatrix}$$

De estos tensores es prácticamente inmediato deducir los momentos principales de inercia. Nótese que las matrices correspondientes al CO_2 y CH_4 están ya completamente diagonalizadas, en tanto que la del H_2O lo está por cajas. Vamos a aplicar, no obstante, el método de Newton para calcular las raíces del polinomio característico, poniendo de manifiesto algunas particularidades de este método en casos patológicos.

Los polinomios característicos se obtienen evaluando el determinante del tensor de inercia restando la variable I en la diagonal principal:

$$P_{CO_2}(I) = -I^3 + 86{,}412375\ I^2 - 1866{,}774656\ I = 0$$

$$P_{H_2O}(I) = -I^3 + 3{,}715780\ I^2 - 4{,}260962\ I + 1{,}503415 = 0$$

$$P_{CH_4}(I) = -I^3 + 9{,}633368\ I^2 - 30{,}933924\ I + 33{,}110873 = 0$$

Comencemos por el H_2O, cuya ecuación iterativa es:

$$I_n = I_{n-1} - \frac{I_{n-1}^3 - 3{,}715780\ I_{n-1}^2 + 4{,}260962\ I_{n-1} - 1{,}503415}{3I_{n-1}^2 - 2 \text{ x } 3{,}715780\ I_{n-1} + 4{,}260962}$$

Las raíces obtenidas son:

$I_0 = 0{,}5$	$I_0 = 1$	$I_0 = 2$
$I_1 = 0{,}636567$	$I_1 = 1{,}244827$	$I_1 = 1{,}888836$
$I_2 = 0{,}688554$	$I_2 = 1{,}162107$	$I_2 = 1{,}859871$
$I_3 = 0{,}696926$	$I_3 = 1{,}160755$	$I_3 = 1{,}857896$
$I^{(3)} = I_4 = 0{,}697139$	$I^{(2)} = I_4 = 1{,}160753$	$I^{(1)} = I_4 = 1{,}857887$

La convergencia es muy rápida y el efecto de los redondeos en los datos de entrada (tensor de inercia) no es importante (compárese $I^{(3)} = 1{,}857887$ con el valor diagonal 1,857890).

Pasemos al CO_2, del cual una solución es $I^{(3)} = 0$. Planteando (con cierta irreflexión) la ecuación iterativa:

$$I_n = I_{n-1} - \frac{I_{n-1}^2 - 86{,}412375\, I_{n-1} + 1866{,}774656}{2I_{n-1} - 86{,}412375}$$

se obtiene:

$I_0 = 30$	...
$I_1 = 36{,}603093$	$I_{12} = 43{,}199557$
$I_2 = 39{,}904637$	$I_{13} = 43{,}199801$
$I_3 = 41{,}555407$ ...	$I^{(2)} = I_{14} = 43{,}199806$

resultado que nos indica multiplicidad dos para la raíz. Utilizando el algoritmo modificado con pendiente mitad:

$$I_n = I_{n-1} - 2\,\frac{I_{n-1}^2 - 86{,}412375\, I_{n-1} + 1866{,}774656}{2I_{n-1} - 86{,}412375}$$

se llega a un comportamiento oscilante:

$$I_0 = 30$$
$$I_1^* = 43{,}206185 \quad ; \quad I_2 = I_0 \quad ; \quad I_3 = I_1^* \quad ; \quad \ldots$$

Esto se debe a que I_1^* es ya la mejor aproximación posible a la raíz y, además, se trata de un punto de tangente prácticamente horizontal. Encontramos que I_1^* hace prácticamente nulos, salvo redondeos, al numerador y denominador de la corrección. Son estos redondeos los responsables de la oscilación. Reescribamos el ciclo iterativo en la forma:

$$I_n = I_{n-1} - \frac{(I_{n-1} - x)^2 + \varepsilon}{(I_{n-1} - x)}$$

donde ε es el «error» (de redondeo) necesario para identificar las dos expresiones $y\ x = (86{,}412375)/2$. Es fácil comprobar que:

$$I_0 = I_2 = I_4 = \ldots \quad ; \quad I_1 = I_3 = I_5 = \ldots$$

Aunque aparentemente similar al caso previo, el cálculo para el CH_4 conduce a una situación drásticamente diferente. El orden es ahora tres:

$$I_n = I_{n-1} - 3\,\frac{I_{n-1}^3 - 9{,}633368\ I_{n-1}^2 + 30{,}933924\ I_{n-1} - 33{,}110873}{3\ I_{n-1}^2 - 2 \times 9{,}633368\ I_{n-1} + 30{,}933924}$$

$I_0 = 4$ ...
$I_1^* = 3{,}211123$ $I_{13} = 1{,}656941$
$I_2 = 2{,}685206$ $I_{14} = 1{,}656842$
$I_3 = 2{,}334596$ ... $I_{15} = 1{,}656842$

Observamos entonces que tanto para el CO_2 como para el CH_4 las primeras estimaciones I_1^* son excelentes y el cálculo debería haberse detenido ahí. Sin embargo, con orden tres la solución iterativa no oscila, sino que converge a un valor erróneo que no verifica la ecuación (un análisis similar al de orden dos es ligeramente tedioso y lo omitimos). Es práctica recomendable, pues, comprobar que el ciclo iterativo no pasa por puntos de tangente quasi-horizontal. Por otra parte, el número de cifras significativas en el cálculo debe ser mantenido tan alto como sea posible, buscando evitar amplificaciones de error por el algoritmo.

Resumiendo, los momentos de inercia esperados y calculados vía método de Newton son:

$I(uma.A^2)$	$I^{(1)}$	$I^{(2)}$	$I^{(3)}$	$I_N^{(1)}$	$I_N^{(2)}$	$I_N^{(3)}$
CO_2	43,206188	43,206188	0	—oscilación—...		0
H_2O	1,857890	1,160753	0,697139	1,857887	1,160753	0,697139
CH_4	3,211123	3,211123	3,211123	—convergencia errónea—...		

Los resultados correctos ilustran los hechos generales siguientes:

- moléculas trompo-simétricas lineales (CO_2): $I^{(1)} = I^{(2)} \neq 0$; $I^{(3)} = 0$
- moléculas trompo-asimétricas (H_2O): $I^{(1)} \neq I^{(2)} \neq I^{(3)}$ no nulos
- moléculas trompo-esféricas (CH_4): $I^{(1)} = I^{(2)} = I^{(3)} \neq 0$

5. Comencemos eliminando a_{12} y a_{21}. Aplicando [56] obtendremos el ángulo de rotación ϕ:

$$\operatorname{tg} 2\phi = \frac{2a_{21}}{a_{22} - a_{11}} = \frac{2}{0} \Rightarrow \phi = 45°$$

La matriz de rotación O_1 a aplicar sobre A será por tanto:

$$O_1 = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

y la transformación a que lleva es: $(O_1^{-1} = O_1^T)$:

$$B = O_1^{-1} A O_1 =$$

$$= \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 2 & 0 & \sqrt{2} \\ 0 & 0 & 0 \\ \sqrt{2} & 0 & 1 \end{pmatrix}$$

El siguiente paso es eliminar b_{13} y b_{31}. Como antes tenemos:

$$\operatorname{tg} 2\phi = -\frac{2b_{13}}{b_{33} - b_{11}} = -\frac{2\sqrt{2}}{1-2} = 2\sqrt{2}$$

Por trigonometría elemental llegamos a:

$$\cos 2\phi = \frac{1}{3} \quad ; \quad \operatorname{sen} \phi = \frac{1}{\sqrt{3}} \quad , \quad \cos \phi = \sqrt{2/3}$$

(cualquier otra elección de signos para sen ϕ, cos ϕ, sólo alteraría el orden en que aparecerían los autovalores y autovectores). Utilizando estos resultados tenemos:

$$C = O_2^{-1} B O_2 =$$

$$= \begin{pmatrix} \sqrt{2/3} & 0 & \sqrt{1/3} \\ 0 & 1 & 0 \\ -\sqrt{1/3} & 0 & \sqrt{2/3} \end{pmatrix} \begin{pmatrix} 2 & 0 & \sqrt{2} \\ 0 & 0 & 0 \\ \sqrt{2} & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2/3} & 0 & -\sqrt{1/3} \\ 0 & 1 & 0 \\ \sqrt{1/3} & 0 & \sqrt{2/3} \end{pmatrix} =$$

$$= \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \Lambda$$

con lo que, en este sencillo ejemplo, hemos alcanzado la diagonalización de A en dos pasos. Los autovalores son:

$$\lambda_1 = 3 \quad ; \quad \lambda_2 = \lambda_3 = 0$$

y vemos que se conserva la traza ($=3$) y el determinante ($=0$), cumpliéndose también la regla para autovalores del test mencionado en el texto.

Resta ahora calcular los vectores propios asociados a los tres autovalores anteriores. Efectuando el producto [58] con $n = 2$ se obtiene la matriz ortogonal global O:

$$O = O_1 \cdot O_2 = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2/3} & 0 & -\sqrt{1/3} \\ 0 & 1 & 0 \\ \sqrt{1/3} & 0 & \sqrt{2/3} \end{pmatrix} =$$

$$= \begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & \sqrt{2/3} \end{pmatrix}$$

Los vectores propios son pues:

$$\bar{x}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix} \quad ; \quad \bar{x}_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} \quad ; \quad \bar{x}_3 = \begin{pmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ \sqrt{2/3} \end{pmatrix}$$

que como puede apreciarse forman un conjunto ortonormal. El lector puede verificar que las parejas $(\lambda_i, \bar{x}_i)$ cumplen la ecuación básica [44].

ESTADISTICA TEORICA

(Temas 6-9)

Tema 6

VARIABLES ALEATORIAS Y DISTRIBUCIONES DE PROBABILIDAD

Este tema se dedica a presentar los conceptos fundamentales sobre *variables aleatorias* y sus *distribuciones de probabilidad.* Se asume que el lector posee los rudimentos básicos de la Teoría de Probabilidades. Primeramente se analiza el concepto de *función de distribución* para una variable independiente, es decir las probabilidades frente a los posibles valores de la variable. En el caso de variables discretas la función de distribución será una tabla numérica, en tanto que para las continuas tendremos expresiones analíticas. La formulación de la función de distribución admite dos versiones relacionadas: la *función integral* (*cumulativa*) y la *función densidad.* Para caracterizar una distribución existen aún otras formas de las que hablaremos posteriormente (*funciones generatriz* y *característica*).

La generalización a dos o más variables es directa y una vez realizada pasamos a estudiar los parámetros que permiten captar de una manera rápida la naturaleza de una distribución. Estos parámetros son llamados, en general, *momentos* y con ellos puede construirse la ya mencionada función generatriz. Estudiadas las funciones generatriz y característica, a continuación nos ocuparemos de la *convolución* que es una operación extremadamente importante cuando se trabaja con distribuciones de probabilidad.

Para concluir el tema se muestra cómo calcular valores medios y dispersiones en Termodinámica Estadística, anticipando resultados posteriores sobre la forma de la función de distribución *gaussiana* que veremos en el próximo tema.

1. FUNCIONES DE DISTRIBUCION PARA UNA VARIABLE ALEATORIA

El concepto de *variable aleatoria* o *estocástica* está en la mente de todos, ya que mediante él se pretende expresar el *azar* o *impredecibilidad* que poseen una gran variedad de sucesos. Para este tipo de sucesos es imposible asegurar con toda certeza si se van a dar o no en un momento dado. Sin embargo, lo que sí puede hacerse es asignarles una medida *a priori* de la posibilidad de que se presenten. A esta medida la denominamos *probabilidad* del suceso.

Supongamos el experimento de arrojar un dado, o el de la lectura de un amperímetro. En ambos casos el resultado del ensayo es variable, pero existe una diferencia fundamental entre ambos. En el primero el conjunto de sucesos posibles es *numerable*, en tanto que en el segundo no lo es. En el caso del dado deberemos numerar los sucesos asignándoles así un valor entero identificativo. Para el amperímetro la lectura es ya su identificación. Designando por X a una variable que recorra los valores asociados a los sucesos la llamaremos variable aleatoria asociada al experimento. Al conjunto de sucesos le denominaremos *espacio de estados*.

Si el espacio de estados es discreto o numerable la variable aleatoria X será discreta, mientras que si tal espacio es no numerable X será continua. Evidentemente puede haber espacios de estado con parte discreta y parte continua, pero no nos fijaremos en ellos aquí.

La cuestión ahora es determinar la probabilidad asociada a los sucesos o estados del experimento. Es éste un asunto en modo alguno trivial. Si la variable X es discreta, $\{x_k\}$, puede obtenerse una estimación de las probabilidades $p_k(x_k)$ utilizando la definición clásica de probabilidad y haciendo un número n de experimentos muy grande:

$$p_k(x_k) = \lim_{n \to \infty} \frac{f(x_k)}{n} \quad ; \quad 0 \leqslant p_k \leqslant 1 \qquad [1]$$

donde $f(x_k)$ es la *frecuencia* o número de veces que se ha presentado x_k en los n ensayos. El *principio de estabilidad de las frecuencias relativas* (Cramér, 1977) permite identificar la probabilidad p_k con el límite expresado en [1]. En otros casos, argumentos de simetría indican los valores p_k (en un dado uniforme cada cara tiene 1/6), y hay situaciones en las que la ley que siguen las p_k se ajusta a algún modelo matemático conocido. De

esta manera se puede construir el esquema probabilista (finito o infinito numerable):

X	x_0	x_1	x_2	x_3	...
$P(X)$	p_0	p_1	p_2	p_3	...

[2]

que debe cumplir la condición de normalización

$$\sum_k p_k = 1$$

Si la variable X es continua, $x_0 \leqslant X \leqslant x_n$, el problema aunque similar es más complicado. Puede discretizarse la variable X y proceder como antes para obtener un esquema del tipo [2]. Lo normal en muchas aplicaciones fisicoquímicas es utilizar leyes deducidas teóricamente que expresan la funcionalidad $p(x)$.

Estas leyes que dan la probabilidad de los sucesos posibles son denominadas *funciones de distribución* y admiten dos formas de las que hablaremos a continuación. Es muy importante notar que las probabilidades una vez fijadas, las supondremos *constantes* para todos los ensayos posteriores sobre un mismo espacio de estados, de ahí que se las califique de probabilidades *a priori*.

1.1. Variables discretas

Sea la variable discreta X que toma los valores x_k con probabilidades p_k:

$$P(X = x_k) = p_k \quad ; \quad k = 0, 1, 2, ... \qquad [3]$$

En sí misma la relación [3] es ya una función de distribución de probabilidades que suele recibir el nombre de *diagrama de probabilidad* (Fig. 1).

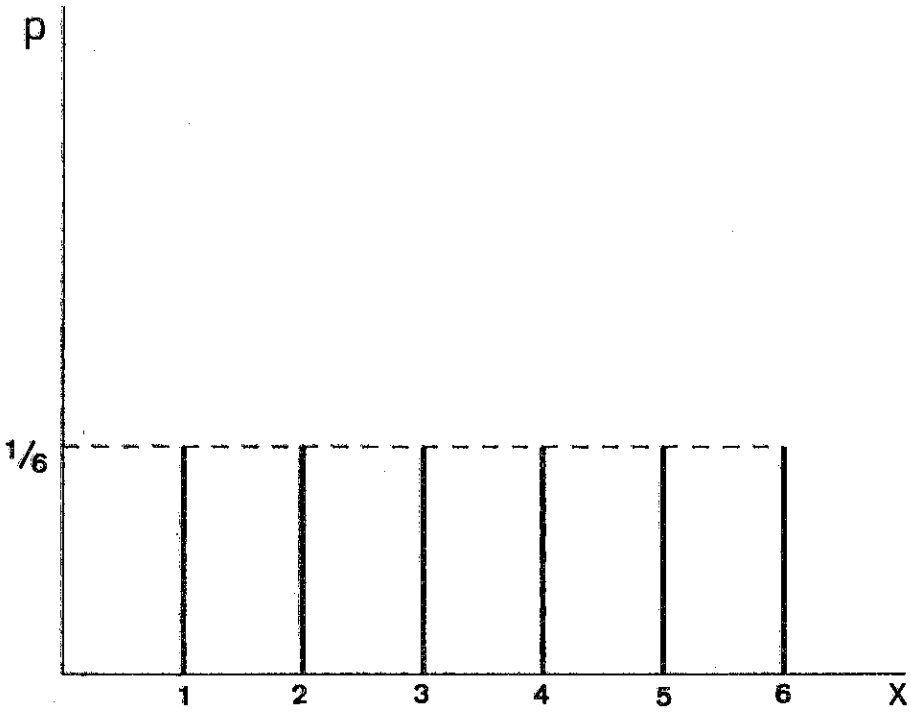


Figura 1.—Diagrama de probabilidad para un dado uniforme.

Además de la función [3] puede construirse una nueva función de distribución, la *función cumulativa* $F(x)$ que se define como:

$$F(X \leqslant x) = F(x) = \sum_{x_k \leqslant x} p_k \qquad [4]$$

extendiéndose la suma sobre todos los puntos x_k que sean menores o iguales al valor x que se esté evaluando. Este tipo de función es escalonada, no decreciente y presenta discontinuidades en cada x_k con un salto p_k (Fig. 2).

Es importante darse cuenta de que $F(X < x_0) = 0$ y que $F(X \geqslant x_n) = 1$, siendo x_n el último valor posible para X.

En este punto resulta atractivo utilizar una analogía mecánica para interpretar la tabla expresada en [3] (Cramér, 1977). Esta tabla puede asimilarse a la distribución de una masa unidad sobre los puntos x_k del eje real en cantidades puntuales p_k. La función cumulativa $F(x)$ de [4] es pues la cantidad de masa acumulada sobre el eje x a medida que nos

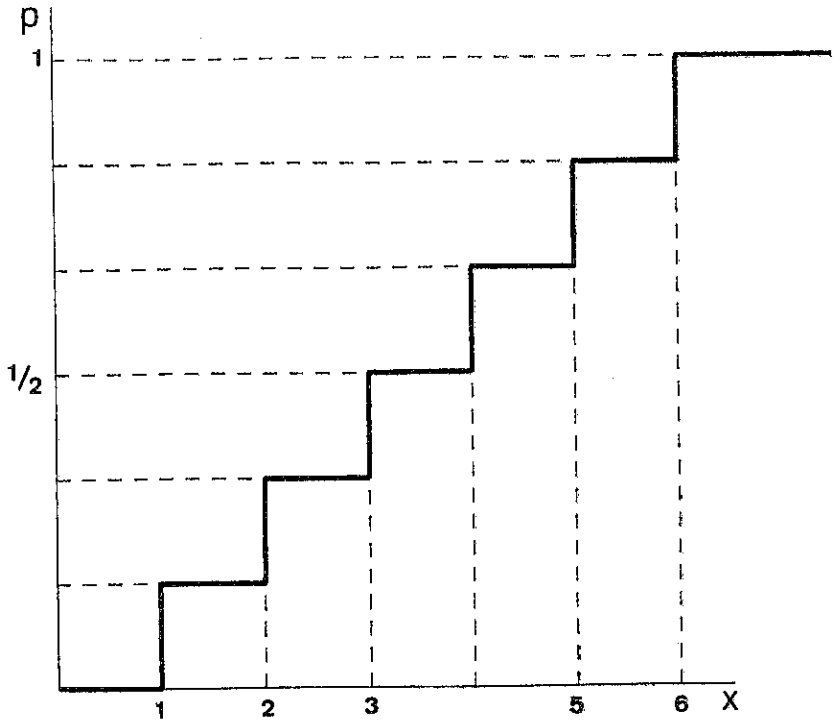


Figura 2.—Función cumulativa para un dado uniforme.

desplazamos hacia la derecha. Esta analogía es útil, pues teoremas mecánicos sobre el sólido rígido (momentos de inercia, etc.) son directamente aplicables en este contexto.

1.2. Variables continuas

Utilizando la analogía mecánica previa, consideremos una distribución de masa continua sobre el eje real que responda a la *función densidad* $f(x)$ $(-\infty < x < +\infty)$. Tal función f debe ser continua, salvo a lo sumo en un conjunto de puntos de discontinuidad de los cuales deberá haber un número finito dentro de cualquier intervalo arbitrario finito de la recta. Dicho en otros términos, el conjunto de puntos de discontinuidad puede ser infinito pero obligatoriamente debe ser numerable (conjunto de *medida* —longitud— *nula*) (Fano, 1968; Rudin, 1974).

Esta función $f(x)$ es justamente la generalización de la función $P(x)$ dada en [3], ya que la cantidad de masa (probabilidad) contenida en un intervalo infinitesimal $(x, x + dx)$ de anchura dx es: $f(x)\,dx$ (masa = densidad × longitud). Notemos que al pasar al continuo hablamos de probabilidades en intervalos. El valor $f(x)$ representa así la densidad de probabilidad de que X tome un valor en $(x, x + dx)$ y, por tanto, se la llama función densidad de probabilidad (o función de frecuencia) para la distribución. La probabilidad infinitesimal de que $x < X < x + dx$ es entonces $f(x)\,dx$.

Aparte de las condiciones de continuidad enunciadas, $f(x)$ debe ser no negativa:

$$f(x) \geqslant 0 \quad ; \quad -\infty < x < +\infty \qquad [5]$$

y evidentemente tal que:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad \text{(normalización)} \qquad [6]$$

resultado que expresa que la probabilidad de que aparezca cualquier valor x es la unidad (masa total = 1). Si se desea averiguar la probabilidad de que X tome un valor dentro del intervalo $a < X \leqslant b$, entonces se calculará la integral:

$$P(a < X \leqslant b) = \int_{a}^{b} f(x)\,dx \qquad [7]$$

es decir el área limitada por $f(x)$ y el eje x entre $x = a$ y $x = b$. De la relación [7] se sigue que:

$$P(X = a) = 0 \quad ; \quad P(a \leqslant X \leqslant b) = P(a < X \leqslant b) = \ldots \qquad [8]$$

que expresada en palabras nos dice que la probabilidad de obtener exactamente un resultado puntual es nula. Esta conclusión choca con lo visto para distribuciones discretas, pero la situación es claramente distinta. Un resultado puntual exacto para un experimento que involucre a una variable continua es *imposible* de obtenerse, ya que exigiría *una precisión de medida infinita* lo que no puede conseguirse en la práctica. Lo más acertado en estos casos es buscar la probabilidad de que el valor de X aparezca dentro de un intervalo pequeño pero medible $(a - \varepsilon, a + \varepsilon)$ a través de:

$$P(a - \varepsilon < X \leqslant a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x)\,dx \qquad [9]$$

Análogamente a lo realizado para el caso discreto puede definirse la *función de distribución cumulativa o integral* $F(x)$:

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt \qquad [10]$$

donde se utiliza como variable de integración t para evitar confusiones. En términos de la función integral la probabilidad calculada en [7] puede ponerse:

$$0 \leqslant P(a < X \leqslant b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a) \qquad [11]$$

de donde concluimos que $F(x)$ es, por definición, *no negativa* y, por tanto, *no decreciente.*

Puesto que la variable X sólo puede tomar valores finitos es inmediato establecer:

$$F(+\infty) = \lim_{x \to +\infty} F(x) = P(X < +\infty) = 1 \qquad [12]$$

$$F(-\infty) = \lim_{x \to -\infty} F(x) = P(X < -\infty) = 0 \qquad [13]$$

Para concluir señalaremos que la función densidad $f(x)$ es la derivada de la función cumulativa $F(x)$

$$f(x) = \frac{d}{dx} F(x) \qquad [14]$$

Conviene aquí que el lector realice el Ejercicio 1. Consideremos a continuación el siguiente ejercicio de aplicación.

Ejercicio de aplicación

Una variable aleatoria continua X responde a la función de distribución $F(x)$. La observación del valor x resultante de un ensayo nos permite

evaluar el valor de una variable Y dependiente linealmente de X: $Y = aX + b$ $(a > 0)$. Determinar la forma de las funciones de distribución integral y densidad a que obedecerá la variable aleatoria Y.

Sean las funciones buscadas $G(y)$, $g(y)$, siendo $g(y) = G'(y)$. Teniendo en cuenta que $a > 0$ será:

$$G(y) = P(Y \leqslant y) = P\left(X \leqslant \frac{y-b}{a}\right) = F\left(\frac{y-b}{a}\right)$$

Por derivación respecto de y se obtendrá la función densidad

$$g(y) = \frac{dG(y)}{dy} = \frac{dG}{dx} \cdot \frac{dx}{dy} = \frac{1}{a} \cdot \frac{dF}{dx}\bigg|_{x=\frac{y-b}{a}} = \frac{1}{a} f\left(\frac{y-b}{a}\right)$$

Existe no obstante una llamativa manera de realizar esta operación utilizando la función δ-Dirac. Con esta herramienta las funciones densidad $f(x)$ y $g(y)$ están relacionadas a través de

$$g(y) = \int_{-\infty}^{\infty} \delta[h(x) - y] f(x)\, dx \qquad [15]$$

cuando las variables aleatorias X e Y lo están por la relación funcional $Y = h(X)$ (Kampen, 1985). No es muy complicado hacerse una idea del sentido de [15]. La «función» δ nos traslada la densidad de masa situada en el punto x sobre el punto imagen $y = h(x)$, repartiéndose la masa total asociada a X sobre la variable Y. La operación [15] es, en definitiva, una redistribución de la probabilidad inicial. El lector puede verificar estas aseveraciones valiéndose de un sencillo ejemplo que involucre a dos variables discretas ($n = m^2$; $m = 0, 1, 2, 3$; $p_m = 1/4$).

En nuestro caso de relación lineal el cálculo de $g(y)$ puede realizarse como se indica:

$$g(y) = \int_{-\infty}^{\infty} \delta(ax + b - y) f(x)\, dx = \int_{-\infty}^{\infty} \delta\left[a\left(x - \frac{y-b}{a}\right)\right] f(x)\, dx =$$

$$= a^{-1} f\left(\frac{y-b}{a}\right)$$

en donde se ha aplicado:

$$\delta(ax) = a^{-1}\delta(x) \quad ; \quad a > 0$$

Por integración deduciríamos $G(y)$. La asimilación correcta de este curioso método es fundamental para abordar con garantías el estudio del Tema 7.(Ejercicio 2).

2. FUNCIONES DE DISTRIBUCION PARA DOS O MAS VARIABLES ALEATORIAS

En muchas ocasiones no será posible expresar el resultado de un experimento aleatorio mediante una única cantidad observada. Supongamos que se necesitan dos cantidades X e Y simultáneamente para caracterizar tal resultado. Si el recorrido de ambas variables es finito o infinito numerable, nos encontraremos en el caso discreto, mientras que en caso contrario estaremos en el continuo. Podemos en ambos definir las funciones de distribución de probabilidad conocidas, pero, ahora, *bidimensionales.*

Para distribuciones bidimensionales discretas la función de probabilidades vendrá dada por:

$$P_{kj} = P(X = x_k,\ Y = y_j) \quad ; \quad k = 0, 1, ..., n, ... \quad ; \quad j = 0, 1, ..., m, ... \quad [16]$$

que nos da la probabilidad de que simultáneamente $X = x_k$, $Y = y_j$. La masa total de esta distribución debe ser la unidad

$$\sum_k \sum_j P_{kj} = 1 \quad ; \quad 0 \leqslant P_{kj} \leqslant 1 \qquad [17]$$

y podemos imaginarla como distribuida en fracciones P_{kj} situadas en los nudos (x_k, y_j) de una malla rectangular.

Preguntémonos por la probabilidad P_k de que $X = x_k$ (o P_j de que $Y = y_j$) independientemente de lo que le ocurra a la variable Y (o a la X). Por ser los sucesos definidos por los pares (x_k, y_j) independientes dos a dos, es elemental establecer:

$$P_k^x = \sum_j P_{kj} \quad ; \quad P_j^y = \sum_k P_{kj} \qquad [18]$$

De esta manera se obtienen dos distribuciones discretas monodimensionales que reciben el nombre de *distribuciones marginales* de X y de Y, respectivamente. Es fácil ver el sentido gráfico de esta operación desde la imagen de la distribución de masas (proyecciones sobre los ejes x e y).

La definición de la función cumulativa $F(x, y)$ asociada a P_{kj} es la generalización de la fórmula [4]:

$$F(x, y) = P(X \leqslant x,\ Y \leqslant y) = \sum_{x_k \leqslant x} \sum_{y_j \leqslant y} P_{kj} \qquad [19]$$

función que comparte con $F(x)$ propiedades análogas.

Pasando ahora al caso continuo tomemos la unidad de masa repartida sobre el plano xy de acuerdo a una función densidad $f(x, y)$ $(-\infty < x < +\infty;\ -\infty < y < +\infty)$. Esto significa que la probabilidad de que en un determinado experimento la variable X caiga en $(x, x + dx)$ y la Y lo haga en $(y, y + dy)$ es:

$$P(x < X < x + dx,\ y < Y < y + dy) = f(x, y)\, dx\, dy \qquad [20]$$

La función $f(x, y)$ debe ser no negativa, normalizable

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} f(x, y)\, dy = 1 \quad ; \quad f(x, y) \geqslant 0 \qquad [21]$$

y continua *casi en todas partes* (con la salvedad de ciertas líneas del plano en donde pueden existir discontinuidades; el área de una línea es ciertamente nula).

Si se desea calcular la probabilidad de que la pareja X, Y caiga dentro de un rectángulo arbitrario $a_1 \leqslant X < a_2$, $b_1 \leqslant Y < b_2$, evaluaremos la integral:

$$P(a_1 \leqslant X < a_2,\ b_1 \leqslant Y < b_2) = \int_{a_1}^{a_2} dx \int_{b_1}^{b_2} f(x, y)\, dy \qquad [22]$$

Las distribuciones marginales de X e Y se definen aquí como:

$$f_x = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad [23]$$

$$f_y = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad [24]$$

en tanto que sus funciones integrales:

$$F_x = P(X \leqslant x) = \int_{-\infty}^{x} du \int_{-\infty}^{\infty} f(u, v)\, dv \qquad [25]$$

$$F_y = P(Y \leqslant y) = \int_{-\infty}^{y} dv \int_{-\infty}^{\infty} f(u, v)\, du \qquad [26]$$

La función integral conjunta es evidentemente:

$$F(x, y) = P(X \leqslant x, Y \leqslant y) = \int_{-\infty}^{x} du \int_{-\infty}^{y} f(u, v)\, dv \qquad [27]$$

Consideremos la situación muy importante en las aplicaciones, de que las variables X y Y sean *independientes*. Esto ocurrirá siempre que al realizar un experimento el resultado para X *no esté relacionado* (*correlacionado*) *de ningún modo* con el resultado para Y (en el Tema 8 profundizaremos más sobre este punto). Si esto es así, entonces utilizando la conocida regla del producto de probabilidades se tiene:

$$P(a_1 \leqslant X < a_2,\ b_1 \leqslant Y < b_2) = P(a_1 \leqslant X < a_2) \cdot P(b_1 \leqslant Y < b_2) \qquad [28]$$

y, por lo tanto, haciendo $a_1,\ b_1 \to -\infty$, resulta:

$$F(x, y) = F_x \cdot F_y \qquad [29]$$

y con $a_1 = x,\ a_2 = x + dx$, etcétera, se obtiene:

$$f(x, y) = f_x \cdot f_y \qquad [30]$$

(en el caso discreto se obtendrían expresiones análogas).

Quedan por considerar, para concluir este epígrafe, las distribuciones conjuntas de n variables aleatorias: $X_1, X_2, ..., X_n$. La generalización del caso previo es directa y la haremos sólo para el caso continuo.

La densidad de probabilidad $f(x_1, x_2, ..., x_n)$ representa la densidad de masa esparcida sobre un espacio n-dimensional, de modo que:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n)\, dx_1 \cdot dx_2 \ldots dx_n = 1 \quad ; \quad f \geqslant 0 \qquad [31]$$

El producto de f por el diferencial de hiper-volumen da la probabilidad de encontrar a cada variable X_i en su elemento diferencial correspondiente $(x_i, x_i + dx_i)$:

$$P(x_1 < X_1 < x_1 + dx_1,\ x_2 < X_2 < x_2 + dx_2, ...) = \\ = f(x_1, x_2, ...)\, dx_1 \cdot dx_2 \ldots \qquad [32]$$

La situación de variables independientes puede ilustrarse con el *gas ideal*. Sea un gas ideal monoatómico compuesto por N partículas. Si designamos por x_i, x_{i+1}, x_{i+2}, las coordenadas (x, y, z) de la partícula i, veamos que todas las x_i son prácticamente independientes unas de otras. Es un hecho físicamente imposible el que las partículas i y j ocupen exactamente la misma posición. Puesto que no hay interacción entre ellas, esta imposibilidad es la única fuente que origina la existencia de correlación o dependencia entre las variables. Como la densidad se asume muy pequeña ($n = N/V \to 0$), tal dependencia resulta despreciable. Tomadas las variables x_i como independientes, la función densidad conjunta se factoriza en el producto de las funciones marginales:

$$f(x_1, x_2, ..., x_n) = f_{x_1} \cdot f_{x_2} \cdot \ldots \cdot f_{x_n} \qquad [33]$$

Naturalmente en gases reales y líquidos estas consideraciones dejan de ser válidas.

3. CARACTERISTICAS DE LAS MAGNITUDES ALEATORIAS

Imaginemos una variable aleatoria X (discreta o continua) con su función de probabilidades asociada y con un gran conjunto de valores x_i para recorrer. Efectivamente, con el formalismo anterior de funciones de distribución queda completamente caracterizado el experimento aleatorio que involucra a X. Podemos además ayudarnos de representaciones gráficas, cuando sea posible, para visualizar el problema. Ahora bien, si nos interesa conocer de tal distribución cuál es el valor más probable, o su valor medio, o cuantificar el grado de «desparrame» de la distribución sobre el eje real, u otras cuestiones, es claro que a poco que la distribución sea complicada la respuesta a estas preguntas no va a ser inmediata. Todo esto se ve agravado si la distribución es multidimensional. No obstante, la información requerida está contenida en la función densidad (la más cómoda para trabajar) y deberemos extraerla de ella. Estudiaremos cómo hacerlo de acuerdo con ciertas reglas que veremos en breve. Así, se podrá mediante un número pequeño de parámetros (3 ó 4) dar las características fundamentales de la distribución y razonar *cualitativamente* con ellos. Antes de pasar a este asunto conviene definir dos conceptos importantes para una magnitud aleatoria: *población* y *muestra.*

3.1. Población y muestra

En las realizaciones prácticas, un experimento aleatorio no podrá ser llevado a cabo un número infinito de veces para fijar con toda exactitud las propiedades del conjunto de todos los posibles sucesos. Habrá que conformarse con un número finito, aunque pueda ser muy grande, de ensayos. Se da el caso en muchos problemas de la Química Física que la expresión de la función de distribución es conocida por argumentos teóricos, pero no puede ser evaluada en la práctica (estudio del estado líquido, por ejemplo (Feynman, 1972)). Deberá aproximarse así la función mediante un número finito de ensayos.

A la vista de lo precedente deberemos distinguir las *magnitudes características* de la variable aleatoria correspondientes a un número infinito de ensayos (características *poblacionales*) de aquellas que, forzosamente, se calculan en la práctica a través de un número limitado de ensayos (características *muestrales*). Las últimas son, pues, la estimación que en muchos casos podemos dar para las características poblacionales.

La diferencia fundamental entre ambos tipos es básica: las características poblacionales *no son* magnitudes aleatorias y tienen valores (conocidos o no) en teoría bien definidos; las características muestrales *sí son*, por contra, variables aleatorias, puesto que dependen del número de ensayos. En lo que seguirá nos ocuparemos únicamente de las características poblacionales al objeto de establecer la teoría básica.

3.2. Esperanza matemática y valores medios

Sea una variable aleatoria X con función de distribución $f(x)$ definida en $-\infty < x < +\infty$ y normalizada. Se denomina *esperanza matemática* de X a:

$$\langle X \rangle = \sum_k p_k x_k \quad \text{(caso discreto)} \qquad [34]$$

$$\langle X \rangle = \int_{-\infty}^{\infty} x f(x)\, dx \quad \text{(caso continuo)} \qquad [35]$$

si y sólo si, bien la suma sobre k o bien la integral son *absolutamente convergentes*. Esto se cumplirá siempre que la suma [34] lo sea sobre un número finito de términos, o que el intervalo de integración en [35] se reduzca a uno finito por ser $f(x)$ nula fuera de él. Estas condiciones, naturalmente, son sólo *suficientes*, por lo que si no se dan deberemos comprobar

$$\sum_k p_k \cdot |x_k| < +\infty \qquad [36]$$

$$\int_{-\infty}^{\infty} |x| \cdot f(x)\, dx < +\infty \qquad [37]$$

para asegurarnos de la existencia de la esperanza matemática.

Recurriendo a la analogía mecánica de la masa repartida sobre el eje x, es directo identificar las expresiones [34]-[35] con la posición del centro de gravedad de la distribución. Al valor $\langle X \rangle$ cuando existe se le denomina *valor medio* y posee las siguientes propiedades:

a) $\langle a \rangle = a \quad ; \quad a = \text{constante}$ [38]

b) $\langle aX \rangle = a\langle X \rangle$ [39]

c) Dadas dos variables aleatorias X e Y, tales que los resultados en Y se determinan a partir de los de X a través de:

$$Y = \varphi(X) \qquad [40]$$

el *valor medio* para Y (esperanza matemática para Y) se calcula como:

$$\langle Y \rangle = \langle \varphi(X) \rangle = \sum_k p_k \cdot \varphi(x_k) \qquad \text{(discreto)} \qquad [41]$$

$$\langle Y \rangle = \langle \varphi(X) \rangle = \int_{-\infty}^{\infty} \varphi(x) f(x)\, dx \qquad \text{(continuo)} \qquad [42]$$

debiendo asegurarnos de la *convergencia absoluta* de la expresión correspondiente. Por ejemplo, consideremos el caso de dos variables continuas sujetas a la relación lineal:

$$Y = aX + b \qquad [43]$$

obteniéndose sin dificultad que:

$$\langle Y \rangle = \int_{-\infty}^{\infty} (ax + b) f(x)\, dx = a \int_{-\infty}^{\infty} x f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx =$$
$$= a\langle X \rangle + b \qquad [44]$$

La generalización a dos o más variables aleatorias es inmediata. Por simplicidad en la notación nos limitaremos a dos variables ligadas por la relación:

$$Z = \varphi(X, Y) \qquad [45]$$

La esperanza matemática, en las condiciones de convergencia ya apuntadas, viene dada por:

$$\langle Z \rangle = \langle \varphi(X, Y) \rangle = \sum_k \sum_j P_{kj} \cdot \varphi(x_k, y_j) \qquad \text{(discreto)} \qquad [46]$$

$$\langle Z \rangle = \langle \varphi(X, Y) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) f(x, y)\, dx\, dy \qquad \text{(continuo)} \qquad [47]$$

Un caso especialmente interesante es cuando

$$Z = X + Y \qquad [48]$$

en donde se obtiene que [46]-[47] son:

$$\langle Z \rangle = \sum_k \sum_j P_{kj}(x_k + y_j) = \sum_k P_k^x x_k + \sum_j P_j^y y_j = \langle X \rangle + \langle Y \rangle$$

(discreto) [49]

$$\langle Z \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y)\, dx\, dy = \int_{-\infty}^{\infty} x f_x\, dx + \int_{-\infty}^{\infty} y f_y\, dy =$$

$$= \langle X \rangle + \langle Y \rangle$$

(continuo) [50]

Este resultado es *general* para n variables aleatorias sean o no independientes:

$$\langle X_1 + X_2 + \ldots + X_n \rangle = \sum_{i=1}^{n} \langle X_i \rangle \qquad [51]$$

Si la relación funcional es la de multiplicación:

$$Z = X_1 \cdot X_2 \cdot X_3 \cdot \ldots \cdot X_n \qquad [52]$$

la obtención del resultado análogo al [51]:

$$\langle Z \rangle = \langle X_1 \rangle \cdot \langle X_2 \rangle \cdot \ldots \cdot \langle X_n \rangle \qquad [53]$$

exige que las variables X_i *sean necesariamente independientes*. La comprobación de este hecho se deja como ejercicio para el lector. Es muy importante insistir en que para aplicar [51] y [53] todos los valores medios $\langle X_i \rangle$ deben existir según el criterio de convergencia absoluta indicado.

3.3. Momentos de una distribución

Dadas dos variables aleatorias X, Y, consideremos la relación:

$$Y = X^m \quad ; \quad m = 1, 2, 3, \ldots \qquad [54]$$

Al valor medio de Y calculado con [41]-[42] se le llama *momento de orden m* (α_m) respecto del origen. Para $m = 1$ este momento coincide con el valor medio de la distribución X.

Igualmente podemos definir momentos respecto del valor medio $\langle X \rangle$:

$$\mu_m = \langle (X - \langle X \rangle)^m \rangle \quad ; \quad m = 1, 2, 3, \ldots \qquad [55]$$

denominados *momentos centrales* de orden m (Ejercicio 3). Es fácil encontrar:

$$\mu_1 = \langle X - \langle X \rangle \rangle = \langle X \rangle - \langle X \rangle = 0 \qquad [56]$$

$$\mu_2 = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 - 2X\langle X \rangle + \langle X \rangle^2 \rangle =$$
$$= \langle X^2 \rangle - 2\langle X \rangle\langle X \rangle + \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2 \qquad [57]$$

y así sucesivamente (nótese que $\langle X \rangle$ es una constante en estas manipulaciones).

Para caracterizar distribuciones resulta muy útil el momento central de segundo orden μ_2, ya que representa el valor medio de una variable no negativa $(X - \langle X \rangle)^2$, siendo pues $\mu_2 \geqslant 0$. Además μ_2 es el mínimo valor que puede alcanzar el momento de segundo orden con respecto a cualquier punto (Ejercicio 4). Más adelante veremos que μ_2 está relacionado con la medida de la *dispersión* de una distribución. En definitiva, siguiendo la analogía mecánica, μ_2 es un *momento de inercia* de la distribución de masa.

3.4. Dispersión y asimetría

Calculado el valor medio $\langle X \rangle$ tenemos una idea de dónde se encuentra centrada la distribución. Nos interesa ahora conocer en qué medida

(*dispersión*) esta distribución está acumulada alrededor de ese valor central y, además, el grado de *asimetría* que presenta.

Antes de seguir adelante indicaremos que existen otras medidas distintas de $\langle X \rangle$ para estimar la posición central de la distribución, tales como la *moda* y la *mediana*. Sin embargo, estas estimaciones no son muy útiles a efectos de cálculo pues no poseen sencillas leyes de operación como las de $\langle X \rangle$. En consecuencia no nos referiremos a ellas aquí, limitando la discusión subsiguiente a la media como posición central.

El valor $\langle X \rangle$ hemos dicho que representa el centro de gravedad de la distribución, por lo que como medida de la dispersión resulta adecuado tomar la magnitud denominada *desviación típica*:

$$\sigma(X) = \sqrt{\mu_2} = +\sqrt{\langle (X - \langle X \rangle)^2 \rangle} \qquad [58]$$

que es la raíz cuadrada (positiva) del momento central de segundo orden μ_2. Esta elección está justificada ya que μ_2 es el momento de inercia de la distribución de masa con respecto a un eje perpendicular al eje x y que pase por el centro de gravedad $\langle X \rangle$. Como μ_2 tiene dimensiones de X^2, la dispersión σ debe ser su raíz cuadrada. Al momento $\mu_2 = \sigma^2$ se le llama *varianza* de la distribución.

Aplicando reglas ya conocidas obtenemos:

$$\sigma^2(aX + b) = a^2\sigma^2(X) \qquad [59]$$

$$\sigma(aX + b) = |a| \cdot \sigma(X) \qquad [60]$$

Una operación muy común y valiosa en este contexto es la *normalización* (*tipificación*) de la variable X. Sea X una variable aleatoria con media $\langle X \rangle$ y desviación típica σ_x. Haciendo el cambio:

$$Z = \frac{X - \langle X \rangle}{\sigma_x} \qquad [61]$$

encontramos una variable Z con media $\langle Z \rangle = 0$ y desviación típica $\sigma_z = 1$. En definitiva Z nos da la desviación de X respecto de $\langle X \rangle$ en unidades de σ_x.

Para profundizar un poco más en el significado de σ como medida de la dispersión de una distribución visualizaremos el *teorema de Tchebycheff* (Cramér, 1977):

Dada una variable aleatoria con valor medio $\langle X \rangle$ y desviación típica σ, siendo t un número real arbitrario y positivo, la probabilidad de que X tome un valor que satisfaga $|X - \langle X \rangle| > t\sigma$ es menor que $1/t^2$:

$$P(|X - \langle X \rangle| > t\sigma) < 1/t^2 \qquad [62]$$

Este teorema se cumple para cualquier tipo de distribución, discreta, continua o mixta. El sentido de la desigualdad [62] se hace más claro notando que $P(|X - \langle X \rangle| > t\sigma)$ representa la masa exterior al intervalo:

$$\langle X \rangle - t\sigma \leqslant X \leqslant \langle X \rangle + t\sigma \qquad [63]$$

y para valores t muy grandes tal masa será muy pequeña.

Dado un conjunto de variables *independientes* $X_1, ..., X_n$, la regla de adición de varianzas es inmediata:

$$\sigma^2(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} \sigma^2(X_i) \qquad [64]$$

(Ejercicios 5-6).

Pasemos ahora a caracterizar la *asimetría* de una distribución. Cualquier medida de este concepto claramente implicará un cierto grado de arbitrariedad. En una distribución completamente simétrica respecto de la media todos los momentos centrales de orden impar μ_{2n+1} son *idénticamente nulos*. Parece pues razonable elegir uno de ellos, normalmente μ_3, para medir la asimetría. Puesto que μ_3 tiene unidades de X^3 se conviene en normalizarlo con σ^3 para dar el *coeficiente de asimetría*:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \qquad [65]$$

resultando que si $\gamma_1 > 0$ la asimetría se dice *positiva* y si $\gamma_1 < 0$ se dirá *negativa* (dependiendo de la bibliografía este convenio aparece justo al revés).

4. FUNCIONES GENERATRIZ Y CARACTERISTICA

Dada una variable aleatoria X con su correspondiente distribución de probabilidades se denomina *función generatriz* al valor medio de la función e^{tX}, siendo t una variable real:

$$\psi(t) = \langle e^{tX} \rangle = \sum_k p_k \cdot e^{tx_k} \qquad \text{(discreto)} \qquad [66]$$

$$\psi(t) = \langle e^{tX} \rangle = \int_{-\infty}^{\infty} e^{tx} f(x) dx \qquad \text{(continuo)} \qquad [67]$$

Estas expresiones recuerdan de alguna manera a la transformación de Fourier. Si en [67] la integral se extiende sólo desde $0 \to +\infty$, y hacemos $t = -s$, la operación recibe el nombre de *transformada de Laplace* de $f(x)$ (Schwartz, 1969). Tanto [66] como [67] son siempre convergentes cuando $t = 0$, pero no tienen por qué serlo para valores $t \neq 0$. Sin embargo, en muchas aplicaciones de interés se da el caso de que $f(x) \to 0$ muy rápidamente cuando $x \to -\infty$, y esto resulta suficiente para garantizar la convergencia, al menos, para algunos valores de t (Cramér, 1977).

Supongamos que [66]-[67] convergen para algún $t \neq 0$. Por derivación con respecto de t obtenemos:

$$\psi^{(m)}(0) = \alpha_m \qquad [68]$$

en tanto los momentos α_m existan. Desarrollando pues en serie $\psi(t)$ en torno a $t = 0$ se tiene entonces:

$$\psi(t) = \sum_{m=0}^{\infty} \frac{\alpha_m}{m!} t^m \qquad [69]$$

Si para algún $t \neq 0$ la serie [69] converge, entonces la sucesión de momentos α_m que definen a $\psi(t)$ representa *unívocamente* a la distribución $f(x)$ o $\{p_k\}$. De ahí que a $\psi(t)$ se la llame función generatriz de la distribución.

Es inmediato comprobar que la función generatriz de una suma de

variables aleatorias *independientes* es igual al producto de las funciones generatrices de cada variable:

$$X = \sum_{i=1}^{n} X_i \Rightarrow \psi(t) = \psi_1(t) \cdot \psi_2(t) \cdot \ldots \cdot \psi_n(t) \qquad [70]$$

La sencillez de esta ley de composición entre funciones generatrices hace que éstas sean ampliamente utilizadas en el trabajo teórico de áreas tales como la Mecánica Estadística (Khinchin, 1949).

Los problemas de convergencia apuntados pueden ser eliminados completamente empleando la *función característica* $\varphi(t)$ de la distribución, que se define como:

$$\varphi(t) = \langle e^{itX} \rangle \quad ; \quad -\infty < t < +\infty \qquad [71]$$

En el caso discreto [71] se especializa en:

$$\varphi(t) = \sum_{k=-\infty}^{\infty} e^{itx_k} \cdot p_k \qquad [72]$$

que es una *serie de Fourier convergente* para todos los valores de t. Para el caso continuo tenemos:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \cdot f(x)\, dx \qquad [73]$$

en donde se reconoce a la *integral de Fourier* de $f(x)$, la cual es igualmente *convergente* para todos los valores reales de t. Claramente la función característica de una suma de variables independientes es el producto de las funciones características de cada una de ellas (Ejercicio 7).

Para concluir consideremos por brevedad el problema de la inversión de [73]. Siempre que $\varphi(t)$ sea *absolutamente integrable*:

$$\int_{-\infty}^{\infty} |\varphi(t)|\, dt < +\infty \qquad [74]$$

la densidad $f(x)$ se obtiene de la transformación inversa (Rozanov, 1975):

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \cdot \varphi(t) \cdot dt \quad ; \quad -\infty < x < +\infty \qquad [75]$$

Según la bibliografía, ya sabemos que el factor preintegral puede aparecer como $(2\pi)^{-1/2}$, etc. En cualquier caso, la normalización de $f(x)$ debe preservarse. En este contexto la convención es la dada en [73] y [75], y la seguiremos de aquí en adelante.

5. CONVOLUCION DE DISTRIBUCIONES DE PROBABILIDAD

Consideremos ahora una operación matemática de gran interés en las aplicaciones, tanto teóricas como experimentales dentro de la Química Física, conocida como *convolución.*

Para centrar ideas sean X_1 y X_2 dos variables continuas e *independientes* con funciones densidad de probabilidad $f_1(x_1)$ y $f_2(x_2)$. Se desea calcular la función densidad $f(x)$ de su suma $X = X_1 + X_2$. Notemos que por ser independientes la función densidad conjunta $f(x_1, x_2)$ para ambas variables es:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \qquad [76]$$

Introduzcamos un cambio de variables $(X_1, X_2) \rightarrow (X, Y)$, tal que realice una transformación *biunívoca* del plano en sí mismo y que tenga Jacobiano unidad:

$$\left.\begin{aligned} X &= X_1 + X_2 \\ Y &= X_2 \end{aligned}\right\} \; ; \quad |J| = \begin{vmatrix} \dfrac{\partial X}{\partial X_1} & \dfrac{\partial X}{\partial X_2} \\ \dfrac{\partial Y}{\partial X_1} & \dfrac{\partial Y}{\partial X_2} \end{vmatrix} = 1 \qquad [77]$$

Repárese en que las nuevas variables X e Y son independientes.

La transformación expresada en [77] transformará un dominio $D(X_1, X_2)$ en otro $D'(X, Y)$. Si un punto (x, y) pertenece a D', implica que el punto imagen correspondiente (x_1, x_2), pertenecerá a D y viceversa. Ambos sucesos son en realidad el mismo y de aquí se sigue que la probabilidad de que el punto genérico (x, y) pertenezca a D' es la misma que la asociada al hecho de que (x_1, x_2) caiga dentro de D:

$$P((x_1, x_2) \in D) = P((x, y) \in D') \qquad [78]$$

o lo que es lo mismo:

$$P((x, y) \in D') = \int_D \int f(x_1, x_2)\, dx_1\, dx_2 = \int_{D'} \int g(x, y)\, dx\, dy \qquad [79]$$

siendo $g(x, y)$ la función densidad conjunta para X e Y.

Con el cambio de variables [77] f y g aparecen relacionadas como:

$$g(x, y) = f(x_1, x_2)/J = f(x - y, y) \qquad [80]$$

Así definida, la función g es no negativa y está normalizada. Consecuentemente puede escribirse:

$$g(x, y) = f_1(x - y) \cdot f_2(y) \qquad [81]$$

La función densidad de la suma $X_1 + X_2$ vendrá dada por integración sobre la variable y en [81], es decir por la distribución marginal de X:

$$f(x) = g_x = \int_{-\infty}^{\infty} f_1(x - y) \cdot f_2(y)\, dy \qquad [82]$$

Un requerimiento elemental de simetría exige que:

$$f(x) = \int_{-\infty}^{\infty} f_1(x - y) f_2(y) \cdot dy = \int_{-\infty}^{\infty} f_2(x - u) f_1(u)\, du \qquad [83]$$

integrales que reciben el nombre de *convolución* de las funciones f_1 y f_2. Esta operación se designa comúnmente mediante un asterisco: $f_1 * f_2$, y resulta fundamental en el análisis de señales (Brigham, 1974).

Por ejemplo, cuando se realiza un espectro sobre una muestra, el registro final que se obtiene (el popularmente llamado espectro) resulta ser la convolución de dos señales: la señal de respuesta de la muestra a la perturbación electromagnética (el verdadero espectro) y la señal debida al ruido aleatorio de fondo del aparato con el que se trabaja. Para obtener la señal real, sin ruido, hay que deconvolucionar el registro final, operación compleja de realizar, si bien necesaria en ciertos casos. Para otras aplicaciones de la convolución puede consultarse la referencia (Schwartz, 1969).

Mencionemos finalmente el teorema que liga la convolución de f_1 y f_2 con sus transformadas de Fourier φ_1 y φ_2 (Schwartz, 1969):

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi_1(t)\cdot\varphi_2(t)\cdot e^{-itx}\,dt=\int_{-\infty}^{\infty}f_1(x-u)\cdot f_2(u)\cdot du=f_1*f_2$$

6. EJEMPLO DE APLICACION: CALCULO DE VALORES MEDIOS Y DISPERSIONES EN TERMODINAMICA ESTADISTICA

Para un gas monoatómico ideal en equilibrio, cuando los efectos cuánticos son despreciables, la distribución de velocidades $\mathbf{v}$ de sus moléculas en ausencia de campos externos se expresa por la ley de Maxwell:

$$f(\mathbf{v})\,dv_x\,dv_y\,dv_z=C\exp\left[-\beta mv^2/2\right]dv_x\,dv_y\,dv_z \qquad [84]$$

en donde β es el inverso del ruido térmico kT, m la masa de las partículas y C una constante de normalización. El sentido físico de $f(\mathbf{v})$ es el de una densidad de probabilidad, y más concretamente $f(\mathbf{v})\cdot dv_x\,dv_y\,dv_z$ representa el número medio de moléculas por unidad de volumen que tienen su velocidad entre los límites $(\mathbf{v}, \mathbf{v}+\mathrm{d}\mathbf{v})$. Dicho en otros términos [84] suministra la densidad media de moléculas del gas que tienen sus componentes de la velocidad tales que:

$$\begin{aligned} v_x &\in (v_x, v_x+dv_x)\\ v_y &\in (v_y, v_y+dv_y)\\ v_z &\in (v_z, v_z+dv_z) \end{aligned} \qquad [85]$$

El primer punto a abordar es la normalización de [84]. Se podría intentar determinar la constante C exigiendo que la integral triple [84] extendida sobre todos los valores posibles de $v_x\,v_y, v_z$, fuese la unidad. Físicamente, sin embargo, esta operación, aunque lícita, no tendría el sentido deseado, ya que [84] da una densidad media de moléculas. Es pues razonable normalizarla de modo que la integral triple nos dé el

número total de moléculas N por unidad de volumen V (densidad media) del gas. La ecuación para determinar C será:

$$C \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left[-\frac{\beta m}{2}(v_x^2 + v_y^2 + v_z^2) \right] dv_x\, dv_y\, dv_z = \frac{N}{V} = n \qquad [86]$$

Notemos entonces que en problemas de interés químico-físico puede resultar más acertado normalizar la función de distribución de probabilidades asociada a un problema a una cantidad sugerida por la propia naturaleza de éste, cantidad que en general será distinta de la unidad. Naturalmente los resultados finales no se verán afectados por el valor de C siempre que se trabaje correctamente.

Utilizando la relación:

$$\int_{-\infty}^{\infty} \exp\left[-\frac{\beta m}{2} v_i^2 \right] dv_i = \left(\frac{2\pi}{\beta m} \right)^{1/2} \quad ; \quad i = x, y, z \qquad [87]$$

resulta el valor

$$C = n\left(\frac{\beta m}{2\pi} \right)^{3/2} \qquad [88]$$

y por tanto

$$f(v_x, v_y, v_z) = n\left(\frac{\beta m}{2\pi} \right)^{3/2} \exp\left[-\frac{\beta m}{2}(v_x^2 + v_y^2 + v_z^2) \right] \qquad [89]$$

Desde bases físicas podemos afirmar que las tres componentes de la velocidad son independientes entre sí. El valor medio de cualquiera de estas variables es idénticamente nulo, como se comprueba sin dificultad. Por ejemplo, para v_x la distribución marginal es:

$$f_x(v_x) = n\left(\frac{\beta m}{2\pi} \right)^{3/2} \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \exp\left[-\frac{\beta m}{2}(v_x^2 + v_y^2 + v_z^2) \right] =$$

$$= n\left(\frac{\beta m}{2\pi} \right)^{1/2} \cdot \exp\left[-\frac{\beta m}{2} v_x^2 \right] \qquad [90]$$

y el valor medio:

$$\langle v_x \rangle = \int_{-\infty}^{\infty} v_x f_x(v_x)\, dv_x \Big/ \int_{-\infty}^{\infty} f_x(v_x)\, dv_x = 0 \qquad [91]$$

puesto que la función a integrar en el numerador es impar y el intervalo de integración es simétrico en torno a $v_x = 0$. Como $f_x(v_x)$ no está normalizada a la unidad debemos incluir el denominador que aparece en [91] en el cálculo.

Igualmente podemos evaluar la dispersión $\sigma(v_x)$ a través de:

$$\sigma^2(v_x) = \langle v_x^2 \rangle - \langle v_x \rangle^2 = \langle v_x^2 \rangle \qquad [92]$$

siendo

$$\langle v_x^2 \rangle = \frac{\int_{-\infty}^{\infty} v_x^2 f_x(v_x)\, dv_x}{\int_{-\infty}^{\infty} f_x(v_x)\, dv_x} = 2\left(\frac{\beta m}{2\pi}\right)^{1/2} \int_0^{\infty} v_x^2 \exp\left[-\frac{\beta m}{2} v_x^2\right] dv_x \qquad [93]$$

en donde se ha hecho uso de la paridad de la función a integrar sobre el intervalo simétrico $-\infty < v_x < +\infty$ y además de:

$$\int_{-\infty}^{\infty} f_x(v_x)\, dv_x = n \qquad [94]$$

Aplicando la relación general:

$$\int_0^{\infty} x^m \exp\left[-ax^2\right] dx = \frac{\Gamma[(m+1)/2]}{2a^{(m+1)/2}} \qquad [95]$$

donde Γ es la función Gamma de Euler:

$$\Gamma\left(\frac{m+1}{2}\right) = \{m = 2i\} = \Gamma\left(i + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2i-1)}{2^i} \sqrt{\pi} \quad ;$$

$$i = 1, 2, 3, \ldots \qquad [96]$$

se obtiene de [93]:

$$\langle v_x^2 \rangle = 2\left(\frac{\beta m}{2\pi}\right)^{1/2} \cdot \frac{\Gamma(1 + 1/2)}{2\left(\frac{\beta m}{2}\right)^{3/2}} = \frac{1}{\beta m} = \frac{kT}{m} \qquad [97]$$

y de ahí:

$$\sigma(v_x) = \sqrt{\frac{kT}{m}} \qquad [98]$$

Este resultado nos indica que la distribución de velocidades en una dimensión, $f_x(v_x)$, está tanto más «picada» sobre el valor medio $v_x = 0$, cuanto más pequeña sea la temperatura absoluta T. La distribución $f_x(v_x)$ dada por [90] es un caso particular de distribución gaussiana monodimensional de la que hablaremos en profundidad en el tema siguiente. Esta distribución es simétrica en torno a su valor medio $\langle v_x \rangle$.

Dentro de este mismo contexto, veamos una distribución que no es simétrica en torno a su valor medio. Para ello preguntémonos por la función de distribución $f(v)$ de los módulos v de las velocidades moleculares. Como antes, exigiremos que el producto $f(v) \cdot dv$ represente la densidad media (N/V) de moléculas con módulo de su velocidad comprendido entre $(v, v + dv)$. Es importante aquí reparar en que $f(\mathbf{v})$ dada por [84] depende únicamente del módulo de la velocidad $v = |\mathbf{v}|$, lo que es lógico dada la inexistencia de campos externos que impongan direcciones privilegiadas. Este hecho implica necesariamente el que $f(\mathbf{v})$ pueda expresarse como $g(v)$.

La densidad media $f(v) \cdot dv$ que buscamos se obtiene con la integración:

$$f(v) \cdot dv = \int_{v<|\mathbf{v}|<v+dv} f(\mathbf{v})\, dv_x\, dv_y\, dv_z \qquad [99]$$

que se extiende sobre todas las velocidades que verifican la condición exigida. Dentro del recinto de integración $f(\mathbf{v}) = g(v) =$ constante y [99] puede ponerse:

$$f(v)\, dv = g(v) \int_{v<|\mathbf{v}|<v+dv} dv_x\, dv_y\, dv_z \qquad [100]$$

representando la integral un «volumen» en el espacio de velocidades. Análogamente al espacio tridimensional introduciremos las coordenadas polares v, θ, φ, de modo que:

$$\begin{aligned} v_x &= v \operatorname{sen} \theta \cos \varphi \\ v_y &= v \operatorname{sen} \theta \operatorname{sen} \varphi \\ v_z &= v \cos \theta \end{aligned} \qquad [101]$$

con lo que el elemento de volumen se transforma en:

$$dv_x\, dv_y\, dv_z = v^2 \operatorname{sen} \theta \cdot dv \cdot d\theta \cdot d\varphi \qquad [102]$$

y la expresión [100] pasa a ser:

$$f(v)\, dv = g(v) \int_v^{v+dv} \int_0^{\pi} \int_0^{2\pi} v^2 \operatorname{sen} \theta \, dv \cdot d\theta \cdot d\varphi \qquad [103]$$

De todo ello concluimos que:

$$f(v) \cdot dv = 4\pi g(v) v^2 \, dv =$$

$$= 4\pi n \left(\frac{\beta m}{2\pi}\right)^{3/2} v^2 \cdot \exp\left[-\frac{\beta m}{2} v^2\right] dv \quad ; \quad 0 \leqslant v < +\infty \qquad [104]$$

Esta distribución *no es simétrica* alrededor de su valor medio:

$$\langle v \rangle = \frac{2}{\sqrt{\pi}} \left(\frac{2kT}{m}\right)^{1/2} \quad ; \quad \Gamma(n+1) = n! \quad ; \quad n = 0, 1, 2, \ldots \qquad [105]$$

pues este valor ni siquiera coincide con el más probable para el módulo de la velocidad $v_{\text{máx}}$ calculable de:

$$\frac{df(v)}{dv} = 0 \Rightarrow v_{\text{máx}} = \left(\frac{2kT}{m}\right)^{1/2} \qquad [106]$$

(Ejercicio 8)

Bibliografía

1. BRIGHAM, E. O., *The Fast Fourier Transform*, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1974, capítulos 4 y 7.

2. CRAMÉR, H., *Elementos de la Teoría de Probabilidades*, Aguilar, Madrid, 1977, capítulo 2 y 5.

3. FANO, G., *Mathematical Methods of Quantum Mechanics*, McGraw-Hill, Nueva York, 1968, capítulo 4.

4. FEYNMAN, R. P., *Statistical Mechanics*, Benjamin, Reading Massachusetts, 1972, capítulo 4.

5. KAMPEN, N. G., *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 1985, capítulo 1.

6. KHINCHIN, A. I., *Mathematical Foundations of Statistical Mechanics*, Dover, Nueva York, 1949.

7. ROZANOV, Y., *Processus Aléatoires*, Mir, Moscú, 1975, capítulos 1 y 2.

8. RUDIN, W., *Principios de Análisis Matemático*, Ediciones del Castillo, Madrid, 1974, capítulo 10.

9. SCHWARTZ, L., *Métodos Matemáticos para las Ciencias Físicas*, Selecciones Científicas, Madrid, 1969, capítulos 3 y 6.

EJERCICIOS DE AUTOCOMPROBACION

1. Una variable aleatoria tiene la función densidad:

$$f(x) = \begin{cases} x + 1/4 & \text{si} \quad 0 < x \leqslant 1 \\ 0 & \text{en otro caso} \end{cases}$$

   a) Normalizarla

   b) Calcular la función cumulativa (integral) $F(x)$

   c) Evaluar la probabilidad $P(1/4 < X < 3/4)$

2. Se tiene un conjunto de osciladores armónicos monodimensionales desacoplados. Todos ellos vibran a la misma frecuencia ω_0 y con la misma amplitud A_0. Sin embargo, sus fases ϕ están uniformemente distribuidas en el intervalo $0 \leqslant \phi < 2\pi$, es decir cualquier valor ϕ es igualmente probable en ese intervalo. Calcular la función densidad $g(x)$ del desplazamiento x de un oscilador en cualquier instante t $(x = A_0 \cos(\omega_0 t + \phi))$.

3. Obtener las relaciones entre los momentos respecto al origen $\alpha_1, \alpha_2, \alpha_3$, y los momentos centrales μ_1, μ_2, μ_3.

4. Demostrar que el momento de segundo orden $\mu_2 = \langle (X - a)^2 \rangle$ con respecto a un punto arbitrario $x = a$ es mínimo cuando $\langle X \rangle = a$.

5. *a*) Dado el conjunto de variables independientes $X_1, X_2, \ldots, X_n$, todas con el mismo valor medio $<X>$ y misma desviación típica σ_x, calcular el valor medio y la desviación para la función:

$$X(n) = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

b) Dos variables aleatorias dependientes X e Y se combinan para dar otra nueva variable aleatoria $Z = X + Y$. Obtener la varianza de Z en función de las de X e Y.

6. Una macromolécula está formada por 1.000 unidades situadas consecutivamente una tras otra. Debido a diversos fenómenos de interacción, la longitud L de cada una de estas unidades está entre 15 Å y 16 Å, no saliéndose nunca de tales límites. Suponiendo, en primera aproximación, que la distribución de probabilidades para la longitud dentro de dicho intervalo es uniforme, y que las unidades son independientes, determinar para la macromolécula la longitud media y su desviación estándar.

7. Dada la densidad rectangular

$$f(x) = \begin{cases} (b-a)^{-1} & a \leqslant x \leqslant b \\ 0 & \text{en otro caso} \end{cases}$$

calcular sus funciones generatriz y característica.

8. Determinar la dispersión y el coeficiente de asimetría γ_1 [65] para la distribución de los módulos de las velocidades $f(v)$ dada en [104]. Discutir el efecto de la temperatura y de la masa de las moléculas sobre tales características.

SOLUCIONES A LOS EJERCICIOS DE AUTOCOMPROBACION

1. *a*) La normalización exige:

$$\int_{-\infty}^{\infty} N\bar{f}(x)\,dx = 1$$

donde N es la constante de normalización:

$$N = \left[\int_{-\infty}^{\infty} \bar{f}(x)\,dx\right]^{-1}$$

que será la unidad si $\bar{f}(x)$ ya está normalizada. Un sencillo cálculo da:

$$\int_{-\infty}^{\infty} \bar{f}(x)\,dx = \int_{0}^{1} \bar{f}(x)\,dx = \int_{0}^{1} (x + 1/4)\,dx = \frac{3}{4} \quad ; \quad N = \frac{4}{3}$$

y por lo tanto encontramos para la función normalizada la expresión:

$$f(x) = \begin{cases} \dfrac{1}{3}(4x + 1) & \text{si} \quad 0 < x \leqslant 1 \\ 0 & \text{en otro caso} \end{cases}$$

b) El cálculo de la función integral $F(x)$ debe hacerse siguiendo los intervalos en los que $f(x)$ esté definida. En este caso $F(x)$ deberá ser calculada en los tres intervalos $(-\infty, 0]$, $(0, 1]$, $(1, +\infty)$. Para el primero $-\infty < x \leqslant 0$ se tiene:

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \{f(x) = 0\} = 0$$

Para el segundo $0 < x \leqslant 1$ es:

$$F(x) = \int_{0}^{x} f(x)\,dx = \int_{0}^{x} \left[\frac{1}{3}(4x + 1)\right] dx = \frac{1}{3}(2x^2 + x)$$

Finalmente para el tercero $1 < x < +\infty$:

$$F(x) = 1$$

La expresión para $F(x)$ será entonces:

$$F(x) = \begin{cases} 0 & \text{si} \quad -\infty < x \leqslant 0 \\ \dfrac{1}{3}(2x^2 + x) & \text{si} \quad 0 < x \leqslant 1 \\ 1 & \text{si} \quad 1 < x < +\infty \end{cases}$$

c) La probabilidad pedida puede evaluarse con ayuda de [11] y resulta:

$$P(1/4 < x < 3/4) = F(3/4) - F(1/4) = \int_{1/4}^{3/4} f(x)\,dx = 1/2$$

2. Comencemos evaluando las funciones de probabilidad asociadas al ángulo ϕ. Nos restringiremos al intervalo de definición $(0,2\pi)$ y asumiremos siempre realizada la transformación de cualquier valor ϕ a ese intervalo fundamental. Por estar ϕ distribuido uniformemente su función densidad será:

$$\bar{f}(\phi) = 1 \quad ; \quad 0 \leqslant \phi < 2\pi$$

que convenientemente normalizada resulta:

$$f(\phi) = 1/2\pi \quad ; \quad 0 \leqslant \phi < 2\pi$$

Vamos a calcular la función densidad $g(x)$ para el desplazamiento x de los osciladores utilizando la δ de Dirac. De la relación entre la variable X y la fase Φ

$$X = A_0 \cos(\omega_0 t + \Phi) \quad ; \quad -A_0 \leqslant X \leqslant A_0$$

haciendo $t = 0$ (elección de origen para Φ) se obtiene la expresión más sencilla:

$$X = A_0 \cos \Phi \quad ; \quad 0 \leqslant \Phi < 2\pi$$

Claramente, cuando se ha dado sólo media vuelta sobre Φ ($0 \leqslant \phi < \pi$) se ha recorrido ya todo el dominio de X ($-A_0 \leqslant x \leqslant A_0$), si bien en sentido negativo: de A_0 a $-A_0$. Completando la vuelta en Φ ($\pi \leqslant \phi < 2\pi$) se recorre de nuevo el dominio de X, pero ahora en sentido contrario al previo. Así una vuelta completa en Φ origina dos recorridos completos del dominio de X (Fig. 3).

Utilizando la ecuación [15] escribimos:

$$g(x) = \int_0^{2\pi} \delta(x - A_0 \cos \phi) \cdot f(\phi)\, d\phi =$$

$$= \int_0^{\pi} \delta(x - A_0 \cos \phi) \cdot f(\phi)\, d\phi + \int_{\pi}^{2\pi} \delta(x - A_0 \cos \phi) \cdot f(\phi) \cdot d\phi$$

Teniendo en cuenta que $f(\phi)$ = constante y las propiedades: $\cos(\phi + \pi) = -\cos(\phi)$; $\delta(x - A_0 \cos \phi) = \delta(-x + A_0 \cos \phi)$, se establece la igualdad de las dos integrales anteriores:

$$g(x) = 2 \int_0^{\pi} \delta(x - A_0 \cos \phi) \cdot f(\phi) \cdot d\phi$$

La integración puede realizarse con ayuda de las propiedades de la δ-Dirac (Epígrafe 5, Tema 4). Con la transformación:

$$\delta(x - A_0 \cos \phi) = (A_0 \cdot \operatorname{sen} \phi)^{-1}\, \delta(\phi - \phi_0) = (A_0 \cdot \operatorname{sen} \phi)^{-1} \cdot \delta(\phi - \arccos \frac{x}{A_0})$$

se obtiene:

$$g(x) = \frac{1}{\pi}\int_0^{\pi} (A_0 \cdot \text{sen } \phi)^{-1} \cdot \delta(\phi - \text{arc cos } \frac{x}{A_0})\, d\phi =$$

$$= \frac{1}{\pi}\int_0^{\pi} (A_0 \sqrt{1 - \cos^2 \phi})^{-1} \cdot \delta(\phi - \text{arc cos } \frac{x}{A_0})\, d\phi =$$

$$= \frac{1}{\pi} \cdot \frac{1}{\sqrt{A_0^2 - x^2}} \quad ; \quad -A_0 \leqslant x \leqslant A_0$$

que es la función densidad pedida. Nótese que el papel jugado aquí por la fase ϕ es análogo al que en coordenadas polares esféricas juega θ ($0 \leqslant \theta \leqslant \pi$, intervalo en el que la función coseno es monótona).

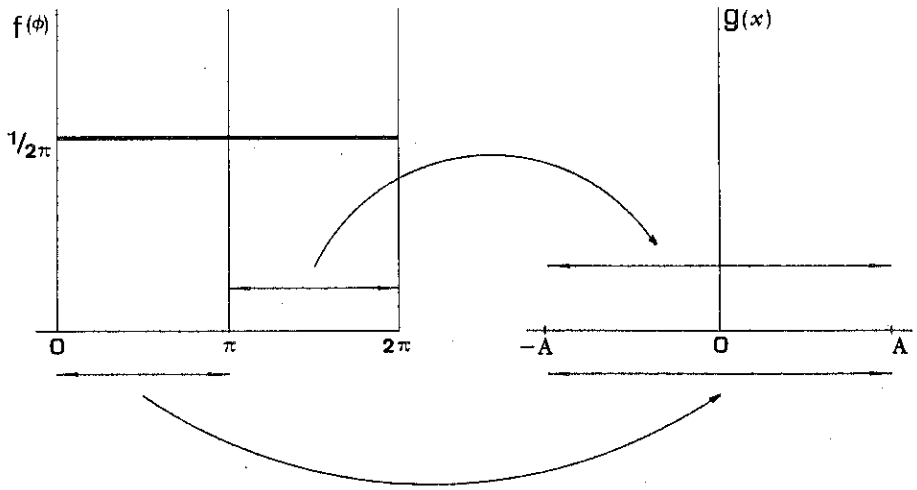


Figura 3.—Diagrama para el ejercicio 2.

3. Las relaciones pedidas son:

$$\mu_1 = 0 \quad ; \quad \alpha_1 = \langle X \rangle$$
$$\mu_2 = \alpha_2 - \alpha_1^2$$
$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3$$

4. Derivando con respecto al parámetro a la expresión:

$$\mu_2 = \langle X^2 \rangle - 2a\langle X \rangle + a^2$$

obtenemos

$$\frac{d\mu_2}{da} = -2\langle X \rangle + 2a = 0 \Rightarrow a = \langle X \rangle$$

que efectivamente es un mínimo:

$$\frac{d^2\mu_2}{da^2} = 2 > 0$$

5. *a*) Aplicando reglas de cálculo ya conocidas:

$$\langle \bar{X} \rangle = \left\langle \frac{X_1 + X_2 + \ldots + X_n}{n} \right\rangle = \langle X \rangle$$

$$\sigma(\bar{X}) = \sigma_x / \sqrt{n}$$

Este es un caso interesante de aplicación del concepto de *muestra* apuntado en el tema. Podemos considerar las variables $X_1, X_2, \ldots, X_n$, como una muestra de ensayos independientes realizados sobre la variable X, con media poblacional $\langle X \rangle$ y desviación típica poblacional σ_x. El valor medio esperado para X coincide con el de cualquiera de las variables X_i. No obstante, a no ser que $n \to \infty$, la estimación numérica $\langle X(n) \rangle'$ obtenida en un experimento con n ensayos será en general diferente del valor esperado $\langle X \rangle$, y por tanto la desviación típica para $X(n)$ no es nula. Es evidente pues que la característica muestral $\langle X(n) \rangle$ es una magnitud aleatoria, y dejará de serlo cuando $n \to \infty$, en cuyo caso $\sigma(\bar{X}(n \to \infty)) = 0$.

b) Sea $f(x, y)$ la función de distribución conjunta para X e Y. El cálculo de la varianza de $Z = X + Y$ se lleva a cabo del modo habitual:

$$\sigma^2(Z) = \langle Z^2 \rangle - \langle Z \rangle^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)^2 f(x, y)\, dx\, dy -$$

$$- \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y)\, dx\, dy \right]^2 =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + 2xy + y^2) f(x, y)\, dx\, dy - \Big(\langle X\rangle + \langle Y\rangle\Big)^2 =$$

$$= \langle X^2\rangle + \langle Y^2\rangle + 2\langle XY\rangle - \Big(\langle X\rangle + \langle Y\rangle\Big)^2 =$$

$$= \sigma^2(X) + \sigma^2(Y) + 2\Big[\langle XY\rangle - \langle X\rangle\langle Y\rangle\Big]$$

en donde la cantidad entre corchetes recibe el nombre de *covarianza* entre las variables X e Y (véase el Tema 8):

$$\text{cov}\,(X, Y) = \langle XY\rangle - \langle X\rangle\ \langle Y\rangle$$

Fácilmente se comprueba que si X e Y son *independientes*, su *covarianza* es nula.

6. Determinemos primero la función de distribución de probabilidades normalizada para la longitud de una unidad:

$$f(L) = \begin{cases} 1 & \text{si} \quad 15 < L < 16 \\ 0 & \text{en otro caso} \end{cases}$$

$$1 = N \int_{-\infty}^{\infty} f(L)\, dL = N \int_{15}^{16} dL = N$$

con lo que la función $f(L)$ escrita arriba ya está normalizada.

Dado que las unidades son independientes, la longitud media de la macromolécula y su varianza serán:

$$\langle L_{1000}\rangle = 1000\ \langle L_1\rangle \quad ; \quad \sigma^2_{1000} = 1000\ \sigma_1^2$$

en donde el subíndice 1 denota magnitudes relativas a una cualquiera de las unidades que forman la cadena. Cálculos simples muestran que:

$$\langle L_1\rangle = \int_{-\infty}^{\infty} L\, f(L)\, dL = \int_{15}^{16} L\, dL = 15.5\ \mathring{A}$$

$$\sigma_1^2 = \langle L_1^2\rangle - \langle L_1\rangle^2 = \int_{-\infty}^{\infty} L^2\, f(\mathrm{L})\, dL - 15.5^2 = \frac{1}{12}\ \mathring{A}^2$$

con lo que se obtiene:

$$\langle L_{1000}\rangle = 15500\ \mathring{A} \quad ; \quad \sigma^2_{1000} = 250/3\ \mathring{A}^2$$

Tomando *sólo* una desviación estándar como error, la longitud de la cadena es entonces:

$$L_{1000} = \langle L_{1000} \rangle \pm \sigma_{1000} = (15500 \pm 9{,}1287...) \simeq (15500 \pm 9)\ \mathring{A}$$

7. La función generatriz vendrá dada por:

$$\psi(t) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x) \cdot dx = \int_a^b e^{tx} \frac{1}{b-a} dx =$$

$$= \frac{e^{bt} - e^{at}}{t(b-a)} \quad ; \quad -\infty < t < +\infty$$

y la característica por:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \cdot f(x)\, dx = \frac{e^{ibt} - e^{iat}}{it(b-a)} \quad ; \quad -\infty < t < +\infty$$

Hay que recordar que el manejo en general de la transformación de Fourier requiere conocimientos de variable compleja y se recomienda precaución al trabajar con ella. Si la función a transformar $f(x)$ es real, entonces el asunto deja de ser tan complicado. En una buena parte de las aplicaciones éste será el caso y podremos operar como hemos hecho aquí.

8. La dispersión de la distribución de módulos v:

$$f(v) = 4\pi n \left(\frac{\beta m}{2\pi}\right)^{3/2} \cdot v^2 \cdot \exp\left[-\frac{\beta m}{2} v^2\right]$$

se evalúa como:

$$\sigma(v) = [\langle v^2 \rangle - \langle v \rangle^2]^{1/2}$$

El cálculo de $\langle v^2 \rangle$ conduce a:

$$\langle v^2 \rangle = \frac{\int_0^{\infty} v^2 \cdot f(v) \cdot dv}{\int_0^{\infty} f(v) \cdot dv} = \frac{4}{\sqrt{\pi}\, m\beta} \Gamma\left(\frac{5}{2}\right) = \frac{3kT}{m}$$

utilizando el resultado [105] se tiene entonces:

$$\sigma(v) = \left[\left(3 - \frac{8}{\pi}\right)\frac{kT}{m}\right]^{1/2}$$

A medida que la temperatura aumente, la distribución tenderá a ensancharse más y más en respuesta a la mayor velocidad que pueden adquirir las moléculas. Una disminución en este parámetro estrechará la distribución. Por el contrario, cuanto más pesadas sean las moléculas, más estrecha será la distribución dada la mayor inercia de aquéllas al movimiento, en tanto que moléculas ligeras poseerán distribuciones anchas. El balance entre ambas magnitudes está contenido en $\sigma(v)$.

El coeficiente de asimetría γ_1 viene dado por:

$$\gamma_1 = \mu_3/\sigma^3$$

siendo μ_3 el momento central de tercer orden:

$$\mu_3 = \langle (v - \langle v \rangle)^3 \rangle = \langle v^3 \rangle - 3\langle v \rangle \cdot \langle v^2 \rangle + 2\langle v \rangle^3$$

Para evaluar γ_1 sólo es necesario calcular $\langle v^3 \rangle$:

$$\langle v^3 \rangle = \frac{\int_0^\infty v^3 f(v) \cdot dv}{\int_0^\infty f(v) \cdot dv} = \{\Gamma(n = \text{entero}) = (n-1)!\} =$$

$$= \frac{4}{\pi^{1/2}}\left(\frac{2kT}{m}\right)^{3/2}$$

El momento central μ_3 es entonces:

$$\mu_3 = \sqrt{\frac{2}{\pi}}\left[\frac{32}{\pi} - 10\right]\left(\frac{kT}{m}\right)^{3/2}$$

y el coeficiente de asimetría γ_1:

$$\gamma_1 = \sqrt{\frac{2}{(3\pi - 8)^3}}(32 - 10\pi) > 0$$

siendo la asimetría de la distribución positiva. Comprobamos que ésta es una magnitud independiente de T y m, como cabría esperar, ya que la forma general de la función $f(v)$ es independiente del valor que tomen tales parámetros.

Tema 7

DISTRIBUCIONES DE MAGNITUDES ALEATORIAS EN UNA DIMENSION

Se estudiarán aquí ejemplos concretos de distribuciones de magnitudes aleatorias en una dimensión, dividiendo la presentación en caso discreto y continuo.

Comenzamos por las distribuciones discretas de tipo *binomial* (*binómica, Bernouilli, multinomial*) que encuentran aplicación en un buen número de situaciones de interés (sistemas de spines, fluctuaciones). Como límite de la distribución binómica está la de *Poisson*, muy útil para estudiar sucesos aleatoriamente distribuidos en el tiempo.

Pasamos a continuación a las distribuciones continuas entre las que ocupa un lugar principal la distribución *gaussiana*. Esta distribución la deduciremos a partir de la de Poisson mediante un paso al límite («*coarse grained*»), y su importancia se verá reflejada en el famoso teorema de Laplace. Otras dos distribuciones continuas, χ^2 y Γ, son introducidas e ilustradas con ejemplos sacados de la teoría cinética de gases. En el desarrollo de estos ejemplos se hace un uso extensivo de las propiedades de la δ de *Dirac* siguiendo lo visto en el tema anterior.

Finalmente, se considera el problema de la generación de *números aleatorios* y su empleo en el cálculo de integrales definidas. La razón de su inclusión aquí es que estos números se obtienen como una distribudión discreta (secuencia), en el intervalo real (0,1), que pretende representar a una distribución continua equiprobable. La utilidad de estas secuencias no está limitada al cálculo de integrales definidas, sino que se extiende a lo que comúnmente se denominan *técnicas de simulación aleatoria Monte Carlo* (Tema 9).

1. DISTRIBUCIONES DISCRETAS

1.1. Distribución binómica y binómica generalizada

Hay una gran cantidad de situaciones en Química Física que llevan a «experimentos» cuyo resultado puede resumirse en dos conocidas posibilidades: éxito si se presenta el suceso deseado; fracaso en caso contrario. De consideraciones teóricas (simetría, mecánica estadística, etc.) o de la realización de un gran número de experimentos es posible obtener las probabilidades de éxito y fracaso. Siguiendo la notación casi universal, denotaremos por p la probabilidad de éxito y por q la de fracaso. Evidentemente al haber sólo estas dos posibilidades: $p + q = 1$.

Para fijar ideas, consideremos el ejemplo de un sistema *ideal* de N spines (1/2). La «orientación de cualquier spin sólo puede ser $+1/2$ ó $-1/2$, que simplificaremos en «arriba» y «abajo», respectivamente. Al haber supuesto la interacción entre spines despreciable, si no hay campos magnéticos externos B actuando sobre el sistema, la probabilidad de que cualquier spin apunte hacia arriba será $p = 1/2$, y su complementaria $q = 1/2$. Este resultado es directo por simetría, pues no existen direcciones privilegiadas en el sistema. Sin embargo, de actuar un campo magnético B los momentos magnéticos de los spines tenderían a alinearse con él, ya que esta disposición sería más estable. En este caso es claro que $p > q$, estando sus valores numéricos dados por la distribución *canónica* de la mecánica estadística.

Determinados los valores p y q es sencillo establecer la *función generatriz* para la distribución que representan:

$$\psi(t) = p \cdot e^t + q \qquad [1]$$

en donde sin pérdida de generalidad hemos asignado el valor (arbitrario) $x = 1$ al suceso «apuntar arriba» y el valor $x = 0$ al «apuntar abajo» para un spin cualquiera.

Seguidamente debemos preguntarnos por la probabilidad p_n de que en tal sistema de N spines se presente una configuración con n momentos «arriba» y $n' = N - n$ momentos «abajo». Por ser los spines independientes, el suceso n spines «arriba» ($0 \leqslant n \leqslant N$) es la suma de lo que le

suceda a cada spin del conjunto total. La función generatriz para la suma X_b de variables independientes X_i sabemos que es:

$$\psi_b(t) = \sum_{n=0}^{N} p_n \cdot e^{nt} = (p \cdot e^t + q)^N = \sum_{n=0}^{N} \binom{N}{n} p^n \cdot q^{N-n} \cdot e^{nt} \qquad [2]$$

de donde la probabilidad buscada p_n resulta:

$$p_n = \binom{N}{n} p^n \cdot q^{N-n} \quad ; \quad 0 \leqslant (X_n = n) \leqslant N \qquad [3]$$

expresión que puede obtenerse alternativamente por argumentos combinatorios (Reif, 1975).

Encontramos así una distribución discreta que llamaremos *binómica* (o *binomial*), cuya masa unidad se distribuye en cantidades p_n sobre los puntos $X = n = 0, 1, 2, ..., N$. Es inmediato comprobar que $\Sigma\, p_n = 1$. Las características de esta distribución se obtienen con facilidad (Ejercicio 1) recordando que X_b es suma de variables independientes. Así su valor medio es:

$$\langle X_b \rangle = Np = \langle n \rangle \qquad [4]$$

su desviación típica:

$$\sigma(X_b) = \sqrt{Npq} \qquad [5]$$

y finalmente su asimetría:

$$\gamma_1(X_b) = \frac{q - p}{\sqrt{Npq}} \qquad [6]$$

Recuérdese que las expresiones [4]-[5] pueden determinarse alternativamente a través de las derivadas de la función generatriz [1] en $t = 0$ (Tema 6, epígrafe 4) representando e^t por su desarrollo en serie de t.

Examinemos a continuación el problema que suministra un gas ideal de N moléculas encerradas en un recipiente con dos compartimentos de volúmenes V_1 y V_2 que están conectados a través de un pequeño orificio. Nos interesa conocer la probabilidad de encontrar n moléculas en V_1.

La probabilidad de que una molécula esté en V_1 es sencillamente:

$$p = V_1/(V_1 + V_2)$$

y por tanto la de que no esté en V_1 será:

$$q = 1 - p = V_2/(V_1 + V_2)$$

La probabilidad pedida para las n moléculas vendrá dada por

$$p_n = \binom{N}{n} \cdot p^n \cdot q^{N-n} = \binom{N}{n} \left(\frac{V_1}{V_1 + V_2}\right)^n \cdot \left(\frac{V_2}{V_1 + V_2}\right)^{N-n} \qquad [7]$$

Si en esta expresión utilizamos el parámetro $\alpha = V_1/V_2$, encontramos:

$$p_n = \binom{N}{n} \alpha^n (1 + \alpha)^{-N} \qquad [8]$$

que es la ley de *distribución de Bernouilli* también conocida como binómica generalizada.

1.2. Ejemplo de aplicación: Sistemas de spines, fluctuaciones y distribución multinomial

Consideremos una red cristalina de N átomos de ^7Li cuyo núcleo en su estado más estable posee un spín nuclear $I = 3/2$. Las $2I + 1 = 4$ posibles proyecciones sobre el eje z: $3\hbar/2$, $\hbar/2$, $-\hbar/2$, $-3\hbar/2$, señalan las cuatro posibles orientaciones del momento magnético de spín μ_z^i $(i = 1, 2, 3, 4)$. Supondremos independientes a los spines nucleares.

En presencia de un campo magnético B en la dirección z y a la temperatura absoluta T, las probabilidades de que μ_z para un núcleo tome cada una de las proyecciones anteriores μ_z^i vienen dadas por (*distribución canónica*):

$$p_i = p(\mu_z^i) = \frac{\exp\left[\beta \mu_z^i B\right]}{\sum_i \exp\left[\beta \mu_z^i B\right]} \quad ; \quad \beta = 1/kT \qquad [9]$$

de donde se deduce que la proyección más probable es aquella que se alinea con B $(3\hbar/2)$, y la menos probable la antiparalela a éste $(-3\hbar/2)$.

Con esta información puede calcularse el valor medio $\langle \mu_z \rangle$ y la dispersión $\sigma(\mu_z)$ de un núcleo cualquiera:

$$\langle \mu_z \rangle = \sum_i p_i \cdot \mu_z^i \qquad [10]$$

$$\sigma(\mu_z) = \left(\sum_i p_i (\mu_z^i)^2 - [\sum_i p_i \cdot \mu_z^i]^2 \right)^{1/2} \qquad [11]$$

La componente media total $\langle M_z \rangle$ del momento magnético M para los N núcleos independientes y su dispersión $\sigma(M_z)$ resultan:

$$\langle M_z \rangle = N \langle \mu_z \rangle \qquad [12]$$

$$\sigma(M_z) = \sqrt{N}\, \sigma(\mu_z) \qquad [13]$$

Exploremos ahora la información contenida en [12]-[13]. Notemos que el valor medio crece proporcionalmente a N y que la dispersión lo hace con la raíz cuadrada de este número ($N^{1/2}$). Resulta natural definir la dispersión en unidades del valor medio, es decir la *dispersión relativa*:

$$\frac{\sigma(M_z)}{\langle M_z \rangle} = \frac{1}{\sqrt{N}} \cdot \frac{\sigma(\mu_z)}{\langle \mu_z \rangle} \qquad [14]$$

Si N es pequeño ($\sim 10^4$), encontramos:

$$\sigma(M_z) \sim 10^{-2} \langle \mathrm{M}_z \rangle \qquad [15]$$

que para campos no muy intensos conduce a:

$$\sigma(M_z) \lesssim \langle M_z \rangle \qquad [16]$$

Esto indica que los valores de M_z están muy desparramados alrededor de su valor medio $\langle M_z \rangle$. El valor $\sigma(M_z)$ es pues una medida de la *fluctuación* que experimenta la variable M_z. La situación considerada conduce pues a grandes fluctuaciones.

Si por el contrario N es un número muy grande ($\sim 10^{24}$), cual es habitual en los sistemas macroscópicos, se tiene:

$$\sigma(M_z) \sim 10^{-12} \langle M_z \rangle \qquad [17]$$

lo que implica un esparcimiento muy pequeño alrededor de $\langle M_z \rangle$:

$$\sigma(M_z) << \langle M_z \rangle \qquad [18]$$

Concluimos así que, aunque la magnitud absoluta de las fluctuaciones en un sistema macroscópico es importante ($\sim N^{1/2}$), su magnitud relativa es en tales sistemas despreciable. Por ello cuando se efectúa la medida del momento magnético M_z el valor que siempre se obtiene es $\langle M_z \rangle$. *Las fluctuaciones quedan enmascaradas como consecuencia del gran número de partículas presentes*, ya que para apreciar aquéllas deberíamos ser capaces de medir M_z con una precisión del orden de 10^{-12}.

Para concluir esta aplicación nos queda por establecer la probabilidad de encontrar una configuración que posea n_1 spines $3\hbar/2\,(p_1)$, n_2 spines $\hbar/2\,(p_2)$, n_3 spines $-\hbar/2\,(p_3)$ y n_4 spines $-3\hbar/2\,(p_4)$. Es un sencillo ejercicio combinatorio encontrar la fórmula (*distribución multinomial*):

$$p(n_1, n_2, n_3, n_4) = \frac{N!}{n_1! n_2! n_3! n_4!} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4} \qquad [19]$$

debiéndose cumplir

$$n_1 + n_2 + n_3 + n_4 = N \qquad [20]$$

$$p_1 + p_2 + p_3 + p_4 = 1 \qquad [21]$$

Repárese en que [19] surge de desarrollar la potencia N-ésima de [21]:

$$(p_1 + p_2 + p_3 + p_4)^N = \sum_{n_1} \sum_{n_2} \sum_{n_3} \sum_{n_4} p(n_1, n_2, n_3, n_4) = 1 \qquad \{n_1 + n_2 + n_3 + n_4 = N\} \qquad [22]$$

lo que no es más que una generalización de la distribución binómica.

1.3. Distribución de Poisson

Recordando la distribución de Bernouilli dada en [8] vamos a obtener mediante un paso al límite la distribución de Poisson. Hagamos en [8] que $V_2 \to \infty$, $N \to \infty$, $N/V_2 = \rho =$ constante:

$$\lim_{N,V_2\to\infty} p_n = \lim_{N,V_2\to\infty} \binom{N}{n}\alpha^n(1+\alpha)^{-N} =$$

$$= \lim_{N,V_2\to\infty}\left[\frac{N!}{(N-n)!n!}\alpha^n\left[\left(1+\frac{1}{V_2/V_1}\right)^{V_2/V_1}\right]^{-NV_1/V_2}\right] =$$

$$= \lim_{N,V_2\to\infty}\left[\frac{(\alpha N)^n}{n!}[1-1/N]\ldots\left[1-\frac{n-1}{N}\right]\left[\left(1+\frac{1}{V_2/V_1}\right)^{V_2/V_1}\right]^{-NV_1/V_2}\right]$$

[23]

Pongamos $\lambda = \rho V_1$, transformándose [23] en:

$$P_n = \lim_{N,V_2\to\infty} p_n = \frac{\lambda^n}{n!}e^{-\lambda} \quad ; \quad X_n = n = 0, 1, 2, \ldots \qquad [24]$$

Esta es la *distribución de Poisson* que como se ve depende únicamente de un parámetro (λ) relacionado con la naturaleza del problema que consideremos. P_n es el límite de la probabilidad binomial p_n cuando el número de ensayos N se hace muy grande y la probabilidad de éxito p se hace muy pequeña. Que las P_n son probabilidades es claro pues son no negativas y cumplen $\Sigma P_n = 1$.

En el caso del gas ideal de N moléculas, P_n representa la distribución probabilista del número n de moléculas independientes en una región V_1 pequeña contenida dentro de (o comunicada con) un almacén infinito (a efectos prácticos, muy grande) de moléculas.

Pasemos al cálculo de las características habituales en distribuciones. Utilizando la *función generatriz*:

$$\psi(t) = e^{-\lambda}\sum_{n=0}^{\infty}\frac{1}{n!}(\lambda e^t)^n = \exp[\lambda(e^t-1)] \qquad [25]$$

se obtienen sin dificultad la media y la varianza a través de los momentos con respecto al origen α_1 y α_2:

$$\alpha_1 = \lambda \quad ; \quad \alpha_2 = \lambda + \lambda^2 \qquad [26]$$

$$\langle X_P\rangle = \langle n\rangle = \alpha_1 = \lambda \quad ; \quad \sigma^2(X_P) = \alpha_2 - \alpha_1^2 = \lambda \qquad [27]$$

Estos resultados son llamativos, puesto que la media y la varianza coinciden en su valor numérico (no en sus unidades), siendo éste justamente el parámetro que define a cada distribución de Poisson en particular:

$$P_n = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle} \quad ; \quad n = 0, 1, 2, \ldots \qquad [28]$$

La importancia de esta nueva distribución es grande, pudiendo reseñar que siguen esta ley:

a) Sucesos aleatoriamente distribuidos en el tiempo: desintegración de sustancias radiactivas, el flujo de llamadas en una línea telefónica, los seguros de accidentes, etc.

b) El número total de veces que se presenta un suceso, que posee una probabilidad muy pequeña, cuando se efectúa un número de experimentos *independientes* muy grande. Este es el caso de la teoría de fluctuaciones para sistemas de partículas en volúmenes pequeños cuyas consecuencias tienen que ver con el color azul del cielo diurno, la dispersión de la luz por muestras líquidas, etc. (McQuarrie, 1976; Sesé y Criado, 1990).

Sin embargo, aunque resulta muy útil para sucesos distribuidos en $n = 0, 1, 2, \ldots$ (enteros positivos), la distribución de Poisson no es tan universal como parece a primera vista. Esto se debe a la hipótesis de sucesos independientes que subyace a la deducción anterior. Además, que la media y la varianza coincidan no es algo precisamente universal. A pesar de ello, el valor de esta distribución puede ser alto incluso en los casos que se escapan de ella, pudiendo constituir una buena hipótesis de prueba (Ejercicio 2).

2. DISTRIBUCIONES CONTINUAS

2.1. Distribución normal (gaussiana)

Abordaremos a continuación el estudio de una distribución continua de capital importancia en las aplicaciones prácticas. Nos referimos a la

distribución normal o *gaussiana*, que representa un caso límite de las dos distribuciones discretas ya vistas. Para obtenerla partiremos de la distribución de Poisson (Chandrasekhar, 1943), aunque igualmente puede determinarse de la binómica (Cramér, 1977). La expresión de la distribución de Poisson:

$$P_n = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle}$$

la reescribiremos tomando logaritmos neperianos (log) como

$$\log P_n = n \log \langle n \rangle - \langle n \rangle - \log n! \qquad [29]$$

Sea $\langle n \rangle$ un número muy grande, lo que sucede frecuentemente en muchas aplicaciones, por lo que el interés se dirigirá a valores de n relativamente próximos a $\langle n \rangle$.

Utilizando la aproximación de Stirling para calcular el logaritmo neperiano del factorial de un número grande:

$$\log n! = (n + 1/2) \log n - n + \frac{1}{2} \log 2\pi + \theta(n^{-1}) \quad ; \quad n \to \infty \qquad [30]$$

en donde $\theta(n^{-1})$ es el término de error de la aproximación, y haciendo

$$n = \langle n \rangle + \Delta \qquad [31]$$

la ecuación [29] se convierte en:

$$\log P_n = -\left(\langle n \rangle + \Delta + \frac{1}{2}\right) \log\left(1 + \frac{\Delta}{\langle n \rangle}\right) + \Delta - \frac{1}{2} \log (2\pi \langle n \rangle) + \theta(n^{-1}) \qquad [32]$$

De acuerdo con lo dicho anteriormente, podemos suponer que $\Delta/\langle n \rangle \ll 1$ y por tanto:

$$\log\left(1 + \frac{\Delta}{\langle n \rangle}\right) \simeq \frac{\Delta}{\langle n \rangle} - \frac{1}{2}\left(\frac{\Delta}{\langle n \rangle}\right)^2 \qquad [33]$$

resultando que [32] es:

$$\log P_n = -\left(\frac{\Delta^2}{2\langle n\rangle}\right) - \frac{1}{2}\log(2\pi\langle n\rangle) \quad ; \quad \langle n\rangle \to \infty \qquad [34]$$

o equivalentemente:

$$P_n = \frac{1}{\sqrt{2\pi\langle n\rangle}} \exp\left[-\frac{(n-\langle n\rangle)^2}{2\langle n\rangle}\right] \qquad (35)$$

operaciones que tienen sentido, insistimos, para Δ y $n \gg 1$. La aproximación realizada recibe el apelativo de «*grano grueso*» del término inglés «*coarse grained*» (Kampen, 1985), mediante la que se representa una distribución discreta por otra continua.

La fórmula [35] es una forma particular de la distribución gaussiana general con función densidad:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{(x-\langle X\rangle)^2}{2\sigma^2}\right] \quad ; \quad -\infty < x < +\infty \qquad [36]$$

en donde σ (>0) es la desviación típica de la distribución. La especialización [36] a [35] es clara sin más que recordar $\sigma^2_{\text{Poisson}} = \langle n\rangle$. Notemos que con esta sustitución la exponencial de [35] actúa sobre una cantidad que es el cociente entre dos números de misma dimensión (orden dos en la variable). El sentido del factor preexponencial allí es justamente el de ser la *constante de normalización* (Ejercicio 3).

La función gaussiana es igualmente el límite al que tiende una distribución binomial cuando el número de ensayos es muy elevado. Se comprueba que cuando $Np, Nq > 5$, la distribución binomial es prácticamente una distribución normal (Spiegel, 1970).

Es común tipificar la distribución normal general dada en [36] haciendo el cambio de variable:

$$Z = \frac{X - \langle X\rangle}{\sigma} \qquad [37]$$

Dado que una distribución normal sólo requiere para su definición dos

parámetros, $\langle X \rangle$ y σ, su forma universal en la variable tipificada Z es (Fig. 1):

$$g(z) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-z^2/2\right] \quad ; \quad -\infty < z < +\infty \qquad [38]$$

ya que $\langle Z \rangle = 0$ y $\sigma(Z) = 1$. Como se observa por simple inspección, $g(x)$ dada en [36], o su equivalente [38], es completamente simétrica en torno al punto medio de la distribución. En consecuencia su coeficiente de asimetría es $\gamma_1 = 0$.

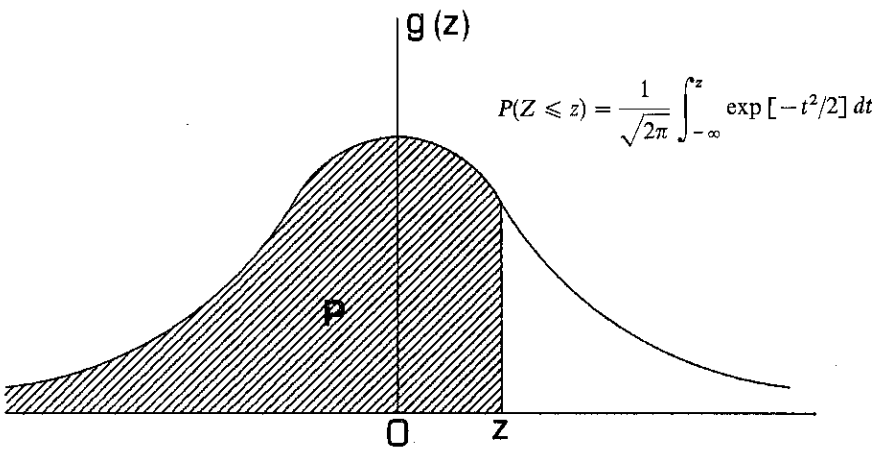


Figura 1.—Distribución normal.

Para la distribución [38] la función integral es claramente:

$$F(z) = P(Z \leqslant z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left[-t^2/2\right] \cdot dt \qquad [39]$$

y, por supuesto, el cálculo de probabilidades de sucesos sigue las reglas dadas en el tema previo. Si deseamos evaluar $P(z_1 \leqslant Z \leqslant z_2)$ se calculará la integral:

$$P(z_1 \leqslant Z \leqslant z_2) = F(Z \leqslant z_2) - F(Z \leqslant z_1) =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} \exp\left[-t^2/2\right] dt \qquad [40]$$

es decir la probabilidad pedida es el área encerrada por la curva normal tipificada entre $Z = z_1$ y $Z = z_2$. La gran ventaja de trabajar con la función tipificada es que los valores de $F(Z \leqslant z)$ están tabulados (Spiegel, 1968; Abramowitz y Stegun, 1972), por lo que la tarea de calcular integrales de los tipos [39]-[40] no es necesaria (Ejercicio 4).

Escribamos a continuación la *función generatriz* de una variable normal X con media $\langle X \rangle$ y desviación típica σ:

$$\psi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x - \langle X \rangle)^2}{2\sigma^2} \right] \cdot e^{tx} \cdot dx \qquad [41]$$

que tras alguna manipulación (Ejercicio 5) se transforma en:

$$\psi(t) = \exp\left[t\langle X \rangle + \frac{1}{2}\sigma^2 t^2 \right] \qquad [42]$$

Esta última expresión resulta extremadamente útil si nos preguntamos por la distribución de la suma X de n variables aleatorias independientes X_i con medias $\langle X_i \rangle$ y desviaciones típicas σ_i:

$$X = X_1 + X_2 + \ldots + X_n \qquad [43]$$

Dada la independencia estadística entre las X_i, la función generatriz será el producto:

$$\psi(t) = \prod_{i=1}^{n} \psi_i(t) = \exp\left[\left(\sum_i \langle X_i \rangle \right) \cdot t + \frac{1}{2}\left(\sum_i \sigma_i^2\right) t^2 \right] \qquad [44]$$

observándose que X se distribuye también normalmente con media:

$$\langle X \rangle = \sum_i \langle X_i \rangle \qquad [45]$$

y varianza:

$$\sigma^2 = \sum_i \sigma_i^2 \qquad [46]$$

En general cualquier combinación lineal de variables normales independientes:

$$X = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b \qquad [47]$$

da origen a una variable aleatoria que se distribuye normalmente.

Todavía podemos ampliar nuestro conocimiento de la distribución normal enunciando sin demostración el famoso *teorema de Laplace*, formulado primeramente en 1812 y generalizado posteriormente por otros autores. A la vista de su enunciado, también conocido como el *Teorema Central del Límite*, se hará patente la tremenda importancia de la distribución normal en las aplicaciones prácticas. El teorema reza como sigue (Cramér, 1977):

En condiciones muy generales relativas a las distribuciones de un número n de variables aleatorias independientes X_i ($\langle X_i \rangle, \sigma_i$), la distribución de su suma X tipificada:

$$Z = \frac{X - \langle X \rangle}{\sigma} \quad ; \quad \langle X \rangle = \sum_i \langle X_i \rangle \quad ; \quad \sigma^2 = \sum_i \sigma_i^2 \qquad [48]$$

es aproximadamente normal para *valores grandes* de n. Si la secuencia $\{X_i\}$ es infinita, se dice entonces que la variable Z es *asintóticamente normal*:

$$\lim_{n \to \infty} P(Z \leqslant z) = F(z) \qquad [49]$$

Es muy importante notar que no se ha hecho referencia alguna a la naturaleza de las distribuciones particulares de cada X_i.

La generalización de este teorema conduce a establecer que no sólo la suma de las X_i es prácticamente normal, sino que funciones más generales de las X_i (suma de potencias m-ésimas de las variables, etc.) también se distribuyen aproximadamente según una gaussiana cuando n es grande.

Claramente, siempre que una variable sea la suma de un gran número de contribuciones independientes, cada una de las cuales contribuya muy poco al resultado final, tal suma se distribuirá de forma aproximadamente normal. El mundo de la Física, Química, Biología, etc., está plagado de situaciones en las que el aserto anterior posee aplicación. Nos encontramos pues ante una distribución de gran valor teórico y práctico, que es al caso continuo lo que la de Poisson al caso discreto.

Ejercicio de aplicación

Antes de pasar a estudiar la distribución χ^2 conviene examinar la aplicación siguiente. Dada una variable aleatoria X normal con función densidad:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad ; \quad -\infty < x < +\infty$$

determinar la función densidad $g(y)$ de la variable $Y = X^2$. Utilizaremos para ello las relaciones:

$$\int_{-\infty}^{\infty} h(x)\delta(x - a)\, dx = h(a)$$

$$\delta(x^2 - y) = \frac{1}{2\sqrt{y}} [\delta(x - \sqrt{y}) + \delta(x + \sqrt{y})] \quad ; \quad y > 0$$

La relación entre $g(y)$ y $f(x)$ a través de las δ-Dirac sabemos ya que es:

$$g(y) = \int_{-\infty}^{\infty} \delta(x^2 - y) f(x)\, dx$$

Es inmediato entonces en nuestro caso establecer:

$$g(y) = \frac{1}{2\sqrt{2\pi y}} \left[\int_{-\infty}^{\infty} \delta(x - \sqrt{y}) e^{-x^2/2}\, dx + \int_{-\infty}^{\infty} \delta(x + \sqrt{y}) e^{-x^2/2}\, dx \right]$$

y aplicando la propiedad básica del operador δ llegamos a:

$$g(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} \cdot y^{-1/2} \cdot e^{-y/2} & ; \quad y \geqslant 0 \\ 0 & ; \quad y < 0 \end{cases} \qquad [50]$$

Nótese que $g(0) \to +\infty$.

2.2. Distribución chi-cuadrado (χ^2)

Fijemos nuestra atención ahora en una variable aleatoria que denominaremos χ^2 y que se construye como la suma de los cuadrados de n variables aleatorias independientes X_j:

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2 \qquad [51]$$

todas ellas normales con la misma distribución:

$$f(x_j) = \frac{1}{\sqrt{2\pi}} e^{-x_j^2/2} \quad ; \quad j = 1, 2, ..., n \qquad [52]$$

Apoyándonos en el ejercicio de aplicación previo calcularemos la función de distribución de χ^2.

Sabemos que cuando X_j es normal la distribución *de* X_j^2 es:

$$g(y_j) = g(x_j^2) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} y_j^{-1/2} \cdot e^{-y_j/2} & ; \quad y_j \geqslant 0 \\ 0 & ; \quad y_j < 0 \end{cases} \qquad [53]$$

por lo que la *función característica* asociada a X_j^2 resulta:

$$\varphi_j(t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{ity_j} \cdot y_j^{-1/2} \cdot e^{-y_j/2} \cdot dy_j \qquad [54]$$

Esta integral se evalúa con facilidad recordando la definición integral de la función Gamma de Euler:

$$\Gamma(n) = \int_0^\infty t^{n-1} \cdot e^{-t} \cdot dt \quad ; \quad n > 0 \qquad [55]$$

Se obtiene así:

$$\varphi_j(t) = \frac{1}{\sqrt{2\pi}} \int_0^\infty y_j^{-1/2} \cdot \exp\left[-(1 - 2ti)y_j/2\right] dy_j =$$

$$= \frac{(1 - 2ti)^{-1/2}}{\sqrt{2\pi}} \cdot 2^{1/2}\Gamma(1/2) = (1 - 2ti)^{-1/2} \qquad [56]$$

Nótese que en este caso el resultado puede obtenerse por manipulación formal de la integral [54], tratando al número complejo $(1 - 2ti)$ como una constante con respecto a la integración sobre la variable y_j. En rigor el cálculo de [54] debería hacerse siguiendo un camino en el plano complejo $z = x + iy$ que evitase la singularidad presente en el integrando $(y_j = 0)$ (Schwartz, 1969).

Por ser las X_i^2 independientes la función característica de su suma χ^2 es simplemente:

$$\varphi(t) = (1 - 2ti)^{-n/2} \quad ; \quad -\infty < t < +\infty \qquad [57]$$

diciéndose que χ^2 tiene entonces n grados de libertad.

La función densidad $f(x)$ asociada a χ^2 puede determinarse con facilidad (Rozanov, 1975). Notemos que si en $g(y_j)$ [53] se tuviera $y_j^{n/2-1}$ en lugar de $y_j^{-1/2}$, entonces haciendo $y = y_j$:

$$N \int_0^\infty y^{n/2-1} \cdot \exp\left[-\frac{(1 - 2it)y}{2}\right] dy = (1 - 2ti)^{-n/2} \qquad [58]$$

siendo N un factor de normalización. Repárese en que la integral en [58] es la Transformada de Fourier de la función densidad:

$$f(y) = Ny^{n/2-1} \cdot e^{-y/2} \quad ; \quad 0 \leqslant y < \infty \qquad [59]$$

por lo que:

$$\varphi(t) = \int_0^\infty e^{ity} \cdot f(y)\, dy \qquad [60]$$

siendo N:

$$N = \left[\int_0^\infty y^{n/2-1} \cdot e^{-y/2} \cdot dy\right]^{-1} = (2^{n/2} \cdot \Gamma(n/2))^{-1} \qquad [61]$$

Encontramos pues que la función densidad para χ^2 con n grados de libertad es:

$$k_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} \cdot x^{n/2-1} \cdot e^{-x/2} \quad ; \quad 0 \leqslant x < +\infty \qquad [62]$$

2.3. Ejemplo de aplicación: Función de distribución del módulo de la velocidad molecular en un gas

Ilustremos cómo con los recursos estadísticos aprendidos se pueden obtener fórmulas de interés físico-químico, deducibles en base a otro tipo de argumentaciones. Siguiendo con la Teoría Cinética de Gases, nos dirigiremos a la distribución χ^2 que se origina de la suma de los cuadrados de las componentes de la velocidad molecular. Esta χ^2 particular representa al cuadrado del módulo de tal velocidad:

$$v^2 = v_x^2 + v_y^2 + v_z^2 \qquad [63]$$

Nos son conocidos ya por la *distribución de Maxwell* la distribución normal de cada componente, su valor medio $\langle v_i \rangle = 0$ y su desviación típica $\sigma_i = (kT/m)^{1/2} = \sigma$.

Para estar dentro completamente de las condiciones de la distribución χ^2 pondremos:

$$\left(\frac{v}{\sigma}\right)^2 = \left(\frac{v_x}{\sigma}\right)^2 + \left(\frac{v_y}{\sigma}\right)^2 + \left(\frac{v_z}{\sigma}\right)^2 \qquad [64]$$

refiriéndonos pues a la variable $(v/\sigma)^2$ = módulo de la velocidad al cuadrado en unidades de σ^2 como variable χ^2. El número de grados de libertad de esta distribución es $n = 3$, por lo que la función densidad para su variable $w = (v/\sigma)^2$ es según [62]:

$$k_3(w) = \frac{1}{\sqrt{2\pi}} w^{1/2} \cdot e^{-w/2} \quad ; \quad 0 \leqslant w < +\infty \qquad [65]$$

La función densidad para la variable v = módulo de la velocidad:

$$v = +\sigma\sqrt{w} \qquad [66]$$

se puede realizar a través de la integral:

$$f(v) = \int_0^\infty \delta(\sigma\sqrt{w} - v) \cdot \frac{1}{\sqrt{2\pi}} \cdot w^{1/2} \cdot e^{-w/2} \cdot dw \qquad [67]$$

Utilizando el cambio de variable admisible (biunívoco o monótono)

$$t = \sigma\sqrt{w} \quad ; \quad w \geqslant 0 \quad , \quad t \geqslant 0 \qquad [68]$$

que conduce a

$$w = t^2/\sigma^2 \quad ; \quad dw = \frac{2t}{\sigma^2}\,dt \qquad [69]$$

y aplicando la propiedad elemental (*convolución*) de la δ-Dirac obtenemos:

$$f(v) = \frac{2}{\sqrt{2\pi}} \int_0^\infty \delta(t - v) \cdot \frac{t^2}{\sigma^3} \cdot \exp\left[-t^2/2\sigma^2\right] dt = \left\{\sigma = \left(\frac{kT}{m}\right)^{1/2}\right\} =$$

$$= \sqrt{\frac{2}{\pi}} \left(\frac{m}{kT}\right)^{3/2} v^2 \cdot \exp\left[-\frac{mv^2}{2kT}\right] \qquad [70]$$

que es la expresión deducida en el tema anterior para $f(v)$ (se ha normalizado aquí con $n = 1$).

Merece la pena resaltar que las manipulaciones algebraicas de cambio de variable en la función $k_3(w)$ de [67] vienen impuestas por el *operador* δ. Fijémonos en que para poder aplicar la propiedad básica (convolución) de δ se ha efectuado la manipulación siguiente:

$$f(v) = \int_0^\infty \delta(\sigma\sqrt{w} - v) k_3(w)\, dw = \int_0^\infty \delta(\sigma\sqrt{w} - v) \bar{k}_3(\sigma\sqrt{w}) d(\sigma\sqrt{w})$$

2.4. Distribución gamma (Γ)

La distribución chi-cuadrado anterior no es sino un caso particular de la llamada *familia de distribuciones gamma* (o tipo III de *Pearson*). Aunque sobra la aclaración, no debe confundirse esta distribución con la integral Γ de Euler. La relación que define a la familia que nos ocupa es:

$$f(x) = \frac{a^v}{\Gamma(v)} \cdot x^{v-1} \cdot e^{-ax} \quad ; \quad a > 0 \quad , \quad v > 0 \quad , \quad 0 \leqslant x < +\infty \qquad [71]$$

y por identificación directa establecemos que $k_n(x)$ [62] es una distribución gamma con parámetros:

$$v = n/2 \quad ; \quad a = 1/2$$

Es fácil comprobar que la integral $\Gamma(v)$ [55] junto con a^v juegan el papel de constante normalizadora para esta distribución.

2.5. Ejemplo de aplicación: función de distribución de las energías moleculares en un gas

Un ejemplo típico de la distribución gamma, y de nuevo extraído de la Teoría Cinética de Gases, es la función de distribución de las energías moleculares E:

$$E = \frac{m}{2}(v_x^2 + v_y^2 + v_z^2) \qquad [72]$$

La relación que nos interesa puede determinarse mediante argumentos de tipo físico como se hace en los cursos de Química Física General. Veamos cómo obtenerla por métodos estadísticos partiendo de la conocida ley de velocidades de Maxwell:

$$f(v_x, v_y, v_z) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{m}{2kT}(v_x^2 + v_y^2 + v_z^2)\right] \qquad [73]$$

que aparece aquí normalizada, de nuevo, a la unidad en vez de a $n = N/V$.

El problema que nos planteamos es transformar $f(v_x, v_y, v_z)$ en $f(E)$, es decir cambiar de una distribución con tres variables a otra distribución sólo con una. Hemos encontrado repetidas veces la expresión que relaciona funciones densidad de dos variables X e Y ligadas por $Y = F(X)$ y en la que aparece el operador δ. Tal relación sigue siendo válida cuando X e Y poseen varias componentes. Para el caso que nos ocupa tenemos:

$$f(E) = \int_{-\infty}^{\infty}\int\int \delta\left(\frac{mv^2}{2} - E\right) \cdot f(v_x, v_y, v_z) \cdot dv_x\, dv_y\, dv_z \qquad [74]$$

que transformada a coordenadas polares v, θ, φ, en el espacio de velocidades es:

$$f(E) = 4\pi \int_0^\infty \delta\left(\frac{mv^2}{2} - E\right) \cdot \left(\frac{m}{2\pi kT}\right)^{3/2} \cdot \exp\left[-\frac{mv^2}{2kT}\right] v^2 \cdot dv \quad [75]$$

Tras un cálculo directo (Ejercicio 6) se obtiene la conocida ecuación:

$$f(E) = \frac{2}{\sqrt{\pi}} \left(\frac{1}{kT}\right)^{3/2} \cdot E^{1/2} \cdot \exp\left[-E/kT\right] \quad [76]$$

3. NUMEROS ALEATORIOS

Este epígrafe podría haberse titulado ¿se puede simular al azar numéricamente? También podríamos preguntarnos el porqué de esta caprichosa operación cuando realizando determinados ensayos (arrojar un dado, etc.) podemos dar imágenes del comportamiento aleatorio. Centremos la discusión sobre los problemas de interés en Química Física.

En primer término puede desearse estudiar un problema sin pretender obtener una ecuación que lo describa completamente pues, con seguridad, la gran cantidad de información contenida en ella no sería útil. En estas circunstancias una buena alternativa la constituye el diseñar un «juego» de azar a través del cual obtengamos una respuesta aproximada al problema.

En segundo término, la situación física pudiera resumirse en una ecuación difícil o imposible de resolver. Un método estadístico, entonces, podría suministrar una solución al problema.

Es muy importante destacar que la respuesta así obtenida será siempre aproximada, como corresponde a haber «simulado» estadísticamente el sistema estudiado. Además, en gran parte de los casos, *esta simulación puede no tener nada que ver con el comportamiento físico real del sistema*, no siendo más que un medio para obtener resultados finales.

Dada la complejidad de los sistemas con que se trabaja, abordar su estudio por simulación (muchas veces el único medio posible) requerirá el uso del cálculo con ordenador. Es aquí donde reside la necesidad de realizar ensayos numéricos que de alguna manera representen un com-

portamiento aleatorio. De acuerdo con ciertas reglas, establecidas por la teoría subyacente al fenómeno, estos ensayos se combinan de modo que se llegue a una respuesta final.

Resulta curioso que estando la Naturaleza llena de fenómenos aleatorios (supuestamente, ya que esta hipótesis funciona), en las situaciones de interés una de las soluciones más eficaces para el cálculo sea simular el azar con ensayos numéricos que no son, en absoluto, al azar, sino que dependen unos de otros (para otros mecanismos véase (Sóbol, 1976)). La producción de *números aleatorios* (*pseudoaleatorios* para ser exactos) se realiza mediante determinados algoritmos generadores (Hammersley y Handscomb, 1983; Press y col., 1988), entre los cuales los más populares son los *uniformes.*

3.1. Familias multiplicativas congruentes

Los algoritmos de generación uniforme dan secuencias $x_1, x_2, ..., x_n, ...$, de números reales dentro del intervalo (0,1), que parecen ser muestreos independientes de la densidad rectangular:

$$f(x) = \begin{cases} 1 & \text{si} \quad 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases} \qquad [77]$$

por lo que tales x_i son *equiprobables.* Siempre se utiliza el intervalo (0, 1) pues la correspondiente transformación lineal lo convertirá en el (a, b) que interese. Los generadores más corrientes de este tipo son los de las *familias multiplicativas congruentes.*

En este grupo se generan enteros positivos $z_1, z_2, ..., z_n, ...$, de acuerdo con recetas numéricas del tipo (Wood, 1975):

$$z_n = \lambda z_{n-1} \,(\text{mód } M) \qquad [78]$$

siendo, por ejemplo, λ y z_1 enteros positivos tales que:

$$\lambda \equiv 5 \,(\text{mód } 8) \qquad [79]$$

$$z_1 \equiv \begin{cases} 1 \\ 5 \end{cases} \,(\text{mód } 8) \qquad [80]$$

y M un entero tal que $M = 2^B$, donde B es otro entero positivo.

La expresión [79] se lee «λ *es congruente con* 5 *en módulo* 8», y significa que λ y 5 dan el mismo resto entero al dividirlos por 8. La secuencia pseudoaleatoria $x_1, x_2, ...$, surge de hacer:

$$x_i = z_i/M \quad ; \quad i = 1, 2, 3, ... \qquad [81]$$

Ejercicio de aplicación

Veamos en un caso sencillo cómo generar una secuencia x_i. Para ello, elegiremos un valor de M ciertamente reducido $M = 32$, y sean $\lambda = 5$ y $z_1 = 1$. La construcción de los números pseudoaleatorios puede esquematizarse como sigue:

$$z_1 = 1 \qquad ; \quad x_1 = z_1/M = 1/32 = 0{,}03125$$
$$z_2 = \lambda z_1 (\text{mód } 32) = 5 \quad ; \quad x_2 = 5/32 = 0{,}15625$$
$$z_3 = \lambda z_2 (\text{mód } 32) = 25 \quad ; \quad x_3 = 25/32 = 0{,}78125$$
$$z_4 = \lambda z_3 (\text{mód } 32) = 125 (\text{mód } 32) = \text{resto entero de } (125/32) = 29$$
$$; \quad x_4 = 29/32 = 0{,}90625$$
$$z_5 = \lambda z_4 (\text{mód } 32) = 145 (\text{mód } 32) = 17$$
$$; \quad x_5 = 17/32 = 0{,}53125$$

y así sucesivamente.

Para realizar el ejercicio anterior hemos deliberadamente elegido unos valores sencillos para λ, z_1 y M. Es fácil darse cuenta de que el granulado o espaciado de la secuencia x_i será tanto más fino cuanto mayor sea M, constante que juega el papel de «normalizadora» para que $0 < x_i < 1$. De haber seguido calculando valores ($x_6, x_7, ...$), habríamos tropezado con un resultado curioso: la secuencia se repite una vez calculados los 8 primeros términos. Es decir $x_9 = x_1$, $x_{10} = x_2$, etc., existiendo pues *periodicidad*. En general, el período de una de estas secuencias es 2^{B-2}. Si tenemos en consideración que en aplicaciones con ordenador $B = 32$, 48, etc. (según sea la arquitectura), vemos que la cantidad de números aleatorios es tan grande que los efectos de *correlación* entre ellos quedan bastante enmascarados (Wood, 1968).

Resta por analizar, no obstante, la calidad de estas secuencias. Determinada, hasta periodicidad, una de ellas, podemos estimar su bondad

comparando la distribución discreta equiprobable que forma con la distribución a la que pretende representar [77]. Así cuánto más próximos estén los momentos de ambas distribuciones tanto mejor será la secuencia ($\langle x_i \rangle \to \langle x \rangle = 0{,}5$, etc.). Procediendo así, claro está, no obtendremos más que una idea aproximada de cómo van las cosas.

Un «test» más poderoso está implícito en el siguiente razonamiento. Si la secuencia x_i debe ser uniforme en (0, 1), entonces las parejas (x_i, x_{i+1}) deben distribuirse igualmente de forma uniforme en el *cuadrado unidad* $(0, 1) \times (0, 1)$. Esto mismo debe suceder para los conjuntos de n-uplas $(x_i, x_{i+1}, ..., x_{i+n-1})$ en sus correspondientes *hipercubos unidad.* La aplicación de este test lleva a obtener para el algoritmo tipo [78]-[80] acumulaciones de puntos n-dimensionales en hiperplanos paralelos (*efecto Marsaglia*). Este efecto pudiera resultar nocivo en cálculos con ordenadores de arquitecturas menores o iguales a 36-bits, sobre todo en problemas de integración numérica. Con arquitecturas superiores, claramente, estos efectos indeseables dejan de ser importantes por la gran cantidad de números distintos que se producen. Existen algoritmos generadores que obvian en parte estas dificultades, pero no nos detendremos en ellos. El lector interesado puede consultar la referencia (James, 1980).

3.2. Ejemplo de aplicación: Cálculo de integrales definidas por el método de Monte Carlo

El uso de un flujo de *números aleatorios* puede resultar muy útil en el muestreo de un conjunto (población), seleccionando al azar un número, comparativamente pequeño sobre el total, de sus elementos. Hecha esta operación es posible estimar las características (muestrales) de ese conjunto (media, varianza, etc.). En definitiva, estas características son valores medios (integrales definidas) con respecto a la función de distribución pertinente. Es fácil imaginar pues que las integrales definidas en general pueden ser aproximadas mediante un método que utilice números aleatorios.

Se denomina método de *Monte Carlo*, en general, a la utilización de un flujo de números aleatorios para resolver un determinado problema. En el cálculo de integrales definidas su aplicación es como sigue.

Sea la integral monodimensional de $h(x)$ entre $x_1 = a$, $x_2 = b$. La aproximaremos por el promedio:

$$\int_a^b h(x)\,dx \simeq \frac{(b-a)}{N} \sum_{i=1}^{N} h(x_i) \qquad [82]$$

siendo x_i puntos *uniformemente* distribuidos en (a, b). Si recordamos la definición de la integral de Riemann

$$\int_a^b h(x)\,dx = \lim_{\Delta x \to 0} \sum_i h(x_i)\,\Delta x \quad ; \quad a < x_i < b \qquad [83]$$

el significado de [82] se aclara todavía más. Haciendo $N \to \infty$ en [82] se tiene:

$$\lim_{N\to\infty} \frac{b-a}{N} \sum_{i=1}^{N} h(x_i) = \lim_{\Delta x \to 0} \sum_i h(x_i)\Delta x = \int_a^b h(x)\,dx \qquad [84]$$

que es el resultado exacto.

En este punto nos surge un problema delicado. Dado que nunca se podrá en una realización práctica hacer $N \to \infty$, ¿cómo dependerá el error cometido, al emplear la discretización, con el número de puntos N utilizado? Sabemos que cuanto mayor sea N tanto mejor será el resultado, pero esto naturalmente no es suficiente. Vamos a encontrar la dependencia explícita haciendo uso de la distribución gaussiana.

Dado el intervalo (a, b), se realizan N ensayos independientes $\{x_i\}$, determinando N números aleatorios uniformemente distribuidos en tal intervalo. El resultado de cada uno de estos ensayos se denotará por $h(x_i)$ siendo h_i la variable aleatoria asociada. Como todas las h_i son idénticas e independientes:

$$\langle h_1 \rangle = \langle h_2 \rangle = \ldots = \langle h_N \rangle = \langle h \rangle = \frac{1}{(b-a)} \int_a^b h(x)\,dx \qquad [85]$$

$$\sigma^2(h_1) = \sigma^2(h_2) = \ldots = \sigma^2(h_N) = \sigma^2(h) \qquad [86]$$

Definimos la variable aleatoria H_N como:

$$H_N = \sum_{i=1}^{N} h_i \qquad [87]$$

por lo que cuando N sea suficientemente grande el *teorema de Laplace* garantizará un comportamiento gaussiano para H_N con parámetros:

$$\langle H_N \rangle = N\langle h \rangle \quad ; \quad \sigma^2(H_N) = N\sigma^2(h) \qquad [88]$$

Sabemos por estadística elemental que la probabilidad de que H_N esté entre los límites $\langle H_N \rangle \pm 3\sigma(H_N)$ es del 99.74 %, por lo que poniendo $\sigma(H_N) = \sigma(h)N^{1/2}$ llegamos a:

$$P\{(N\langle h\rangle - 3\sqrt{N}\sigma(h)) < H_N < (N\langle h\rangle + 3\sqrt{N}\sigma(h))\} =$$

$$= P\left\{\left|\frac{H_N}{N} - \langle h\rangle\right| < 3\sigma(h)/\sqrt{N}\right\} =$$

$$= P\left\{\left|\sum_{i=1}^{N} h(x_i)/N - \langle h\rangle\right| < 3\sigma(h)/\sqrt{N}\right\} = 0{,}9974 \qquad [89]$$

Esta última ecuación indica que la proximidad entre el valor real y el estimado

$$\langle h\rangle = \frac{1}{b-a}\int_a^b h(x)\,dx \simeq \frac{1}{N}\sum_{i=1}^{N} h(x_i)$$

crece con la raíz cuadrada del número N. Esto implica que multiplicar por diez la precisión, es decir, obtener un nuevo dígito significativo, requerirá realizar 100 veces más ensayos ($100N$). La relación apuntada es *general* para todos los métodos Monte Carlo, siendo su principal limitación, pues convierte a esta técnica en una muy cara.

La generalización para varias dimensiones es inmediata:

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} h(x_1, x_2, ..., x_n)\,dx_1\cdot dx_2\cdot ...\cdot dx_n \simeq$$

$$\simeq \frac{\prod_{i=1}^{n}(b_i - a_i)}{N^n}\sum_{i=1}^{N}\cdots\sum_{k=1}^{N} h(x_{1i}, ..., x_{nk}) \qquad [90]$$

en donde hemos tomado, entre otras posibilidades, un número N igual de puntos sobre las variables x_n. Los puntos $x_{1i}, x_{2j}, ...$, se generan como números aleatorios uniformemente distribuidos en (a_1, b_1), (a_2, b_2), ..., y así sucesivamente.

En el caso de que la región a integrar fuese complicada, el artificio a seguir consiste en encerrar tal región $\mathfrak{R}$ dentro de un hipercubo y sortear N_T puntos aleatoriamente dentro de él de manera análoga a lo ya visto. Una fracción de tales puntos, $N_{\mathfrak{R}}$, caerán dentro de $\mathfrak{R}$ y contribuirán efectivamente a la integral, en tanto que los externos a $\mathfrak{R}$, $(N_T - N_{\mathfrak{R}})$, no son útiles y se desechan. Claramente, cuanto más ajustado esté el hipercubo a $\mathfrak{R}$ tanto más eficiente será el muestreo, pues el número de puntos rechazados será menor. La evaluación de la integral se lleva a cabo entonces de la forma siguiente:

$$\int_{\mathfrak{R}} \cdots \int h(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n \simeq \frac{\vartheta_{\mathfrak{R}}}{N_{\mathfrak{R}}} \sum_{i=1}^{N_{\mathfrak{R}}} h(x_1, x_2, \ldots, x_n)_i \qquad [91]$$

en donde $\vartheta_{\mathfrak{R}}$ es el volumen de la región $\mathfrak{R}$. Este dato puede siempre estimarse utilizando los resultados del muestreo:

$$\vartheta_{\mathfrak{R}} \simeq \frac{N_{\mathfrak{R}}}{N_T} \text{ x Volumen del hipercubo} \qquad [92]$$

Estimaciones del error cometido se pueden calcular con la fórmula:

$$|E_{\mathrm{MC}}| = \vartheta_{\mathfrak{R}} \sqrt{\frac{\langle h^2 \rangle - \langle h \rangle^2}{N_{\mathfrak{R}}}} \qquad [93]$$

Siendo:

$$\langle h \rangle = \frac{1}{N_{\mathfrak{R}}} \sum_{i=1}^{N_{\mathfrak{R}}} h(x_1, x_2, \ldots, x_n)_i \quad ; \quad \langle h^2 \rangle = \frac{1}{N_{\mathfrak{R}}} \sum_{i=1}^{N_{\mathfrak{R}}} h^2\,(x_1, x_2, \ldots, x_n)_i \qquad [94]$$

En el Tema 2 vimos que existen algoritmos de integración aproximada (Gauss-Legendre, Simpson, etc.) que pueden suministrar muy buenas soluciones para el cálculo numérico de integrales definidas en una dimensión. ¿Hasta qué punto resulta más adecuado utilizar estas técnicas de integración convencional, adaptadas a mayores dimensionalidades, que la de Monte *Carlo y viceversa?*

La respuesta a esta problemática viene dada por la velocidad de cálculo en cada una de ellas. Es importante señalar que así como el caso monodimensional ha sido extensamente estudiado, para dimensiones

mayores la situación no es tan clara (Stroud, 1974). Cuando los dominios de integración son rectangulares la extensión de fórmulas de cuadratura a D dimensiones es inmediata. Por ejemplo, para $D = 2$ la regla del trapecio con dos particiones (N puntos) iguales en x e y sería:

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} g(x, y)\,dx\,dy = \int_{a_1}^{b_1} dx \cdot \frac{h_2}{2}\,[g(x, y_1) + 2g(x, y_2) + \ldots + g(x, y_N)] =$$

$$= \frac{h_1 h_2}{4}\,[g(x_1, y_1) + 2g(x_2, y_1) + \ldots + g(x_N, y_1) +$$

$$+ 2[g(x_1, y_2) + 2g(x_2, y_2) + \ldots + g(x_N, y_2)] + \ldots]$$

Este tipo de reglas de composición, llamadas *reglas-producto*, preservan las propiedades de las reglas monodimensionales, pero el número de puntos crece con D (en el caso anterior N^2) en forma potencial. En una dimensión con N puntos, si se necesitan N evaluaciones funcionales, en dos dimensiones serán necesarias N^2 evaluaciones, en tres dimensiones N^3, etc. En general en D dimensiones se tendrán N^D evaluaciones funcionales. Como consecuencia, la *convergencia efectiva* de estas reglas producto es peor a medida que D crece. Bajo esta perspectiva, se introduce un factor $1/D$ en el exponente de la convergencia efectiva de los algoritmos convencionales. Para la regla del trapecio la convergencia efectiva depende del número de puntos N en la forma $N^{-2/D}$; en el caso de la regla de Simpson se encuentra $N^{-4/D}$.

Por el contrario, la convergencia del método de Monte Carlo ($N^{-1/2}$) *es independiente de la dimensionalidad* D del problema, por lo que cabe esperar que a partir de un valor D dado este método convergerá mucho más rápidamente que las reglas-producto. Por otra parte, tales reglas pueden resultar imposibles de aplicar (regla del trapecio para $N = 20$ y $D = 30$, por ejemplo). En la referencia (James, 1980) se da una gráfica comparativa de la convergencia del método de Monte Carlo frente a la de otros algoritmos. En un diagrama $x = N$, $y = D$, se ve que Monte Carlo es más rápido dentro de la región sombreada, comprendida entre $N = 3$ y $D = 3{,}4N - 0{,}2$, de la figura 2.

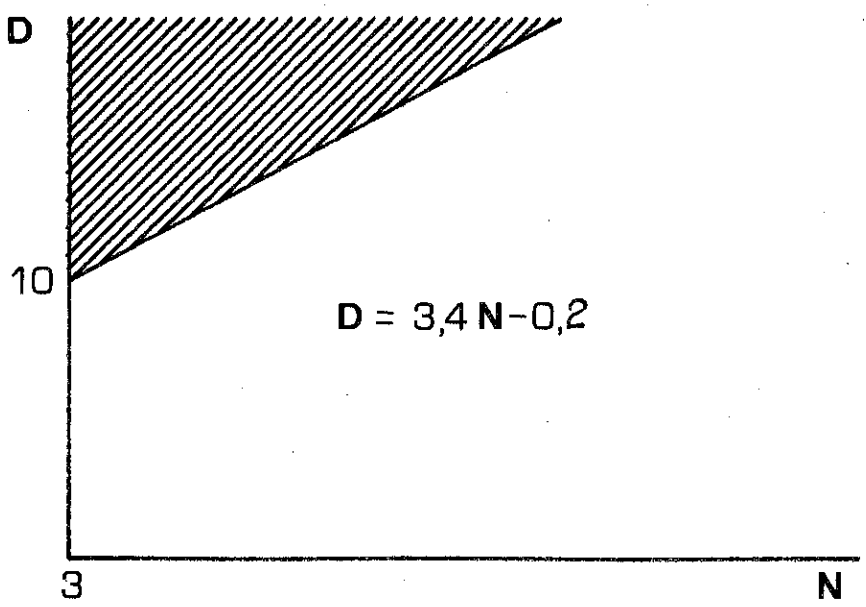


Figura 2.—Zona (rayada) de eficiencia de la integración de Monte Carlo.

El cálculo de integrales vía Monte Carlo puede ser optimizado con el uso de números aleatorios no uniformemente distribuidos, sino adaptados a la función a integrar (Sóbol, 1976). Es fácil comprender que este proceso mejorado puede ser altamente complejo por lo que salvo en ocasiones sencillas no suele utilizarse (Ejercicio 7).

Bibliografía

1. ABRAMOWITZ, M., y STEGUN, I. A., *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, Wiley, Nueva York*, 1972.
2. CHANDRASEKHAR, S., *Rev. Mod. Phys.*, 15, 1, 1943.
3. CRAMÉR, H., *Elementos de la Teoría de Probabilidades*, Aguilar, Madrid, 1977, capítulos 6 y 7.
4. HAMMERSLEY, J. M., y HANDSCOMB, D. C., *Monte Carlo Methods*, Chapman and Hall, Londres, 1983, capítulo 3.
5. JAMES, F., *Monte Carlo Theory and Practice*, CERN Data Handling Division, February, 1980.
6. KAMPEN, N. G. van, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 1985, capítulo 1.
7. MCQUARRIE, D. A., *Statistical Mechanics*, Harper & Row, Nueva York, 1976, capítulo 3.
8. PRESS, W. H.; FLANNERY, B. P.; TEUKOLSKY, S. A., y VETTERLING, W. T., *Numerical Recipes*, Cambridge University Press, Cambridge, 1988, capítulo 7.
9. REIF, F., *Física Estadística*, Reverté, Barcelona, 1975, capítulo 2.
10. ROZANOV, Y., *Processus Aléatoires*, Mir, Moscú, 1975, capítulo 2.
11. SCHWARTZ, L., *Métodos Matemáticos para las Ciencias Físicas*, Selecciones Científicas, Madrid, 1969, capítulo 5.
12. SESÉ, L. M., y CRIADO, M., *Termodinámica Química Molecular*, Universidad Nacional de Educación a Distancia, Madrid, 1990, capítulos 5 y 8.
13. SOBOL, I. M., *El Método de Monte Carlo*, Mir, Moscú, 1976.
14. SPIEGEL, M. R., *Manual de Fórmulas y Tablas Matemáticas*, McGraw-Hill, México, 1968.
15. SPIEGEL, M. R., *Estadística*, McGraw-Hill, México, 1970, capítulo 7.
16. STROUD, A. H., *Numerical Quadrature and Solution of Ordinary Differential Equations*, Springer-Verlag, Nueva York, 1974, capítulo 3.
17. WOOD, W. W., *J. Chem. Phys.*, 48, 415, 1968.
18. WOOD, W. W., en *Fundamental Problems in Statistical Mechanics*, vol. 3, «Computer Studies on fluid systems of hard-core particles», Editor E. D. G. Cohen, North-Holland, Amsterdam, 1975.

EJERCICIOS DE AUTOCOMPROBACION

1. Obtener las expresiones [4]-[5]-[6] para la media, la desviación típica y el coeficiente de asimetría de una distribución binómica.

2. Determinar la función generatriz de la suma $n = n_1 + n_2$ siendo n_1 y n_2 variables independientes distribuidas de acuerdo con la ley de Poisson con parámetros $\langle n_1 \rangle$ y $\langle n_2 \rangle$.

3. Dar una prueba de la normalización de la distribución P_n dada por [35], habida cuenta que $\langle n \rangle \to \infty$ y $n = 0, 1, 2, ..., \infty$.

4. La función densidad de probabilidad para un oscilador armónico cuántico monodimensional en su estado fundamental es:

$$f(x) = |\phi_0(x)|^2 = \left(\frac{\alpha^2}{\pi}\right)^{1/2} \exp(-\alpha^2 x^2) \quad ; \quad -\infty < x < +\infty$$

Evaluar la magnitud del *efecto túnel* como la probabilidad de que el oscilador se encuentre fuera de los puntos de retorno clásico que se defi-

nen por la igualdad entre la energía potencial $V = (kx^2)/2$ y la energía del estado $E_0 = (h\tilde{\nu}_e)/2$.

(Datos: $\alpha^2 = \left(\frac{m\,k}{\hbar^2}\right)^{1/2}$; $\tilde{\nu}_e = \frac{1}{2\pi}\left(\frac{k}{m}\right)^{1/2}$

m = masa del oscilador ; k = constante de fuerza ; $\tilde{\nu}_e$ = frecuencia clásica de vibración).

5. Obtener la expresión [42] a partir de la [41].

6. Obtener la ecuación [76] por integración de [75].

7. Normalizar mediante el método de Monte Carlo la función de onda de una partícula monodimensional que se encuentra en el estado fundamental de una caja de potencial con longitud $l = 1$:

$$\psi(x) = A \operatorname{sen} \pi x$$

Utilizar un flujo de números aleatorios en (0, 1).

a) uniformemente distribuidos

b) linealmente distribuidos

$$p(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

c) distribuidos según

$$p(x) = \begin{cases} 0 & x < 0 \\ 2x & 0 < x < 1/2 \\ -2x + 2 & 1/2 < x < 1 \\ 0 & 1 < x \end{cases}$$

(emplear como partida 10 números aleatorios de acuerdo con lo visto en el apartado 3.1 y con $M = 64$, $\lambda = 5$, $z_1 = 1$).

SOLUCIONES A LOS EJERCICIOS DE AUTOCOMPROBACION

1. Dado que

$$X_b = X_1 + X_2 + \ldots + X_N$$

se tiene

$$\langle X_b \rangle = \langle X_1 \rangle + \langle X_2 \rangle + \ldots + \langle X_N \rangle = N\langle X_1 \rangle$$

por ser todas las X_i idénticas entre sí. Sea X_1 tal que:

$X_1 = 0$ apuntar abajo, probabilidad q
$X_1 = 1$ apuntar arriba, probabilidad p

entonces

$$\langle X_1 \rangle = q \cdot 0 + p \cdot 1 = p$$

y de aquí:

$$\langle X_b \rangle = N \cdot p$$

Por la independencia sabemos que:

$$\sigma^2(X_b) = N\sigma^2(X_1)$$

y siendo

$$\sigma^2(X_1) = \langle X_1^2 \rangle - \langle X_1 \rangle^2 = p(1-p) = p \cdot q$$

resulta

$$\sigma(X_b) = \sqrt{Npq}$$

Finalmente se obtiene sin dificultad:

$$\gamma_1(X_b) = \frac{q-p}{\sqrt{Npq}}$$

2. La función generatriz pedida es el producto:

$$\psi(t) = \psi(t_1) \cdot \psi(t_2) = \exp\left[(\lambda_1 + \lambda_2)(e^t - 1)\right]$$

3. Debemos comprobar:

$$\sum_{n=0}^{\infty} P_n = \frac{1}{\sqrt{2\pi\langle n \rangle}} \cdot \sum_{n=0}^{\infty} \exp\left[-\frac{(n - \langle n \rangle)^2}{2\langle n \rangle} \right] = 1$$

Para simplificar la notación introduciremos la variable auxiliar z_n:

$$z_n = \frac{n - \langle n \rangle}{\sqrt{2\langle n \rangle}}$$

teniendo entonces:

$$\frac{1}{\sqrt{2\pi\langle n \rangle}} \cdot \sum_{n=0}^{\infty} \exp\left[-z_n^2\right] = \sum_{n=0}^{\infty} \frac{\exp\left[-z_n^2\right]}{\sqrt{2\pi\langle n \rangle}} = \sum_{n=0}^{\infty} f(n)$$

Como $\langle n \rangle$ es un número muy grande, es evidente que el cambio experimentado por $f(n)$ es muy pequeño cuando n cambia en una

unidad. La variable de sumación n puede considerarse entonces como una variable continua de modo que:

$$\sum_{n=0}^{\infty} f(n) = \int_0^{\infty} f(n)\,dn = \left\{ z = \frac{n - \langle n \rangle}{\sqrt{2\langle n \rangle}} \quad ; \quad dn = \sqrt{2\langle n \rangle}\,dz \right\} =$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left[-z^2\right] dz = 1$$

con lo que se comprueba la normalización. Nótese que el límite inferior de la integral anterior es un número negativo de gran valor absoluto, por lo que, a efectos prácticos, puede reemplazarse por $-\infty$ dadas las características de la función exponencial.

4. El problema se reduce a trabajar con una distribución gaussiana y determinar el área fuera de los puntos de retorno clásicos. Utilizando la condición dada para identificar estos puntos ($\tilde{\nu}_e$ en unidades de s^{-1}) se tiene:

$$E_0 = h\tilde{\nu}_e/2 = kx^2/2 = V \qquad x_{\text{ret. clás.}} = \pm \left(h\tilde{\nu}_e/k\right)^{1/2}$$

Debemos ahora tipificar la gaussiana del enunciado, para lo que hay que notar dos hechos:

- está centrada en el origen $\langle X \rangle = 0$
- su parámetro σ se relaciona con α del modo: $\sigma^2 = (2\alpha^2)^{-1}$

La variable tipificada Z resulta entonces:

$$Z = \frac{X - \langle X \rangle}{\sigma} = \alpha \sqrt{2} \cdot X$$

y empleando las relaciones del enunciado, los puntos de retorno clásico $Z_{\pm}$ son:

$$Z_{\pm} = \pm\, \alpha \sqrt{\frac{2\,h\,\tilde{\nu}_e}{k}} = \pm\, 2^{1/2}$$

Haciendo uso de las tablas de la distribución gaussiana se calcula fácilmente la probabilidad pedida:

$$P(\textit{efecto túnel}) = P(Z \leqslant -2^{1/2}) + P(Z \geqslant 2^{1/2}) \simeq 2\,.\,0{,}078695 = 0{,}15739$$

Es decir, la magnitud del efecto túnel resulta ser bastante apreciable $\simeq 16\,\%$ de encontrar la partícula fuera de los límites que cabría esperar desde un punto de vista clásico.

5. Introduzcamos el cambio de variable:

$$Z = \frac{X - \langle X \rangle}{\sigma}$$

con lo que la función generatriz [41] se convierte en:

$$\psi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-z^2/2\right] \exp\left[t\sigma z + t\langle X \rangle\right] dz$$

Completando cuadrados en z se tiene:

$$-\frac{z^2}{2} + t\sigma z + t\langle X \rangle = -\left(\frac{z}{\sqrt{2}} - \frac{\sigma t}{\sqrt{2}}\right)^2 + \frac{\sigma^2 t^2}{2} + t\langle X \rangle$$

y $\psi(t)$ pasa a:

$$\psi(t) = \exp\left[t\langle X \rangle + \frac{1}{2}\sigma^2 t^2\right] \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\left(\frac{z - \sigma t}{\sqrt{2}}\right)^2\right] dz =$$

$$= \exp\left[t\langle X \rangle + \frac{1}{2}\sigma^2 t^2\right]$$

6. La integración de [75] en este caso es particularmente simple debido a la forma del integrando.

Empecemos haciendo el cambio de variables (monótono):

$$\frac{1}{2}mv^2 = x \quad ; \quad v > 0 \quad , \quad x > 0$$

de donde obtenemos:

$$mv \cdot dv = dx \quad ; \quad v = + \sqrt{\frac{2x}{m}}$$

Con estos cambios $f(E)$ se escribe como:

$$f(E) = 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \delta\left(\frac{mv^2}{2} - E\right) \exp\left[-\frac{mv^2}{2kT}\right] \cdot v^2 \cdot dv =$$

$$= 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \delta(x - E) \cdot \exp\left[-x/kT\right] \cdot \left(\frac{2x}{m^3}\right)^{1/2} \cdot dx$$

y aplicando la propiedad elemental del operador δ se llega a:

$$f(E) = \frac{2}{\sqrt{\pi}} (kT)^{-3/2} \cdot E^{1/2} \cdot \exp\left[-E/kT\right]$$

7. *a*) La constante de normalización A viene dada por:

$$A = \left[\int_0^1 \operatorname{sen}^2 \pi x \, dx\right]^{-1/2} = \sqrt{2}$$

Evaluaremos la integral entre corchetes por el método de Monte Carlo con $N = 10$ y los valores x_i:

$$x_1 = 0{,}015625 \qquad x_6 = 0{,}828125$$
$$x_2 = 0{,}078125 \qquad x_7 = 0{,}140625$$
$$x_3 = 0{,}390625 \qquad x_8 = 0{,}703125$$
$$x_4 = 0{,}953125 \qquad x_9 = 0{,}515625$$
$$x_5 = 0{,}765625 \qquad x_{10} = 0{,}578125$$

Estos números se combinarán en la fórmula:

$$\int_0^1 \operatorname{sen}^2 \pi x \cdot dx \simeq \frac{1}{10} \sum_{i=1}^{10} \operatorname{sen}^2 \pi x_i = \frac{1}{10}(4{,}451274)$$

para estimar el valor de la integral. La constante A resulta ser:

$$A = 1{,}4988...$$

que no es muy mala aproximación si tenemos en cuenta el bajo número N y la pobre calidad de la secuencia empleada (período = 16).

b) Vamos a ensayar una variante de muestreo diferente. La integral a realizar puede reescribirse utilizando una función densidad $p(x)$ como:

$$\int_0^1 \operatorname{sen}^2 \pi x \, dx = \int_0^1 \left[\frac{\operatorname{sen}^2 \pi x}{p(x)}\right] \cdot p(x) \, dx$$

Evidentemente el valor medio de $Y(x) = \operatorname{sen}^2 \pi x / p(x)$, respecto de $p(x)$ es justamente el valor de la integral buscada:

$$\langle Y \rangle = \left\langle \frac{\operatorname{sen}^2 \pi x}{p(x)} \right\rangle_p = \int_0^1 \operatorname{sen}^2 \pi x \, dx$$

La elección óptima para $p(x)$ es aquélla en la que ésta sea proporcional a la función valor absoluto de la función a integrar $|\operatorname{sen}^2 \pi x| = \operatorname{sen}^2 \pi x$ (en el dominio de integración considerado) (Sóbol, 1976). Para no complicar el muestreo elijamos una $p(x)$ lineal en (0, 1):

$$p(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Repitiendo razonamientos ya utilizados encontramos:

$$\int_0^1 \operatorname{sen}^2 \pi x \, dx \simeq \frac{1}{N} \sum_{i=1}^{N} \frac{\operatorname{sen}^2 \pi x_i}{p(x_i)} = \frac{1}{20} \sum_{i=1}^{10} \frac{\operatorname{sen}^2 \pi x_i}{x_i}$$

y deberemos determinar puntos x_i distribuidos de acuerdo con $p(x)$. En este caso esto va a resultar sencillo empleando números uniformemente distribuidos ξ_i. La distribución cumulativa asociada a $p(x)$ es:

$$\xi = F(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 < x < 1 \\ 1 & 1 < x \end{cases}$$

cuya función inversa es fácil de calcular $(0 < F(x) < 1)$:

$$x = F^{-1}(\xi) = +\sqrt{\xi} \quad ; \quad 0 < x < 1$$

por lo que utilizando los ξ_i uniformemente distribuidos en (0, 1), su raíz cuadrada nos dará los x_i que nos interesan ahora. Tomando los ξ_i del apartado anterior obtenemos:

$$\begin{aligned} x_1 &= 0{,}125000 & x_6 &= 0{,}910014 \\ x_2 &= 0{,}279508 & x_7 &= 0{,}375000 \\ x_3 &= 0{,}625000 & x_8 &= 0{,}838525 \\ x_4 &= 0{,}976281 & x_9 &= 0{,}718070 \\ x_5 &= 0{,}875000 & x_{10} &= 0{,}760345 \end{aligned}$$

que llevan a un valor para la integral:

$$\int_0^1 \operatorname{sen}^2 \pi x \cdot dx \simeq \frac{1}{20} \sum_{i=1}^{10} \frac{\operatorname{sen}^2 \pi x_i}{x_i} = \frac{1}{20}(8{,}921988)$$

y para la constante A:

$$A = 1{,}4972...$$

No hemos obtenido pues un resultado significativamente mejor que en *a*).

c) Calculemos el resultado empleando una distribución más próxima a la función $\operatorname{sen}^2 \pi x$ en el sentido de proporcionalidad comentado. Convenientemente normalizada $p(x)$ es:

$$p(x) = \begin{cases} 0 & ; \quad x < 0 \\ 4x & ; \quad 0 < x < 1/2 \\ -4x + 4 & ; \quad 1/2 < x < 1 \\ 0 & ; \quad 1 < x \end{cases}$$

con función integral o cumulativa:

$$\xi = F(x) = \begin{cases} 0 & ; \quad x < 0 \\ 2x^2 & ; \quad 0 < x < 1/2 \Rightarrow 0 < \xi < 1/2 \\ -2x^2 + 4x - 1 & ; \quad 1/2 < x < 1 \Rightarrow 1/2 < \xi < 1 \\ 1 & ; \quad x > 1 \end{cases}$$

Recuérdese que al calcular la función $F(x)$ en $1/2 < x < 1$ hay que sumar el valor acumulado del fragmento $0 < x < 1/2$, que es $2x^2 = 2(1/2)^2 = 1/2$, al resultado de la integral correspondiente.

Aproximaremos la integral por:

$$\int_0^1 \operatorname{sen}^2 \pi x \cdot dx \simeq \frac{1}{10} \sum_{i=1}^{10} \frac{\operatorname{sen}^2 \pi x_i}{p(x_i)}$$

si bien los cálculos serán ahora más laboriosos. Los puntos x_i se determinarán de acuerdo con:

$$x_i = \begin{cases} \sqrt{\dfrac{\xi_i}{2}} & ; \quad 0 < \xi_i < 1/2 \quad ; \quad 0 < x_i < 1/2 \\ 1 - \sqrt{1 - (1 + \xi_i)/2} & ; \quad 1/2 < \xi_i < 1 \quad ; \quad 1/2 < x_i < 1 \end{cases}$$

Tomando los valores ξ_i uniformemente distribuidos calculados en a), se tienen los x_i siguientes:

$$\begin{aligned} x_1 &= 0{,}088388 & x_6 &= 0{,}706849 \\ x_2 &= 0{,}197642 & x_7 &= 0{,}265165 \\ x_3 &= 0{,}441942 & x_8 &= 0{,}614724 \\ x_4 &= 0{,}846907 & x_9 &= 0{,}507874 \\ x_5 &= 0{,}657673 & x_{10} &= 0{,}540721 \end{aligned}$$

Conviene reescribir la aproximación a la integral como:

$$\int_0^1 \operatorname{sen}^2 \pi x \cdot dx \simeq \frac{1}{10}\left[\sum_{0<x_i<1/2} \frac{\operatorname{sen}^2 \pi x_i}{4x_i} + \sum_{1/2<x_i<1} \frac{\operatorname{sen}^2 \pi x_i}{4 - 4x_i}\right]$$

en donde hemos separado las contribuciones de las dos regiones $0 < x < 1/2$ y $1/2 < x < 1$. Realizando estas operaciones llegamos a:

$$\int_0^1 \operatorname{sen}^2 \pi x \cdot dx \simeq \frac{1}{10}[1{,}703995 + 3{,}066732]$$

resultando el valor de la constante de normalización A bastante mejor que en los casos anteriores:

$$A \simeq 1{,}44779...$$

mostrando un error la mitad del de aquéllos.

Tema 8

DISTRIBUCIONES DE MAGNITUDES ALEATORIAS EN DOS DIMENSIONES Y CORRELACION

Abordaremos ahora la generalización de los conceptos monodimensionales previos a dos dimensiones.

Es fundamental en el contexto de varias variables la noción de *correlación*. En dos dimensiones, a partir de los momentos generalizados, puede definirse una cantidad que mida la correlación entre las dos variables aleatorias. Nos referimos al *coeficiente de correlación*, que el lector asociará con el método de los mínimos cuadrados. Desde una perspectiva estadística, este método está englobado dentro del estudio de la *regresión*, a la que dedicaremos una atención especial.

Como ejemplo de distribución continua en dos dimensiones nos limitaremos a presentar, en razón de su importancia, a la distribución gaussiana. Esta distribución nos servirá como modelo para presentar un caso interesante de regresión lineal entre variables aleatorias. Entre las múltiples aplicaciones que encuentra la función gaussiana, nos detendremos en el papel que juega en la teoría de *fluctuaciones termodinámicas* y en los *cálculos moleculares* de química cuántica (aquí ya en tres dimensiones).

La generalización a más dimensiones es directa y no nos detendremos en ella. Resulta más provechoso, para concluir, analizar brevemente la noción de *función de correlación* en varias variables ($\geqslant 2$), herramienta matemática fundamental en el estudio por mecánica estadística de sistemas macroscópicos.

1. MOMENTOS Y COEFICIENTES DE CORRELACION

Dada la variable aleatoria $Z = f(X, Y)$, se llama *momento central de orden i respecto X y orden j respecto Y* a la cantidad:

$$\mu_{ij} = \langle (X - \langle X \rangle)^i \cdot (Y - \langle Y \rangle)^j \rangle \qquad [1]$$

que especializada al caso discreto es:

$$\mu_{ij} = \sum_n \sum_m P_{nm}(x_n - \langle X \rangle)^i \cdot (y_m - \langle Y \rangle)^j \qquad [2]$$

y en el continuo:

$$\mu_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle X \rangle)^i \cdot (y - \langle Y \rangle)^j \cdot f(x, y) \cdot dx\, dy \qquad [3]$$

No es complicado obtener a partir de aquí:

$$\begin{aligned} \mu_{01} &= \mu_{10} = 0 \\ \mu_{20} &= \sigma^2(X) \\ \mu_{02} &= \sigma^2(Y) \\ \mu_{11} &= \alpha_{11} - \langle X \rangle \langle Y \rangle \end{aligned} \qquad [4]$$

en donde α_{11} es el *momento respecto al origen*:

$$\alpha_{11} = \langle X^i \cdot Y^j \rangle_{i=j=1} = \langle X \cdot Y \rangle \qquad [5]$$

Estudiemos ahora el valor medio de la variable no negativa:

$$Z = [\mu_{02}(X - \langle X \rangle) - \mu_{11}(Y - \langle Y \rangle)]^2 \qquad [6]$$

que operando (Ejercicio 1) resulta:

$$\langle Z \rangle = \mu_{20} \cdot \mu_{02}^2 - \mu_{02} \cdot \mu_{11}^2 \geqslant 0 \qquad [7]$$

de donde se obtiene:

$$\mu_{11}^2 \leqslant \mu_{20} \cdot \mu_{02} \qquad [8]$$

De esta desigualdad puede definirse el denominado *coeficiente de correlación* r para las variables X e Y:

$$r = \frac{\mu_{11}}{\sigma(X) \cdot \sigma(Y)} \qquad [9]$$

que claramente verifica la condición:

$$-1 \leqslant r \leqslant +1 \qquad [10]$$

Se supone que en [9] $\sigma(X), \sigma(Y) \neq 0$. La cantidad μ_{11} se denomina *covarianza* de X e Y.

El coeficiente de correlación r toma los valores extremos ± 1, si y sólo si la relación funcional entre X e Y es *lineal*:

$$Y = aX + b \qquad [11]$$

es decir, cuando la masa total de la distribución está concentrada en una recta del plano XY (Cramér, 1977). Esta es una situación ciertamente muy particular. En general $r \neq \pm 1$ y la distribución se repartirá en el plano de acuerdo a una ley no lineal. No obstante, cuanto más próximo sea r a esos valores límites, cabe pensarse que la distribución se situará más y más en las cercanías de alguna recta. Es r, entonces, una medida de la dependencia lineal entre X e Y.

Cuando X e Y son independientes es trivial verificar que su covarianza $\mu_{11} = \text{cov}(X, Y) = 0$ y por lo tanto $r = 0$. *Esto expresa que dos variables independientes nunca están correlacionadas. Por el contrario, dos variables no correlacionadas podrían ser dependientes una de otra. Es muy importante, pues, no confundir los conceptos de correlación y de dependencia.*

2. REGRESION DE MINIMOS CUADRADOS

Teniendo en cuenta los resultados anteriores interesa ahora determinar qué recta del plano aproxima o representa mejor a una distribución bidimensional, bien sea ésta una nube discreta de puntos bien una continua (línea o superficie). Las bases para resolver este problema se dieron en el Tema 4, y seguiremos como allí *el criterio de mínimos*

cuadrados para calcular la mejor línea recta. Este principio conduce a planteamientos tremendamente operativos y de ahí, en parte, su popularidad. Deberemos pues encontrar las rectas:

$$Y = aX + b \quad ; \quad X = a'Y + b' \qquad [12]$$

que hagan mínimo el valor medio del cuadrado de la distancia entre los puntos Y (o X) estimados vía [12] y los iniciales Y (o X) de la distribución. La existencia de dos líneas de ajuste o regresión obedece a que tanto X como Y pueden tomarse como variables independientes a la hora de efectuar el ajuste.

Fijémonos en la línea de regresión de Y sobre X que es la primera de [12]. Todo lo que digamos para ella será directamente traducible a su compañera de X sobre Y. La primera ecuación debe ser tal que haga:

$$M = \langle (Y - aX - b)^2 \rangle = \text{mínimo} \qquad [13]$$

Es rápido calcular los parámetros a y b:

$$a = r \cdot \frac{\sigma(Y)}{\sigma(X)} \quad ; \quad b = \langle Y \rangle - a\langle X \rangle \qquad [14]$$

resultando la ecuación lineal:

$$\frac{Y - \langle Y \rangle}{\sigma(Y)} = r \cdot \frac{X - \langle X \rangle}{\sigma(X)} \qquad [15]$$

cuyo valor mínimo M es:

$$M = \sigma^2(Y)(1 - r^2) \qquad [16]$$

Análogamente la regresión de X sobre Y es:

$$\frac{Y - \langle Y \rangle}{\sigma(Y)} = \frac{1}{r} \cdot \frac{X - \langle X \rangle}{\sigma(X)} \qquad [17]$$

con un valor M asociado:

$$M = \langle (X - a'Y - b')^2 \rangle = \sigma^2(X) \cdot (1 - r^2) \qquad [18]$$

De las ecuaciones [15]-[18] observamos que:

a) las dos rectas de regresión pasan por el centro de gravedad de la distribución;

b) a medida que $r \to \pm 1$ la distribución es tanto más cercana a una recta, siendo r una buena medida de la dependencia lineal entre X e Y; cuando $|r| = 1$ las dos rectas [12] coinciden.

c) la pendiente de las rectas tiene el signo de r.

Evidentemente, podríamos haber planteado un problema más general, buscando la función no lineal que representara a la distribución con el criterio de mínimos cuadrados (Ejercicio 2). Sin embargo la recta es, con mucho, la funcionalidad más útil. Aparte de su sencillez, en bastantes casos, determinados cambios de variable reducen a forma lineal funcionalidades de diverso tipo como vamos a comprobar en el próximo epígrafe.

3. CAMBIOS DE VARIABLE UTILES Y PROBLEMATICA DEL AJUSTE

Por simplicidad, vamos a ocuparnos aquí de encontrar una relación funcional lineal para representar conjuntos de puntos discretos y equiprobables (x_k, y_k) en el plano xy. Conviene antes de lanzarse al ajuste directo:

$$\{(x_k, y_k)\} \to Y = aX + b \qquad [19]$$

investigar varias posibilidades. De no hacerlo así es muy probable que la representación [19] sea francamente mala. Esto puede ponerse de manifiesto utilizando simplemente un método gráfico. Se vería en los casos desfavorables que los puntos en cuestión quedan muy apartados de la recta calculada, resultando la suma de los cuadrados de las desviaciones $(y_k - ax_k - b)$ sospechosamente elevada. El cálculo del coeficiente de correlación r arrojaría, en consecuencia, un valor alejado de los extremos ± 1.

Esta situación es claramente un indicio de que la mejor relación entre X e Y no es lineal sino de otro tipo (exponencial, parabólico, etc.). La

relación $f(X, Y) = 0$ suele en muchos casos ser reducible a una forma lineal involucrando a dos nuevas variables $\tilde{X}$ e $\tilde{Y}$ definidas como:

$$\begin{aligned} \tilde{X} &= f_1(X, Y) \\ \tilde{Y} &= f_2(X, Y) \end{aligned} \qquad [20]$$

de modo que los puntos transformados $\{(\tilde{x}_k, \tilde{y}_k)\} = \{(f_1(x_k, y_k), f_2(x_k, y_k))\}$ se ajusten a una recta en el plano $\tilde{x}\tilde{y}$:

$$\tilde{Y} = \tilde{a}\tilde{X} + \tilde{b} \qquad [21]$$

El requerimiento matemático para realizar esta transformación es que [20] sea *biunívoca*, es decir que a cada punto xy le corresponda un punto $\tilde{x}\tilde{y}$ y sólo uno, y viceversa.

La determinación de [20] es un asunto complicado dada la gran variedad de relaciones funcionales $f(X, Y)$ que pueden postularse *a priori*. Sin embargo, en las situaciones de interés en Química Física hay tres relaciones fundamentales:

$$Y = aX + b \quad \text{(lineal)} \qquad [22]$$

$$Y = aX^m \quad \text{(doble logarítmica)} \qquad [23]$$

$$Y = am^X \quad \text{(semilogarítmica)} \qquad [24]$$

Es inmediato establecer que para [23]-[24] los cambios de variable son sencillamente:

$$\left.\begin{aligned} \tilde{Y} &= \log Y \\ \tilde{X} &= \log X \end{aligned}\right\} \rightarrow \log Y = \log a + m \cdot \log X \quad ; \quad a > 0 \qquad [25]$$

$$\left.\begin{aligned} \tilde{Y} &= \log Y \\ \tilde{X} &= X \end{aligned}\right\} \rightarrow \log Y = \log a + X \cdot \log m \quad ; \quad a, m > 0 \qquad [26]$$

Para decidir cuál de entre las tres formas es la mejor en nuestro problema concreto podemos primeramente utilizar el método gráfico y decidir con él la forma de la mejor aproximación. Hecho esto y emplean-

do las ecuaciones deducidas en el Tema 4 calcularemos los parámetros a y b de [21] con N puntos:

$$\tilde{a} = \frac{N(\sum_k \tilde{x}_k \tilde{y}_k) - (\sum_k \tilde{x}_k) \cdot (\sum_k \tilde{y}_k)}{N(\sum_k \tilde{x}_k^2) - (\sum_k \tilde{x}_k)^2} \qquad [27]$$

$$\tilde{b} = \frac{(\sum_k \tilde{y}_k) \cdot (\sum_k \tilde{x}_k^2) - (\sum_k \tilde{x}_k) \cdot (\sum_k \tilde{x}_k \cdot \tilde{y}_k)}{N(\sum_k \tilde{x}_k^2) - (\sum_k \tilde{x}_k)^2} \qquad [28]$$

Los errores en estos parámetros vienen dados por (Spiridonov y Lopatkin, 1973):

$$\sigma\,(\tilde{a}) = \sigma \cdot \sqrt{\frac{N}{N \sum_k \tilde{x}_k^2 - (\sum_k \tilde{x}_k)^2}} \qquad [29]$$

$$\sigma\,(\tilde{b}) = \sigma \cdot \sqrt{\frac{\sum_k \tilde{x}_k^2}{N \sum_k \tilde{x}_k^2 - (\sum_k \tilde{x}_k)^2}} \qquad [30]$$

siendo

$$\sigma = \sqrt{\frac{\sum_k (\tilde{y}_k - \tilde{Y}_k)^2}{N - 2}} \qquad [31]$$

Del mismo modo se calcula la recta de regresión de $\tilde{X}$ sobre $\tilde{Y}$. No hay que olvidar que la ecuación buscada es no lineal por lo que los resultados [27]-[28] deben reconvertirse a los valores significativos del problema (a, m) (Ejercicio 3).

Naturalmente si se dispone, como es lo normal, de facilidades de cálculo puede resultar más conveniente realizar los ajustes de mínimos cuadrados sin pasar por la representación gráfica. En este caso, siempre más fiable, el *coeficiente de correlación lineal*:

$$r = \frac{N(\sum_k \tilde{x}_k \cdot \tilde{y}_k) - (\sum_k \tilde{x}_k) \cdot (\sum_k \tilde{y}_k)}{\sqrt{[N \sum_k \tilde{x}_k^2 - (\sum_k \tilde{x}_k)^2] \cdot [N \sum_k \tilde{y}_k^2 - (\sum_k \tilde{y}_k)^2]}} \qquad [32]$$

será el criterio para decidir la mejor relación (nótese la simetría entre x e y).

Si el coeficiente de correlación lineal r así calculado es en valor absoluto igual a la unidad, nuestras pesquisas habrán concluido con

éxito. No obstante, ésta será una circunstancia excepcional pues en estos casos favorables, normalmente, los valores x_k, y_k no estarán dados con precisión infinita. El valor r será entonces distinto de ± 1, aunque pudiera ser cercano a ellos. En estas condiciones, ¿cómo podemos asegurar que la representación elegida es suficientemente buena para el conjunto de puntos (x_k, y_k)? ¿No habrá alguna otra forma funcional que lo haga todavía mejor? Por otra parte, si el mejor r obtenido en estos ensayos previos está alejado de ± 1, las dudas anteriores cobran aún un valor mucho más crítico.

La respuesta a las preguntas planteadas no es fácil. En líneas generales cabe decir que, salvo conocimiento de la relación $Y = f(X)$ exacta por argumentos ajenos a la Estadística, la relación que obtengamos entre los valores «experimentales» x_k, y_k, será buena en tanto nos conduzca a resultados significativos para nuestro problema concreto. Si esto no es así, deberemos ensayar otras posibilidades. Es conveniente recordar que, en general, las relaciones funcionales obtenidas de esta manera tienen un valor meramente *empírico* y por tanto *interpolatorio*. A pesar de ello, hay que tener presente que muchas leyes físicas fueron establecidas a partir de este tipo de planteamientos.

A continuación especificamos algunas otras funcionalidades con los cambios de variable que las convierten en lineales:

$$a) \quad Y = \frac{a}{X} + b \quad ; \quad \left\{ \begin{matrix} \tilde{X} = X \\ \tilde{Y} = X \cdot Y \end{matrix} \right\} \Rightarrow \tilde{Y} = a + b\tilde{X} \qquad [33]$$

$$b) \quad Y = \frac{1}{aX + b} \quad ; \quad \left\{ \begin{matrix} \tilde{X} = X \\ \tilde{Y} = 1/Y \end{matrix} \right\} \Rightarrow \tilde{Y} = a\tilde{X} + b \qquad [34]$$

$$c) \quad Y = aX^m + b \quad ; \quad \left\{ \begin{matrix} \tilde{X} = \log X \\ \tilde{Y} = \log (Y - b) \end{matrix} \right\} \Rightarrow$$

$$\Rightarrow \tilde{Y} = m\tilde{X} + \log a \quad ; \quad a > 0 \qquad [35]$$

$$d) \quad Y = aX^2 + bX + c \quad ; \quad \left\{ \begin{matrix} \tilde{X} = X - x_0 \\ \tilde{Y} = \dfrac{Y - y_0}{X - x_0} \end{matrix} \right\} \Rightarrow \tilde{Y} = a\tilde{X} + d \qquad [36]$$

en donde (x_0, y_0) es uno de los puntos (x_k, y_k) y el parámetro d viene dado por:

$$d = b + 2ax_0 \qquad [37]$$

Todos los cambios que hagan lineal una relación $Y = f(X)$ serán

considerados bajo el criterio de un mejor coeficiente de correlación lineal r a la hora de seleccionar el óptimo. Naturalmente, existen otras muchas formas funcionales no contempladas en este epígrafe y criterios de selección no basados en r (Demidowistch y col., 1980).

4. DISTRIBUCIONES CONDICIONALES Y REGRESION DE LA MEDIA

Puede presentarse la circunstancia de que la regresión lineal, dirigida a representar una distribución de probabilidades con funcionalidad arbitraria $Y = f(X)$, arroje unos resultados fuera de lo esperado, debido a que tal función no sea ni aproximadamente lineal. El coeficiente de correlación lineal r no sirve pues para indicar la proximidad de la distribución a una recta, sino justo para lo contrario. Sabremos entonces que no hay proximidad a una recta y desconoceremos sobre qué curva del plano xy se aglomera la distribución. En estos casos es útil emplear como referencia de aglomeración las denominadas *curvas de regresión de las medias*, y, como medida de proximidad, las llamadas *razones de correlación* de una variable sobre otra. Para abordar el estudio de las curvas de regresión de las medias es necesario introducir antes las *distribuciones condicionales*. Nos referiremos al caso discreto dejando para el lector la tarea de traducir lo que digamos al caso continuo.

Sea la distribución bidimensional discreta definida por:

$$P_{nm} = P(X = x_n \quad ; \quad Y = y_m) \quad ; \quad \begin{matrix} m = 0, 1, 2, \ldots \\ n = 0, 1, 2, \ldots \end{matrix} \qquad [38]$$

con distribuciones marginales asociadas:

$$P_n^x = \sum_m P_{nm} \quad ; \quad P_m^y = \sum_n P_{nm} \quad ; \quad \begin{matrix} n = 0, 1, 2, \ldots \\ m = 0, 1, 2, \ldots \end{matrix} \qquad [39]$$

La *probabilidad condicional* de que $Y = y_k$, habiéndose dado $X = x_j$, es:

$$P(Y = y_k \mid X = x_j) = \frac{P(X = x_j \ ; \ Y = y_k)}{P(X = x_j)} = \frac{P_{jk}}{P_j^x} \qquad [40]$$

La función de frecuencia [40] está normalizada para el recorrido de la variable Y:

$$\sum_k P(Y = y_k \mid X = x_j) = \frac{\sum_k P_{jk}}{P_j^x} = 1 \qquad [41]$$

y esto es cierto para todos los valores $X = x_j (j = 0, 1, 2, ...)$. Dado que [40] es no negativa, se tiene una serie de distribuciones condicionales de Y con respecto a los posibles valores $X = x_j$. Empleando [40] se calcula el valor medio condicional de Y cuando $X = x_j$:

$$\langle Y \mid X = x_j \rangle = \sum_k y_k P(Y = y_k \mid X = x_j) = \frac{\sum_k y_k P_{jk}}{P_j^x} \qquad [42]$$

Si ahora retomamos la analogía mecánica de la distribución de masas, el valor medio que acabamos de calcular es el centro de gravedad de la distribución de masa en la vertical $X = x_j$ (Fig. 1).

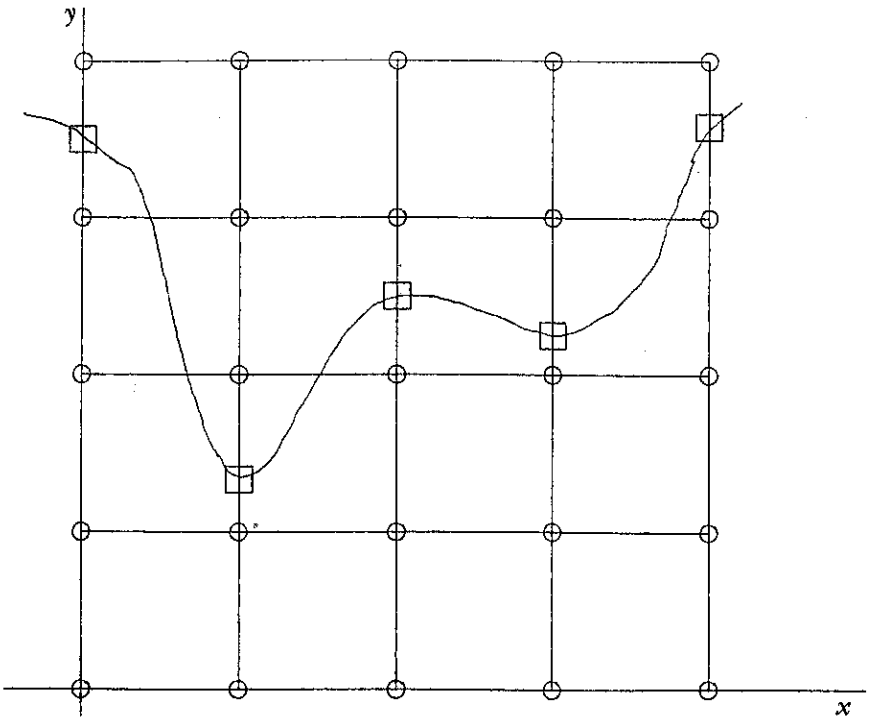


Figura 1.—Regresión de la media.

El proceso expresado en [42] puede repetirse para el resto de los valores x_n, calculándose todos los valores medios condicionales de Y con respecto a X. El lugar geométrico de los puntos $(x_j, \langle Y \mid X = x_j \rangle)$ se llama

curva de regresión de la media de Y. Invirtiendo los papeles de X e Y se determina la curva de regresión de la media de X, como el lugar geométrico de los puntos $(\langle X \mid Y = y_k \rangle, y_k)$.

Surge la duda en el caso discreto de seleccionar la curva de regresión de la media, pues ésta no es única. Parece razonable exigirle requisitos de sencillez. En este sentido podemos optar por trazos rectos de colocación a los valores medios condicionales anteriores. Desestimamos un posible ajuste por mínimos cuadrados, puesto que éste no necesariamente coloca los puntos de entrada. Para ver esto planteemos en nuestro caso discreto el valor medio del cuadrado de la distancia entre Y (o X) y una función de ajuste arbitraria $g(x)$ (o $g(y)$):

$$\langle (Y - g(X))^2 \rangle = \sum_n \sum_m (y_m - g(x_n))^2 P_{nm} \qquad [43]$$

Según el criterio de mínimos cuadrados $g(x)$ será aquella que haga mínima [43] que reescribiremos en la forma:

$$\langle (Y - g(X))^2 \rangle = \sum_n \sum_m (y_m - g(x_n))^2 P_n^x P(y_m \mid x_n) \qquad [44]$$

en donde hemos hecho uso de [40] simplificando la notación de la probabilidad condicional. La expresión [44] es:

$$\langle (Y - g(X))^2 \rangle = \sum_n P_n^x [\sum_m (y_m - g(x_n))^2 \cdot P(y_m \mid x_n)] \qquad [45]$$

en donde apreciamos que para cada n el término entre corchetes asociado es el momento de segundo orden de la distribución condicional de Y con respecto X y relativo al punto $g(x_n)$. En un tema anterior vimos que estos momentos son mínimos (Ejercicio 4 del Tema 6) cuando el punto arbitrario $a = g(x_n)$ coincide con el valor medio (condicional) de Y:

$$g(x_n) = \langle Y \mid X = x_n \rangle \qquad [46]$$

Es claro entonces que *la función de colocación* a $(x_n, \langle Y \mid X = x_n \rangle)$ *es la curva de regresión buscada.* Obviamente, si la curva de regresión de la media es una línea recta, ésta coincide con la línea de regresión de mínimos cuadrados [15].

Analizaremos ahora una medida de la aglomeración de la masa de la distribución sobre la curva de regresión $g(x)$. La *varianza* de Y se reduce a la expresión (Ejercicio 4):

$$\sigma^2(Y) = \langle (Y - \langle Y \rangle)^2 \rangle = \langle (Y - g(X))^2 \rangle + \langle (g(X) - \langle Y \rangle)^2 \rangle \quad [47]$$

Definiendo la cantidad no negativa:

$$R^2_{YX} = \frac{1}{\sigma^2(Y)} \cdot \langle (g(X) - \langle Y \rangle)^2 \rangle \quad [48]$$

resulta que:

$$R^2_{YX} = 1 - \frac{1}{\sigma^2(Y)} \cdot \langle (Y - g(X))^2 \rangle \quad [49]$$

de donde se concluye:

$$0 \leqslant R^2_{YX} \leqslant 1 \quad [50]$$

El coeficiente R^2_{YX} se llama *razón de correlación* de Y sobre X y su significado es fácil de interpretar observando sus valores extremos. Cuando $R^2_{YX} = 1$, la masa de la distribución está concentrada sobre $g(x)$. Cuando $R^2_{YX} = 0$, tenemos que esto se debe a que $g(x) = \langle Y \rangle$, siendo $g(x)$ una recta horizontal. Intercambiando los papeles de X e Y se pueden definir análogamente la curva de regresión de la media de X y la razón de correlación R^2_{XY}.

5. EJEMPLO DE APLICACION: LEY DE LAMBERT-BEER

Ayudándonos de la técnica de los mínimos cuadrados vamos a determinar la concentración de una sustancia disuelta utilizando espectroscopía visible-ultravioleta. En estos casos la muestra se ilumina con luz (monocromática) a una cierta longitud de onda λ y con una intensidad incidente I_o. La célula en la que está contenida la muestra posee un espesor l. El soluto absorbe parte de la radiación incidente en una cantidad que depende de la concentración c. En consecuencia, la radia-

ción emergente tras este proceso posee, naturalmente, una intensidad I_e menor que la inicial I_o. La magnitud $T = I_e/I_o$, denominada *transmisión*, posee una relación sencilla con las variables del problema c y l cuando la concentración de la muestra no es muy elevada. Obtengamos primeramente esta ley a partir del ejemplo siguiente.

Para una disolución de Br_2 en CCl_4 se ha obtenido la tabla de transmisiones en función de la concentración de bromo:

$T = I_e/I_o$	0,640	0,570	0,510	0,455	0,405
$c \times 10^3$ (mol/l)	1,00	1,25	1,50	1,75	2,00

con una $\lambda = 4.360$ A y $l = 1$ cm, no absorbiendo a esa λ el tetracloruro de carbono.

El primer paso será fijar la forma aproximada de la relación funcional $T = T(c)$. Es un sencillo ejercicio de representaciones gráficas establecer que la relación es del tipo semilogarítmico, pues es la forma que mejor se ajusta a una recta:

$$\ln T = -\chi c l \qquad [51]$$

Fijada esta relación ajustaremos por mínimos cuadrados $\ln(T)$ frente a c, siendo la pendiente de esta recta $-\chi l$, de donde extraeremos el llamado *coeficiente de extinción neperiano* χ (concentración · longitud)$^{-1}$, que es constante a una λ fija.

La nueva tabla con la que trabajar es:

$\tilde{T} = \ln T$	−0,446	−0,562	−0,673	−0,787	−0,904
$c \times 10^3$	1,00	1,25	1,50	1,75	2,00

y ajustaremos una recta de la forma $\tilde{T} = \tilde{a}c + \tilde{b}$. Los valores de las constantes son:

$$\tilde{a} = -4{,}564 \times 10^2 \quad ; \quad \tilde{b} = 0{,}010 \quad ; \quad r = -0{,}9999$$

de donde:

$$\chi = -\tilde{a}/l \simeq 456(\text{cm} \cdot \text{mol/l})^{-1}$$

El ajuste es bastante bueno aunque la recta no pase por el origen ($\tilde{b} \neq 0$). Un análisis más fino requeriría estudiar los errores en los parámetros, en lo que no nos vamos a detener por ser algo común en prácticas de laboratorio. La ecuación que buscamos tiene la forma general (Lambert-Beer):

$$I_e = I_o \exp[-\chi c l] \qquad [52]$$

Las desviaciones que se suelen obtener respecto a este resultado teórico pueden tener que ver con la magnitud de la concentración o con que la fuente de iluminación no sea completamente monocromática. Esto último se consigue con luz láser.

Conocido este resultado podemos averiguar la concentración del soluto en disoluciones problema contenidas en células de espesor diferente y disponer así de un método de análisis no destructivo (χ = constante a λ constante). Por ejemplo, consideremos una célula de 2 cm de espesor con una disolución de Br_2 en CCl_4 que a la λ anterior absorbe el 60 % de la radiación incidente. La concentración de Br_2 en esa muestra es sencillamente:

$$c = -\frac{1}{\chi l} \cdot \ln \frac{I_e}{I_o} = -\frac{1}{2 \times 456} \ln \frac{0{,}4 I_o}{I_o} = 1{,}00 \times 10^{-3} \text{ mol/l}$$

6. DISTRIBUCION NORMAL EN DOS DIMENSIONES

Sean dos variables aleatorias independientes $\tilde{X}$ e $\tilde{Y}$ distribuidas normalmente con medias $\langle \tilde{X} \rangle = \langle \tilde{Y} \rangle = 0$, y varianzas $\sigma^2(\tilde{X}) = \sigma^2(\tilde{Y}) = 1$. Su función densidad es:

$$\tilde{f}(\tilde{x}, \tilde{y}) = \tilde{f}_1(\tilde{x}) \cdot \tilde{f}_2(\tilde{y}) = \frac{1}{2\pi} \cdot \exp\left[-\frac{1}{2}(\tilde{x}^2 + \tilde{y}^2) \right] \qquad [53]$$

que no es sino un caso particular de forma gaussiana.

Definimos a continuación dos variables X e Y:

$$\begin{aligned} X &= \langle X \rangle + \sigma(X)\tilde{X} \\ Y &= \langle Y \rangle + r\sigma(Y)\tilde{X} + (1 - r^2)^{1/2}\sigma(Y)\tilde{Y} \end{aligned} \qquad [54]$$

con valores medios $\langle X\rangle$, $\langle Y\rangle$, desviaciones típicas $\sigma(X)$, $\sigma(Y)$, y coeficiente de correlación de X e Y r. La comprobación de estos detalles se deja como ejercicio para el lector.

En virtud de lo que conocemos sobre la ley gaussiana, tanto X como Y se distribuyen de acuerdo a tal ley con sus respectivas medias y varianzas, siendo ambas no independientes. La función densidad gaussiana conjunta para ambas nuevas variables será pues la forma más general de esta ley de distribución en dos dimensiones. Despejando $\tilde{X}$ e $\tilde{Y}$ de [54] en función de X e Y la función densidad conjunta $f(x, y)$ se obtiene vía:

$$\tilde{f}_1(\tilde{x}) \cdot \tilde{f}_2(\tilde{y}) = f(x, y) \cdot \left| \frac{\partial(\tilde{x}, \tilde{y})}{\partial(x, y)} \right| \qquad [55]$$

en donde aparece el Jacobiano, en valor absoluto, de la transformación [54]. La función densidad general que se obtiene para dos variables es de la forma:

$$f(x, y) = \frac{\exp\left[-\frac{1}{2(1-r^2)} \left\{ \frac{(X-\langle X\rangle)^2}{(\sigma(X))^2} + \frac{(Y-\langle Y\rangle)^2}{(\sigma(Y))^2} - \frac{2r(X-\langle X\rangle)\cdot(Y-\langle Y\rangle)}{\sigma(X)\cdot\sigma(Y)} \right\} \right]}{2\pi\sigma(X)\sigma(Y)\sqrt{1-r^2}} \qquad [56]$$

Para concluir mencionemos tres propiedades características de la distribución normal:

a) Las curvas de regresión de las medias de Y y de X son líneas rectas (Ejercicio 5). La distribución normal en dos dimensiones nos da así un ejemplo típico de regresión lineal (continuo) en ambas variables.

b) Si dos variables X e Y poseen una función densidad conjunta normal, dos funciones lineales cualesquiera Z_1, Z_2, de ambas también la tendrán siempre y cuando estas últimas sean linealmente independientes:

$$\left.\begin{aligned} Z_1 &= a_1X + b_1Y + c_1 \\ Z_2 &= a_2X + b_2Y + c_2 \end{aligned}\right\} \quad ; \quad \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0 \qquad [57]$$

c) Dos variables con densidad conjunta gaussiana son independientes, si y sólo si no están correlacionadas, es decir cuando $r = 0$.

Por inspección de [56] es fácil comprobar este punto que en general no es cierto para cualquier par de variables.

7. EJEMPLOS DE APLICACION

7.1. Fluctuaciones termodinámicas

La distribución gaussiana tiene un papel preponderante en la teoría de las *fluctuaciones termodinámicas*. Por razonamientos de Mecánica Estadística (Landau y Lifshitz, 1969) se establece para un sistema aislado con variables $X_1, X_2, ..., X_n$ (que no sean magnitudes conservadas como la energía) que la función densidad conjunta para ellas es:

$$w(x_1, x_2, ..., x_n) = N \cdot \exp\left[S(x_1, x_2, ..., x_n)\right] \qquad [58]$$

a temperaturas no muy bajas y con variaciones en cada X_i no muy rápidas. N es la constante de normalización y $S(X_1, ..., X_n)$ la entropía del sistema total. La fluctuación en cada variable la mediremos a través de la desviación típica:

$$[\sigma(X_i)]^2 = [\tilde{\Delta} x_i]^2 = \langle X_i^2 \rangle - \langle X_i \rangle^2 \quad ; \quad i = 1, 2, ..., n \qquad [59]$$

y por comodidad tomaremos $\langle X_i \rangle = 0$ (un simple cambio de origen).

Se sabe que la entropía toma su valor máximo cuando $X_1 = \langle X_1 \rangle, ..., X_n = \langle X_n \rangle$, por lo que desarrollando en serie de Taylor S en torno a $(0, 0, ..., 0)$, se tiene, suponiendo pequeños valores para $X_i = x_i$, que:

$$S(x_1, x_2, ..., x_n) = S(0, 0, ..., 0) + \sum_{i=1}^{n} \left(\frac{\partial S}{\partial x_i}\right)_0 x_i + \frac{1}{2} \sum_i \sum_j \left(\frac{\partial^2 S}{\partial x_i \partial x_j}\right)_0 x_i x_j + ...$$

$$\simeq S(0, 0, ..., 0) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial^2 S}{\partial x_i \partial x_j}\right)_0 x_i x_j =$$

$$= S(0, 0, ..., 0) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \quad ; \quad a_{ij} = a_{ji} \qquad [60]$$

Esta expresión sustituida en [58] nos conduce a:

$$w(x_1, x_2, ..., x_n) = N \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \right] \qquad [61]$$

que es una densidad gaussiana multidimensional cuyo exponente es una forma cuadrática definida negativa.

Si en [61] hacemos $n = 2$ encontramos una expresión similar a [56]:

$$w(x_1, x_2) = N \exp\left[-\frac{1}{2} a_{11} x_1^2 - a_{12} x_1 x_2 - \frac{1}{2} a_{22} x_2^2 \right] \qquad [62]$$

Aunque hemos obtenido [61]-[62] con la hipótesis de pequeños desplazamientos en cada variable ($X_i - \langle X_i \rangle = x_i$), la gran rapidez con que la función exponencial decae cuando x_i crece en valor absoluto permite con muy buena aproximación normalizar $w(x_1, x_2, ..., x_n)$ asumiendo los rangos de variación $-\infty < x_i < +\infty$. Por ejemplo, con $n = 2$ obtenemos:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(x_1, x_2)\, dx_1\, dx_2 =$$

$$= N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1\, dx_2 \exp\left[-\frac{a_{11}}{2} x_1^2 - a_{12} x_1 x_2 - \frac{a_{22}}{2} x_2^2 \right] = 1 \qquad [63]$$

lo que lleva a (Ejercicio 6):

$$N = \frac{|a_{11} a_{22} - a_{12}^2|^{1/2}}{2\pi} \qquad [64]$$

Es ahora por fin posible identificar completamente [56] con [62] y determinar así el significado estadístico de los coeficientes a_{ij}.

Como consecuencia de su universalidad la distribución gaussiana ha sido extensamente estudiada. Existen pues un gran número de relaciones matemáticas útiles que facilitan el trabajar con ella. Restringiéndonos a $n = 2$ se tiene:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1\, dx_2 \cdot \exp\left[\sum_i t_i x_i - \frac{1}{2} \sum_i \sum_j a_{ij} x_i x_j \right] =$$

$$= \frac{2\pi}{|A|^{1/2}} \cdot \exp\left[-\frac{1}{2} \sum_i \sum_j (A^{-1})_{ij} t_i t_j \right] \quad ; \quad |A| = \det(A) \qquad [65]$$

pudiéndose calcular los momentos μ_{nm}:

$$\mu_{nm} = \langle x_1^n \cdot x_2^m \rangle \qquad [66]$$

derivando secuencialmente ambos miembros de [65] n veces con respecto a t_1 y m veces con respecto a t_2, para hacer finalmente $t_1 = t_2 = 0$. El lector reconocerá en [65] a la *función generatriz* de la distribución $w(x_1, x_2)$. Se supone la convergencia de la integral de arriba para al menos dos valores $t_1 \neq 0$ y $t_2 \neq 0$. En particular, la *covarianza* entre x_1 y x_2 será (nótese el lugar que ocupa la constante $N = |A|^{1/2}/2\pi$):

$$\mu_{11} = \langle x_1 \cdot x_2 \rangle =$$

$$= \left[\frac{\partial}{\partial t_2} \left(\frac{\partial}{\partial t_1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1\, dx_2 \cdot \frac{|A|^{1/2}}{2\pi} \cdot \right. \right.$$

$$\left. \left. \cdot \exp\left[\sum_i t_i x_i - \frac{1}{2} \sum_i \sum_j a_{ij} x_i x_j \right] \right) \right]_{t_1 = t_2 = 0} =$$

$$= \left[\frac{\partial}{\partial t_2} \left(\frac{\partial}{\partial t_1} \exp\left[-\frac{1}{2} \sum_i \sum_j (A^{-1})_{ij} t_i t_j \right] \right) \right]_{t_1 = t_2 = 0}$$

resultando el valor:

$$\mu_{11} = \langle x_1 \cdot x_2 \rangle = (A^{-1})_{12} = \frac{a_{12}}{a_{12}^2 - a_{11} a_{22}} \qquad [67]$$

Es sencillo deducir tanto de [62] como de [67] que si $a_{12} = 0$, las variables x_1 y x_2 son independientes. En el primer caso porque $w(x_1, x_2)$ se factoriza en el producto de dos distribuciones independientes. En el segundo debido a que $\mu_{11} = 0$, lo que implica $r = 0$ y, *siendo la distribución conjunta gaussiana*, se concluye forzosamente la independencia de x_1 y x_2.

Para ilustrar lo anterior calcularemos las fluctuaciones de la temperatura T y del volumen V para una porción pequeña de un sistema aislado. El número de partículas dentro de esta porción (subsistema) debe ser suficientemente grande como para que la descripción estadística tenga sentido.

La densidad de probabilidad de una fluctuación vendrá dada por:

$$w = N' \exp [\Delta S'/k] \quad ; \quad k = \text{constante de Boltzmann} \qquad [68]$$

en donde $\Delta S'$ es la variación en la entropía total del sistema debida a la fluctuación en el subsistema. Esta variación puede ponerse en términos del trabajo mínimo necesario para realizar *reversiblemente* el cambio en las magnitudes termodinámicas que acontece con la fluctuación. A presión P y temperatura T constantes (valores medios del sistema total) el trabajo buscado es la variación en la energía libre de Gibbs:

$$W_{\text{mín}} = \Delta E - T\Delta S + P\Delta V \qquad [69]$$

del subsistema que ha experimentado variaciones en su energía interna, su entropía y su volumen. Cuando las fluctuaciones son pequeñas se puede desarrollar en serie ΔE y se llega a (Landau y Lifshitz, 1969):

$$\Delta E = \left(\frac{\partial E}{\partial S}\right)_V \cdot \Delta S + \left(\frac{\partial E}{\partial V}\right)_S \cdot \Delta V + \\ + \frac{1}{2}\left[\left(\frac{\partial^2 E}{\partial S^2}\right)_V \cdot (\Delta S)^2 + 2\left(\frac{\partial^2 E}{\partial S \partial V}\right)\Delta S \cdot \Delta V + \left(\frac{\partial^2 E}{\partial V^2}\right)_S \cdot (\Delta V)^2\right] \qquad [70]$$

Con esto [68] se transforma en:

$$w = N' \exp [(\Delta P \cdot \Delta V - \Delta T \cdot \Delta S)/2kT] \qquad [71]$$

Fijando como variables independientes T y V se establece por simple cálculo:

$$\Delta S = \left(\frac{\partial S}{\partial T}\right)_V \cdot \Delta T + \left(\frac{\partial S}{\partial V}\right)_T \cdot \Delta V \qquad [72]$$

$$\Delta P = \left(\frac{\partial P}{\partial T}\right)_V \cdot \Delta T + \left(\frac{\partial P}{\partial V}\right)_T \cdot \Delta V \qquad [73]$$

quedando w como:

$$w = N' \exp\left[-\frac{1}{2kT}\left\{\left(\frac{\partial S}{\partial T}\right)_V (\Delta T)^2 - \left[\left(\frac{\partial P}{\partial T}\right)_V - \left(\frac{\partial S}{\partial V}\right)_T\right]\Delta T \cdot \Delta V - \right.\right.$$

$$\left.\left. - \left(\frac{\partial P}{\partial V}\right)_T \cdot (\Delta V)^2 \right\}\right] \qquad [74]$$

que es una gaussiana en las variables ΔT y ΔV (las fluctuaciones). Teniendo en cuenta que:

$$\left(\frac{\partial P}{\partial T}\right)_V = \left(\frac{\partial S}{\partial V}\right)_T \qquad [75]$$

la distribución w se simplifica a:

$$w(\Delta T, \Delta V) = N' \exp\left[-\frac{1}{2kT}\left\{\left(\frac{\partial S}{\partial T}\right)_V \cdot (\Delta T)^2 - \left(\frac{\partial P}{\partial V}\right)_T \cdot (\Delta V)^2 \right\}\right] \qquad [76]$$

Por simple inspección se identifica rápidamente:

$$\langle \Delta V \cdot \Delta T \rangle = 0$$

$$\langle (\Delta T)^2 \rangle = kT \Big/ \left(\frac{\partial S}{\partial T}\right)_V = \sigma^2(T)$$

$$\langle (\Delta V)^2 \rangle = -kT \Big/ \left(\frac{\partial P}{\partial V}\right)_T = \sigma^2(V)$$

Como era de esperar, dado que V y T son variables independientes $\langle V \cdot T \rangle = 0$, también sus fluctuaciones lo son.

7.2. La función gaussiana en Química Cuántica

Para concluir con la función gaussiana exploraremos someramente su utilidad dentro del dominio de la Química Cuántica. Cuando se desea

aquí resolver la ecuación electrónica de Schrödinger para determinar la función de onda ψ de una molécula:

$$H\psi = E\psi \qquad [77]$$

un problema fundamental es el del cálculo de integrales de atracción núcleo-electrón y repulsión electrón-electrón entre las *funciones de base* χ con las que se contruye ψ.

Por ejemplo, dentro de la aproximación RHF (*Restricted-Hartree-Fock*) se asigna a ψ la forma de un *determinante de Slater* que para moléculas con todos sus electrones apareados (capa cerrada) se escribe:

$$\psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = \begin{vmatrix} \phi_1(1) & \bar{\phi}_1(1) & \cdots & \phi_{N/2}(1) & \bar{\phi}_{N/2}(1) \\ \phi_1(2) & \bar{\phi}_1(2) & \cdots & \phi_{N/2}(2) & \bar{\phi}_{N/2}(2) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \phi_1(N) & \bar{\phi}_1(N) & \cdots & \phi_{N/2}(N) & \bar{\phi}_{N/2}(N) \end{vmatrix} \qquad [78]$$

en donde $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N$, representan las coordenadas de cada electrón, las funciones monoelectrónicas ϕ son orbitales moleculares de spin electrónico $+1/2$ y las funciones $\bar{\phi}$ son orbitales moleculares idénticos a los anteriores en la parte espacial pero con spin $-1/2$. Estos orbitales se desarrollan en serie de funciones de base atómicas χ_r:

$$\phi_i = \sum_r c_{ri}\chi_r = \sum_A \sum_{r \in A} c_{ri}\chi_r(A) \qquad [79]$$

en donde A son los átomos de la molécula en cuestión. Como se aprecia en [79], cada átomo A aporta una serie de funciones de base $\chi_r(A)$ para construir los orbitales moleculares. Estas funciones de base describen el comportamiento de los electrones que A aporta a la molécula antes de que ésta se forme; son por tanto orbitales atómicos centrados en su origen natural: el núcleo de A.

La solución de [77] se suele realizar convencionalmente a través del *método variacional* (Löwdin, 1955), ya que no es posible resolverla exactamente (excepción hecha del H_2^+). El proceso variacional mezcla las funciones de base para dar las combinaciones óptimas (mejores c_{ri} en [79]) que denominamos *orbitales moleculares autoconsistentes*. Tras alguna manipulación se obtienen las propiedades moleculares tales como la energía, el momento dipolar, etc. (Pople y Beveridge, 1970).

Para poder realizar todo este proceso es necesario evaluar integrales de tipo electrón μ-núcleo A:

$$-Z_A\langle\chi_r(\mu)\left|\frac{1}{r_{\mu A}}\right|\chi_s(\mu)\rangle = \int \chi_r^*(\mu)\cdot\frac{-Z_A}{r_{\mu A}}\cdot\chi_s(\mu)\,d\mu \qquad [80]$$

electrón μ-electrón ν:

$$\langle\chi_r(\mu)\chi_t(\nu)\left|\frac{1}{r_{\mu\nu}}\right|\chi_v(\nu)\chi_s(\mu)\rangle = \int \chi_r^*(\mu)\chi_t^*(\nu)\frac{1}{r_{\mu\nu}}\chi_v(\nu)\chi_s(\mu)\cdot d\mu\cdot d\nu \qquad [81]$$

e integrales de tipo cinético que involucran al operador ∇^2. En las expresiones anteriores Z_A es la carga nuclear del átomo A, r simboliza la distancia (módulo) entre las partículas a que se refiere, y las integraciones cubren todo el espacio accesible a los electrones μ y ν (que pueden pertenecer cada uno a cualquier átomo de la molécula).

La elección del conjunto de funciones de base $\{\chi_r\}$ para llevar a cabo estos cálculos plantea no pocos problemas (Carsky y Urban, 1980). Podemos optar por usar una base con sentido físico: el conjunto formado por los *orbitales de Slater* para cada átomo A. Este tipo de orbitales (STO) describen a un electrón del átomo A según la forma (sin incluir spin):

$$\varphi_A(r,\theta,\phi)_{n,l,m} = R_{nl}(r)\cdot Y_{lm}(\theta,\phi) \qquad [82]$$

donde n, l, m, son los habituales números cuánticos. La parte radial viene dada por:

$$R_{nl}(r) = (2\xi)^{n+1/2}\cdot[(2n)!]^{-1/2}\cdot r^{n-1}\cdot\exp[-\xi r] \qquad [83]$$

(ξ = exponente orbital) y $Y_{lm}(\theta,\phi)$ simboliza al *armónico esférico*:

$$Y_{lm}(\theta,\phi) = P_l^m(\theta)\cdot e^{-im\phi} \qquad [84]$$

con P_l^m como la función asociada de Legendre *lm*.

Trabajar con este conjunto de funciones de base STO plantea, a la hora de calcular las integrales anteriores, unos problemas formidables

por su dificultad. En este punto es donde, entre otras posibilidades, pueden emplearse funciones gaussianas para *representar de forma aproximada* a la parte radial de los orbitales [82]. Dicho con más detalle, se emplea una base secundaria gaussiana $\{G_i(r_A)\}$ para representar a la base inicial STO-φ que es la que interviene en el desarrollo [79]:

$$\varphi_A(r, \theta, \phi)_{n,l,m} = [\sum_i c_i^{n,l,m} \cdot G_i(r_A)] Y_{lm}(\theta, \phi) \qquad [85]$$

Dadas las ventajas matemáticas que reporta la función gaussiana es comprensible que se emplee aquí. Los coeficientes $c_i^{n,l,m}$ pueden determinarse por mínimos cuadrados u otros métodos (Shavitt, 1963). La expresión de $G_i(r_A)$ es la de una gaussiana centrada en el átomo A:

$$G_i(r_A) = \exp[-a_i\{(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2\}] \qquad [86]$$

que contiene a las coordenadas x, y, z del electrón considerado.

Introduciendo estos desarrollos en las integrales originales en STO [80]-[81], surgen integrales en la nueva base gaussiana. Ahora el cálculo es mucho más simple pues se hace uso de relaciones matemáticas bien conocidas para las funciones G. Por ejemplo, el producto de dos gaussianas centradas una en A y otra en $B \neq A$ es otra gaussiana centrada en un punto C intermedio a los primeros:

$$G_i(r_A) \cdot G_j(r_B) = NG_k(r_c) \qquad [87]$$

o con más detalle (Ejercicio 7):

$$\exp[-a_i r_A^2] \cdot \exp[-a_j r_B^2] = \exp\left[-\frac{a_i a_j}{a_i + a_j} \cdot r_{AB}^2\right] \cdot \exp[-(a_i + a_j) \cdot r_C^2] \qquad [88]$$

siendo el primer factor una constante y estando situado el punto C en:

$$x_C = \frac{a_i x_A + a_j x_B}{a_i + a_j} \quad ; \quad y_C = \frac{a_i y_A + a_j y_B}{a_i + a_j} \quad ; \quad z_C = \frac{a_i z_A + a_j z_B}{a_i + a_j} \qquad [89]$$

que no es más que el *centro de gravedad* de A y B con pesos a_i en A y a_j en B. Por aplicación reiterada de este resultado se tiene que en general el

producto de n gaussianas es una nueva gaussiana afectada de un factor constante y centrada en el «centro de gravedad» de los centros de las gaussianas originales.

No obstante sus ventajas, las bases gaussianas ofrecen ciertas dificultades conceptuales (ver referencias citadas) que no han sido suficientes para desbancarlas del lugar privilegiado que ocupan dentro de los cálculos químico-cuánticos actuales.

8. FUNCIONES DE CORRELACION EN VARIAS VARIABLES

La teoría de la correlación entre variables vista anteriormente puede generalizarse con facilidad al caso de n dimensiones (Cramér, 1977). Sin embargo, desde el punto de vista de las aplicaciones es muy conveniente introducir las llamadas *funciones de correlación* $g(x_1, x_2, ..., x_n)$ que suministran una medida simple de la correlación entre las variables $X_1, X_2, ..., X_n$.

Sabemos que si las variables $X_1, ..., X_n$, son independientes, la función densidad conjunta se factoriza en el producto de funciones particulares (marginales):

$$f(x_1, x_2, ..., x_n) = f_1(x_1) \cdot f_2(x_2) \ldots f_n(x_n) \qquad [90]$$

y no existe correlación entre estas variables. En el caso general las n variables pudieran ser dependientes y, por tanto, estar correlacionadas. Para poner esto de manifiesto se escribe la función densidad conjunta en la forma:

$$f(x_1, x_2, ..., x_n) = f_1(x_1) \cdot f_2(x_2) \ldots f_n(x_n) + g'_n(x_1, x_2, ..., x_n) \qquad [91]$$

siendo g'_n *la función de correlación* que nos mide la separación entre $f(x_1, ..., x_n)$ y el producto de las funciones marginales independientes.

Pero todavía g'_n puede ser poco explícita por lo que respecta a informarnos sobre las correlaciones. Conviene analizar todas las posibles situaciones que un conjunto de n variables puede ofrecer. Se recurre así a

dividir el conjunto de n variables en subconjuntos («*clusters*») de s elementos, asociándoles la representación siguiente (Kampen, 1985):

$$\begin{aligned} h_1(x_1) &= h_1(x_1) \\ h_2(x_1, x_2) &= f_1(x_1) \cdot f_2(x_2) + g_2(x_1, x_2) \\ h_3(x_1, x_2, x_3) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) + f_1(x_1) \cdot g_2(x_2, x_3) + \\ &\quad + f_2(x_2) \cdot g_2(x_1, x_3) + f_3(x_3) g_2(x_1, x_2) + g_3(x_1, x_2, x_3) \\ &\ldots\ldots\ldots\ldots \end{aligned} \qquad [92]$$

y así sucesivamente. Las funciones marginales de s variables se obtienen a través de integración:

$$h_s(x_1, x_2, ..., x_s) = \int \ldots \int dx_{s+1} \ldots dx_n \cdot f(x_1, x_2, ..., x_n) \qquad [93]$$

Veamos, por ejemplo, el significado del desarrollo de f_3. El primer término es la situación de variables independientes. Los términos segundo, tercero y cuarto, simbolizan la situaciones en que una variable es independiente de las otras dos que a su vez sí están correlacionadas. El último término da idea de la correlación conjunta entre las tres variables. Nótese que ocasionalmente puede ser $h_1 = f_1$ (independencia).

9. EJEMPLO DE APLICACION: FUNCIONES DE CORRELACION EN MECANICA ESTADISTICA

Una de las principales razones para introducir en los problemas físicos las funciones g_s es que si las variables son casi independientes se espera que estas g_s tiendan rápidamente a cero y sea suficiente tener un conocimiento no completo de ellas. Esta es, claramente, una hipótesis de trabajo que pudiera no cumplirse. En la mecánica estadística de sistemas de partículas clásicas se utiliza esta herramienta con profusión. La razón se halla en que las *funciones dinámicas* $d(x_1, ..., x_n)$ del sistema, asociadas a propiedades de éste (como la energía, etc.) pueden generalmente representarse con muy buena aproximación en la forma simplificada (recuérdese que n es del orden de 10^{23}):

$$d(x_1, x_2, ..., x_n) \simeq d_0 + \sum_{i=1}^{n} d_1(x_i) + \sum_{i<j} d_2(x_i, x_j) \qquad [94]$$

El valor medio de la propiedad $\langle D \rangle$ asociada puede expresarse como:

$$\langle D \rangle \simeq \int dx_1 \ldots dx_n \left[d_0 + \sum_{i=1}^{n} d_1(x_i) + \sum_{i<j} d_2(x_i, x_j) \right] \cdot f(x_1, x_2, ..., x_n) =$$

$$= \int d_0 \cdot f(x_1, ..., x_n)\, dx_1 \ldots dx_n + \sum_{i=1}^{n} \int d_1(x_1) f(x_1, x_2, ..., x_n)\, dx_1 \ldots dx_n +$$

$$+ \sum_{i<j} \int d_2(x_i, x_j) \cdot f(x_1, x_2, ..., x_n) \cdot dx_1 \cdot dx_2 \cdot \ldots \cdot dx_n \qquad [95]$$

que vemos es fundamentalmente una suma de «pocos» términos: una constante, contribuciones debidas a una partícula, y contribuciones de dos partículas. La magnitud del problema inicial se ve así grandemente reducida, máxime si utilizamos las funciones marginales y las de correlación:

$$\langle D \rangle = d_0 + \sum_{i=1}^{n} \int d_1(x_i) h_i(x_i) \cdot dx_i +$$

$$+ \sum_{i<j} \int d_2(x_i, x_j) [f_i(x_i) f_j(x_j) + g_2(x_i, x_j)] \cdot dx_i \cdot dx_j \qquad [96]$$

Se advierte al lector que en obras de Mecánica Estadística suelen utilizarse factores de normalización sobre número de partículas y no sobre la unidad como aquí (Balescu, 1975). Como ya sabemos, esto no afecta para nada a los resultados finales.

Bibliografía

1. BALESCU, R., *Equilibrium and Nonequilibrium Statistical Mechanics*, Wiley, Nueva York, 1975, capítulo 3.

2. CARSKY, P., y URBAN, M., *Ab-initio Calculations*, Springer-Verlag, Berlín, 1980.

3. CRAMÉR, H., *Elementos de la Teoría de Probabilidades*, Aguilar, Madrid, 1977, capítulos 9 y 10.

4. DEMIDOWITSCH, B. P.; MARON, I. A., y SCHUWALOWA, E. S., *Métodos Numéricos del Análisis*, Paraninfo, Madrid, 1980, capítulo 2.

5. KAMPEN, N. G., *Stochastic Processes in Physics and Chemistry*, Elsevier, Ámsterdam, 1985, capítulo 2.

6. LANDAU, L. D., y LIFSHITZ, E. M., *Física Estadística*, Reverté, Barcelona, 1969, capítulo 12.

7. LÖWDIN, P. O., *Adv. Chem. Phys.*, 2, 207 (1959).

8. POPLE, J. A., y BEVERIDGE, D. L., *Approximate Molecular Orbital Theory*, McGraw-Hill, Nueva York, 1970.

9. SHAVITT, I., en *Methods in Computational Physics*, vol. 2, «The gaussian function in calculations of statistical mechanics and quantum mechanics», editores B. Alder, S. Fernbach y M. Rotenberg, Academic Press, Nueva York, 1963.

10. SPIRIDONOV, V. P., y LOPATKIN, A. A., *Tratamiento matemático de datos físico-químicos*, Mir, Moscú, 1973, capítulo 6.

EJERCICIOS DE AUTOCOMPROBACION

1. Efectuar las operaciones pertinentes para establecer la igualdad [7].

2. Obtener la parábola de mínimos cuadrados que represente a una distribución en el plano xy.

3. Deducir aplicando el método de mínimos cuadrados la ecuación $P = P(V)$ de un gas a partir de la siguiente tabla presión/volumen:

P(atm.)	10	20	30	40
V(l.)	87,45	58,17	45,83	38,69

4. Demostrar la relación [47].

5. Demostrar que la distribución normal bidimensional [56] da regresión lineal en ambas variables estudiando las curvas de regresión de las medias.

6. Normalizar la expresión de $w(x_1, x_2)$ contenida en [63] aplicando una transformación lineal a las variables x_1, x_2, que reduzca la forma cuadrática $a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$ a una suma simple $\tilde{x}_1^2 + \tilde{x}_2^2$.

7. Verificar [88] haciendo uso de las relaciones [89].

SOLUCIONES A LOS EJERCICIOS DE AUTOCOMPROBACION

1. El valor medio de una variable no negativa es evidentemente no negativo:

$$\langle Z\rangle = \langle(\mu_{02}(X - \langle X\rangle) - \mu_{11}(Y - \langle Y\rangle))^2\rangle \geqslant 0$$

teniéndose pues:

$$\langle Z\rangle = \mu_{02}^2[\langle X^2\rangle - \langle X\rangle^2] - 2\mu_{11}\mu_{02}[\langle XY\rangle - \langle X\rangle\langle Y\rangle] + \\ + \mu_{11}^2[\langle Y^2\rangle - \langle Y\rangle^2]$$

de donde aplicando las relaciones [5]-[6] llegamos a la igualdad pedida:

$$\langle Z\rangle = \mu_{02}^2 \cdot \mu_{20} - 2\mu_{11}^2 \cdot \mu_{02} + \mu_{11}^2 \cdot \mu_{02} = \mu_{02}^2 \cdot \mu_{20} - \mu_{11}^2 \cdot \mu_{02}$$

2. Sea la parábola $Y = aX^2 + bX + c$. Se trata de hacer mínimo:

$$M = \langle(Y - aX^2 - bX - c)^2\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y - ax^2 - bx - c)^2 f(x, y)\,dx\,dy$$

con respecto a los parámetros a, b, c. Por derivación parcial se llega al sistema de ecuaciones:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yx^2 f(x,y)\,dx\,dy = a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^4 f(x,y)\,dx\,dy +$$

$$+ b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^3 f(x,y)\,dx\,dy + c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\,dx\,dy$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yx f(x,y)\,dx\,dy = a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^3 f(x,y)\,dx\,dy +$$

$$+ b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\,dx\,dy + c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dx\,dy = a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\,dx\,dy +$$

$$+ b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy + c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

que convenientemente resuelto da los valores óptimos de a, b, c.

3. Estamos ante un caso de relación doble logarítmica:

$$\log P + a \log V = b$$

que requeriría ajustar la tabla transformada:

$\log P$	1	1,3010	1,4771	1,6021
$\log V$	1,9418	1,7647	1,6611	1,5876

obteniéndose la relación:

$$P \cdot V^{1,7} \simeq 20\,000$$

4. Basta con comprobar que

$$A = \langle (Y - g(X)) \cdot (g(X) - \langle Y \rangle) \rangle = 0$$

Desarrollando este promedio se encuentra:

$$A = \sum_n \sum_m (y_m - g(x_n)) \cdot (g(x_n) - \langle Y \rangle) P_{nm} =$$

$$= \sum_n (g(x_n) - \langle Y \rangle) P_n^x \left[\sum_m (y_m - g(x_n)) P(y_m \mid x_n) \right] =$$

$$= \{ g(x_n) = \sum_m y_m P(y_m \mid x_n) \} = 0$$

quedando probada la relación pedida (recuérdese [41]).

5. Comencemos evaluando las funciones densidad marginales asociadas a $f(x, y)$ [56]:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \frac{1}{\sigma(X)\sqrt{2\pi}} \cdot \exp\left[- \frac{(x - \langle X \rangle)^2}{2\sigma^2(X)} \right]$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \frac{1}{\sigma(Y)\sqrt{2\pi}} \cdot \exp\left[- \frac{(y - \langle Y \rangle)^2}{2\sigma^2(Y)} \right]$$

Por generalización directa de la expresión [40] la probabilidad condicional de que $y < Y < y + dy$, cuando $x < X < x + dx$, será:

$$P(y < Y < y + dy \mid x < X < x + dx) = \frac{f(x, y)}{f_1(x)} \cdot dy = f(y \mid x) \cdot dy$$

(nótese que debemos multiplicar por dy al estar en caso continuo). En tanto $f_1(x) \neq 0$, $f(y \mid x)$ se comporta como una función densidad normalizada para Y condicional a X. Esta función resulta ser:

$$f(y \mid x) = \frac{f(x, y)}{f_1(x)} =$$

$$= \frac{\exp\left[- \frac{1}{2\sigma^2(Y)(1 - r^2)} \cdot \left\{ y - \langle Y \rangle - \frac{r\sigma(Y)}{\sigma(X)} (x - \langle X \rangle) \right\}^2 \right]}{\sqrt{2\pi} \cdot \sqrt{1 - r^2} \cdot \sigma(Y)}$$

que es una distribución gaussiana con desviación típica independiente de X:

$$\sigma = \sigma(Y) \cdot \sqrt{1 - r^2}$$

Los valores medios de Y condicionales a X son:

$$\langle Y \mid X = x \rangle = \int_{-\infty}^{\infty} y \cdot f(y \mid x)\, dy = \langle Y \rangle + \frac{r\sigma(Y)}{\sigma(X)} \cdot (x - \langle X \rangle)$$

que es la ecuación de una recta si dejamos variar libremente a x. Esta recta coincide con la de regresión de Y sobre X, y dada la simetría entre ambas variables se obtendría el resultado análogo para la curva de regresión de la media de X con respecto a Y. Queda comprobado entonces que la distribución gaussiana bidimensional da un ejemplo de regresión lineal en ambas variables.

6. Sea la pareja de transformaciones (inversa una de otra):

$$\left.\begin{aligned} x_1 &= c_{11}\tilde{x}_1 + c_{12}\tilde{x}_2 \\ x_2 &= c_{21}\tilde{x}_1 + c_{22}\tilde{x}_2 \end{aligned}\right\} \qquad \left.\begin{aligned} \tilde{x}_1 &= d_{11}x_1 + d_{12}x_2 \\ \tilde{x}_2 &= d_{21}x_1 + d_{22}x_2 \end{aligned}\right\}$$

tales que la forma cuadrática original se reduzca a:

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = \tilde{x}_1^2 + \tilde{x}_2^2$$

El problema es equivalente al de diagonalizar la matriz simétrica A que aparece en la expresión matricial:

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = (x_1, x_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

problema que sabemos tiene siempre solución.

Los coeficientes c_{ij}, d_{ij}, pueden obtenerse resolviendo un sencillo sistema de ecuaciones. Lo importante aquí es darse cuenta de que

una vez hecha la transformación a las variables $\tilde{x}_1, \tilde{x}_2$, la integral de w [63] es muy fácil de calcular pues:

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1\, dx_2\, w(x_1, x_2) =$$

$$= N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\tilde{x}_1\, d\tilde{x}_2 \cdot \left| \frac{\partial(x_1, x_2)}{\partial(\tilde{x}_1, \tilde{x}_2)} \right| \cdot \exp\left[-\frac{1}{2}(\tilde{x}_1^2 + \tilde{x}_2^2) \right] =$$

$$= N \cdot |C| \cdot \left[\int_{-\infty}^{\infty} d\tilde{x}_1 \cdot \exp\left[-\frac{1}{2}\tilde{x}_1^2 \right] \right]^2 = N \cdot |C| \cdot 2\pi$$

en donde $|C|$ es el valor absoluto del Jacobiano de la transformación $(x_1, x_2) \to (\tilde{x}_1, \tilde{x}_2)$, es decir:

$$|C| = |c_{11} \cdot c_{22} - c_{12} \cdot c_{21}|$$

Ahora bien el valor de $|C|$ está relacionado con el valor del determinante de la matriz A. Para ver cómo utilizaremos una notación más compacta.

La reducción de la forma cuadrática original puede escribirse:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \sum_i \sum_j a_{ij} \cdot \left[\sum_{l=1}^{n} c_{il} \tilde{x}_l \right] \cdot \left[\sum_{m=1}^{n} c_{jm} \tilde{x}_m \right] =$$

$$= \sum_l \sum_m \left[\sum_i \sum_j a_{ij} \cdot c_{il} \cdot c_{jm} \right] \tilde{x}_l \tilde{x}_m = \sum_l \sum_m \tilde{x}_l \cdot \tilde{x}_m \cdot \delta_{lm}$$

donde δ_{lm} es la δ-Kroenecker. Notemos que todo lo que digamos a continuación será válido para cualquier valor de n, en particular para $n = 2$. Para que se verifique la última igualdad es preciso que:

$$\sum_i \sum_j a_{ij} \cdot c_{il} \cdot c_{jm} = \delta_{lm} = \begin{cases} 1 & l = m \\ 0 & l \neq m \end{cases}$$

Si utilizamos la matriz transpuesta de C, C^T, definida:

$$(C^T)_{li} = (C)_{il}$$

tendremos:

$$(C^T \cdot A \cdot C)_{lm} = \sum_i \sum_j c_{li}^T \cdot a_{ij} \cdot c_{jm} = \delta_{lm}$$

lo que indica que:

$$I = C^T A C = \text{matriz unidad}$$

Como el determinante de un producto de matrices es el producto de los determinantes individuales y, además, el determinante de una matriz es idéntico al de su transpuesta, encontramos:

$$|C^T \cdot A \cdot C| = |C^T| \cdot |A| \cdot |C| = |A| \cdot |C|^2 = 1$$

y por tanto:

$$|C| = |A|^{-1/2}$$

$|A|$ es mayor que cero por ser su forma cuadrática definida positiva.

De todo ello resulta, finalmente, el valor para la constante de normalización:

$$N = \frac{1}{2\pi \cdot |C|} = \frac{|A|^{1/2}}{2\pi} = \frac{|a_{11}a_{22} - a_{12}^2|^{1/2}}{2\pi}$$

7. Si el punto C cae entre A y B según [89] (Fig. 2) y siendo:

$$r_C = [(x - x_C)^2 + (y - y_C)^2 + (z - z_C)^2]^{1/2}$$
$$r_A = [(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2]^{1/2}$$
$$r_B = [(x - x_B)^2 + (y - y_B)^2 + (z - z_B)^2]^{1/2}$$

las distancias $|\mathbf{AC}|$ y $|\mathbf{CB}|$ son:

$$a = |\mathbf{AC}| = \frac{a_j}{a_i + a_j} \cdot |\mathbf{AB}| \quad ; \quad b = |\mathbf{CB}| = \frac{a_i}{a_i + a_j} \cdot |\mathbf{AB}|$$

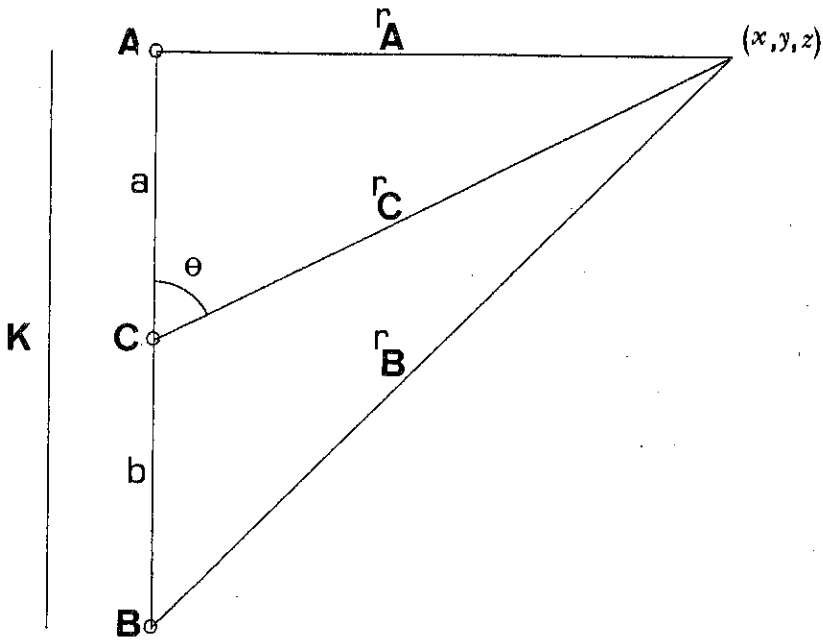


Figura 2.—Diagrama para el producto de gaussianas.

Por trigonometría elemental tenemos:

$$r_A^2 = a^2 + r_C^2 - 2ar_C \cdot \cos\theta$$
$$r_B^2 = b^2 + r_C^2 + 2br_C \cdot \cos\theta$$

y de ahí:

$$br_A^2 + ar_B^2 = |\mathbf{AB}|(ab + r_C^2)$$

$$a_i r_A^2 + a_j r_B^2 = \frac{a_i a_j}{a_i + a_j}|\mathbf{AB}|^2 + (a_i + a_j)r_C^2$$

quedando probada la relación [88].

Tema 9

PROCESOS ESTOCASTICOS Y CADENAS DE MARKOV DISCRETAS

Como colofón al apartado de Estadística centraremos aquí la atención en los *procesos estocásticos*, que permiten diseñar aplicaciones de gran valor para el estudio de muchos fenómenos: movimiento browniano, caminos aleatorios, estudio de líquidos, reactividad, etc.

Tras introducir la nomenclatura usual y las condiciones matemáticas asociadas a los procesos estocásticos en general, especializaremos este concepto al caso *estacionario*. Aparecerán entonces interesantes cuestiones como la formulación de la *hipótesis ergódica*, las *funciones de autocorrelación* y la *densidad espectral*.

Una clase de procesos estacionarios importantísima es la de *Markov*, que se caracteriza por poseer memoria de un paso. En estos procesos, el resultado de un ensayo depende únicamente del resultado previo. Como aplicación se estudiará el comportamiento de la velocidad de una partícula browniana monodimensional.

Dentro de la clase de los procesos de Markov nos fijaremos, con especial detalle, en las *cadenas de Markov finitas en tiempo discreto*, las cuales recorren *espacios de estado* finitos. Todas las propiedades del proceso se encuentran contenidas en la *matriz de transición a un paso*. Atendiendo a sus características, los estados y las cadenas pueden clasificarse en diferentes categorías. La situación más interesante es la de cadenas que conduzcan a una *distribución estacionaria* de probabilidades para los estados del proceso (*ergodicidad*), pues sirven para simular el comportamiento de sistemas físicos en determinadas condiciones (equilibrio). Como aplicación se describen someramente los fundamentos de la llamada *simulación Monte Carlo* de líquidos.

La restricción de Markov relativa a memoria de un paso es relajada, finalmente, para dar cabida a procesos de Markov más generales. Encontraremos así procesos de *orden superior* que guardan memoria de dos, tres, y más pasos. Un sencillo artificio los reduce a la clase de Markov simple, con lo que su estudio detallado se reduce a lo ya visto.

1. GENERALIDADES

Sea dada una variable aleatoria X que nos informa de alguna característica de un sistema y que posee la *densidad de probabilidad* $f(x)$. Cualquier función de X define, como sabemos, una nueva variable aleatoria Y. Si consideramos ahora la posible evolución temporal de los resultados particulares y de Y, denominaremos *proceso estocástico* (*aleatorio*) a la función aleatoria general:

$$Y_X(t) = F(X, t) \qquad [1]$$

Queda, así, el tiempo, incluido para describir la dinámica de las funciones de probabilidad. No obstante, t puede ser tanto *continua como discreta*, siendo muchas veces un artificio para medir etapas dentro de un proceso aleatorio. Es claro pues que esta variable temporal podría no tener nada que ver con el tiempo real.

El proceso estocástico $Y(t)$ es entonces *el conjunto de realizaciones o trayectorias posibles*:

$$Y_X(t) = \{Y_x(t) = F(x, t)\} \qquad [2]$$

que recorren el espacio de estados asociado (conjunto de valores posibles $Y(t)$). La dependencia temporal con t lo es en un sentido probabilista y no de manera exactamente determinista. Precisemos esto un poco más.

El proceso $Y(t)$ se describe a través de una colección de distribuciones de probabilidad en lo que se llama una *jerarquía* de distribuciones. La primera función de distribución a considerar es:

$$w_1(y, t) = \int \delta(y - Y_x(t)) f(x)\, dx \qquad [3]$$

que nos da la densidad de probabilidad de que $Y_X(t)$ tome el valor y en el instante t, conocida la distribución $f(x)$. La segunda función de distribución se expresa:

$$w_2(y_1, t_1; y_2, t_2) = \int \delta(y_1 - Y_x(t_1)) \cdot \delta(y_2 - Y_x(t_2)) \cdot f(x)\, dx \qquad [4]$$

y representa la densidad conjunta de que $Y_X(t)$ tome el valor y_1 en t_1 y el valor y_2 en t_2. En general, la n-ésima función densidad vendría dada por:

$$w_n(y_1, t_1; y_2, t_2; \ldots; y_n, t_n) = \int \prod_{i=1}^{n} \delta(y_i - Y_x(t_i)) \cdot f(x)\, dx \qquad [5]$$

calculándose la probabilidad de que $Y_X(t)$ caiga entre $(y_i, y_i + dy_i)$ en los instantes t_i como el producto habitual:

$$P(Y_X(t_1) = y_1; \ldots; Y_X(t_n) = y_n) = w_n \cdot dy_1 \cdot dy_2 \cdot \ldots \cdot dy_n \qquad [6]$$

Se obtiene así una jerarquía infinita de funciones de distribución $\{w_n\}$ que caracteriza completamente al proceso aleatorio. Las funciones w_n deben satisfacer las *condiciones de consistencia de Kolmogoroff* (Kampen, 1985):

i) $w_n \geqslant 0 \quad ; \quad n = 1, 2, 3, \ldots$

ii) $w_n(y_1, t_1; \ldots; y_n, t_n)$ es *simétrica* bajo la permutación de pares $(y_k, t_k) \leftrightarrow (y_j, t_j)$

iii) $$w_k(y_1, t_1; y_2, t_2; \ldots; y_k, t_k) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} dy_{k+1} \ldots dy_n \cdot w_n(y_1, t_1; \ldots; y_n, t_n) \qquad n > k$$

iv) $$\int_{-\infty}^{\infty} w_1(y, t)\, dy = 1 \qquad [7]$$

A la inversa, todavía podemos afirmar que cualquier conjunto de funciones que obedezca i)-iv) define un proceso estocástico.

La utilización de las w_n permite calcular valores medios que, en

muchos casos, poseen un importante significado físico. Así el valor medio de $Y(t)$ es:

$$\langle Y(t)\rangle = \int_{-\infty}^{\infty} Y_X(t)\cdot f(x)\,dx = \int_{-\infty}^{\infty} y\cdot w_1(y,t)\,dy \qquad [8]$$

como se comprueba inmediatamente aplicando la propiedad fundamental del operador δ (Ejercicio 1). La siguiente magnitud de interés es el momento de orden dos:

$$\langle Y(t_1)\cdot Y(t_2)\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1\cdot y_2\cdot w_2(y_1,t_1;y_2,t_2)\cdot dy_1\cdot dy_2 \qquad [9]$$

con ayuda del cual se contruye la *función de autocorrelación*:

$$R(t_1,t_2) = \langle Y(t_1)\cdot Y(t_2)\rangle - \langle Y(t_1)\rangle\cdot\langle Y(t_2)\rangle \qquad [10]$$

que cuando $t_1 = t_2$ se reduce a la varianza dependiente del tiempo $\sigma_Y^2(t)$. En procesos con media nula [9] y [10] coinciden. Mediante esta mecánica cualquier tipo de momento o de valor medio puede ser expresado formalmente.

Cabe hacer todavía una definición más. Un proceso aleatorio se dice *puramente aleatorio* cuando las funciones de la jerarquía w_n se factorizan en la forma:

$$w_n(y_1,t_1;y_2,t_2;\ldots;y_n,t_n) = \prod_{i=1}^{n} w_1(y_i,t_i) \qquad [11]$$

en respuesta a que los valores y_i en instantes diferentes están completamente descorrelacionados.

2. PROCESOS ESTACIONARIOS

No obstante su elegancia formal, el conjunto $\{w_n\}$, en la práctica, no puede ser determinado. Solamente fijar la función w_1 requeriría, desde un punto de vista experimental, un gran esfuerzo. Deberíamos observar un número muy grande de sistemas, preparados similarmente, durante el

tiempo preciso para confeccionar la función $w_1(y, t)$. Por ello, la búsqueda de argumentos que permitan simplificar este esquema en situaciones de interés resulta de gran valor.

La consideración más usual es asumir que el proceso estocástico en cuestión es *estacionario*, lo que significa que la forma de las funciones de distribución no se verá afectada por un cambio en el origen de tiempos. Expresado en términos más intuitivos, lo anterior equivale a suponer que el mecanismo básico que produce los resultados del proceso aleatorio no cambia con el tiempo. La simplificación que se introduce en la jerarquía $\{w_n\}$ es entonces notable:

$$\begin{aligned} w_1(y, t) &\to w_1(y) \\ w_2(y_1, t_1; y_2, t_2) &\to w_2(y_1; y_2; t_2 - t_1) \\ w_3(y_1, t_1; y_2, t_2; y_3, t_3) &\to w_3(y_1, y_2, y_3; t_2 - t_1; t_3 - t_1) \\ &\cdots\cdots \end{aligned} \qquad [12]$$

Las propiedades del proceso dependerán de intervalos de tiempo transcurridos, pero nunca del origen de tiempos. Esta característica hace muy útiles a estos procesos para representar fenómenos y sistemas de interés físico-químico (Kampen, 1985). La simplificación de estacionariedad da una aproximación, en muchos casos muy buena, a los procesos fenomenológicos (equilibrio). Sin embargo, en la naturaleza no existen procesos estrictamente estacionarios, aunque puede obtenerse una imagen aproximada de éstos dejando que el tiempo de observación sea mucho mayor que el tiempo característico del fenómeno físico bajo estudio. Veamos algunas características de la simplificación de *estacionariedad.*

i) Por lo que respecta a las w_n su determinación en procesos estacionarios es mucho más cómoda que en los que no lo son. La razón se encuentra en la llamada *hipótesis ergódica* según la cual, en procesos estacionarios, un número muy grande de observaciones sobre un único sistema en N instantes arbitrarios equivale a observar simultáneamente en un conjunto grande de sistemas (similares al anterior) a N de ellos.

De este modo, la tarea se reduce a observar un único sistema durante un período muy grande de tiempo. Este período puede dividirse en intervalos de longitud $2T$, asimilando cada uno de ellos a una observación sobre un sistema del conjunto global. La única condición es que $2T$ sea mucho mayor que cualquier periodicidad que pudiere tener lugar en el proceso. Los valores medios del conjunto de N observaciones sobre un único sistema son promedios temporales que denotaremos por $\bar{Y}$:

$$\bar{Y} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} y(t)\, dt \qquad [13]$$

Los valores medios de las N observaciones sobre el conjunto (*colectivo* en sentido amplio) de sistemas similares son promedios en un instante dado:

$$\langle Y(t)\rangle = \int_{-\infty}^{\infty} y w_1(y)\, dy \qquad [14]$$

estableciendo la hipótesis ergódica que en procesos estacionarios:

$$\bar{Y} = \langle Y(t)\rangle \qquad [15]$$

ii) Hemos obtenido en [15] que el valor medio de un proceso estacionario es independiente del tiempo. Para los momentos de orden superior la dependencia temporal es tal que éstos no se ven afectados por un corrimiento en el origen de tiempos, como consecuencia de [12]:

$$\langle Y(t_1 + \tau)\cdot Y(t_2 + \tau) \ldots Y(t_n + \tau)\rangle = \langle Y(t_1)\cdot Y(t_2) \ldots Y(t_n)\rangle \qquad [16]$$

Es particularmente útil la función de autocorrelación $\langle Y(t_1)\cdot Y(t_2)\rangle$, que en procesos con media nula se reduce a:

$$R(t_1, t_2) = \langle Y(t_1)\cdot Y(t_2)\rangle = R(|t_1 - t_2|) = R(\tau) \qquad [17]$$

Este tipo de función aparece en las aplicaciones de la *Teoría de la Respuesta Lineal* (a perturbaciones externas sobre un sistema) y encuentra utilidad en el estudio de la espectroscopía en fase líquida, fenómenos de transporte, etc. (Balescu, 1975; McQuarrie, 1976).

iii) Se define el *tiempo de autocorrelación* τ_c como aquél tal que $R(t_1, t_2) \to 0$, cuando $|t_1 - t_2| > \tau_c$. Esto indica que a medida que pasa el tiempo, desde una situación inicial dada (t_1), se pierde la correlación (dependencia) con dicha situación. Se dice que el proceso pierde la «memoria» de cómo era en t_1. Todo ello tiene su paralelismo con los denominados *tiempos de relajación* en los fenómenos físicos.

iv) El concepto de función de autocorrelación admite una generalización para estudiar la correlación entre dos procesos estocásticos $Y_i(t)$ e

$Y_j(t)$. Surge de aquí la *función de correlación cruzada* $R_{ij}(t_1, t_2)$, que en procesos estacionarios con media nula es:

$$R_{ij}(t_1, t_2) = \langle Y_i(t_1) \cdot Y_j(t_2) \rangle = \langle Y_i(0) \cdot Y_j(|t_2 - t_1|) \rangle \qquad [18]$$

En particular, cuando la variable $Y(t)$ sea de carácter vectorial y se especifique, por tanto, a través de sus componentes $Y(t) = \{Y_i(t)\}$, aparecerá este tipo de correlación que conviene presentar en forma de matriz. Los términos diagonales serán las funciones de autocorrelación de cada componente, en tanto que los no diagonales serán las correlaciones cruzadas. Haciendo uso de la hipótesis ergódica, escribiremos para la función de correlación temporal general estacionaria:

$$R_{ij}(t_1, t_2) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} y_i(t) \cdot y_j(t + |t_1 - t_2|)\, dt =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_{i1} \cdot y_{j2} \cdot w_2^{ij}(y_{i1}, t_1; y_{j2}, t_2)\, dy_{i1} \cdot dy_{j2} =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_{i1} \cdot y_{j2} \cdot w_2^{ij}(y_{i1}; y_{j2}, |t_1 - t_2|)\, dy_{i1}\, dy_{j2} \qquad [19]$$

Nótese, finalmente, que $R_{ij}(\tau) = R_{ji}(-\tau)$ (no paridad).

3. CORRELACION Y DENSIDAD ESPECTRAL

Dado un proceso estocástico continuo y estacionario hay dos funciones que permiten caracterizarlo con gran precisión. Una de ellas es la función (o funciones) de correlación temporal, de la que hemos hablado anteriormente. La segunda es la *densidad espectral*. Por simplicidad trataremos con procesos estacionarios con una componente.

La función de autocorrelación para un $Y(t)$ estacionario es:

$$R(|t_1 - t_2|) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} y(t) \cdot y(t + |t_1 - t_2|)\, dt =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 \cdot y_2 \cdot w_2(y_1, y_2; |t_1 - t_2|)\, dy_1 \cdot dy_2 \qquad [20]$$

Si no hubiese correlación entre y_1 e y_2, entonces [11-12]:

$$w_2(y_1, y_2; |t_1 - t_2|) = w_1(y_1) \cdot w_1(y_2) \qquad [21]$$

y para procesos con media nula se concluye $R(|t_1 - t_2|) = 0$. Es evidente, por su definición, que R es una función par de $t_1 - t_2$.

Conviene analizar el comportamiento de R en función de la evolución temporal. Intuitivamente, se espera que, caso de existir correlación entre $y_1(t_1)$ e $y_2(t_2)$, esta correlación decrezca al crecer el incremento temporal $|t_1 - t_2|$. Si se parte de la desigualdad (McQuarrie, 1975):

$$[y(t) \pm y(t + |t_1 - t_2|)]^2 \geqslant 0 \qquad [22]$$

se llega a:

$$y^2(t) + y^2(t + |t_1 - t_2|) \geqslant \pm 2y(t) \cdot y(t + |t_1 - t_2|) \qquad [23]$$

y tomando el promedio temporal:

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \{y^2(t) + y^2(t + |t_1 - t_2|)\}\, dt \geqslant$$

$$\geqslant \pm 2 \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} y(t) \cdot y(t + |t_1 - t_2|)\, dt \qquad [24]$$

es inmediato establecer:

$$2R(0) \geqslant \pm 2R(|t_1 - t_2|) \Rightarrow R(0) \geqslant |R(|t_1 - t_2|)| \qquad [25]$$

de donde se concluye que la función de autocorrelación R está acotada superiormente. La velocidad de caída se tratará en el epígrafe 5.

Ocupémonos ahora de la *densidad espectral* de un proceso estacionario localizado en el intervalo $-T \leqslant t \leqslant T$. Este proceso es una función de t y puede ser analizado vía integral de Fourier definiendo:

$$Y_T(t) = \begin{cases} Y(t) & -T \leqslant t \leqslant T \\ 0 & \text{en otro caso} \end{cases} \qquad [26]$$

La transformada de Fourier de $Y_T(t)$ es:

$$A(\omega) = \int_{-\infty}^{\infty} Y_T(t) e^{-i\omega t} \cdot dt = \int_{-T}^{T} Y(t) e^{-i\omega t} \cdot dt \qquad [27]$$

y por consiguiente su inversión conduce a:

$$Y_T(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} A(\omega) e^{i\omega t} \cdot d\omega \qquad [28]$$

La *densidad espectral* (*espectro de potencias*) del proceso $Y(t)$ se define como:

$$S(\omega) = \lim_{T \to \infty} \frac{1}{2T} |A(\omega)|^2 \qquad [29]$$

Dado que $Y(t)$ es real $A^*(\omega) = A(-\omega)$ y $S(\omega)$ resulta ser una función par de ω.

La función de autocorrelación R y la densidad espectral S contienen la misma información sobre el proceso estocástico, ya que según el *teorema de Wiener-Khinchin* (Kampen, 1985) ambas forman una pareja de transformados de Fourier (en coseno). Designemos $\tau = t_1 - t_2$ y sea la función de autocorrelación reducida al intervalo $[-T, T]$:

$$R_T(\tau) = \frac{1}{2T} \int_{-T}^{T} Y_T(t) \cdot Y_T(t + \tau)\, dt \qquad [30]$$

cuyo límite cuando $T \to \infty$ es justamente $R(\tau)$:

$$R(\tau) = \lim_{T \to \infty} R_T(\tau) \qquad [31]$$

La transformada de Fourier de [30] es (por paridad):

$$\int_{-\infty}^{\infty} R_T(\tau) \cdot e^{-i\omega\tau} \cdot d\tau = 2 \int_{0}^{\infty} R_T(\tau) \cdot \cos \omega\tau \cdot d\tau \qquad [32]$$

Introduciendo ahora [30] en la integral izquierda de [32] encontramos:

$$2\int_0^{\infty} R_T(\tau)\cdot\cos\omega\tau\cdot d\tau = \frac{1}{2T}|A(\omega)|^2 = \frac{A(\omega)A(-\omega)}{2T} \qquad [33]$$

y tomando límites cuando $T\to\infty$ se establece:

$$2\int_0^{\infty} R_T(\tau)\cdot\cos\omega\tau\cdot d\tau = S(\omega) \qquad [34]$$

y recíprocamente:

$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\cdot\cos\omega\tau\cdot d\omega \qquad [35]$$

4. PROCESOS DE MARKOV

Dentro de los procesos aleatorios con aplicación directa al estudio de situaciones químico-físicas ocupan un lugar preponderante los *procesos de Markov*. Estos procesos no son puramente aleatorios, sino que guardan en parte «memoria de la evolución temporal».

Ante un proceso estocástico cualquiera, podemos definir la densidad de probabilidad condicional de que en t_n esté $Y(t_n)$ entre $(y_n, y_n + dy_n)$, habiéndose dado la sucesión previa:

$$y_0 < Y(t_0) < y_0 + dy_0; \ldots; y_{n-1} < Y(t_{n-1}) < y_{n-1} + dy_{n-1} \qquad [36]$$

y escribiendo tal probabilidad en la forma:

$$P_n(y_n, t_n \mid y_0, t_0; y_1, t_1; \ldots; y_{n-1}, t_{n-1}) \qquad [37]$$

Se llama proceso de Markov a aquel en que:

$$P_n(y_n, t_n \mid y_0, t_0; \ldots; y_{n-1}, t_{n-1}) = p_2(y_n, t_n \mid y_{n-1}, t_{n-1}) \qquad [38]$$

es decir que guarda «*memoria de sólo un paso*, olvidando la historia del proceso previa al instante t_{n-1}. La probabilidad p_2 resulta ser:

$$p_2(y_n, t_n \mid y_{n-1}, t_{n-1}) = \frac{w_2(y_n, t_n; y_{n-1}, t_{n-1})}{w_1(y_{n-1}, t_{n-1})} \qquad [39]$$

y *caracteriza completamente al proceso*. Si éste es estacionario, p_2 cumplirá:

i) $p_2(y_2, t \mid y_1) \geqslant 0 \quad ; \quad t = t_2 - t_1$

ii) $\int_{-\infty}^{\infty} p_2(y_2, t \mid y_1)\, dy_2 = 1$

iii) $w_1(y_2) = \int_{-\infty}^{\infty} w_1(y_1) \cdot p_2(y_2, t \mid y_1) \cdot dy_1$

iv) $p_2(y_2, t \mid y_1) = \int_{-\infty}^{\infty} p_2(y_2; t - t_0 \mid y) \cdot p_2(y; t_0 \mid y_1) \cdot dy \quad ; \quad 0 \leqslant t_0 \leqslant t$

[40]

Se conoce a iv) como la condición de *Smoluchowski-Chapman-Kolmogoroff* y su significado es el siguiente. La densidad de probabilidad de que el proceso esté en $(y_2, y_2 + dy_2)$, pasado un tiempo t desde que estuvo en $(y_1, y_1 + dy_1)$, se obtiene integrando el producto de las densidades de probabilidad asociadas al paso intermedio (y, t_0) sobre todos estos posibles pasos intermedios. El orden temporal es básico $0 \leqslant t_0 \leqslant t$. La condición que nos ocupa puede ser desarrollada considerando más pasos intermedios: $y_1 \to (y_0, t_0) \to (y_{01}, t_{01}) \to \ldots \to (y_2, t)$, incluyendo las densidades apropiadas y las integraciones sobre $y_0, y_{01}, \ldots$, etc. La ecuación iv) es fundamental en la teoría de los procesos de Markov y puede usarse para definirlos.

Notemos finalmente la similitud entre la regla de composición de probabilidades que expresa iv) y el producto convencional de matrices. La diferencia fundamental es que se ha considerado al proceso como una variable continua y, por tanto, el elemento $(y_2, t; y_1)$ es la integral (suma continua) sobre el índice y de «columna» de los productos $(y_2, t - t_0 \mid y) \cdot (y, t_0 \mid y_1)$.

5. EJEMPLO DE APLICACION: MOVIMIENTO BROWNIANO EN UNA DIMENSION

Para describir el comportamiento estocástico de la velocidad de una partícula monodimensional se propuso en 1930 (Uhlenbeck y Ornstein, 1930) el proceso de Markov, *estacionario y gaussiano*, siguiente:

$$w_1(v_1) = \frac{1}{\sqrt{2\pi}} \exp\left[-v_1^2/2\right]$$

$$p_2(v_2, t \mid v_1) = \frac{1}{\sqrt{2\pi(1 - e^{-2t})}} \cdot \exp\left[- \frac{(v_2 - v_1 e^{-t})^2}{2(1 - e^{-2t})} \right]$$

Comencemos verificando las condiciones de consistencia de Kolmogoroff [40]:

i) Es trivial por ser p_2 intrínsecamente no negativa.

ii) Una simple integración conduce a:

$$\frac{1}{\sqrt{2\pi(1 - e^{-2t})}} \cdot \int_{-\infty}^{\infty} \exp\left[- \frac{v_2^2 - 2 \cdot v_2 \cdot v_1 \cdot e^{-t} + v_1^2 e^{-2t}}{2(1 - e^{-2t})} \right] dv_2 =$$

$$= \frac{\sqrt{2\pi(1 - e^{-2t})}}{\sqrt{2\pi(1 - e^{-2t})}} \cdot \exp\left[\frac{1 - e^{-2t}}{2} \cdot \left\{ \frac{v_1^2 e^{-2t}}{(1 - e^{-2t})^2} - \frac{4 v_1^2 e^{-2t}}{4(1 - e^{-2t})^2} \right\} \right] = 1$$

resultado que no debe extrañar pues, partiendo de un valor v_1 dado, la suma (integral) sobre todos los posibles resultados finales v_2 debe, necesariamente, ser la unidad.

iii) A la vista de $w_1(v_1)$ podemos anticipar la forma de $w_1(v_2)$:

$$w_1(v_2) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-v_2^2/2\right]$$

que es, ciertamente, el resultado de la integración:

$$\int_{-\infty}^{\infty} w_1(v_1) \cdot p_2(v_2, t \mid v_1)\, dv_1 =$$

$$= \frac{1}{2\pi\sqrt{1 - e^{-2t}}} \int_{-\infty}^{\infty} dv_1 \cdot \exp\left[-\frac{v_1^2}{2} - \frac{(v_2 - v_1 e^{-t})^2}{2(1 - e^{-2t})} \right] =$$

$$= \frac{1}{2\pi\sqrt{1 - e^{-2t}}} \int_{-\infty}^{\infty} dv_1 \cdot \exp\left[-\left\{ \frac{v_1^2 - 2v_1 v_2 e^{-t} + v_2^2}{2(1 - e^{-2t})} \right\} \right] =$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot v_2^2 \right] = w_1(v_2)$$

iv) Utilizando propiedades de la gaussiana se verifica la condición de Smoluchowski-Chapman-Kolmogoroff:

$$\int_{-\infty}^{\infty} dv \cdot \frac{1}{\sqrt{2\pi(1 - e^{-2(t-t_0)})}} \cdot \frac{1}{\sqrt{2\pi(1 - e^{-2t_0})}} \cdot$$

$$\cdot \exp\left[-\frac{(v_2 - ve^{-(t-t_0)})^2}{2(1 - e^{-2(t-t_0)})} \right] \cdot \exp\left[-\frac{(v - v_1 e^{-t_0})^2}{2(1 - e^{-2t_0})} \right] = p_2(v_2, t \mid v_1)$$

El proceso propuesto es evidente que posee media nula:

$$\langle v \rangle = 0$$

Veamos ahora la forma de la función de autocorrelación de velocidades. Según [20], para procesos estacionarios con media nula, tenemos ($\tau > 0$):

$$R(\tau) = \langle v_1(0) \cdot v_2(\tau) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_1 \cdot v_2 \cdot w_2(v_2, \tau; v_1) \cdot dv_1 \cdot dv_2 =$$

$$= \frac{1}{2\pi\sqrt{1 - e^{-2\tau}}} \int_{-\infty}^{\infty} v_1 \cdot e^{-v_1^2/2} \cdot dv_1 \cdot \int_{-\infty}^{\infty} v_2 \cdot \exp\left[-\frac{(v_2 - v_1 e^{-\tau})^2}{2(1 - e^{-2\tau})} \right] \cdot dv_2 =$$

$$= \frac{\sqrt{2\pi}}{2\pi\sqrt{1 - e^{-2\tau}}} \cdot \sqrt{1 - e^{-2\tau}} \cdot e^{-\tau} \cdot \int_{-\infty}^{\infty} v_1^2 \cdot \exp\left[-v_1^2/2\right] dv_1 =$$

$$= \frac{e^{-\tau}}{\sqrt{2\pi}} \cdot \frac{\Gamma(3/2)}{(1/2)^{3/2}} = e^{-\tau} \equiv e^{-|\tau|}$$

resultado que indica una *caída exponencial*, con el tiempo τ, de la correlación en la velocidad. La pérdida de la «memoria» inicial sigue pues una ley decreciente de tipo exponencial. Este comportamiento es bastante común a una gran variedad de funciones de autocorrelación (velocidades, dipolos eléctricos moleculares, etc.), si bien su comportamiento asintótico a tiempos largos plantea una serie de interesantes problemas conceptuales (Balescu, 1975).

Finalmente, la densidad espectral asociada es por el teorema de Wiener-Khinchin:

$$S(\omega) = 2 \int_0^{\infty} e^{-\tau} \cdot \cos \omega\tau \cdot d\tau = \frac{2}{1 + \omega^2}$$

que posee la forma de una distribución de Cauchy.

En general, la función $S(\omega)$, como transformada de Fourier de una función de autocorrelación, representa una *susceptibilidad generalizada* para el sistema físico al que se refiere el proceso estacionario. En definitiva, $S(\omega)$ es la respuesta del sistema físico a una perturbación externa débil (teoría de la respuesta lineal). Por ejemplo, en espectroscopía, la forma de una banda de absorción en el infrarrojo para una disolución muy diluida de una molécula activa en un disolvente inerte sería, en esencia, la transformada de Fourier de la función de autocorrelación dipolar de la molécula activa (McQuarrie, 1976):

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \cdot e^{-i\omega t} \cdot \langle \boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}(t) \rangle$$

6. CADENAS DE MARKOV ESTACIONARIAS EN TIEMPO DISCRETO

Vamos a centrarnos ahora en el estudio de las *cadenas de Markov estacionarias y finitas en tiempo discreto.* Sus propiedades han sido bien establecidas dado que son los procesos de Markov más simples que retienen la mayor parte de las características generales básicas. Dentro del dominio de la Química Física gozan de una gran popularidad en el estudio de fases condensadas.

6.1. Espacio de estados, matriz de transición y vector de probabilidad

Para abordar la presentación de las cadenas de Markov es preceptivo introducir varios conceptos: *espacio de estados, matriz de transición a un paso* y *vector de probabilidad.* La variable temporal t será aquí discreta, tomando valores enteros $t = ..., -2, -1, 0, 1, 2, ...$, que generalmente no guardan relación alguna con el tiempo físico real. Como de costumbre llamaremos $Y(t)$ al proceso estocástico descrito por la cadena.

a) *Espacio de estados*

Es el rango de la variable $Y(t)$, es decir el conjunto discreto de estados $\{Y(t_0), Y(t_1), ..., Y(t_n)...\}$ en los que en algún «paso» t_j puede encontrarse el proceso. Si este conjunto es finito, como sucede en muchas aplicaciones de interés, la cadena se dice finita. Conviene resaltar que, aunque el espacio de estados sea finito, su número de elementos puede ser muy elevado. De esta forma procesos con un espacio de estados infinito (numerable o no) pueden ser bien representados por procesos con espacios finitos. La operación de discretización mencionada suele ir asociada al uso de cálculo con ordenador.

b) *Matriz de transición*

Como el proceso se supone estacionario, las probabilidades de transición de un estado i a otro estado j dependerán únicamente del incremento de tiempo. Dado que los procesos de Markov vienen caracterizados

por [38], en este caso quedará definido por el *conjunto de probabilidades de transición a un solo paso* p_{ij}. El conjunto de estas probabilidades recorre todas las posibilidades de que estando en cualquier estado i se pueda efectuar en el instante siguiente un cambio a cualquier estado j (i incluido). La disposición de este conjunto en forma de matriz es, como veremos, especialmente adecuada para la manipulación algebraica de sus cantidades. Se tiene así la matriz de transición P a un paso que para un espacio finito escribiremos:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} \qquad [41]$$

matriz cuadrada en la que todos sus elementos deben ser no negativos:

$$p_{ij} = P(Y(t_j = j) \mid Y(t_i = i)) \geqslant 0 \quad ; \quad i,j = 1, 2, ..., N \qquad [42]$$

Existe aún una condición más que, al menos, deben cumplir las p_{ij}:

$$\sum_j p_{ij} = 1 \quad ; \quad i,j = 1, 2, ..., N \qquad [43]$$

es decir la *normalización por filas* que expresa el suceso seguro de que partiendo de cualquier i se alcance uno cualquiera de los estados j. Por razones de claridad vamos a ceñirnos al caso finito de aquí en adelante (N estados).

Las matrices que cumplen [41]-[43] se denominan *matrices estocásticas* y han sido muy estudiadas desde el punto de vista matemático. Si además de la normalización por filas tenemos la normalización por columnas:

$$\sum_i p_{ij} = 1 \quad ; \quad i,j = 1, 2, ..., N \qquad [44]$$

la matriz P recibe el nombre de *biestocástica*.

En una etapa o paso cualquiera del proceso estaremos interesados en conocer *a priori* cuál es la distribución de probabilidades asociada al espacio de estados, que es lo que llamaremos luego *vector de probabilida-*

des. Para poder determinarla hay que reparar, primero, en que el juego de probabilidades que se pide está condicionado por los pasos previos y, segundo, que necesitamos un conjunto de $N \times N$ probabilidades de transición. Veamos cómo se obtiene de modo constructivo.

A un paso, la matriz P [41] ya contiene la información pedida, los elementos p_{ij} cuyos valores procederán de argumentos físicos concernientes al problema bajo estudio. La mayor o menor bondad de los valores p_{ij} así obtenidos influirá críticamente en los resultados finales que obtengamos con la aproximación al fenómeno natural vía modelo de Markov.

Para obtener la matriz a dos pasos, compuesta por los elementos $p_{ij}^{(2)}$ que representan la probabilidad de que partiendo de i se alcance j en dos pasos, basta con aplicar las reglas del cálculo de probabilidades elemental. Notemos que la citada transición puede esquematizarse como se muestra en la figura 1 y, por tanto, encontramos:

$$p_{ij}^{(2)} = \sum_{k=1}^{N} p_{ik} \cdot p_{kj} \qquad [45]$$

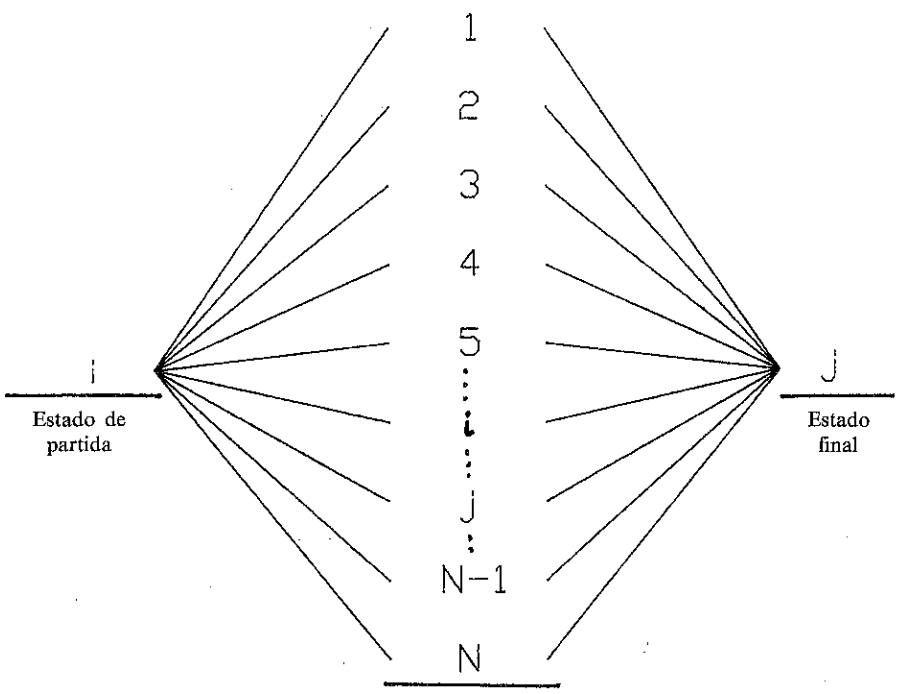


Figura 1.—Diagrama de transiciones a dos pasos.

Puesta en forma matricial [45] se escribirá:

$$P^{(2)} = P \times P = P^2 \qquad [46]$$

es decir, la matriz de transición a dos pasos es justamente igual al cuadrado de la matriz de transición a un paso. En general, se tiene que a n pasos:

$$P^{(n)} = \overbrace{P \times P \times \ldots \times P}^{n\text{-veces}} = P^n \qquad [47]$$

coincidiendo la matriz $P^{(n)}$ con la potencia n-ésima de P. A partir de aquí es simple obtener la versión de la ecuación iv) de [40]:

$$P^{(n+m)} = P^n \times P^m \qquad [48]$$

que recibe el nombre de relación de Chapman-Kolmogoroff.

c) *Vector de probabilidades*

Con lo discutido hasta aquí tenemos a mano muchas de las características globales de un proceso de Markov en tiempo discreto. Estas características están relacionadas con las potencias sucesivas de la matriz de transición a un paso P. Sin embargo, en la *realización* del proceso de Markov (*cadena*) se recorre una sucesión de estados posibles:

$$Y(t_0) \to Y(t_1) \to \ldots \to Y(t_n) \to \ldots \qquad [49]$$

que evoluciona a partir de un estado inicial según sea la matriz P. Toda la información relativa a las probabilidades absolutas de que los estados aparezcan en cada una de las etapas conviene recogerla en un *vector fila* (*vector de probabilidad*):

$$v^{(n)} = (p_1^{(n)}, p_2^{(n)}, \ldots, p_N^{(n)}); \quad n = 0, 1, 2, \ldots \qquad [50]$$

que debe estar, naturalmente, normalizado:

$$\sum_{i=1}^{N} p_i^{(n)} = 1 \qquad [51]$$

siendo sus componentes no negativas.

Para fijar ideas supongamos que el proceso parte del estado $Y(t_0) = 1$. El vector de probabilidad asociado es claramente:

$$v^{(0)} = (1, 0, 0, ..., 0) \qquad [52]$$

Dados este vector $v^{(0)}$ y una matriz P, nos preguntamos por el conjunto de probabilidades de aparición a un paso $p_i^{(1)}$ de los estados $i = 1, 2, ..., N$. Aplicando un razonamiento similar al que llevó a [45] encontramos que:

$$v^{(1)} = (p_1^{(1)}, p_2^{(1)}, ..., p_N^{(1)}) = (1, 0, ..., 0) \cdot \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} =$$

$$= (p_{11}, p_{12}, ..., p_{1N}) \qquad [53]$$

que es el resultado trivial. Consideremos lo que sucederá a dos pasos:

$$v^{(2)} = (p_1^{(2)}, p_2^{(2)}, ..., p_N^{(2)}) = (p_{11}, p_{12}, ..., p_{1N}) \cdot \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} \qquad [54]$$

siendo entonces la probabilidad de aparición de j:

$$p_j^{(2)} = \sum_{k=1}^{N} p_{1k} \cdot p_{kj} \qquad [55]$$

En general, encontramos que el vector de probabilidades (la distribución) en el paso n es:

$$v^{(n)} = v^{(n-1)} \cdot P = v^{(n-2)} \cdot P^2 = \dots = v^{(0)} \cdot P^n \qquad [56]$$

Conviene en este punto realizar los ejercicios 2-3.

Para concluir queda por establecer la probabilidad de una sucesión particular de estados. Sea la cadena $1 \to 3 \to 5 \to 2 \to N \to 4 \to ...$ que tendrá una probabilidad:

$$1 \cdot p_{13} \cdot p_{35} \cdot p_{52} \cdot p_{2N} \cdot p_{N4} \cdot ...$$

6.2. Clasificación de estados

En el ejercicio 3, al analizar cuestiones relativas a la probabilidad de aparición de estados, encontramos dos tipos característicos: *estados transitorios* y *estados ergódicos*. Por los primeros el proceso pasa a lo sumo un número finito de veces, en tanto que en los segundos el proceso entra en ellos quedando atrapado por este conjunto sin poder escapar para ningún valor del paso n. El ejemplo allí discutido puede ampliarse sin dificultad para incluir a varios conjuntos de estos tipos de estados.

En la figura 2 se pretende ilustrar las situaciones más generales. El convenio utilizado para representar transiciones implica que cuando entre dos estados i, j aparecen dos flechas conectoras entonces $p_{ij}, p_{ji} \neq 0$, mientras que si sólo hay una flecha $i \to j$ significa que $p_{ij} \neq 0$ y $p_{ji} = 0$.

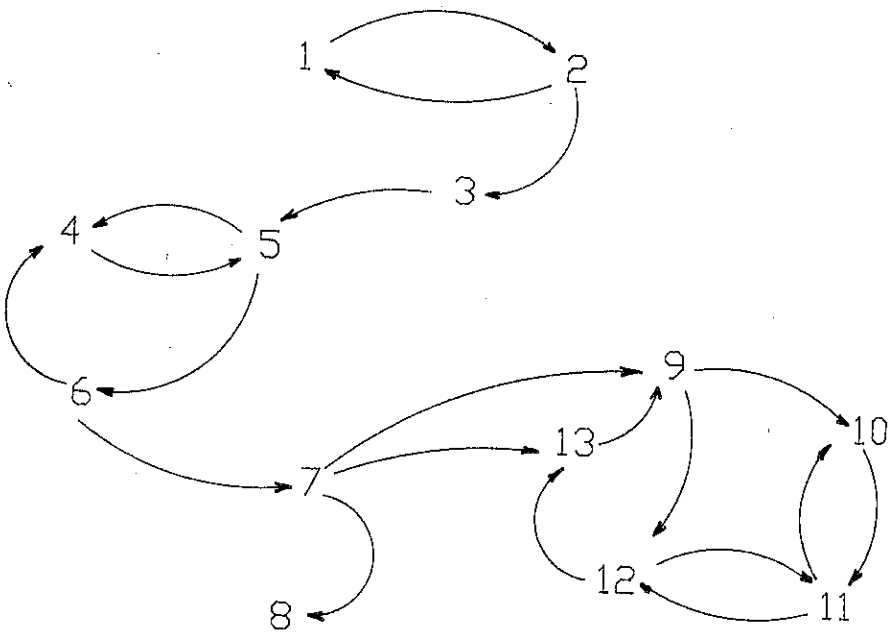


Figura 2.—Grafo de transiciones.

De acuerdo con el grafo de la figura 2 podemos clasificar los estados del proceso en los conjuntos:

$$T_1 = \{1, 2\}$$
$$T_2 = \{3\}$$
$$T_3 = \{4, 5, 6\}$$
$$T_4 = \{7\}$$
$$A = \{8\} = E_1$$
$$E_2 = \{9, 10, 11, 12, 13\}$$

Todos los conjuntos T son *transitorios* y se caracterizan por: *a*) todos sus estados deben poder ser alcanzados desde cualquier otro de ese conjunto en un número finito de pasos; *b*) existe al menos una ruta de escape que sitúa al proceso fuera de ese conjunto sin posibilidad alguna de retorno. Es importante darse cuenta al aplicar *a*) para identificar conjuntos que, si bien dos estados pueden no estar directamente conectados por dos flechas (4 y 6 ó 6 y 5 en T_3), basta que puedan comunicarse a través de estados intermedios ($4 \rightarrow 5 \rightarrow 6$; $6 \rightarrow 4 \rightarrow 5$).

Los conjuntos E_1 y E_2 *son ergódicos.* Sus características son: *a*) todos sus estados están doblemente conectados entre sí a través de un número finito de pasos; *b*) no existe ruta de escape que conduzca al proceso fuera de ese conjunto una vez se ingresa en él, desarrollándose dicho proceso dentro del conjunto ergódico. Un caso especialmente interesante es cuando un conjunto ergódico está constituido por un solo estado, como es el caso de E_1. En este caso el estado en cuestión actúa como una «trampa» para el proceso y recibe el nombre gráfico de *estado absorbente.*

Podemos imaginar todavía situaciones más complejas en cuanto a un mayor número de conjuntos E y demás, pero básicamente la discusión se reduciría, por lo que a los conceptos se refiere, a la presentada arriba. Como resultado general mencionaremos el siguiente teorema relativo a cadenas de Markov finitas (Turner, 1981): «Toda cadena de Markov finita posee al menos un conjunto ergódico». Se concluye así que la existencia de conjuntos transitorios no es necesaria para definir una cadena de Markov.

Clasificación avanzada de estados

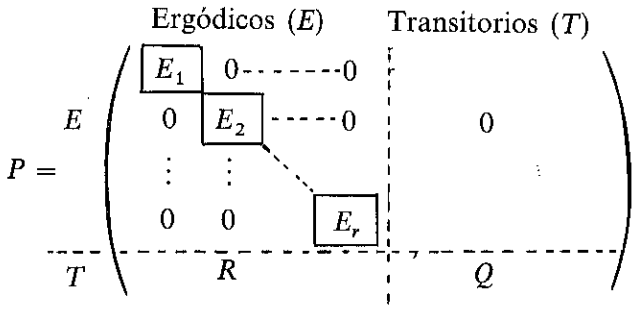
La clasificación de estados y conjuntos presentada (Turner, 1981) no es la más completa desde un punto de vista matemático. Tiene, no obstante, la gran ventaja de ser tremendamente operativa a la hora de realizar aplicaciones, reteniendo además una nomenclatura muy utilizada en problemas de interés fisicoquímico (Wood, 1975).

Clasificaciones más avanzadas (Rozanov, 1975; Vélez, 1977) dividen a los estados en *transitorios* y *recurrentes*. El paralelismo conceptual con los estados transitorios y ergódicos anteriores es completo. Los estados recurrentes se dividen a su vez en *recurrentes positivos* (*ergódicos* y *absorbentes*) y *recurrentes nulos* (*ergódicos*). En el primer caso, un estado j será recurrente positivo cuando el número medio de pasos necesarios para partiendo de j retornar a j sea finito. Por el contrario, j será recurrente nulo cuando tal número medio de pasos sea infinito. Esta última situación se puede presentar sólo en cadenas con espacio de estados infinito.

Dentro de esta nueva perspectiva, los estados de un espacio se agruparán en dos conjuntos: uno *transitorio* y otro *recurrente* (*Teorema de descomposición*). Ambos conjuntos son *disjuntos* pues no poseen elementos comunes. Cada uno de ellos puede a su vez estar formado por subconjuntos del mismo carácter. De hecho, para los estados recurrentes es posible definir una relación de equivalencia entre ellos en base a la exigencia de que (en un número finito de pasos) todos los estados que formen una clase de equivalencia estén conectados en los dos sentidos (análogo al conjunto ergódico anterior). Las clases de equivalencia son, entonces, cada uno de estos subconjuntos.

Por lo que respecta a los conjuntos de estados, éstos se clasifican en dos tipos no excluyentes uno del otro: *cerrados e irreducibles*. Se entiende por *cerrado* a un subconjunto C del espacio de estados U tal que todos sus estados $i \in C$ comunican exclusivamente con estados $j \in C$. No hay comunicaciones con estados del *complementario* $U - C$ de C. Por *irreducibles* se conocen a los subconjuntos C de U tales que dos a dos todos sus estados están comunicados en los dos sentidos ($i \leftrightarrows j$). Un subconjunto cerrado y, además, irreducible es lo que habíamos denominado anteriormente conjunto ergódico. Estos conjuntos son fundamentales en las aplicaciones prácticas ya que, de alguna manera que precisaremos más adelante, son una posible situación límite ($n \to \infty$) de un proceso estocástico (Ejercicio 4).

En lo que sigue nos atendremos a la clasificación más intuitiva de estados: ergódicos, absorbentes y transitorios. Utilizando la descomposición del espacio de estados U en subconjuntos transitorios T_i y ergódicos E_i, resulta muy conveniente para manejar cadenas de Markov finitas reordenar la matriz de transición a un paso P *en la forma canónica* siguiente:



[57]

Aparecen cuatro regiones *EE*, *ET*, *TE* y *TT*. La *EE* está dividida en cajas cuadradas, con ceros fuera de ellas, que dan las probabilidades de transición dentro de cada subconjunto ergódico. La *ET* está compuesta exclusivamente de ceros. La región *TE* es una matriz no cuadrada que reúne las probabilidades de paso desde estados transitorios a ergódicos. Finalmente *TT* es de nuevo una matriz cuadrada relativa a las probabilidades de paso entre estados transitorios.

6.3. Clasificación de cadenas

La clasificación más sencilla de las cadenas de Markov es la basada en la distinción entre aquellas que poseen estados transitorios y aquellas que no los poseen. Distinguiremos dos tipos elementales: a) *cadenas absorbentes*; b) *cadenas ergódicas.*

a) Una *cadena absorbente* está compuesta exclusivamente de estados transitorios y absorbentes.

b) Una *cadena ergódica* consta de un único conjunto ergódico y puede ser:

   b1) *cíclica*, si se entra en cada estado a intervalos periódicos (número de pasos) fijos;

   b2) *regular*, cuando es no cíclica, resultando ser la más importante en las aplicaciones.

La situación general será una combinación de a) y b). Su estudio se reduce al de estos casos más sencillos. Para clarificar ideas veamos algunos ejemplos.

Una *cadena absorbente* sería la representada, por ejemplo, en el grafo de la figura 3, siendo el conjunto de estados transitorios $T = \{1, 2, 3\}$ y los conjuntos ergódicos absorbentes $A_1 = \{4\}$, $A_2 = \{5\}$.

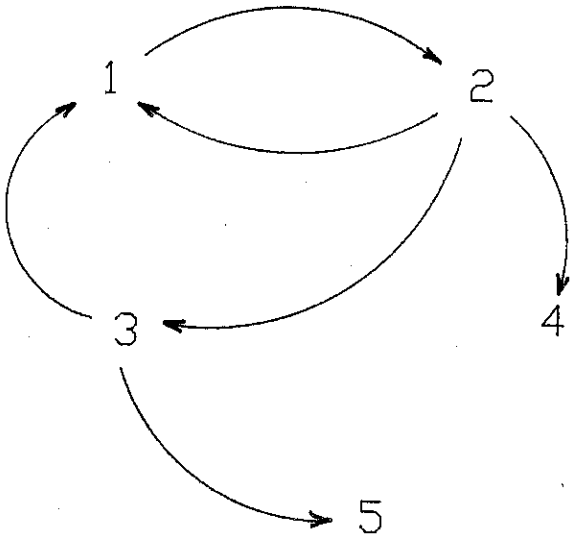


Figura 3.—Grafo de cadena absorbente.

Una *cadena ergódica cíclica* vendría esquematizada por la figura 4 con las probabilidades de transición que se indican. Su período es evidentemente 5, pues éste es el número de pasos para retornar a un estado de salida (siempre se da la misma ruta).

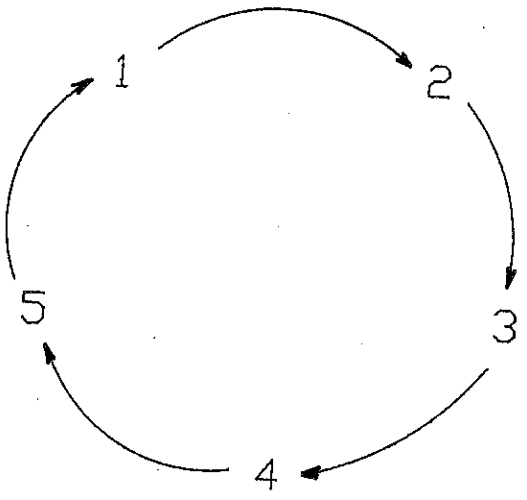


Figura 4.—Grafo de una cadena ergódica cíclica.

Para obtener una *cadena ergódica regular* basta con hacer distinta de cero una $p_{ij} = 0$ del ejemplo previo. Sea pues $p_{21} \neq 0$, lo que obliga a $p_{23} \neq 1$, de modo que $p_{23} + p_{21} = 1$ (Fig. 5). Esta cadena ya no es cíclica, dado que el proceso puede quedar retenido entre 1 y 2 un número n indeterminado de pasos. Efectivamente, para valores n grandes el proceso acabará saliendo por la ruta $2 \rightarrow 3$, pero la periodicidad (fija) queda rota.

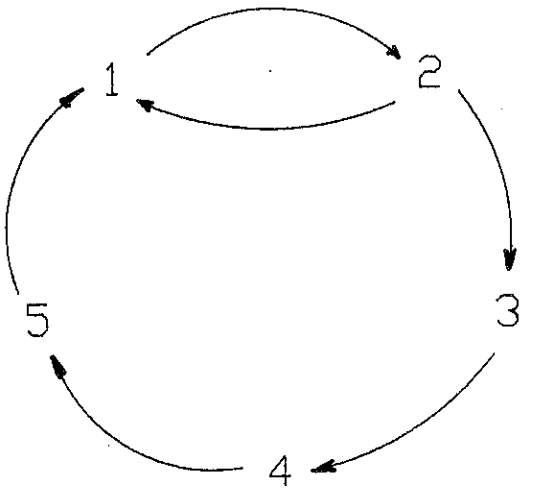


Figura 5

6.4. Ergodicidad y distribuciones estacionarias

En la materia discutida anteriormente hemos planteado en algunas ocasiones la situación límite a que conducirá la cadena cuando el número de pasos $n \rightarrow \infty$. Este problema de determinar la distribución (*vector de probabilidades*) límite posee la máxima importancia. En las aplicaciones prácticas tales distribuciones estacionarias sirven para simular el comportamiento del sistema estudiado en condiciones de interés (equilibrio termodinámico, etc.).

Centraremos el análisis de esta cuestión en las cadenas regulares. El teorema fundamental aquí (Vélez, 1977) establece que, independientemente del estado inicial $v^{(0)}$, el vector de probabilidades $v^{(n)}$ tiende a un vector constante Π cuando $n \rightarrow \infty$:

$$P = \text{regular} \Leftrightarrow \lim_{n \rightarrow \infty} v^{(n)} = \lim_{n \rightarrow \infty} [v^{(0)} P^n] =$$

$$= \Pi(p_1, p_2, ..., p_N) \quad ; \quad \text{para todo } v^{(0)} \qquad [58]$$

(recuérdese el ejercicio 3). Aplicando relaciones conocidas, podemos reescribir [58] como:

$$p_j = \lim_{n\to\infty} p_j^{(n)} = \lim_{n\to\infty} \sum_i p_i^{(n-1)} \cdot p_{ij} \quad ; \quad i,j = 1,2,3,... \qquad [59]$$

Tomando límites en [59] es claro que al existir Π:

$$p_j = \sum_i p_i \cdot p_{ij} \quad ; \quad i,j = 1,2,3,... \qquad [60]$$

ecuación que nos determina la distribución estacionaria, pues se trata de un sistema lineal en las incógnitas p_i con coeficientes dados por los elementos de la matriz a un paso P. En forma matricial [60] se expresa:

$$\Pi = \Pi P \qquad [61]$$

pudiendo calificar a Π como el *vector propio por la izquierda* (vector fila) asociado al autovalor $\lambda = 1$ de la matriz a un paso P.

Las consecuencias de formular el cálculo de Π en lenguaje matricial son de gran trascendencia.

i) El cálculo de Π se convierte en un asunto puramente algebraico, evitándose la complejidad inherente a [58]. Para abordar el problema a través de [58] deberían realizarse las siguientes operaciones:

   a) calcular $\lim_{n\to\infty} P^n$, para lo cual lo más apropiado sería expresar P con ayuda de la *descomposición canónica de Jordan* (Ralston, 1970):

$$P = RJR^{-1} \qquad [62]$$

   que es similar a una diagonalización, pero más compleja en el caso general. Si P fuera simétrica, [62] sería justamente su diagonalización. El límite a calcular quedaría entonces reducido al de J^n:

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} (RJR^{-1})^n = \lim_{n\to\infty} RJ^nR^{-1} \qquad [63]$$

   b) comprobar que *para todos* los vectores iniciales $v^{(0)}$, el resultado Π es siempre el mismo, lo que resulta una tarea desbordante.

ii) Las matrices estocásticas regulares tienen un autovalor (*único*) igual a la unidad y una *única* distribución estacionaria Π (Ejercicio 5).

iii) De existir Π, la matriz P^n se convierte al tomar límites en:

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \Pi \\ \Pi \\ \vdots \\ \Pi \end{pmatrix} = \begin{pmatrix} p_1 & p_2 & \dots & p_N \\ p_1 & p_2 & \dots & p_N \\ \dots & \dots & \dots & \dots \\ p_1 & p_2 & \dots & p_N \end{pmatrix} \qquad [64]$$

como se demuestra en el ejercicio 6.

iv) Los autovalores de P diferentes de la unidad determinan la velocidad de aproximación, cuando n crece, a la distribución Π. Todos estos autovalores (pueden ser números complejos) son de módulo menor que la unidad. Cuanto menores sean estos módulos tanto más rápida será la convergencia hacia Π.

v) La matriz P es de período T si posee T autovalores que sean las T raíces complejas de la unidad.

vi) Si la matriz P admite una *reordenación canónica*:

$$P = \left(\begin{array}{c|c} E & O \\ \hline TE & TT \end{array}\right)$$

y $TE \neq 0$ (=matriz rectangular nula), los autovalores de TT son todos de módulo *menor* que la unidad.

Los puntos iv)-v)-vi) pueden verse ilustrados en el ejercicio 6, y un criterio alternativo de convergencia a Π se presenta en el ejercicio 7.

6.5. Ejemplo de aplicación: Matrices biestocásticas y sistemas uniformes

Consideremos el ejemplo que suministra un gas formado por N_0 moléculas uniformemente distribuidas en un recipiente de volumen V_0 a

una cierta temperatura T. La única interacción posible entre sus moléculas asumiremos es por choque. Dividamos imaginariamente el recipiente en cuatro subvolúmenes iguales (Fig. 6). Como resultado de los choques una molécula puede, o no, cambiar de «caja» en una cierta unidad de

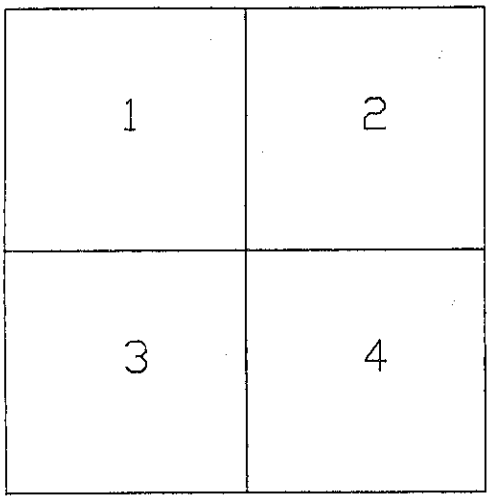


Figura 6.

tiempo. Para describir este proceso vamos a adoptar *a priori* la siguiente matriz de transición:

$$P = \begin{pmatrix} 1/3 & 1/4 & 1/6 & 1/4 \\ 1/4 & 1/3 & 1/4 & 1/6 \\ 1/6 & 1/4 & 1/3 & 1/4 \\ 1/4 & 1/6 & 1/4 & 1/3 \end{pmatrix}$$

La matriz P así dada será aceptable, en principio, si conduce a una distribución estacionaria uniforme para los cuatro estados (cajas).

Calculemos el vector estacionario Π:

$$\Pi P = \Pi \Rightarrow (p_1, p_2, p_3, p_4) \begin{pmatrix} 1/3 & 1/4 & 1/6 & 1/4 \\ 1/4 & 1/3 & 1/4 & 1/6 \\ 1/6 & 1/4 & 1/3 & 1/4 \\ 1/4 & 1/6 & 1/4 & 1/3 \end{pmatrix} = (p_1, p_2, p_3, p_4)$$

que tras las correspondientes manipulaciones resulta ser:

$$\Pi = (1/4, 1/4, 1/4, 1/4)$$

es decir una *distribución estacionaria uniforme* entre los cuatro estados. El número medio de moléculas dentro de cada subvolumen, como era de esperar, es sencillamente $\langle N \rangle = N_0/4 = N_0 \cdot p_i$. Concluimos entonces que P es, en principio, buena.

Si observamos la forma de P, vemos que posee una simetría subyacente que la convierte en una *matriz biestocástica*. Tanto sus filas como sus columnas están normalizadas a la unidad. El resultado final para Π ha sido una distribución uniforme y afirmamos ahora que esto será cierto para cualquier matriz biestocástica, sea ésta simétrica o no. Esto se expresa mediante el *teorema*: «*Una matriz es biestocástica si y sólo si su distribución estacionaria es uniforme.*» Probemos a continuación este teorema.

Deberemos para ello establecer la doble implicación:

$$P = \text{biestocástica} \Leftrightarrow \Pi = \text{uniforme} = (1/N, 1/N, ..., 1/N) \qquad [65]$$

Comencemos por la condición $\Leftarrow$: si Π es uniforme, entonces P es biestocástica:

$$\Pi P = \Pi \Rightarrow (1/N, 1/N, ..., 1/N) \cdot \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} =$$

$$= (1/N, 1/N, ..., 1/N) \qquad [66]$$

de donde para todas las columnas j se tiene:

$$\frac{1}{N} = \frac{1}{N} \sum_{i=1}^{N} p_{ij} \Rightarrow \sum_{i=1}^{N} p_{ij} = 1 \quad ; \quad j = 1, 2, ..., N \qquad [67]$$

que es la condición buscada.

Para demostrar la condición $\Rightarrow$: si P es biestocástica, entonces Π es uniforme, hay que reparar en un sencillo resultado. Este es que si P es biesto-

cástica, entonces sus sucesivas potencias P^n también lo son. La demostración es trivial y se deja a cargo del lector. Aplicando razonamientos ya conocidos:

$$\lim_{n\to\infty} P^n = A = \begin{pmatrix} p_1 & p_2 & \dots & p_N \\ p_1 & p_2 & \dots & p_N \\ \dots & \dots & \dots & \dots \\ p_1 & p_2 & \dots & p_N \end{pmatrix} = \text{biestocástica} \qquad [68]$$

y por lo tanto:

$$1 = Np_j \Rightarrow p_j = 1/N \quad ; \quad j = 1, 2, ..., N \qquad [69]$$

con lo que $\Pi = (1/N, 1/N, ..., 1/N)$, quedando completada la demostración.

El papel de las matrices biestocásticas en las aplicaciones es amplio, pues representan una buena primera aproximación para el estudio de sistemas de los que no se conocen en detalle todas sus características relevantes. Cabe mencionar como aplicación interesante la del estudio de la fase líquida por el método de Monte Carlo con potenciales de interacción de esferas rígidas (Sesé y Criado, 1990).

6.6. Cadenas absorbentes

En lo precedente nos hemos dedicado a determinar la distribución estacionaria para una cadena regular simple. Normalmente, estos entes matemáticos muestran los tres tipos de estados considerados, que a su vez se pueden agrupar en distintos conjuntos. Es un asunto de importancia pues dar respuesta a cuestiones tales como la del momento en que se alcanza un conjunto ergódico a partir de un cierto estado transitorio y con qué probabilidad, etc. Dado que un estado absorbente es una situación límite de un conjunto ergódico, estudiaremos las cadenas absorbentes. Visto a la recíproca, un conjunto ergódico es, en definitiva, un «macroestado absorbente» y la generalización de lo que se presentará a continuación será directa.

Una cadena absorbente es aquella que sólo posee estados transitorios

(t) y estados absorbentes (a). Puesta en *forma canónica* la matriz de transición a un paso P se esquematiza en:

$$P = \begin{array}{c|c} \overbrace{\quad}^{a} & \overbrace{\quad}^{t} \\ \left.\begin{array}{c|c} I & O \\ \hline R & T \end{array}\right. & \begin{array}{l} \}a \\ \}t \end{array} \end{array} \qquad [70]$$

siendo I = matriz unidad ($a \times a$); O = matriz nula ($a \times t$); R = matriz rectangular ($t \times a$); T = matriz ($t \times t$). Se define la *matriz fundamental* de la cadena como:

$$F = (I' - T)^{-1} \quad ; \quad I' = \text{matriz unidad } (t \times t) \qquad [71]$$

Daremos a continuación sin demostrar (Turner, 1981) las expresiones necesarias para calcular algunos parámetros de interés. En las fórmulas siguientes [72] y [76] se contabiliza la posición original.

i) La *media* y la *varianza* del número total de veces n_{t_i} que la cadena pasa por el estado transitorio t_i, sabiendo que se parte del transitorio t_j (podría ser $t_j = t_i$), vienen dados por los elementos ji de las matrices siguientes:

$$\langle n_{t_i} \rangle_{t_j} = F_{ji} \qquad [72]$$

$$\text{var}\,(n_{t_i})_{t_j} = (F(2F_d - I') - F_c)_{ji} \qquad [73]$$

en donde la matriz F_d es tal que:

$$(F_d)_{kl} = F_{kl} \cdot \delta_{kl} \qquad [74]$$

y la matriz F_c:

$$(F_c)_{kl} = F_{kl}^2 \qquad [75]$$

ii) La *media* y la *varianza* del número total n_t de veces que la cadena, partiendo del transitorio t_i, evoluciona entre los estados transitorios antes de caer en un estado absorbente, son los elementos i de las matrices columna:

$$\langle n_t \rangle_{t_i} = (Fu)_i = \tau_i \qquad [76]$$

$$\text{var}\,(n_t)_{t_i} = [(2F - I')\tau - \tau_c]_i \qquad [77]$$

en donde u es un vector columna $(t \times 1)$ compuesto por unos, τ es un vector columna $(t \times 1)$ tal que $\tau = Fu$, y τ_c es también un vector columna tal que:

$$(\tau_c)_i = \tau_i^2 \quad ; \quad i = 1, 2, ..., t \qquad [78]$$

iii) La probabilidad de que partiendo de un estado transitorio t_i la cadena quede atrapada en el estado absorbente a_k se calcula mediante el elemento ik de la matriz:

$$p(t_i \to a_k) = (F \cdot R)_{ik} \qquad [79]$$

En el ejercicio 8 se puede encontrar una realización práctica de estos cálculos.

7. EJEMPLO DE APLICACION: SIMULACION MONTE CARLO DE LIQUIDOS

El estudio de la materia en fase condensada ocupa un lugar importante en la investigación en Física y Química. Centrándonos en el estudio de la fase líquida, la descripción de las partículas que la componen en términos de Mecánica Clásica, junto con las pertinentes leyes estadísticas, derivadas del gran número de partículas involucradas ($N \sim 10^{23}$), ha llevado a una amplia comprensión de buena parte de las propiedades de esta fase. La Mecánica Estadística Clásica no es, por supuesto, aplicable en situaciones en las que los *efectos cuánticos* juegan un papel importante, como en bajas temperaturas, sistemas con partículas ligeras (He, Ne, etc.). No obstante, para nuestros fines aquí, el nivel de aproximación clásico

será suficiente, de modo que en lo siguiente *tomaremos los efectos cuánticos, siempre presentes, como despreciables.*

Dentro del marco conceptual reseñado describimos a un sistema a través del concepto de *colectivo* (Gibbs, 1902; Balescu, 1975). Un *colectivo* es una colección muy grande de sistemas idénticos a nivel macroscópico (un observador humano no podría distinguir uno de otro), pero diferentes microscópicamente (el «*diablillo de Maxwell*» sí sería capaz de distinguirlos). Estos colectivos quedan definidos por su correspondiente función de distribución (densidad) en el *espacio fásico* asociado al sistema. Entendemos por *espacio fásico*, al espacio (ejes ortogonales) definido por todas las coordenadas e impulsos de las N partículas del sistema. Resulta así un espacio cartesiano $6N$-dimensional:

$$(\mathbf{q}^N, \mathbf{p}^N) = (\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_N;\ \mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N) =$$
$$= (q_{1_x}, q_{1_y}, q_{1_z}, ..., q_{N_x}, q_{N_y}, q_{N_z};\ p_{1_x}, p_{1_y}, p_{1_z}, ..., p_{N_x}, p_{N_y}, p_{N_z}) \qquad [80]$$

Entre las diversas posibilidades para describir sistemas por medio de colectivos vamos a elegir al *colectivo canónico* que, desde un punto de vista matemático, es el más cómodo. Este colectivo se representa como (N, V, T) y simboliza a una colección muy grande de sistemas *cerrados e idénticos* (macroscópicamente) en *equilibrio térmico* con su entorno. Cada uno de estos sistemas tiene, pues, fijos los valores N = número de partículas, V = volumen, T = temperatura. Su función de distribución involucra al factor de Boltzmann $\exp[-E/kT]$, y viene dada por la expresión:

$$f(\mathbf{q}^N, \mathbf{p}^N) = \frac{\exp[-E(\mathbf{q}^N, \mathbf{p}^N)/kT]}{\int\cdots\int \exp[-E(\mathbf{q}^N, \mathbf{p}^N)/kT] d\mathbf{q}^N \cdot d\mathbf{p}^N} \qquad [81]$$

en donde E es la energía interna del sistema, y las integraciones cubren todo el espacio fásico accesible. Como de costumbre, el sentido físico de $f(\mathbf{q}^N, \mathbf{p}^N)$ es:

$f(\mathbf{q}^N, \mathbf{p}^N) d\mathbf{q}^N \cdot d\mathbf{p}^N$ = probabilidad de que el sistema tenga a su partícula i $(i = 1, 2, ..., N)$ con posición entre $\mathbf{q}_i$ y $\mathbf{q}_i + d\mathbf{q}_i$, y con momento entre $\mathbf{p}_i$ y $\mathbf{p}_i + d\mathbf{p}_i$ [82]

Es interesante señalar que fijar T con toda precisión implica *necesariamente* que E no poseerá un valor perfectamente definido en el

equilibrio, sino que *E fluctuará* alrededor de un valor medio $\langle E \rangle$. Esto es una consecuencia de la teoría de fluctuaciones que da la relación de indeterminación (Sesé y Criado, 1990):

$$\Delta(1/T) \cdot \Delta E \geqslant k \qquad [83]$$

cuya similitud con las de Heisenberg es sorprendente y de ninguna manera fruto de la casualidad como se sabe hoy por la *teoría de los flujos-K* (Primas, 1981).

Antes de abordar el estudio del sistema fluido, recordemos algunos hechos fundamentales. Es conocido por Termodinámica que un sistema en equilibrio a T y V constantes posee un *mínimo* en su energía *Helmholtz A*. Para ser más precisos diremos que es el valor medio de la energía *Helmholtz* quien es un mínimo, y lo mismo entenderemos para el resto de las variables que surjan entrecomillando los calificativos necesarios. Sabemos además que $A = E - TS$ (S = entropía), por lo que el «mínimo» en A es un balance entre el «mínimo» en E y el «máximo» en S compatibles con el estado de equilibrio del sistema. La determinación del «mínimo» de A es un problema arduo, ya que ésta es una propiedad que involucra al *espacio fásico* en su totalidad. La relación entre A y la función de partición canónica $Z(N, V, T)$ es:

$$A = -kT \ln Z(N, V, T) \qquad [84]$$

En ausencia de campos externos, Z se factoriza en el producto de una parte *translacional*, constante a T constante, y una parte *configuracional* que depende, entre otros factores, de las interacciones entre partículas. Formalmente pondremos:

$$E(\mathbf{q}^N, \mathbf{p}^N) = E_c(\mathbf{p}^N) + U(\mathbf{q}^N) \qquad [85]$$

$$Z(N, V, T) = \text{cte} \int \cdots \int \exp\left[-E(\mathbf{q}^N, \mathbf{p}^N)/kT\right] d\mathbf{q}^N \cdot d\mathbf{p}^N \qquad [86]$$

y, por tanto, la *energía Helmholtz* resulta:

$$A = -kT \ln Z_{\text{trans.}} - kT \ln Q(N, V, T) \qquad [87]$$

siendo $Q(N, V, T)$ la *función de partición configuracional:*

$$Q(N, V, T) = \int \cdots \int \exp\left[-U(\mathbf{q}^N)/kT\right] d\mathbf{q}^N \qquad [88]$$

y la consecución de una A «mínima» se logrará cuando $Q(N, V, T)$ sea máxima.

A pesar de esta reducción, el problema dista de ser trivial, pues Q no puede calcularse analíticamente, ni siquiera en casos idealmente simples como el de esferas rígidas (Feynman, 1972).

Sin embargo, cuando se está interesado en la determinación de propiedades mecánicas, como la energía o la presión, una solución aproximada bastante eficaz consiste en hacer un muestreo del espacio configuracional en busca de la región de bajas energías configuracionales $U(\mathbf{q}^N)$, pues su contribución a la integral será *más importante* que la de aquellas regiones en las que $U(\mathbf{q}^N)$ sea alta. Localizada esta región (zona sombreada en la Figura 8a), un muestreo estadístico sólo sobre ella de las propiedades mecánicas arrojará estimaciones para las propiedades globales del fluido. Este tipo de proceso, insistimos, no podría aplicarse para determinar propiedades globales del espacio fásico como son las magnitudes termomecánicas (A, S, etc.), pues todas las regiones contribuyen de modo no despreciable a la propiedad.

Estudiemos con un poco de detalle cómo realizar la operación de búsqueda descrita.

El primer paso será calcular la *función de distribución marginal* asociada al espacio configuracional $\mathbf{q}^N$. Integrando sobre los momentos se obtiene:

$$f(\mathbf{q}^N) = \int \cdots \int f(\mathbf{q}^N, \mathbf{p}^N)\, d\mathbf{p}^N = \frac{\exp\left[-U(\mathbf{q}^N)/kT\right]}{\int \cdots \int \exp\left[-U(\mathbf{q}^N)/kT\right] d\mathbf{q}^N} \qquad [89]$$

Como aproximación razonable para evaluar $U(\mathbf{q}^N)$ tomaremos la llamada *aproximación par-aditiva*, que construye la energía potencial de interacción total como la suma de las interacciones entre todas las

parejas de partículas (moléculas). Si estudiamos un líquido compuesto por partículas esféricas idénticas (líquido monoatómico), escribiremos:

$$U(\mathbf{q}^N) = \sum_{i<j} u(r_{ij}) \qquad [90]$$

siendo $u(r_{ij})$ el potencial par (esferas rígidas, Lennard-Jonnes, etc.) entre las partículas i y j. Por la simetría de la interacción u sólo dependerá de la distancia entre las partículas. En lo que sigue nos fijaremos en estos sistemas.

El conocimiento de potenciales pares u es, para sistemas sencillos, muy completo (McDonald y Singer, 1972). Esto no sucede cuando las partículas son moléculas con un número de átomos grande o con interacciones fuertes (puentes de hidrógeno). Para líquidos monoatómicos el potencial de Lennard-Jones es muy empleado:

$$u(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \qquad [91]$$

y como se ve consta de una parte repulsiva (potencia 12) que domina a distancias r_{ij} cortas, y otra parte atractiva (potencia 6) que lo hace a r_{ij} largas.

Para determinar *la región de muestreo* sombreada en la Figura 8a hay que sortear todavía un buen número de dificultades. Podemos mencionar: la determinación del tamaño de la muestra con la que trabajar, de modo que con un número pequeño de partículas ($N_s \sim 10^2$, 10^3) se obtengan buenas estimaciones de las propiedades del sistema global; los diferentes esquemas de muestreo con *condiciones de contorno periódicas;* etc. No es éste el lugar adecuado para discutirlas en detalle (Sesé y Criado, 1990), por lo que pasaremos sobre ellas brevemente.

Simular el comportamiento de un sistema con $N \sim 10^{23}$ partículas es una tarea imposible, incluso, con los ordenadores más potentes. Se impone una elección de una *muestra* reducida (N_s) que adecuadamente tratada conduzca a estimaciones fiables para las magnitudes termodinámicas del líquido. La muestra se encierra en una caja (cúbica) de dimensiones acordes a la densidad del líquido $n = N/V = N_s/V_s$. Esta caja (Fig. 7, 8b) se reproduce a lo largo, ancho y alto del espacio un número infinito de veces. Todo lo que suceda en la caja central, sucederá automáticamente en todas las imágenes periódicas (condiciones de contorno periódicas).

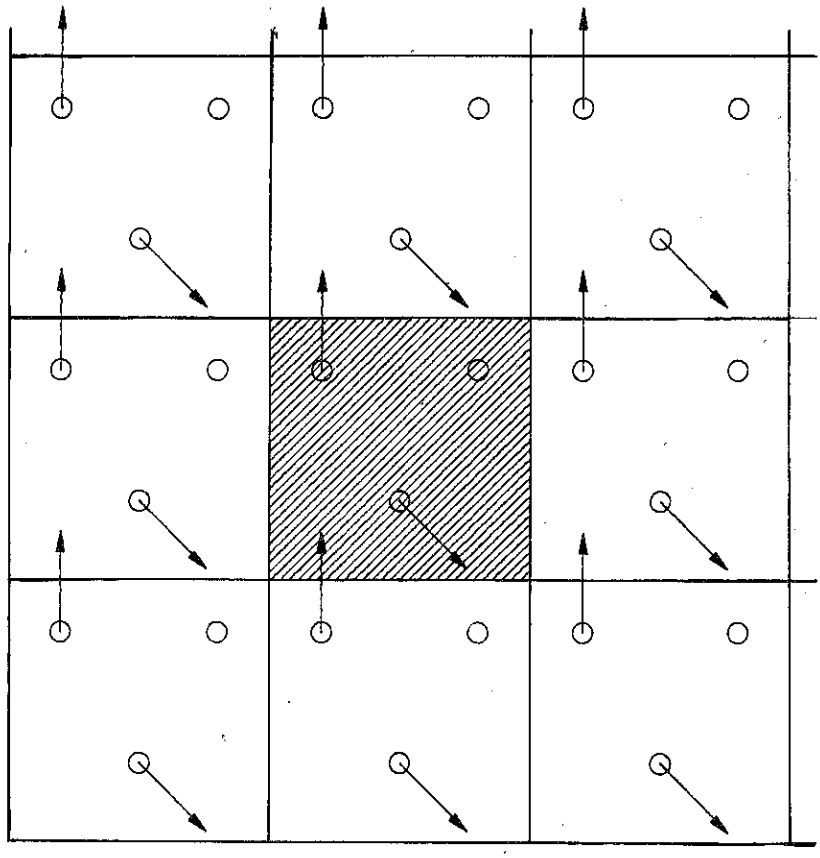


Figura 7.—Representación esquemática de las condiciones de contorno periódicas.

Las partículas pueden abandonar su caja para pasar a otra vecina, y pueden interaccionar con las de cajas distintas de la propia.

El método más eficaz para localizar la región de muestreo es el denominado de *Metropolis* (Metropolis y col., 1953), que está basado en la construcción de una *cadena de Markov* que busque tal región en el espacio $\mathbf{q}^{N_s}$ de acuerdo con la ley [89]. El espacio configuracional restringido, asociado a la caja central, se discretiza en celdas sobre las cuales se realiza la búsqueda aleatoria (Fig. 8a). Utilizando la caja central como base y repitiendo periódicamente lo que acontezca en ella, se pretende que a partir de un estado inicial *transitorio* (fuera de la región rayada) la cadena evolucione hacia el *conjunto ergódico* (región rayada) que simulará el estado de equilibrio macroscópico. Esta situación de estabilización se reconocerá por las fluctuaciones en U.

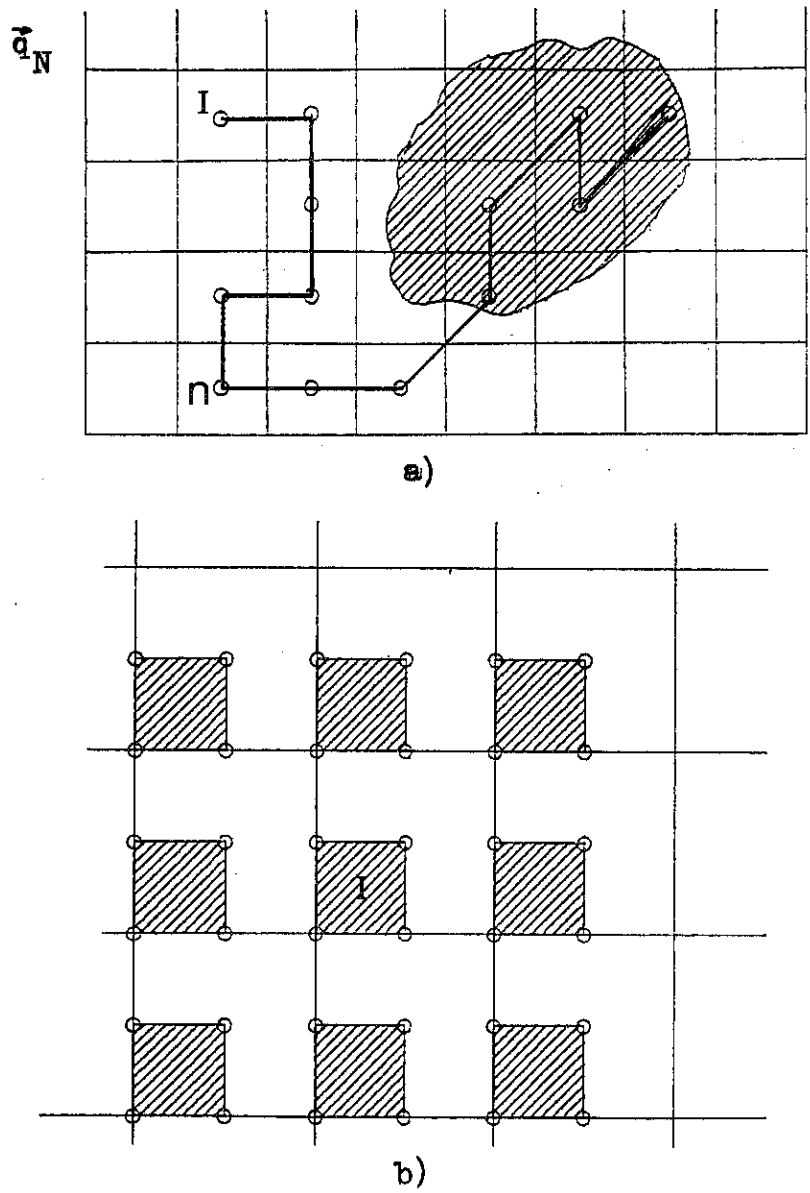


Figura 8.—a) Cadena de Markov en el espacio configuracional. b) Espacio físico, estado inicial de la cadena (1).

Alcanzado el conjunto ergódico, promediaremos sobre él las magnitudes termodinámicas que nos interesen. Veamos cómo construir una cadena de Markov sencilla que nos lleve a la región ergódica. Por simplicidad en la notación haremos $N_s = N$.

i) Estado inicial de la cadena. La configuración inicial $\mathbf{q}^N(0)$ suele tomarse como aquella disposición espacial de las N_s partículas en la caja central (y en sus imágenes) dada por la red del sólido correspondiente.

ii) Se elige al azar, uniformemente, una partícula j entre las N con ayuda de un número aleatorio ξ entre 0 y 1, convenientemente transformado al intervalo $(1, N)$.

iii) Se generan tres números aleatorios independientes ξ_1, ξ_2, ξ_3 dentro del intervalo (0, 1). Con ellos se calcula la posición de prueba de j en el paso n $(=1)$:

$$x_{ji} = x_{ji}(n-1) + (2\xi_i - 1)\delta \quad ; \quad i = 1,2,3 \qquad [92]$$

donde δ es un número real *a optimizar* durante la simulación.

iv) Se calcula la energía configuracional de este nuevo estado $U(\mathbf{r}^N(n))$ y se compara con la del estado $n-1$:

$$\Delta U = U(\mathbf{r}^N(n)) - U(\mathbf{q}^N(n-1)) \qquad [93]$$

Si la diferencia $\Delta U \leqslant 0$, se admite el estado $\mathbf{r}^N(n)$ como siguiente de la cadena, $\mathbf{q}^N(n) = \mathbf{r}^N(n)$. Si por el contrario $\Delta U > 0$, se genera uniformemente un nuevo ξ en (0, 1) y:

a) si $\xi \geqslant \exp(-\Delta U/kT)$, $\mathbf{r}^N(n)$ se rechaza y se hace $\mathbf{q}^N(n) = \mathbf{q}^N(n-1)$;

b) si $\xi < \exp(-\Delta U/kT)$, $\mathbf{r}^N(n)$ se acepta como siguiente estado $\mathbf{q}^N(n) = \mathbf{r}^N(n)$.

v) Se repite el proceso a partir de $\mathbf{q}^N(n)$ y empezando en ii).

Se obtiene así una cadena de Markov:

$$\mathbf{q}^N(0), \mathbf{q}^N(1), ..., \mathbf{q}^N(n), ... \qquad [94]$$

que es ergódica, aperiódica y acorde a $f(\mathbf{q}^N)$ (Metropolis y col., 1953; Balescu, 1975). El parámetro δ está directamente relacionado con el número de pasos necesarios para alcanzar la distribución estacionaria Π, y además con la calidad del muestreo. Su valor se ajusta a lo largo de la simulación de tal manera que la tasa en aceptación de configuraciones

sea de un 50 %, aproximadamente. De esta manera se consigue que los estados (configuraciones) tengan probabilidades a un paso $p_{ii} \neq 0$.

El esquema propuesto constará de dos etapas: alcanzar la región ergódica («equilibrio») y realizar el muestreo de las magnitudes mecánicas del sistema (las propiedades termomecánicas como la entropía caen fuera de estas evaluaciones) (Wood, 1975; Valleau y Whittington, 1976). Las fluctuaciones en U indican que se ha alcanzado el «equilibrio», por lo que se descartan las configuraciones iniciales (*estados transitorios*) y se continúa con el esquema ii)-v) dentro del *conjunto ergódico*, evaluando las propiedades que interesen (energía interna, estructura, presión, etc.). Para evitar correlaciones entre estados se utilizan estimaciones por «bloques», que dan los valores medios buscados para *las magnitudes intensivas* (las realmente significativas) del sistema modelo y, se espera, del sistema real.

8. PROCESOS DE MARKOV DE ORDEN SUPERIOR

Evidentemente, la clase de procesos de Markov es una entre una gran variedad de diferentes procesos estocásticos estacionarios. Si en un proceso de Markov relajáramos la restricción de memoria a un paso, obtendríamos un tipo de proceso más general. Podemos definir así un proceso estocástico $Y(t)$ de modo que la probabilidad de que en $n + 1$ se produzca un cierto resultado dependa de lo sucedido en los pasos previos n y $n - 1$:

$$P_{n+1}(y_{n+1}, t_{n+1}|y_0, t_0;\ y_1, t_1; \ldots; y_{n-1}, t_{n-1}; y_n, t_n) =$$
$$= P(y_{n+1}, t_{n+1}|y_{n-1}, t_{n-1}; y_n, t_n) = p_3(y_{n+1}, t_{n+1}|y_{n-1}, t_{n-1}; y_n, t_n) \quad [95]$$

guardándose memoria de dos pasos.

Este proceso no es de Markov propiamente, pero tal carácter puede ser recuperado *redefiniendo* $Y(t)$ *como un proceso bicomponente* (Y_1, Y_2). En él, Y_1 es la situación del proceso en n e Y_2 la situación del proceso en $n - 1$. Con la introducción de una variable adicional se consigue restaurar el carácter Markov del proceso [95] y se le denomina *proceso de Markov de segundo orden.* Análogamente se definirán procesos de orden superior.

Ejercicio de aplicación

Como ilustración consideremos el problema del *camino aleatorio con persistencia* (Kampen, 1985). Una partícula puede moverse a saltos de longitud unidad sobre el eje real. Supongamos que después de un paso a la derecha la probabilidad de ir de nuevo a la derecha es p y la de ir a la izquierda es $q = 1 - p$. Por otra parte, después de un paso a la izquierda la probabilidad de repetir a la izquierda es p y la de volver a la derecha $q = 1 - p$.

Este proceso no es de Markov, pero redefiniéndolo como $(Y_1, Y_2) =$ $=$ (posición de la partícula en t_n, posición de la partícula en t_{n-1}) tenemos un proceso de Markov de segundo orden. El punto de interés es determinar la matriz de transición asociada. Notemos previamente que el espacio de estados es infinito ..., -2, -1, 0, 1, 2, ..., y que la forma de la matriz de transición a un paso para el camino sin persistencia (de Markov) sería de la forma:

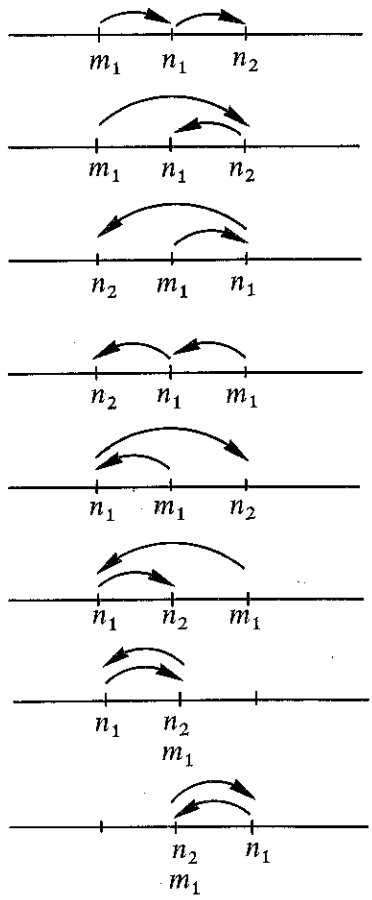
$$P = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & q & 0 & p & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & q & 0 & p & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & q & 0 & p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

si la probabilidad de ir desde un punto a la izquierda es q y la de ir a la derecha $p = 1 - q$. En el proceso bicomponente la forma de la matriz de transición es algo más complicada y nos serviremos de una serie de diagramas para obtenerla.

Comencemos escribiendo:

$$\begin{aligned} &P(Y(t_{n+1})|Y(t_n), Y(t_{n-1})) = \\ &= P(Y(t_{n+1});\ Y(t_n)|Y(t_n);\ Y(t_{n-1})) = \\ &= p(n_2, m_2|n_1, m_1) \end{aligned} \qquad [96]$$

siendo necesario que $n_1 = m_2$, lo que introducirá un factor $\delta_{n_1 m_2}$ al calcular las probabilidades. Para el cálculo de esta probabilidad aparecen ocho situaciones posibles que esquematizamos a continuación junto con sus probabilidades asociadas:



$$p \cdot \delta_{n_2, n_1+1} \cdot \delta_{n_1, m_1+1} \cdot \delta_{m_2 n_1}$$

$$0$$

$$0$$

$$p \cdot \delta_{m_2 n_1} \cdot \delta_{n_2, n_1-1} \cdot \delta_{n_1, m_1-1}$$

$$0$$

$$0$$

$$q \cdot \delta_{m_2 n_1} \cdot \delta_{n_2, n_1+1} \cdot \delta_{n_1, m_1-1}$$

$$q \cdot \delta_{m_2 n_1} \cdot \delta_{n_2, n_1-1} \cdot \delta_{n_1, m_1+1}$$

Como los ocho sucesos son independientes, la probabilidad de transición que buscamos será la suma de todas estas contribuciones obteniéndose la expresión final:

$$p(n_2, m_2 | n_1, m_1) = \delta_{m_2 n_1} [\delta_{n_2, n_1+1} \{p \cdot \delta_{n_1, m_1+1} + q \cdot \delta_{n_1, m_1-1}\} + \\ + \delta_{n_2, n_1-1} \{q \cdot \delta_{n_1, m_1+1} + p \cdot \delta_{n_1, m_1-1}\}] \qquad [97]$$

Bibliografía

1. BALESCU, R., *Equilibrium and Nonequilibrium Statistical Mechanics*, Wiley, Nueva York, 1975, capítulos 2, 4 y 21.
2. FEYNMAN, R. P., *Statistical Mechanics*, Benjamin, Reading, Massachusetts, 1972, capítulo 4.
3. GIBBS, J. W., *Elementary Principles of Statistical Mechanics*, Yale University Press, 1902.
4. KAMPEN, N. G. van, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 1985, capítulos 3 y 4.
5. MCDONALD, I. R. y SINGER, K., *Mol. Phys.*, 23, 29 (1972).
6. MCQUARRIE, D. A., en *Physical Chemistry an Advanced Treatise*, vol. XI B, «Probability theory and stochastic processes», editores H. Eyring, D. Henderson y W. Jost, Academic Press, Nueva York, 1975.
7. MCQUARRIE, D. A., *Statistical Mechanics*, Harper & Row, Nueva York, 1976, capítulos 21 y 22.
8. METROPOLIS, N. A., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., y TELLER, E., *J. Chem. Phys.*, 21, 1087 (1953).
9. PRIMAS, H., *Chemistry, Quantum Mechanics and Reductionism*, Springer-Verlag, Berlín, 1981, capítulo 2.
10. RALSTON, A., *Introducción al Análisis Numérico*, Limusa-Wiley, México, 1970, capítulo 10.
11. ROZANOV, Y., *Processus Aléatoires*, Mir, Moscú, 1975, capítulo 3.
12. SESÉ, L. M. y CRIADO, M., *Termodinámica Química Molecular*, Universidad Nacional de Educación a Distancia, Madrid, 1990, temas 5 y 9.
13. TURNER, J. C., *Matemática Moderna Aplicada*, Alianza, Madrid, 1981.
14. UHLENBECK, G. E. y ORNSTEIN, L. S., *Phys. Rev.*, 36, 823 (1930).

15. VALLEAU, J. P. y WHITTINGTON, S. G., en *Modern Theoretical Chemistry*, vol. 5, «A guide to Monte Carlo for statistical mechanics (1. Highways)», editor B. J. Berne, Plenum, Nueva York, 1976.

16. VÉLEZ, R., *Procesos Estocásticos*, Universidad Nacional de Educación a *Distancia*, *Madrid*, 1977, *Unidad Didáctica* 1.

EJERCICIOS DE AUTOCOMPROBACION

1. Comprobar la igualdad [8].

2. Demostrar que si una matriz cuadrada P es estocástica, sus sucesivas potencias también lo son.

3. Una partícula puede moverse mediante desplazamientos aleatorios entre los cuatro vértices (1, 2, 3, 4) de un cuadrado de acuerdo con la matriz de transición con un paso P:

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 1/4 & 1/2 \end{pmatrix}$$

   a) Dibujar un diagrama que represente a P.

   b) ¿Es P biestocástica?

   c) Si la partícula parte del vértice 1, ¿cuál es la probabilidad de que alcance el 3 en tres pasos?

   d) Calcular $p_1^{(10^6)}$.

e) A la vista de la forma que toma P, ¿es posible anticipar el vector de probabilidades $v^{(n)}$ cuando $n \to \infty$?

4. Representar a través de sendos grafos un ejemplo de subconjunto cerrado y otro de irreducible. ¿Los conjuntos cerrados son ergódicos? ¿Los subconjuntos irreducibles pueden ser transitorios?

5. Dada una matriz de transición regular P, para un espacio finito, demostrar que en el límite $n \to \infty$ P^n tiende a una matriz cuadrada A cuyas filas son todas idénticas al vector estacionario $\Pi = (p_1, p_2, ...)$. Demostrar también que Π es único.

6. Para las dos matrices a un paso siguientes:

$$\text{a)}\quad P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{b)}\quad P = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$$

identificar el tipo de cadena, determinar la distribución estacionaria cuando sea posible, y calcular sus autovalores y la forma límite de P^n cuando $n \to \infty$.

7. Sea una cadena de Markov formada por estados recurrentes. Se denomina *coeficiente de ergodicidad* $k(n_0)$ de la cadena en el paso n_0 a la cantidad positiva (Rozanov, 1975):

$$k(n_0) = 1 - \frac{1}{2} \sup_{i,j} \sum_{m} \left| p_{im}(n_0) - p_{jm}(n_0) \right|$$

donde sup simboliza el supremo de las sumas expresadas, haciendo variar i, j, m sobre todos los estados y empleando la matriz P^{n_0}. Si para un cierto n_0 el coeficiente $k(n_0)$ *es positivo*, existe distribución estacionaria $\Pi = (p_1, p_2, ..., p_m, ...)$. La velocidad de convergencia con el número de pasos n a esta distribución límite puede estimarse con la relación:

$$\sup_{p_i^0} \left| p_j(n) - p_j \right| \leqslant \frac{1}{1 - k(n_0)} \cdot \exp\left[-\frac{n}{n_0} \ln \frac{1}{1 - k(n_0)} \right]$$

en donde se da una cota superior para el caso más desfavorable de proximidad entre el vector de probabilidades en el paso n y la distribución Π. Esta cota es independiente del vector de probabilidad inicial $\{p_i^0\}$.

Dadas las matrices de transición a un paso:

$$a)\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix} \qquad b)\begin{pmatrix}0 & 1/2 & 1/2\\ 1/2 & 0 & 1/2\\ 1/2 & 1/2 & 0\end{pmatrix} \qquad c)\begin{pmatrix}1/2 & 1/4 & 1/4\\ 1/4 & 1/2 & 1/4\\ 1/4 & 1/4 & 1/2\end{pmatrix}$$

estudiar si poseerán distribución estacionaria y determinar cuál convergerá más rápidamente utilizando los conceptos anteriores.

8. Una partícula se mueve entre los cinco vértices de un pentágono de acuerdo con la siguiente matriz de transición a un paso:

$$P=\begin{pmatrix}0 & 1/3 & 0 & 1/3 & 1/3\\ 0 & 1/2 & 1/2 & 0 & 0\\ 0 & 3/4 & 1/4 & 0 & 0\\ 1/2 & 0 & 1/2 & 0 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}$$

a) Reescribir P en forma canónica, identificando los estados.

b) Calcular la media y la varianza del número de veces que la partícula pasa por los estados 1 y 4, partiendo en ambos casos de 1 y 4.

c) Calcular la media y la varianza del número de veces que la partícula evoluciona entre los estados transitorios, partiendo de cualquiera de ellos, antes de pasar a los conjuntos absorbentes.

d) Calcular la probabilidad de que partiendo de 1 se llegue a cualquiera de los dos conjuntos absorbentes. Lo mismo partiendo de 4.

SOLUCIONES A LOS EJERCICIOS DE AUTOCOMPROBACION

1. Un cálculo directo nos lleva a:

$$\int_{-\infty}^{\infty} y w_1(y,t)\,dy = \int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} \delta(y - Y_x(t)) f(x)\,dx\right] dy =$$

$$= \int_{-\infty}^{\infty} f(x)\left[\int_{-\infty}^{\infty} y\delta(y - Y_x(t))\,dy\right] dx = \int_{-\infty}^{\infty} Y_x(t)\cdot f(x)\cdot dx$$

que es la relación pedida.

2. Recordemos que P es estocástica cuando todas sus filas están normalizadas a la unidad:

$$\sum_j p_{ij} = 1 \quad ; \quad p_{ij} \geqslant 0 \quad ; \quad i,j = 1,2,3,...$$

La matriz de transición a dos pasos $P^{(2)} = P^2$ tiene sus elementos $p_{ij}^{(2)}$:

$$P^{(2)} = P \times P$$

$$p_{ij}^{(2)} = \sum_k p_{ik}\cdot p_{kj} \quad ; \quad i,j,k = 1,2,3,...$$

por lo que si sumamos sobre columnas j:

$$\sum_j p_{ij}^{(2)} = \sum_j \sum_k p_{ik} \cdot p_{kj} = \sum_k \left(\sum_j p_{kj} \right) \cdot p_{ik} =$$

$$= \sum_k 1 \cdot p_{ik} = 1 \quad ; \quad i = 1, 2, 3, \ldots$$

quedando demostrado que $P^{(2)}$ es también estocástica.

Supongamos comprobado que $P^{(n)}$ es estocástica. La demostración pedida quedará completada por *inducción* si probamos que $P^{(n+1)}$ es estocástica. Para ello escribamos:

$$P^{n+1} = P^n \cdot P$$

cuyos elementos $p_{ij}^{(n+1)}$ son:

$$p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} \cdot p_{kj} \quad ; \quad i, j, k = 1, 2, 3, \ldots$$

y sumando sobre columnas:

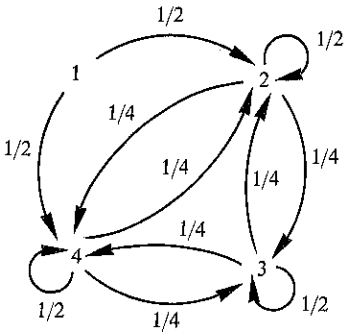
$$\sum_j p_{ij}^{(n+1)} = \sum_j \sum_k p_{ik}^{(n)} \cdot p_{kj} =$$

$$= \sum_k \left(\sum_j p_{kj} \right) \cdot p_{ik}^{(n)} = 1 \quad ; \quad i = 1, 2, 3, \ldots$$

dado que $P^{(n)}$ y P son estocásticas, Queda así probado que $P^{(n+1)}$ es estocástica y, por tanto, la propiedad pedida.

Notemos que si las potencias P^n no fuesen matrices estocásticas la receta [56] suministraría vectores de probabilidad $v^{(n)}$ no normalizados, no siendo las P^n útiles para describir transiciones a n-pasos.

3. a) Resulta muy conveniente en este tipo de problemas dibujar un diagrama como el siguiente:



b) No. Unicamente es estocástica, pues sus filas están normalizadas a la unidad y, en cambio, sus columnas no lo están.

c) Si se parte del vértice 1, el vector de probabilidad inicial es:

$$v^{(0)} = (1, 0, 0, 0)$$

La probabilidad de llegar al vértice 3 en n pasos es la tercera componente $p_3^{(n)}$ del vector $v^{(n)}$. Para $n = 3$:

$$v^{(3)} = (p_1^{(3)},\ p_2^{(3)},\ p_3^{(3)},\ p_4^{(3)})$$

y la probabilidad buscada es $p_3^{(3)}$. Efectuaremos recursivamente el cálculo aplicando $v^{(n)} = v^{(n-1)}P$:

$$v^{(1)} = (1,0,0,0)\begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 1/4 & 1/2 \end{pmatrix} = (0, 1/2, 0, 1/2)$$

$$v^{(2)} = (0,1/2,0,1/2)\begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 1/4 & 1/2 \end{pmatrix} = (0, 3/8, 1/4, 3/8)$$

$$v^{(3)} = (0,3/8,1/4,3/8)\begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/4 & 1/4 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 1/4 & 1/4 & 1/2 \end{pmatrix} = (0, 11/32, 10/32, 11/32)$$

Por tanto, a tres pasos la probabilidad pedida es:

$$p_3^{(3)} = \frac{10}{32}$$

d) La probabilidad $p_1^{(10^6)}$ es nula (esto es cierto para todo valor de n):

$$p_1^{(10^6)} = 0$$

pues como se ve en el diagrama a) no hay rutas que tengan al vértice 1 por punto de llegada. Este es un estado de los que llamaremos *transitorios*.

e) No es muy difícil anticipar el vector límite $v^{(n)}$ $(n \to \infty)$ dada la simetría de la matriz P. En ella hay un estado (1) transitorio que, a lo sumo, interviene en las etapas iniciales del proceso. Además hay tres estados (2, 3, 4) equivalentes entre sí. Es razonable pensar entonces que las probabilidades de aparición cuando $n \to \infty$ se repartirán uniformemente entre los tres estados *ergódicos*:

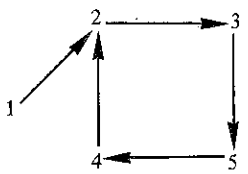
$$\lim_{n \to \infty} v^{(n)} = (0,\ 1/3,\ 1/3,\ 1/3)$$

Observando c) se ve que a tres pasos la influencia del estado inicial (1) está ya muy disipada, pues aunque groseramente:

$$11/32 \simeq 10/32 \approx 1/3$$

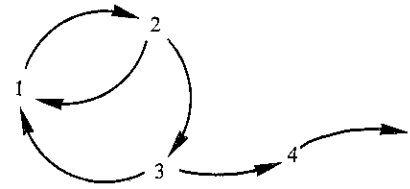
El vector $v^{(n \to \infty)}$ constituye la llamada *distribución estacionaria* del proceso. La velocidad con que se alcanza depende de P y del estado inicial seleccionado para iniciar la cadena.

4. Un subconjunto cerrado podría ser:



y como se comprueba no es ergódico, pues el estado 1 es transitorio y no es alcanzable por el resto de los estados.

Un subconjunto irreducible (1, 2, 3) respondería a un grafo del tipo:



que es fácil identificar que contiene un subconjunto transitorio.

5. Sea

$$\lim_{n\to\infty} P^n = A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \hline a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

y recordemos que A debe ser estocástica. Por tanto:

$$\lim_{n\to\infty} P^{n-1} \cdot P = A \quad ; \quad \lim_{n\to\infty} P^{n-1} = A$$

Esto conduce a establecer:

$$A = AP$$

o en forma matricial:

$$a_{ij} = \sum_{k=1}^{N} a_{ik} \cdot p_{kj} \quad ; \quad i,j = 1,2,...,N$$

Estamos interesados en mostrar que todas las filas $a_i = (a_{i1}, a_{i2}, ..., a_{iN})$ de A son iguales. Notemos que una fila cualquiera (i) se calcula del sistema de ecuaciones:

$$\begin{cases} a_{i1} = a_{i1} \cdot p_{11} + a_{i2} \cdot p_{21} + \cdots + a_{iN} \cdot p_{N1} \\ a_{i2} = a_{i1} \cdot p_{12} + a_{i2} \cdot p_{22} + \cdots + a_{iN} \cdot p_{N2} \\ \hline a_{iN} = a_{i1} \cdot p_{1N} + a_{i2} \cdot p_{2N} + \cdots + a_{iN} \cdot p_{NN} \end{cases}$$

y que otra fila $j \neq i$, $\mathbf{a}_j = (a_{j1}, a_{j2}, ..., a_{jN})$ surge de un sistema de ecuaciones idéntico, ya que:

$$\begin{cases} a_{j1} = a_{j1} \cdot p_{11} + a_{j2} \cdot p_{21} + \cdots + a_{jN} \cdot p_{N1} \\ a_{j2} = a_{j1} \cdot p_{12} + a_{j2} \cdot p_{22} + \cdots + a_{jN} \cdot p_{N2} \\ \text{- -} \\ a_{jN} = a_{j1} \cdot p_{1N} + a_{j2} \cdot p_{2N} + \cdots + a_{jN} \cdot p_{NN} \end{cases}$$

Ambos sistemas son homogéneos y el determinante de sus coeficientes $(p_{ij} - \delta_{ij})$ es nulo, por lo que en ambos sobra una ecuación que será reemplazada respectivamente por las condiciones:

$$a_{i1} + a_{i2} + \cdots + a_{iN} = 1 \quad ; \quad a_{j1} + a_{j2} + \cdots + a_{jN} = 1$$

Evidentemente, los vectores $\mathbf{a}_i$ y $\mathbf{a}_j$ son idénticos y, consecuentemente, todas las filas de A son iguales:

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_N \\ a_1 & a_2 & \cdots & a_N \\ \text{- - - - - - - - -} \\ a_1 & a_2 & \cdots & a_N \end{pmatrix}$$

Es ahora inmediato escribir que

$$AP = A \Rightarrow \mathbf{a}P = \mathbf{a} \quad ; \quad \mathbf{a} = (a_1, a_2, ..., a_N)$$

de donde podríamos, recordando la definición $\Pi = \Pi \cdot P$, identificar $\mathbf{a} = \Pi$, siempre que aseguremos que el vector $\mathbf{a}$ es único.

Para comprobar esta última cuestión, supongamos que existiera otro vector de probabilidad $\mathbf{b} = (b_1, b_2, ..., b_N)$ tal que:

$$\mathbf{b} \cdot P = \mathbf{b}$$

lo que conduciría a:

$$\begin{aligned} \mathbf{b}P^2 &= \mathbf{b} \cdot P \cdot P = \mathbf{b} \cdot P = \mathbf{b} \\ \mathbf{b}P^3 &= \text{- - - - - - - - - - -} = \mathbf{b} \\ &\text{- - - - - - - - - - - - - - - - -} \\ \mathbf{b}P^n &= \text{- - - - - - - - - - -} = \mathbf{b} \end{aligned}$$

y tomando límites:

$$\lim_{n\to\infty} \mathbf{b}\cdot\mathbf{P}^n = \mathbf{b} \Rightarrow \mathbf{b}A = \mathbf{b}$$

Una pequeña manipulación algebraica identifica, componente a componente, los vectores **a** y **b**:

$$b_k = a_k(b_1 + b_2 + \cdots + b_N) = \left\{\sum_i b_i = 1\right\} = a_k \quad ; \quad k = 1, 2, ..., N$$

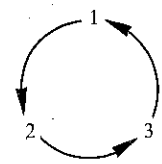
y por consiguiente:

$$\Pi = \mathbf{a} = \mathbf{b}$$

quedando demostrada la proposición pedida.

6. a) El grafo correspondiente a la matriz es:

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$



y se ve que la situación es ergódica cíclica. No existe una distribución estacionaria en el sentido $\Pi = \Pi \cdot P$ con $\Pi = \lim v^{(0)}P^n$ $(n \to \infty)$, dado el carácter oscilatorio que se presenta. Sabemos que si existe Π (regular):

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \cdots & \pi_N \\ \cdots & \cdots & \cdots & \cdots \\ \pi_1 & \pi_2 & \cdots & \pi_N \end{pmatrix}$$

en tanto que para la matriz dada:

$$P^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad ; \quad P^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

repitiéndose a partir de aquí la secuencia P, P^2, P^3:

$$P^4 = P \quad ; \quad P^5 = P^2 \quad ; \quad P^6 = P^3 \quad ; \quad \cdots$$

Si estos tres estados con su matriz P formaran parte de una cadena más grande, con más estados, y hubiese alguna ruta desde el exterior que condujera a alguno de ellos, entonces el proceso quedaría atrapado cíclicamente dentro de este subconjunto. Pero no habría distribución estacionaria asociada a ellos. La forma de P^n para n grandes sería una de las tres P, P^2, P^3, según el valor de n.

Calculemos a continuación los autovalores de P:

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 1 & 0 & -\lambda \end{vmatrix} = -\lambda^3 + 1 = 0$$

Aplicando aritmética compleja obtenemos las tres raíces de la unidad:

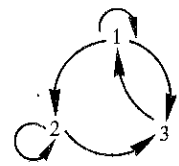
$$\lambda_n = \cos\frac{2\pi n}{3} + i\,\mathrm{sen}\frac{2\pi n}{3} = e^{i2\pi n/3} \quad ; \quad n = 0, 1, 2$$

$$\lambda_1 = 1 \quad ; \quad \lambda_2 = \cos\frac{2\pi}{3} + i\,\mathrm{sen}\frac{2\pi}{3} \quad ; \quad \lambda_3 = \cos\frac{4\pi}{3} + i\,\mathrm{sen}\frac{4\pi}{3}$$

comprobando la observación hecha en el texto (v).

b) En este caso la situación es diferente. El grafo asociado a la matriz P es:

$$P = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$$



y es rápido comprobar que los tres estados están interconectados en un número finito de pasos. La cadena es ergódica regular. Su distribución estacionaria Π vendrá dada por:

$$\Pi \cdot P = \Pi \Rightarrow (p_1, p_2, p_3) = (p_1, p_2, p_3)\begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$$

de donde se deduce el sistema de ecuaciones:

$$\begin{cases} p_1 = \frac{1}{2}p_1 + p_3 \\ p_2 = \frac{1}{3}p_1 + \frac{1}{2}p_2 \\ p_3 = \frac{1}{6}p_1 + \frac{1}{2}p_2 \end{cases} \Rightarrow \begin{cases} -\frac{1}{2}p_1 + p_3 = 0 \\ \frac{1}{3}p_1 - \frac{1}{2}p_2 = 0 \\ \frac{1}{6}p_1 + \frac{1}{2}p_2 - p_3 = 0 \end{cases}$$

que es un sistema homogéneo. Para resolverlo nos quedaremos con las dos primeras ecuaciones y añadiremos una tercera con la condición de normalización:

$$\begin{cases} -\frac{1}{2}\,p_1 + p_3 = 0 \\ -\frac{1}{3}\,p_1 - \frac{1}{2}\,p_2 = 0 \\ p_1 + p_2 + p_3 = 1 \end{cases} \Rightarrow \quad p_1 = 6/13;\ p_2 = 4/13;\ p_3 = 3/13$$

y de ahí surge la distribución estacionaria:

$$\Pi = (6/13, 4/13, 3/13)$$

Calculemos los autovalores de P:

$$\begin{vmatrix} 1/2 - \lambda & 1/3 & 1/6 \\ 0 & 1/2 - \lambda & 1/2 \\ 1 & 0 & -\lambda \end{vmatrix} = 0 \Rightarrow 12\lambda^3 - 12\lambda^2 + \lambda - 1 = 0$$

$$\lambda_1 = 1 \quad ; \quad \lambda_2 = i/\sqrt{12} \quad ; \quad \lambda_3 = -i/\sqrt{12}$$

de donde efectivamente comprobamos:

$$|\lambda_1| > |\lambda_2|,\ |\lambda_3|$$

como se indicaba en la observación iv).

Cuando $n \to \infty$, la matriz P^n pasa a ser:

$$\lim_{n\to\infty} P^n = \begin{pmatrix} 6/13 & 4/13 & 3/13 \\ 6/13 & 4/13 & 3/13 \\ 6/13 & 4/13 & 3/13 \end{pmatrix}$$

7. *a*) En este caso ya sabemos (Ejercicio 6a) que no existe distribución estacionaria en el sentido $\Pi P = \Pi$, pues la matriz $P(a)$ y sus potencias describen transiciones cíclicas entre los estados:

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = P^4 = P^7 = \ldots$$

$$P^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = P^5 = P^8 = \ldots$$

$$P^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = P^6 = P^9 = \ldots$$

Comprobemos que no existe ningún n_0 tal que $k(n_0) > 0$. Para ello cualquiera de las matrices P^n nos sirve (sea $n = n_0 = 1$):

$$i=1, j=2\ ,\ m=1,2,3 : |p_{11}-p_{21}| + |p_{12}-p_{22}| + |p_{13}-p_{23}| = 0+1+1=2$$

$$i=1, j=3\ ,\ m=1,2,3 : |p_{11}-p_{31}| + |p_{12}-p_{32}| + |p_{13}-p_{33}| = 1+1+0=2$$

$$i=2, j=3\ ,\ m=1,2,3 : |p_{21}-p_{31}| + |p_{22}-p_{32}| + |p_{23}-p_{33}| = 1+0+1=2$$

(si $i = j$ las sumas son idénticamente nulas). El valor máximo sup de las sumas sobre $m = 1, 2, 3$ es 2 para cualquier n y por tanto: $_{ij}$

$$k(n_0) = 1 - \frac{1}{2} \sup_{i,j} \sum_m |\, p_{im}(n_0) - p_{jm}(n_0)\,| = 1 - \frac{1}{2}\cdot 2 = 0$$

En base a este resultado no puede asegurarse la existencia de distribución estacionaria.

b) Tomemos $n_0 = 1$. Procediendo como en el apartado anterior el coeficiente de ergodicidad es:

$$k(n_0 = 1) = 1 - \frac{1}{2} \sup_{i,j} \sum_m |p_{im}(1) - p_{jm}(1)| = 1 - \frac{1}{2} \cdot 1 = \frac{1}{2} > 0$$

y la cota superior que informa de la velocidad de convergencia:

$$\sup_{p_i^0} |p_j(n) - p_j| \leqslant 2\, e^{-n \ln 2} = 2^{1-n} \qquad (n > 1)$$

Esta matriz $P(b)$ poseerá distribución estacionaria, ya que $k(n_0) > 0$, y la velocidad de aproximación a ésta va como 2^{1-n} en el caso más desfavorable posible.

c) Igualmente tomemos $n_0 = 1$. Se encuentra que:

$$k(n_0 = 1) = 1 - \frac{1}{2} \sup_{i,j} \sum_m |p_{im}(1) - p_{jm}(1)| = 1 - \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4} > 0$$

$$\sup_{p_i^0} |p_j(n) - p_j| \leqslant 4\, e^{-n \ln 4} = 4^{1-n} \qquad (n > 1)$$

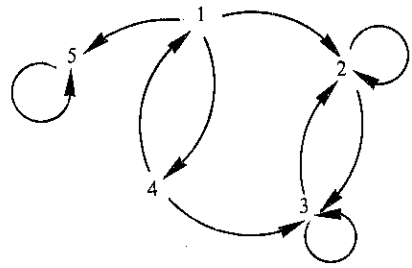
La matriz $P(c)$ poseerá también distribución estacionaria y su velocidad de aproximación va como 4^{1-n} en el caso más desfavorable.

Claramente, la cadena con matriz $P(c)$ convergerá con mayor rapidez que la de $P(b)$ (en tanto se emplee el mismo vector probabilidad inicial), pues los máximos errores absolutos se mantienen:

$$2^{1-n} > 4^{1-n} \quad ; \quad n > 1 \quad , \quad n_0 = 1$$

Las matrices $P(b)$ y $P(c)$ son biestocásticas y además $P(c) = P(b)^2$, por lo que el resultado alcanzado era de esperar. Conviene hacer hincapié en que tales cotas de error dependen del paso n_0 en el que $k(n_0) > 0$. Notemos que la influencia directa de n_0 sobre la convergencia es tal que a mayor n_0 la cota es, en principio, mayor (menor velocidad). Este efecto podría verse compensado con la influencia indirecta de n_0 contenida en los elementos $p_{im}(n_0)$ de la matriz de transición a n_0 pasos.

8. a) El grafo de la matriz P es:

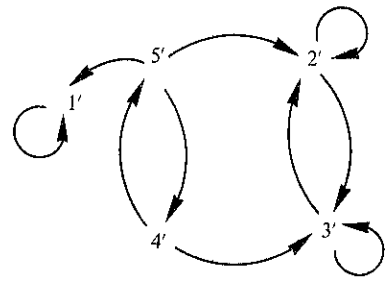


y los estados se clasifican en las clases:

— absorbentes $A = E_1 = \{5\}$
— ergódicos $E_2 = \{2, 3\}$
— transitorios $T = \{1, 4\}$

Para escribir P en forma canónica renumeraremos los estados $5 \to 1'$; $2 \to 2'$; $3 \to 3'$; $4 \to 4'$; $1 \to 5'$, con lo que

$$P = \left(\begin{array}{c|cc|cc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 3/4 & 1/4 & 0 & 0 \\ \hline 0 & 0 & 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \end{array}\right)$$

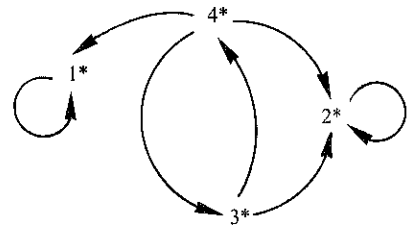


Todavía haremos una reducción más, convirtiendo el conjunto ergódico en un «macroestado absorbente». Es un asunto trivial ver que la matriz a un paso entre los cuatro «estados» 1', 4', 5', {2', 3'}, haciendo:

$$1' \to 1^* \quad ; \quad \{2', 3'\} \to 2^* \quad ; \quad 4' \to 3^* \quad ; \quad 5' \to 4^*$$

es:

$$P = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 & 0 \end{array}\right)$$



quedando ya en condiciones de responder a las siguientes preguntas del ejercicio.

b) Deberemos calcular las matrices:

$$F = (I' - T)^{-1} \quad ; \quad F_2 = F(2F_d - I') - F_c$$

siendo:

$$I' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad ; \quad T = \begin{pmatrix} 0 & 1/2 \\ 1/3 & 0 \end{pmatrix}$$

La matriz F resulta:

$$F = \begin{pmatrix} 1 & -1/2 \\ -1/3 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 6/5 & 3/5 \\ 2/5 & 6/5 \end{pmatrix}$$

donde la primera fila se refiere al estado transitorio de partida 3* y la segunda al 4* (según los gráficos $1 \to 4^*$; $4 \to 3^*$).

El número medio de veces que partiendo de 3* (4) se pasa por 3* es:

$$\langle n_{3*} \rangle_{3*} = 6/5$$

y por 4*:

$$\langle n_{4*} \rangle_{3*} = 3/5$$

Partiendo de 4* (1) se pasa por 3* un número medio:

$$\langle n_{3*} \rangle_{4*} = 2/5$$

y por 4*:

$$\langle n_{4*} \rangle_{4*} = 6/5$$

Las varianzas deben calcularse con F_2:

$$F_2 = \begin{pmatrix} 6/5 & 3/5 \\ 2/5 & 6/5 \end{pmatrix} \left[\begin{pmatrix} 12/5 & 0 \\ 0 & 12/5 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] -$$

$$- \begin{pmatrix} 36/25 & 9/25 \\ 4/25 & 36/25 \end{pmatrix} = \begin{pmatrix} 6/25 & 12/25 \\ 10/25 & 6/25 \end{pmatrix}$$

siendo sus valores los elementos de F_2:

$$\text{var}\,(n_{3*})_{3*} = 6/25 \quad ; \quad \text{var}\,(n_{3*})_{4*} = 10/25$$
$$\text{var}\,(n_{4*})_{3*} = 12/25 \quad ; \quad \text{var}\,(n_{4*})_{4*} = 6/25$$

c) Las matrices columna τ y τ_2 se evalúan sin dificultad:

$$\tau = Fu = \begin{pmatrix} 6/5 & 3/5 \\ 2/5 & 6/5 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 9/5 \\ 8/5 \end{pmatrix}$$

$$\tau_2 = [(2F - I')\tau - \tau_c] =$$

$$= \left[\begin{pmatrix} 12/5 & 6/5 \\ 4/5 & 12/5 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{pmatrix} 9/5 \\ 8/5 \end{pmatrix} - \begin{pmatrix} 81/25 \\ 64/25 \end{pmatrix} = \begin{pmatrix} 30/25 \\ 28/25 \end{pmatrix}$$

y los valores pedidos son:

$$\langle n_t \rangle_{3*} = 9/5 \quad ; \quad \text{var}\,(n_t)_{3*} = 6/5$$
$$\langle n_t \rangle_{4*} = 8/5 \quad ; \quad \text{var}\,(n_t)_{4*} = 28/25$$

d) La probabilidad de ser absorbido es:

$p(t_i \to a_k) = (F \cdot R)_{ik}$

Calculando el producto de matrices:

$$F \cdot R = \begin{pmatrix} 6/5 & 3/5 \\ 2/5 & 6/5 \end{pmatrix}\begin{pmatrix} 0 & 1/2 \\ 1/3 & 1/3 \end{pmatrix} = \begin{pmatrix} 1/5 & 4/5 \\ 2/5 & 3/5 \end{pmatrix}$$

vemos que partiendo de 3* (4) la probabilidad de:

— caer en $1^* = A(E_1)$ es $= 1/5$

— caer en $2^* = E_2$ es $= 4/5$

en tanto que para 4* (1) la probabilidad de:

— caer en $1^* = A$ es $= 2/5$

— caer en $2^* = E_2$ es $= 3/5$

ÍNDICE VOLÚMEN I

ANÁLISIS NUMÉRICO

Pág.

TEMA 1. La aproximación polinómica de colocación

1. Fundamentos matemáticos 18
2. El polinomio de colocación y su error 20
3. Cálculo con diferencias finitas 22
   - 3.1. Diferencias de avance 22
   - 3.2. Carácter operacional de Δ 24
   - 3.3. Propagación de errores en una tabla de diferencias 24
4. Polinomios factoriales y números de Stirling 26
5. Operadores para argumentos igualmente espaciados 28
   - 5.1. Operador Δ y polinomio de avance de Newton 28
   - 5.2. Operadores E, ∇, μ, δ 31
   - 5.3. Utilidad de los diferentes polinomios de colocación 34
6. Polinomios de colocación para argumentos no equiespaciados 35
   - 6.1. Polinomio de Lagrange 36
   - 6.2. Diferencias divididas 37

Bibliografía 41

Pág.

Ejercicios de autocomprobación 43

Soluciones a los ejercicios de autocomprobación 45

TEMA 2. Aplicaciones de la aproximación polinómica

1. Interpolación y extrapolación 58

   1.1. Interpolación directa 58
   1.2. Subtabulación 61
   1.3. Interpolación inversa 62
   1.4. El error en interpolación 62
   1.5. Extrapolación 65

2. Ejemplo de aplicación: calor específico de sólidos a baja temperatura 65

3. Diferenciación numérica 68

4. Ejemplo de aplicación: Ecuación diferencial de los polinomios de Hermite 71

5. Integración numérica 73

   5.1. Regla del trapecio 75
   5.2. Regla de Simpson 77
   5.3. Método de los coeficientes indeterminados 78
   5.4. Integración gaussiana 79
   5.5. Integrales singulares 81
   5.6. Integración numérica de funciones oscilantes. Fórmula de Filon 84

6. Ejemplo de aplicación: cálculo de la integral de Debye 86

Bibliografía 89

Ejercicios de autocomprobación 91

Soluciones a los ejercicios de autocomprobación 95

TEMA 3. Resolución numérica de ecuaciones diferenciales ordinarias

1. Estabilidad y error 110

2. Ecuaciones de primer orden 111

Pág.

2.1. Método de Euler 111
2.2. Serie de Taylor 112
2.3. Método de Runge-Kutta 115
2.4. Método predictor-corrector de Euler 117

3. Ejemplo de aplicación: cinética de isomerización cis-trans 118

4. Ecuaciones de orden superior y sistemas de ecuaciones diferenciales 119

4.1. Método de Runge-Kutta 121
4.2. Predictor-corrector de Milne 122
4.3. Problemas de contorno 124

5. Ejemplo de aplicación: el método de la Dinámica Molecular para estudios de la fase líquida 127

Bibliografía 133

Ejercicios de autocomprobación 135

Soluciones a los ejercicios de autocomprobación 139

TEMA 4. Aproximación por mínimos cuadrados y desarrollos en serie de funciones ortogonales

1. El principio de mínimos cuadrados. Problemática 151

2. Polinomios ortogonales en caso discreto (Gram-Tchebycheff) 154

3. Polinomios ortogonales en caso continuo 157

3.1. Generalidades 157
3.2. Polinomios de Legendre 161
3.3. Polinomios de Laguerre 164
3.4. Ejemplo de aplicación: la vibración molecular y los polinomios de Hermite 165

4. Análisis de Fourier 168

4.1. Series de Fourier 169
4.2. Ejemplo de aplicación: evolución temporal de una función de onda 174
4.3. Transformación de Fourier 177
4.4. Sumas trigonométricas (caso discreto) 180

5. La «función» δ de Dirac 182

Pág.

6. Ejemplo de aplicación: visualización del principio de incertidumbre 188

Bibliografía 191

Ejercicios de autocomprobación 193

Soluciones a los ejercicios de autocomprobación 197

TEMA 5. Resolución numérica de ecuaciones no lineales y diagonalización de matrices

1. Ecuaciones no lineales 214

   1.1. Separación de raíces en un intervalo 214
   1.2. Método de bisección 216
   1.3. Regula falsi 217
   1.4. Método de iteración 219
   1.5. Método de Newton 220
   1.6. Raíces múltiples 223
   1.7. Ejemplo de aplicación: Posición del máximo de la radiación del cuerpo negro 225

2. Sistemas de ecuaciones 228

   2.1. Método de Newton-Raphson 228
   2.2. Método de Gauss para sistemas lineales 229

3. Problema de valores propios y diagonalización de matrices . 230

   3.1. Polinomio característico 231
   3.2. Método de Jacobi 233
   3.3. Ejemplo de aplicación: cálculo de orbitales moleculares de Hückel para el radical ciclopropenilo 237

Bibliografía 247

Ejercicios de autocomprobación 249

Soluciones a los ejercicios de autocomprobación 251

ESTADÍSTICA TEÓRICA

TEMA 6. Variables aleatorias y distribuciones de probabilidad

1. Funciones de distribución para una variable aleatoria 268

Pág.

1.1. Variables discretas 269

1.2. Variables continuas 271

2. Funciones de distribución para dos o más variables aleatorias 275

3. Características de las magnitudes aleatorias 279

3.1. Población y muestra 279
3.2. Esperanza matemática y valores medios 280
3.3. Momentos de una distribución 283
3.4. Dispersión y asimetría 283

4. Funciones generatriz y característica 286

5. Convolución de distribuciones de probabilidad 288

6. Ejemplo de aplicación: cálculo de valores medios y dispersiones en Termodinámica Estadística 290

Bibliografía 295

Ejercicios de autocomprobación 297

Soluciones a los ejercicios de autocomprobación 299

TEMA 7. Distribuciones de magnitudes aleatorias en una dimensión

1. Distribuciones discretas 312

1.1. Distribuciones binómica y binómica generalizada 312
1.2. Ejemplo de aplicación: sistemas de spines, fluctuaciones y distribución multinomial 314
1.3. Distribución de Poisson 316

2. Distribuciones continuas 318

2.1. Distribución normal (gaussiana) 318
2.2. Distribución chi-cuadrado (χ^2) 325
2.3. Ejemplo de aplicación: función de distribución del módulo de la velocidad molecular en un gas 327
2.4. Distribución gamma (Γ) 328
2.5. Ejemplo de aplicación: función de distribución de las energías moleculares en un gas 329

3. Números aleatorios 330

3.1. Familias multiplicativas congruentes 331

Pág.

3.2. Ejemplo de aplicación: Cálculo de integrales definidas por el método de Monte Carlo 333

Bibliografía 339

Ejercicios de autocomprobación 341

Soluciones a los ejercicios de autocomprobación 343

TEMA 8. Distribuciones de magnitudes aleatorias en dos dimensiones y correlación

1. Momentos y coeficientes de correlación 356
2. Regresión de mínimos cuadrados 357
3. Cambios de variable útiles y problemática del ajuste 359
4. Distribuciones condicionales y regresión de la media 363
5. Ejemplo de aplicación: Ley de Lambert-Beer 366
6. Distribución normal en dos dimensiones 368
7. Ejemplos de aplicación 370
   - 7.1. Fluctuaciones termodinámicas 370
   - 7.2. La función gaussiana en Química Cuántica 374
8. Funciones de correlación en varias variables 378
9. Ejemplo de aplicación: funciones de correlación en Mecánica Estadística 379

Bibliografía 381

Ejercicios de autocomprobación 383

Soluciones a los ejercicios de autocomprobación 385

TEMA 9. Procesos estocásticos y cadenas de Markov discretas

1. Generalidades 396
2. Procesos estacionarios 398
3. Correlación y densidad espectral 401
4. Procesos de Markov 404

Pág.

5. Ejemplo de aplicación: movimiento browniano en una dimensión 406
6. Cadenas de Markov estacionarias en tiempo discreto 409
   6.1. Espacio de estados, matriz de transición y vector de probabilidad 409
   6.2. Clasificación de estados 414
   6.3. Clasificación de cadenas 417
   6.4. Ergodicidad y distribuciones estacionarias 419
   6.5. Ejemplo de aplicación: matrices biestocásticas y sistemas uniformes 421
   6.6. Cadenas absorbentes 424
7. Ejemplo de aplicación: simulación Monte Carlo de líquidos 426
8. Procesos de Markov de orden superior 434

Bibliografía 437

Ejercicios de autocomprobación 439

Soluciones a los ejercicios de autocomprobación 443