

Chapter 6

Método de Monte Carlo

6.1 Introducción

Los sistemas con muchos grados de libertad son de gran interés en fisico-química. Entre éstos podemos citar la gran cantidad de átomos presentes en una muestra macroscópica de cualquier sustancia, los electrones en un metal o aún en un solo átomo, el número de aminoácidos en una proteína o de bases nitrogenadas en un ácido nucleico. La descripción de tales sistemas a menudo involucra (o puede ser reducida a) la evaluación de integrales multidimensionales. Por ejemplo la función de partición clásica de un gas monoatómico de N átomos, a una temperatura T interactuando a través de un potencial de pares v es proporcional a la integral 3N-dimensional

$$Z = \int dr_1 \cdots dr_N \exp(-\beta \sum_{i<j}^N v(r_{ij})) \quad (6.1)$$

La evaluación de esta integral por cualquiera de los métodos de cuadratura estudiados en los primeros capítulos puede realizarse sin inconvenientes sólo para valores pequeños de N . Para ver por qué, supongamos que la cuadratura permite que cada coordenada pueda tomar 10 valores diferentes (lo cual no es una discretización muy fina), por lo tanto la integral debe ser evaluada en 10^{3N} puntos. Para un valor modesto de N , digamos $N = 20$ y una computadora muy rápida capaz de hacer unas 10^7 evaluaciones por segundo, la tarea completa podría tomar alrededor de 10^{53} segundos, o sea unas 10^{34} veces la edad del universo! Por supuesto que existen algunos trucos basados en las propiedades de simetría del integrando que permiten reducir el número

de evaluaciones considerablemente, pero aún así el empleo de técnicas de cuadratura resulta impracticable.

Los métodos de Monte Carlo discutidos en este capítulo permiten evaluar eficientemente integrales multidimensionales [1, 4]. El nombre Monte Carlo proviene del carácter aleatorio del método y del famoso casino de Mónaco. La idea esencial es no evaluar el integrando en todos y cada uno de los puntos de cuadratura, sino en una muestra aleatoria **representativa** de las abscisas. Esto es análogo a la predicción de los resultados de una elección en base al análisis de un conjunto pequeño de votantes. Aunque no resulta obvio que algo sensible pueda provenir de la generación de números aleatorios en una computadora, el método de Monte Carlo se ha vuelto una herramienta extremadamente útil en el tratamiento de una amplia clase de problemas en mecánica estadística y mecánica cuántica. Presentaciones más detalladas del método pueden encontrarse en [5, 6, 7, ?]

6.2 La estrategia básica de Monte Carlo

Aún cuando la utilidad del método de Monte Carlo está en la evaluación de integrales multidimensionales, es más fácil ilustrar las ideas básicas en una situación unidimensional. Supongamos que deseamos evaluar la integral

$$I = \int_0^1 f(x)dx \tag{6.2}$$

para alguna función particular f . En los primeros capítulos discutimos varias fórmulas de cuadratura que empleaban valores de f para valores muy particulares de la variable x (por ejemplo puntos equiespaciados). Sin embargo una forma alternativa de evaluar I es multiplicar y dividir el integrando por una función $w(x)$ que cumple con las condiciones necesarias para ser una

función de densidad de probabilidad ¹

$$I = \int_0^1 f(x) \left(\frac{w(x)}{w(x)} \right) dx$$

recordando que si x sigue una densidad $w(x)$ entonces el valor medio de cualquier función de x , digamos $g(x)$, vendrá dado por $\langle g \rangle = \int w(x)g(x)dx$, donde la integral se extiende a todo el rango de la variable x ; la integral I que nos interesa es exactamente el valor medio de la función $\frac{f(x)}{w(x)}$ si la variable de integración sigue la densidad de probabilidad w .

$$I = \left\langle \frac{f}{w} \right\rangle_w$$

el subíndice w destaca que el promedio es calculado según w , así por ejemplo si w es una densidad uniforme en el intervalo $[0, 1]$, esto es $w = 1$, $I = \langle f \rangle$ y una fórmula de cuadratura apropiada será

$$I \approx \frac{1}{N} \sum_{i=1}^N f(x_i) + error \quad (6.3)$$

Para evaluar la misma deben generarse N valores de x , aleatoria y uniformemente distribuidos (o sea con igual probabilidad) en el intervalo $[0, 1]$, evaluar la función f en cada uno de ellos y sacar el promedio correspondiente. Más adelante discutiremos la forma en que se generan estos números “aleatorios”, por ahora basta con suponer que existe una función de la computadora que provee cada uno de ellos uno después del otro.

La forma de evaluar la integral monodimensional I descrita en los párrafos anteriores puede generalizarse trivialmente a integrales multidimensionales. El procedimiento general se conoce como “importance sampling” o muestreo significativo ya que eligiendo adecuadamente la función w se puede muestrear preferencialmente los zonas del dominio de f que más contribuyen a la integral.

¹Para ser una densidad de probabilidad $w(x)$ debe ser continua de a pedazos y cumplir con las siguientes condiciones:

- $w(x) \geq 0$
- $\int w(x)dx = 1$

Sobre todo el rango de la variable x .

En este punto debemos preguntarnos cómo generalizar el algoritmo anterior para una densidad de probabilidad arbitraria $w(x)$, y cual es la elección mas conveniente en cuanto a la forma funcional de $w(x)$ dada la correspondiente forma funcional de $f(x)$. La primera pregunta tiene una respuesta rápida: Se generan N valores de x , digamos $\{x_1, x_2 \cdots x_N\}$ según la densidad de probabilidad $w(x)$, para cada uno de ellos se evalúa el cociente $\frac{f(x_i)}{w(x_i)}$ y se calcula el valor medio de los mismos según

$$\left\langle \frac{f}{w} \right\rangle = \frac{1}{N} \sum_i^N \frac{f(x_i)}{w(x_i)} + error \quad (6.4)$$

Nuevamente, la forma de generar el conjunto $\{x_1, x_2 \cdots x_N\}$ conforme a $w(x)$ será abordada más adelante.

Para responder a nuestra segunda pregunta: Cómo elegir adecuadamente la función de densidad $w(x)$?, debemos hacer una breve disquisición respecto al cálculo del término de error en las ecuaciones 6.14 y 6.15. Parafraseando el teorema central del límite (uno de los más importantes de la estadística) podemos decir que el valor medio I , de N cantidades $\{g_1 \cdots g_N\}$ aleatorias e independientes

$$I = \frac{1}{N}(g_1 + g_2 + \cdots + g_N)$$

sigue una distribución gaussiana con media I y desviación estándar $\sigma_I = \frac{\sigma_g}{\sqrt{N}}$ para grandes valores de N , donde σ_g es la desviación estándar de la variable g^2 . Por otro lado podemos considerar que el término de error en las ecuaciones 6.14 y 6.15 proviene de un error aleatorio e independiente que se comete en la evaluación de cada término de la suma, con lo cual establecemos la siguiente igualdad $g_i = \frac{f(x_i)}{w(x_i)}$. Con lo cual el cálculo del error en la integral I estará dado por

$$error = \sigma_I = \frac{\sigma_{f/w}}{\sqrt{N}}$$

$$\sigma_I = \frac{1}{\sqrt{N}} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{f_i}{w_i} \right)^2 - \left(\frac{1}{N} \sum_{i=1}^N \frac{f_i}{w_i} \right)^2 \right]^{1/2} \quad (6.5)$$

²Recordemos que la desviación estándar de g se define como $\sigma_g = [\frac{1}{N} \sum_{i=1}^N (g_i - \langle g \rangle)^2]^{1/2}$ y mide la extensión en la cual la variable g se desvia de su media $\langle g \rangle$.

y las ecuaciones 6.14 y 6.15 pueden reescribirse como

$$I \approx \frac{1}{N} \sum_{i=1}^N f(x_i) + O\left(\frac{cte}{\sqrt{N}}\right)$$

$$\left\langle \frac{f}{w} \right\rangle = \frac{1}{N} \sum_i \frac{f(x_i)}{w(x_i)} + O\left(\frac{cte}{\sqrt{N}}\right)$$

La ecuación 6.16 revela dos aspectos muy importantes del método de Monte Carlo. Primero, la incerteza en la estimación de la integral, σ_I , disminuye con $N^{-1/2}$. Por lo tanto mientras más puntos son usados más preciso será nuestro cálculo, aunque el error disminuye muy lentamente con el número de puntos (se necesita cuatro veces más trabajo numérico para disminuir a la mitad la incerteza en el resultado). El segundo punto importante que puede extraerse de la ecuación 6.16 es que la precisión en la integral es mayor mientras más pequeña sea $\sigma_{f/w}$ es decir mientras más suave sea la función $\frac{f}{w}(x)$. El caso límite es cuando $\frac{f}{w}(x) = cte.$, para el cual necesitamos conocer el valor de la función en un solo punto para poder calcular el promedio. A partir de esto vemos que la función $w(x)$ debe elegirse de forma tal que el cociente $\frac{f}{w}(x)$ se mantenga más o menos constante (suave) en todo el intervalo de integración, en otras palabras, los valores $\{x_1 \cdots x_N\}$ generados acorde a la distribución $w(x)$ deben concentrarse fundamentalmente en las zonas donde $f(x)$ es grande y por lo tanto su contribución a la integral es **importante**, mientras que la densidad de los mismos debe ser baja en las zonas donde $f(x)$ es chica y su contribución a la integral es **poco importante**.

- **Ejemplo:** Evaluar la integral

$$\frac{1}{\sqrt{2\pi}} \int_{-b}^{+b} \exp\left(-\frac{x^2}{2}\right) dx$$

1. Una posibilidad es evaluar esta integral usando una distribución uniforme en el intervalo $[-b, +b]$, o sea que la función de peso es $w(x) = \frac{1}{2b}$. Genero N números aleatorios y evalúo la cantidad

$$I = 2b * \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right)$$

2. Otra posibilidad es usar como función densidad una función que sea igual al integrando en el intervalo $[-b, +b]$ y cero fuera del mismo.

$$w(x) = \Theta(x + b) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \Theta(b - x)$$

donde $\Theta(y)$ es una función escalón. Con lo cual

$$I = \int_{-\infty}^{+\infty} \Theta(x + b) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \Theta(b - x) dx$$

y $I \approx \frac{M}{N}$, donde N es el número total de puntos generados y M el número de puntos que caen dentro del intervalo $[-b, +b]$

Todo lo dicho hasta aquí puede extenderse fácilmente al caso en el que \vec{X} es un vector en un espacio N -dimensional, así la integral a evaluar será

$$I = \int \dots \int f(\vec{X}) dX^N$$

y la correspondiente función de peso $w(\vec{X})$.

6.2.1 Ejemplos de aplicación y algunos algoritmos:

Considere el cálculo de la integral $\int_0^1 \frac{dx}{1+x^2} = \frac{\pi}{4}$. El siguiente programa en FORTRAN calcula esta integral para N valores de x junto con una estimación de la precisión de la cuadratura.

```

INTEGER SEED
DATA SEED/987654321/
DATA EXACT/.78540/
FUNC(X)=1./(1.+X**2)
10 PRINT *, ' Enter number of points (0 to stop)'
   READ *, N
   IF (N .EQ. 0) STOP
   SUMF=0.
   SUMF2=0.
   DO 20 IX=1,N
     FX=FUNC(RAN(SEED))

```

```

        SUMF=SUMF+FX
        SUMF2=SUMF2+FX**2
20     CONTINUE
        FAVE=SUMF/N
        F2AVE=SUMF2/N
        SIGMA=SQRT((F2AVE-FAVE**2)/N)
        PRINT *, ' integral = ',FAVE,' +- ',SIGMA,
& 'error = ',EXACT-FAVE
        GOTO 10
    END

```

Como ejemplo de cómo el uso de una función de peso afecta al cálculo de la integral considere la integral anterior y elijamos como función de peso

$$w(x) = \frac{1}{3}(4 - 2x)$$

que es positiva definida y decrece monótonamente sobre todo el rango de integración (como lo hace f) y está normalizada adecuadamente. Por lo tanto $f/w = 3/4$ en $x = 0$ y $x = 1$, con lo cual w aproxima muy bien el comportamiento de f . El correspondiente código FORTRAN es

```

    INTEGER SEED
    DATA SEED/987654321/
    DATA EXACT/.78540/
    FUNC(X)=1./(1.+X**2)
    WEIGHT(X)=(4.-2*X)/3.
    XX(Y)=2.-SQRT(4.-3.*Y)

10     PRINT *, ' Enter number of points (0 to stop)'
        READ *, N
        IF (N .EQ. 0) STOP

        SUMF=0.
        SUMF2=0.
        DO 20 IX=1,N
            Y=RAN(SEED)
            X=XX(Y)
            FX=FUNC(X)/WEIGHT(X)

```

```

        SUMF=SUMF+FX
        SUMF2=SUMF2+FX**2
20     CONTINUE
        FAVE=SUMF/N
        F2AVE=SUMF2/N
        SIGMA=SQRT((F2AVE-FAVE**2)/N)
        PRINT *, ' integral = ',FAVE, ' +- ',SIGMA,
& ' error = ',EXACT-FAVE

        GOTO 10
    END

```

El problema que debemos atacar ahora es cómo generar números aleatorios según la densidad de probabilidad $w(x)$ o $w(\vec{X})$ según sea el caso.

6.3 Generación de números aleatorios con una distribución específica

La discusión anterior muestra que el método de Monte Carlo involucra dos operaciones básicas: generación de abscisas aleatoriamente distribuidas sobre el volumen de integración con una distribución específica $w(x)$ y la evaluación de la función $\frac{f}{w}$ en esos puntos. La segunda operación es muy sencilla, pero no es obvio cómo generar números “aleatorios” en una computadora determinista. En esta sección vamos a revisar algunos de los métodos más usados.

La generación de números aleatorios uniformemente distribuidos es la operación que subyace a cualquier tratamiento de distribuciones más complicadas. Hay numerosos métodos para realizar esta tarea básica y de chequear si los resultados obtenidos corresponden a la distribución uniforme requerida. Uno de los algoritmos más comunes es el método congruencial lineal que genera una secuencia completa de números aleatorios a partir de un número inicial llamado “semilla”. El número actual de la secuencia es multiplicado por un primer número “mágico” incrementado por un segundo número mágico y a esta suma se le toma el módulo con un tercer número mágico. Así

$$x_{i+1} = \text{mod}(ax_i + c, m) \quad (6.6)$$

donde

$$\text{mod}(z_1, z_2) = z_1 - \text{int}\left(\frac{z_1}{z_2}\right)z_2$$

i es la etiqueta del i -ésimo miembro de la secuencia y a , c y m son los tres números mágicos. Note que los valores de x_i no pueden ser verdaderamente aleatorios ya que ellos provienen de la semilla en una forma bien definida (algoritmo determinista) de manera tal que dos secuencias que crezcan a partir de la misma semilla serán exactamente iguales. Por esta razón estos números son a veces llamados “pseudo aleatorios”.

Una vez que uno tiene una función como la ecuación 6.17 capaz de generar números aleatorios uniformemente distribuidos, la generación de variables aleatorias con distribuciones más complicadas puede efectuarse de muy distintas maneras. Un caso particular es aquel en el que existe una relación uno a uno entre la variable que podemos generar (por ej. uniformemente distribuidas) y aquellas que queremos generar. Para mayor claridad, supongamos que tenemos una función de computadora capaz de generar las variables $\{x\}$ según la densidad de probabilidad $p(x)$, y que a partir de éstas deseamos generar las variables $\{y(x)\}$ distribuidas según la densidad $q(y)$. Que relación existe entre $p(x)$ y $q(y)$?

Teniendo en cuenta que

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \quad (6.7)$$

y que $x = x(y)$, dado que suponemos que la función $y(x)$ es invertible

$$\int_{-\infty}^{+\infty} p(x(y)) \left| \frac{dx}{dy} \right| dy = 1 \quad (6.8)$$

donde hemos usado la siguiente relación: $dx = \left| \frac{dx}{dy} \right| dy$. Comparando las ecuaciones 6.18 y 6.19 concluimos que

$$q(y) = p(x(y)) \left| \frac{dx}{dy} \right| \quad (6.9)$$

Resolviendo esta ecuación diferencial encontraremos la relación funcional entre x e y . Si $p(x) = 1$, entonces la ecuación 6.20 se transforma en

$$q(y) = \left| \frac{dx}{dy} \right|$$

de donde

$$x(y) = \int_0^y q(y') dy' \quad y \in [0, \infty] \quad (6.10)$$

a esta integral se la denomina integral incompleta y la fórmula resultante debe ser invertida para obtener y en función de x .

- **Ejemplo:** Cómo generar números aleatorios y distribuidos acorde a una densidad de probabilidad exponencial en el intervalo $[0, \infty)$, a partir de números aleatorios x , distribuidos uniformemente en el intervalo $[0, 1]$?

$$q(y) = A \exp(-Ay) \quad y \in [0, \infty)$$

$$p(x) = 1 \quad x \in [0, 1]$$

segun la ecuación 6.20

$$A \exp(-Ay) = 1 \left| \frac{dx}{dy} \right|$$

resolviendo la ecuación diferencial obtenemos

$$x(y) = \int_0^y A \exp(-Ay') dy' = 1 - \exp(-Ay)$$

$$y = -\frac{1}{A} \ln(1 - x) \quad x \in [0, 1]$$

- **Ejercicio:** A partir de números uniformemente distribuidos en el intervalo $[0, 1]$ generar números aleatorios y que sigan una distribución de Lorenz

$$q(y) = \frac{1}{\pi} \frac{1}{(1 + y^2)} \quad y \in (-\infty, +\infty)$$

Respuesta: $y = \tan[\pi(x - \frac{1}{2})]$ $x \in [0, 1]$.

Como puede verse en el ejemplo y el ejercicio anterior, una vez resuelta la ecuación diferencial debemos ser capaces de invertir la función $x(y)$ para obtener y en función de x . Mientras que esto puede hacerse en forma analítica sólo para una restringida clase de funciones, es posible llevar a cabo la integración y la inversión en forma numérica. Supongamos que tabulamos los

6.3. GENERACIÓN DE NÚMEROS ALEATORIOS CON UNA DISTRIBUCIÓN ESPECÍFICA 19

valores de $y^{(j)}$ para los cuales la integral incompleta de $q(y)$ se calcula en una serie de valores equiespaciados de x , $x^{(j)} \equiv j/M$ con $j = 0, 1 \dots M$, tales que los mismos cubran el intervalo $[0, 1]$ entonces la ecuación 6.21 se transforma en

$$x^{(j)} = \frac{j}{M} = \int_0^{y^{(j)}} q(y') dy'$$

Recordemos que los valores de x están distribuidos uniformemente, de manera que para nuestro caso discreto el procedimiento análogo consiste en elegir con igual probabilidad algún valor de j de entre los M posibles. Los correspondientes valores de $y^{(j)}$ estarán distribuidos de acuerdo a $q(y)$. El problema es por supuesto cómo generar los $y^{(j)}$. Para esto integro numéricamente la ecuación diferencial

$$q(y) = \left| \frac{dx}{dy} \right|$$

$$\frac{x^{(j+1)} - x^{(j)}}{y^{(j+1)} - y^{(j)}} = q(y^{(j)})$$

ya que $x^{(j+1)} - x^{(j)} = \frac{1}{M}$

$$y^{(j+1)} = \frac{1}{M} \frac{1}{q(y^{(j)})} + y^{(j)}$$

la cual puede ser integrada usando $y^{(0)} = 0$. En otras palabras, elijo un valor de j al azar e integro hasta obtener $y^{(j)}$.

Otra forma de generar variables aleatorias uni o multidimensionales es el método de rechazo (rejection) de von Newman.

Supongamos que estamos interesados en generar números aleatorios x en el intervalo $[0, 1]$ con distribución $p(x)$ y sea $p'(x)$ una función positiva con la propiedad de que $p'(x) > p(x)$ sobre la región de integración. Note que esto significa que la integral definida de $p'(x)$ es mayor que uno. Una forma conveniente de elegir $p'(x)$ es como una constante mayor que el máximo valor de la función $p(x)$ en el intervalo de integración. Si generamos puntos en dos dimensiones que llenen uniformemente el área bajo la curva $p'(x)$ y entonces aceptamos sólo aquellos que caigan bajo la curva $p(x)$, entonces los puntos aceptados estarán distribuidos de acuerdo a $p(x)$. En la práctica lo que se hace es elegir un número aleatorio x de la distribución $p'(x)$ y otro y , uniformemente distribuido en el intervalo $[0, p'(x)]$. El valor de x es aceptado

si el correspondiente valor de y es menor que $p(x)$, es decir si el punto (x,y) se encuentra bajo la curva $p(x)$. Esta técnica es eficiente si $p'(x)$ es próximo a $p(x)$ durante todo el intervalo de integración.

6.4 El algoritmo de Box-Muller

La distribución gaussiana (o normal) de media nula y varianza unitaria

$$g(x) = (2\pi)^{-1/2} \exp(-x^2/2)$$

juega un papel fundamental en estadística y teoría de probabilidad, por lo tanto es importante tener un algoritmo eficiente para la generación de números aleatorios acorde a esta densidad de probabilidad. Uno de los métodos más eficientes para generar variables aleatorias gaussianas empieza considerando una distribución gaussiana en dos dimensiones (x_1, x_2) para la cual el número de puntos en un área diferencial es proporcional a

$$\exp(-(x_1^2 + x_2^2)/2) dx_1 dx_2$$

Pasando a coordenadas polares

$$r = (x_1^2 + x_2^2)^{1/2}$$

$$\theta = \arctan\left(\frac{x_2}{x_1}\right)$$

La distribución es

$$\exp\left(-\frac{1}{2}r^2\right) r dr d\theta$$

o si $u = r^2/2$, la distribución es

$$\exp(-u) du d\theta$$

Por lo tanto si generamos u entre 0 e infinito con una distribución exponencial y θ uniformemente entre 0 y 2π , entonces los correspondientes valores de

$$x_1 = (2u)^{1/2} \cos(\theta)$$

$$x_2 = (2u)^{1/2} \sin(\theta)$$

estarán distribuidos normalmente. Este algoritmo es conocido como algoritmo de Box-Muller.

6.5 El algoritmo de Metrópolis et al.

Aunque los métodos descritos anteriormente generan eficientemente variables aleatorias para densidades de probabilidad específicas, es muy difícil o imposible generalizar los mismos a funciones de peso complicadas en espacios multidimensionales. Una forma muy general de producir variables aleatorias con distribución de probabilidad de forma arbitraria es conocida como algoritmo de Metrópolis et al. Ya que para implementarlo sólo se necesita poder evaluar la función peso para un dado valor de las variables de integración, el algoritmo ha sido utilizado ampliamente en problemas de mecánica estadística.

Así por ejemplo en el colectivo canónico deseamos evaluar integrales del tipo

$$\langle A \rangle_{NVT} = \int A(X) \rho_{NVT}(X) dX \quad (6.11)$$

donde X representa un vector en un espacio de N dimensiones.

$$\rho_{NVT}(X) = \frac{\exp(-\frac{H(X)}{kT})}{Z_{NVT}}$$

y

$$Z_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int dX \exp(-\frac{H(X)}{kT})$$

Supongamos que elegimos para evaluar la integral 6.22 la función de peso ρ , entonces

$$\langle A \rangle_{NVT} = \left\langle \frac{A \rho_{NVT}}{\rho} \right\rangle_{\rho}$$

por lo general $A(X)$ será significativa donde $\rho_{NVT}(X)$ lo sea de manera tal que una buena elección para ρ es $\rho = \rho_{NVT}$, con lo cual

$$\langle A \rangle_{NVT} = \langle A \rangle_{\rho_{NVT}}$$

Para evaluar este promedio debemos ser capaces de generar vectores X distribuidos según ρ_{NVT} , y esto es lo que se hace eficientemente mediante el algoritmo de Metrópolis et al..

La secuencia de vectores generada $X_0, X_1 \dots X_N$ puede ser vista como los sucesivos puntos visitados por un caminante que se mueve al azar en el espacio de las fases, mientras más larga es la caminata los puntos que esta

conecta se aproximan más estrechamente a la distribución deseada (ρ_{NVT} por ejemplo).

Las reglas por las cuales la caminata al azar procede a través del espacio de las fases son las siguientes. Supongamos que el caminante está en el punto X_n de la secuencia. Para generar X_{n+1} , éste hace un movimiento de prueba a un nuevo punto X_t . Este nuevo punto puede ser elegido de cualquier manera conveniente, por ejemplo aleatoria y uniformemente en el interior de un cubo multidimensional de lado L y centrado en X_n , así $X_t = X_n + \delta X$. La movida de prueba es aceptada o rechazada de acuerdo a la razón

$$r = \frac{w(X_t)}{w(X_n)}$$

para el caso de la distribución canónica ρ_{NVT} el cociente anterior toma la forma

$$r = \frac{\rho_{NVT}(X_t)}{\rho_{NVT}(X_n)} = \exp\left(-\frac{\Delta E}{kT}\right)$$

donde ΔE es el cambio de energía del sistema asociado a la movida de prueba. Si $r > 1$, entonces la movida es aceptada (es decir que debemos poner $X_{n+1} = X_t$), mientras que si $r < 1$ el paso es aceptado con probabilidad r . Esto último se hace comparando r con un número aleatorio η uniformemente distribuido en el intervalo $[0, 1]$ y aceptando el movimiento si $\eta < r$. Si el paso de prueba no es aceptado entonces ésta es rechazada y ponemos $X_{n+1} = X_n$. Este procedimiento genera X_{n+1} y a partir de éste podemos proceder a generar X_{n+2} por el mismo procedimiento, haciendo una movida de prueba a partir de X_{n+1} . Cualquier punto arbitrario puede servir como punto de partida X_0 para la caminata al azar.

Mostremos ahora por qué el algoritmo de Metrópolis genera una secuencia de estados que converge a la función de distribución $w(X)$. Consideremos un gran número de caminantes que se mueven en el espacio de las fases independientemente unos de otros, y sea $N_n(X)$ el número de dichos caminantes en el punto X en el paso n -ésimo. Entonces el número de caminantes que pasan del punto X al punto Y en el próximo paso es

$$\Delta N(X) = N_n(X)P(X \rightarrow Y) - N_n(Y)P(Y \rightarrow X)$$

$$\Delta N = N_n(Y)P(X \rightarrow Y) \left[\frac{N_n(X)}{N_n(Y)} - \frac{P(Y \rightarrow X)}{P(X \rightarrow Y)} \right]$$

donde $P(X \rightarrow Y)$ es la probabilidad de transición de X a Y en tiempo n . Esta ecuación muestra que el equilibrio se obtiene (no cambio neto en las poblaciones) si $\Delta N = 0$ con lo cual

$$\frac{N_n(X)}{N_n(Y)} = \frac{N_e(X)}{N_e(Y)} = \frac{P(Y \rightarrow X)}{P(X \rightarrow Y)}$$

de manera tal que para un gran número de pasos las poblaciones de los caminantes habrán disminuido a sus valores de equilibrio.

Debemos mostrar ahora que las probabilidades de transición del algoritmo de Metrópolis conducen a la distribución de equilibrio $N_e(X) \sim w(X)$. La probabilidad de hacer un paso de X a Y puede ser escrita como un producto de probabilidades

$$P(X \rightarrow Y) = A(X \rightarrow Y)T(X \rightarrow Y)$$

donde $T(X \rightarrow Y)$ es la probabilidad de generar un paso de prueba que conduzca de X a Y y siempre puede ser elegida de manera tal que

$$T(X \rightarrow Y) = T(Y \rightarrow X)$$

Por otro lado $A(X \rightarrow Y)$ es la probabilidad de aceptación del cambio de X a Y , de manera tal que la distribución de equilibrio satisface

$$\frac{N_e(X)}{N_e(Y)} = \frac{A(Y \rightarrow X)}{A(X \rightarrow Y)}$$

si $w(X) > w(Y)$ entonces $A(Y \rightarrow X) = 1$ ya que la probabilidad de estar en X es mayor que la de estar en Y y

$$A(X \rightarrow Y) = \frac{w(Y)}{w(X)}$$

si $w(X) < w(Y)$ entonces $A(X \rightarrow Y) = 1$ y

$$A(Y \rightarrow X) = \frac{w(X)}{w(Y)}$$

De manera que en ambos casos la población de equilibrio satisfará

$$\frac{N_e(X)}{N_e(Y)} = \frac{w(X)}{w(Y)}$$

6.6 El modelo de Ising 1D

Consideremos un sistema de N espines interactuantes dispuestos regularmente en forma de red. En la presencia de un campo magnético, B , el Hamiltoniano del sistema puede ser escrito en una primera aproximación como:

$$H = -B\mu \sum_{i=1}^N s_i - J \sum_{\langle i,j \rangle} s_i s_j \quad (6.12)$$

donde $s_i = \pm 1$, μs_i , es el momento magnético en la dirección del campo y J , es la constante de acoplamiento espín-espín. El símbolo $\langle ij \rangle$ indica que la suma se lleva a cabo sobre pares de espines ubicados en sitios vecinos próximos.

Debe notarse que cuando $J > 0$, la configuración energéticamente más favorable corresponde a una en la cual todos los espines se encuentran alineados. Por lo tanto podemos anticipar que por debajo de una cierta temperatura el sistema presentará un momento magnético resultante distinto de cero. Una medida del mismo viene dada por la magnetización, M , cuyo valor instantáneo se define como:

$$M = \sum_{i=1}^N \mu s_i \quad (6.13)$$

Ejemplo 1: Si $J > 0$ y $B = 0$, entonces la energía más baja para el modelo de Ising unidimensional es $H_0 = -NJ$ y la magnetización correspondiente será $M = N\mu$. Las propiedades macroscópicas del modelo de Ising pueden ser calculadas fácilmente a partir de la función de partición, $Q(\beta, N, B)$, en donde las variables independientes β , N y B resultan la elección más natural para este sistema.

$$Q(\beta, N, B) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_N = \pm 1} \exp(+\beta B \mu \sum_{i=1}^N s_i + \beta J \sum_{\langle i,j \rangle} s_i s_j) \quad (6.14)$$

Recordemos que la función de partición es una suma sobre todos los estados posibles y, como es de esperar, en la mayor parte de los casos resulta muy difícil, si no imposible, encontrar una expresión cerrada para la misma. En cuanto a los modelos de Ising, sólo se han podido resolver el sistema unidimensional (1D) en forma general, y el bidimensional (2D) sólo en ausencia

de campo magnético, es decir $B = 0$. Esto último fue llevado a cabo por Lars Onsager en sus épocas de estudiante de doctorado y constituyó uno de los grandes logros de la ciencia del siglo XX, de hecho, el modelo de Ising 2D en sus distintas variantes es investigado activamente en la actualidad ya que es uno de los sistemas más simples que presenta una transición de fase orden-desorden.

En estas notas nos concentraremos en el modelo de Ising 1D y alentamos al lector interesado en profundizar sus conocimientos en la mecánica estadística de sistemas reticulares a recurrir a la bibliografía citada al final. Consideremos una red unidimensional del tipo:

↓	↑	↑	...	↓	↓	↑	...	↓	↓	↑
1	2	3	...	i-1	i	i+1	...	N-2	N-1	N

El término del hamiltoniano correspondiente a la interacción entre los espines puede ser escrito como una suma que corre sobre un solo índice

$$J \sum_{i=1}^N s_i s_{i+1}$$

Debe notarse que hemos impuesto condiciones de contorno periódicas, es decir que el espín ubicado en el sitio 1 interactúa con el ubicado en el sitio N , de lo contrario la suma debería correr desde $i = 1$ hasta $i = N - 1$. En base a lo dicho anteriormente y haciendo el cambio de variables $h = \beta B \mu$ y $K = \beta J$ la función de partición del sistema luce de la siguiente forma

$$Q(h, N, K) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_N = \pm 1} \exp\left(h \sum_{i=1}^N s_i + K \sum_i s_i s_{i+1}\right) \quad (6.15)$$

Intentaremos ahora encontrar una expresión cerrada para esta suma múltiple, para lo cual recurriremos a dos métodos: el primero puede ser bautizado como método explícito o directo y es fácilmente aplicable si $h = 0$, es decir, en ausencia de campo externo. El segundo se conoce como método de la matriz de transferencia y tiene la ventaja de que permite evaluar la función de partición 1D para valores arbitrarios de h . Por otro lado, ésta fue la técnica utilizada por Onsager para resolver el sistema 2D y puede ser aplicada en algunos otros problemas de mecánica estadística, como por ejemplo el cálculo de entropías de cristales a una dada temperatura.

6.6.1 Método directo:

Consideremos el caso particular en el que $h = 0$ y el sistema no posee condiciones de contorno periódicas, en este caso la función de partición puede escribirse como:

$$Q(0, N, K) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_N = \pm 1} \exp(K(s_1 s_2 + s_2 s_3 + \cdots + s_{N-1} s_N)) \quad (6.16)$$

Sumando sobre s_N obtenemos

$$Q(0, N, K) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_{N-1} = \pm 1} \exp(K(s_1 s_2 + s_2 s_3 + \cdots + s_{N-1})) + \sum_{s_1} \sum_{s_2} \cdots \sum_{s_{N-1} = \pm 1} \exp(K(s_1 s_2 + s_2 s_3 + \cdots - s_{N-1}))$$

$$Q(0, N, K) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_{N-1} = \pm 1} \exp(K(s_1 s_2 + s_2 s_3 + \cdots + s_{N-2} s_{N-1})) \times [\exp(-K s_{N-1}) + \exp(K s_{N-1})]$$

Sumando sobre s_{N-1}

$$Q(0, N, K) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_{N-2} = \pm 1} \{ \exp(K(s_1 s_2 + s_2 s_3 + \cdots + s_{N-2})) [2 \cosh K] + \exp(K(s_1 s_2 + s_2 s_3 + \cdots - s_{N-2})) [2 \cosh K] \}$$

$$Q(0, N, K) = 2 \cosh K \sum_{s_1} \sum_{s_2} \cdots \sum_{s_{N-2} = \pm 1} [\exp(K(s_1 s_2 + s_2 s_3 + \cdots + s_{N-2})) + \exp(K(s_1 s_2 + s_2 s_3 + \cdots - s_{N-2}))]$$

$$Q(0, N, K) = 2 \cosh K \sum_{s_1} \sum_{s_2} \cdots \sum_{s_{N-2} = \pm 1} (\exp(K(s_1 s_2 + s_2 s_3 + \cdots + s_{N-2})) \times [\exp(-K s_{N-2}) + \exp(K s_{N-2})])$$

Sumando sobre s_{N-2}

$$Q(0, N, K) = 2 \cosh K \sum_{s_1} \sum_{s_2} \cdots \sum_{s_{N-3} = \pm 1} [\exp(K(s_1 s_2 + s_2 s_3 + \cdots + s_{N-3})) 2 \cosh K + \exp(K(s_1 s_2 + s_2 s_3 + \cdots - s_{N-3})) 2 \cosh K]$$

Sumando sobre $s_{N-3} \dots s_{N-2}$, obtenemos:

$$Q(0, N, K) = [2 \cosh K]^{N-2} \sum_{s_1 = \pm 1} \exp(K s_1) + \exp(-K s_1)$$

sumando ahora sobre s_1

$$Q(0, N, K) = 2[2 \cosh K]^{N-1} \quad (6.17)$$

Para el sistema con condiciones de contorno periódicas y usando un método similar se obtiene:

$$Q(0, N, K) = [2 \cosh K]^N \quad (6.18)$$

6.6.2 Método de la matriz de transferencia:

El método directo visto en la sección anterior es difícil de aplicar cuando $h \neq 0$. El método de la matriz de transferencia generaliza al anterior y puede ser aplicado fácilmente al cálculo de la función de partición y de funciones de correlación cuando $h \neq 0$. Consideremos un sistema de N espines con condiciones de contorno periódicas, esto es $s_{N+1} = s_1$, en tal caso la función de partición viene dada por la ecuación 6.15. La exponencial puede factorizarse como

$$Q(h, N, K) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_N = \pm 1} [\exp(\frac{h}{2}(s_1 + s_2) + K s_1 s_2)] [\exp(\frac{h}{2}(s_2 + s_3) + K s_2 s_3)] \cdots [\exp(\frac{h}{2}(s_N + s_1) + K s_N s_1)]$$

Podemos pensar que cada término del producto es un elemento de una matriz T de dimensiones (2×2) , esta es la llamada matriz de transferencia

$$T_{s_1 s_2} = \exp[\frac{h}{2}(s_1 + s_2) + K s_1 s_2]$$

Debe notarse que s_1 y s_2 son los índices de los elementos de matriz, así:

$$T = \begin{bmatrix} T_{11} & T_{1-1} \\ T_{-11} & T_{-1-1} \end{bmatrix} = \begin{bmatrix} e^{h+K} & e^{-K} \\ e^{-K} & e^{-h+K} \end{bmatrix}$$

Entonces la función de partición puede escribirse como:

$$Q(h, N, K) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_N = \pm 1} T_{s_1 s_2} T_{s_2 s_3} \cdots T_{s_N s_1} \quad (6.19)$$

Recordando que la potencia de orden n de una matriz T se define como:

$$T^n = \prod_{i=1}^n T \quad (6.20)$$

y que $(T^2)_{i,j} = \sum_k T_{i,k} T_{k,j}$, entonces la suma desde s_2 hasta s_N de la ecuación 6.19 es equivalente a T^N , recordando también que la traza de una matriz se

define como $Tr(T) = \sum_i T_{ii}$, vemos que la suma sobre s_1 es equivalente a tomar la traza de la matriz de transferencia, con lo cual:

$$Q(h, N, K) = Tr(T^N)$$

El problema con el que nos enfrentamos ahora es encontrar la potencia N -ésima de T , lo cual no es tarea fácil aún tratándose de una matriz de (2×2) . Cómo calcular entonces T^N ? Para responder a esta pregunta recordemos las siguientes propiedades de las matrices hermíticas y de su traza:

1. Toda matriz $A(n \times n)$ hermítica (en nuestro caso real y simétrica), es semejante a la matriz de sus autovalores, es decir que representan la misma transformación lineal. Además

$$\Lambda = C^{-1}AC$$

donde $\Lambda = diag(\lambda_1, \dots, \lambda_n)$, donde C es una matriz no singular cuya inversa es igual a su adjunta

2. Usando la propiedad de invarianza de la traza:

$$Tr(AB) = Tr(BA)$$

tenemos $Tr(\Lambda) = Tr(C^{-1}AC) = Tr(ACC^{-1}) = Tr(A)$, con lo cual

$$Tr(\Lambda) = \sum_i^n \lambda_i$$

A partir de estas dos propiedades de las matrices debería quedar claro cuál es el camino a seguir para evaluar la función de partición: Haciendo uso de la propiedad uno, se diagonaliza T' : $S^{-1}TS$

$$T' = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

donde λ_1 y λ_2 son los autovalores de la matriz de transferencia. Por lo tanto

$$Q(h, N, K) = \lambda_1^N + \lambda_2^N \tag{6.21}$$

Consideremos $\lambda_1 > \lambda_2$

$$Q(h, N, K) = \lambda_1^N \left(1 + \frac{\lambda_2^N}{\lambda_1^N}\right)$$

$$\lim_{N \rightarrow \infty} Q(h, N, K) \cong \lambda_1^N \quad (6.22)$$

Con lo cual el límite termodinámico sólo el mayor autovalor es importante. La energía libre de Helmholtz está dada entonces por

$$F(N, H, T) = -kT \ln(Q(h, N, K)) = -kTN \ln(\lambda_1) \quad (6.23)$$

Debemos dar ahora una forma explícita a λ_1 , es decir que debemos calcular los autovalores de la matriz de transferencia, lo cual se logra fácilmente a partir de la condición

$$\det \begin{vmatrix} e^{h+K} - \lambda & e^{-K} \\ e^{-K} & e^{-h+K} - \lambda \end{vmatrix} = 0$$

resolviendo esto se obtiene

$$\lambda_{1,2} = e^K (\cosh(h) \pm \sqrt{\sinh^2(h) + e^{-4K}})$$

y

$$\frac{F}{N} = -J - kT \ln(\cosh(h) \pm \sqrt{\sinh^2(h) + e^{-4K}}) \quad (6.24)$$

Resultado general para el modelo de Ising 1D en presencia de un campo magnético externo.

6.6.3 Propiedades Termodinámicas:

Comencemos fijando $h = 0$, en este caso

$$\lambda_1 = e^K (1 + e^{-2K}) = 2 \cosh(K)$$

ya que en límite termodinámico el autovalor dominante es λ_1 , la energía libre puede expresarse como

$$F = -kTN [K + \ln(1 + e^{-2K})] \quad (6.25)$$

A bajas temperaturas, $T \rightarrow 0$, $K \rightarrow \infty$

$$\frac{F}{N} = -J$$

lo cual indica que el comportamiento está dominado por la energía, mientras que a altas temperaturas $T \rightarrow \infty$, $K \rightarrow 0$ el compartamiento es puramente entrópico:

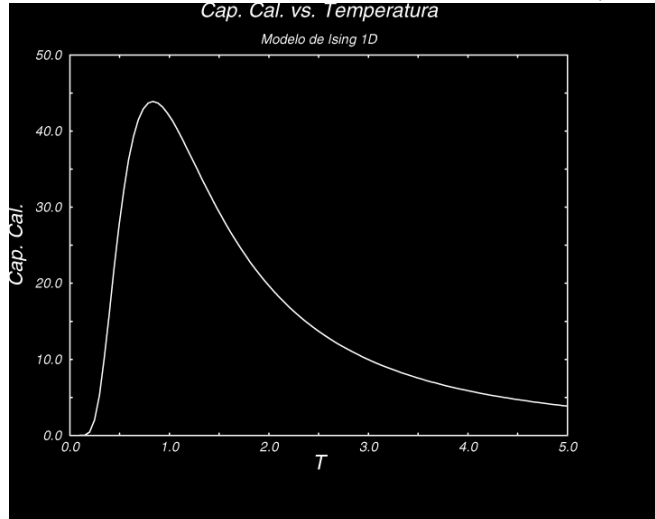
$$\frac{F}{N} = -kT \ln(2)$$

La energía interna y la capacidad calorífica se obtienen fácilmente según lo dicta la estadística en el colectivo canónico:

$$E = -\frac{\partial \ln(Q)}{\partial \beta} = -N \frac{\partial}{\partial \beta} (\ln[2 \cosh(\beta J)]) = -NJ \tanh(\beta J) \quad (6.26)$$

$$\begin{aligned} C &= \frac{dE}{dT} = -\frac{1}{kT^2} \frac{dE}{d\beta} \\ &= \frac{NJ^2}{kT^2} \operatorname{sech}^2(\beta J) \end{aligned} \quad (6.27)$$

La capacidad calorífica no exhibe ninguna singularidad ($\operatorname{sech}(x) = 1/\cosh(x) = \frac{2}{e^x + e^{-x}}$), aunque sí presenta un máximo en las proximidades de $J \sim kT$, lo cual es conocido como **anomalía de Schottky** (Figura ??).



Para calcular la magnetización escribimos la energía libre por espín en la forma

$$\frac{F}{N} = -J - kT \ln(\cosh^2(h) + \sqrt{\sinh^2(h) + w^2})$$

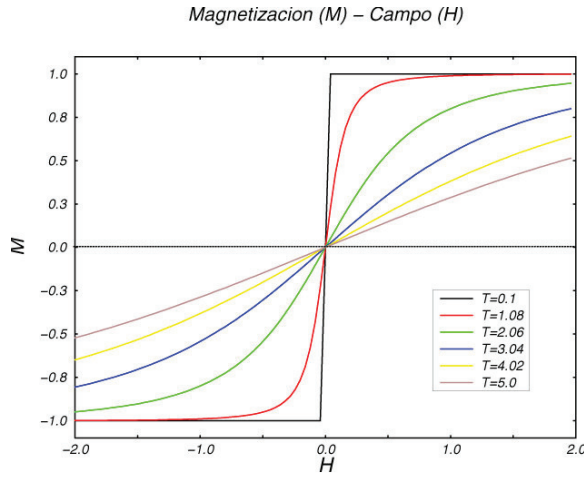


Figure 6.1:

donde $w = e^{-4K}$ es la probabilidad relativa de dos configuraciones que difieren en un espín $\uparrow\uparrow\uparrow\uparrow$ y $\uparrow\uparrow\downarrow\uparrow\uparrow$. Así

$$\begin{aligned}
 M &= -\frac{1}{N} \frac{\partial F}{\partial H} = -\frac{1}{NkT} \frac{\partial F}{\partial h} \\
 &= \frac{\partial}{\partial h} \ln[\cosh(h) + \sqrt{\sinh^2(h) + w^2}] \\
 &= \frac{\sinh(h)}{\sqrt{\sinh^2(h) + w^2}}
 \end{aligned} \tag{6.28}$$

La magnetización en función de H para distintos valores de T se muestra en la figura ??

Por último la susceptibilidad isotérmica χ_T describe cómo cambia la magnetización en respuesta a un campo externo:

$$\chi_T = \frac{\partial M}{\partial H} = \frac{\cosh(H\beta)\beta e^{-4J\beta}}{(\cosh^2(H\beta) - 1 + e^{-4J\beta})^{3/2}} \tag{6.29}$$

Los límites de la susceptibilidad magnética son los siguientes (Figura 6.2):

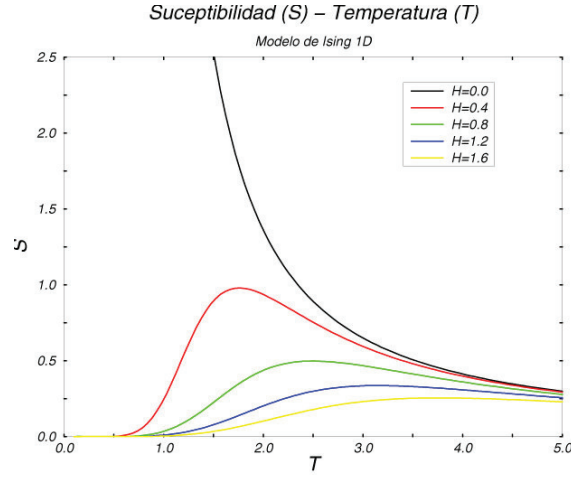


Figure 6.2:

$$\chi_T = \begin{cases} \frac{1}{kT} & T \rightarrow \infty \text{ Ley : de : Curie} \\ e^{(2J/kT)}/kT & T \rightarrow 0 \end{cases}$$

6.6.4 Correlaciones espaciales:

La función de correlación de dos puntos se define como:

$$G(i, j) = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle \quad (6.30)$$

A campo cero ($h = 0$) se cumple $G(i, j) = \langle s_i s_j \rangle$ ya que $\langle s_i \rangle = 0$. La función de correlación de dos puntos puede ser escrita de manera más transparente de la siguiente forma:

$$G(i, j) = \langle (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \rangle$$

mostrando que G mide la correlación en las fluctuaciones de los espines ubicados en diferentes sitios. $G(i, j)$ está también relacionado a la probabilidad $P_{i,j}$ de que los espines i y j tengan el mismo valor.

$$P_{ij} = \langle \delta_{s_i s_j} \rangle = \left\langle \frac{1}{2}(1 + s_i s_j) \right\rangle$$

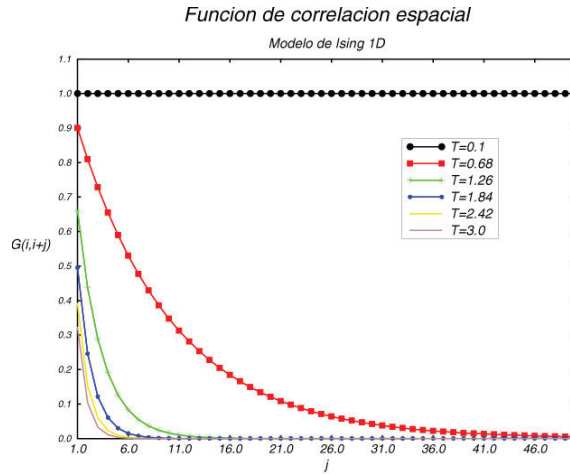


Figure 6.3:

$$= \frac{1}{2} + \frac{1}{2}[G(i, j) + \langle s_i \rangle \langle s_j \rangle]$$

Omitiremos aquí el cálculo de $G(i, j)$ para expresar únicamente el resultado

$$G(i, i + j) = (\tanh(K))^j \quad (6.31)$$

Qué es lo que podemos aprender de la función de correlación de dos puntos?. Como es de esperar existen dos comportamientos límites $T \rightarrow 0$ ($K \rightarrow \infty$), en donde $\tanh(K) = 1$ y

$$G(i, j) = 1 : \forall : j$$

Así a temperatura cero la probabilidad de que $s_i = s_j$ es igual a uno para todo j . Esto representa un estado perfectamente correlacionado y que exhibe orden de largo alcance. Hay dos estados consistentes con esta situación, en uno todos los espines apuntan hacia arriba y en el otro todos los espines apuntan hacia abajo. La función de correlación espacial para diferentes T se muestra en la figura 6.3.

Para $T \neq 0$ tenemos que $\tanh(K) < 1$ y $\coth(k) > 1$ con lo cual $\ln(\coth(K)) > 0$. Por lo tanto para $j > 0$ podemos escribir

$$G(i, i + j) = e^{-j \ln(\coth(K))} \quad (6.32)$$

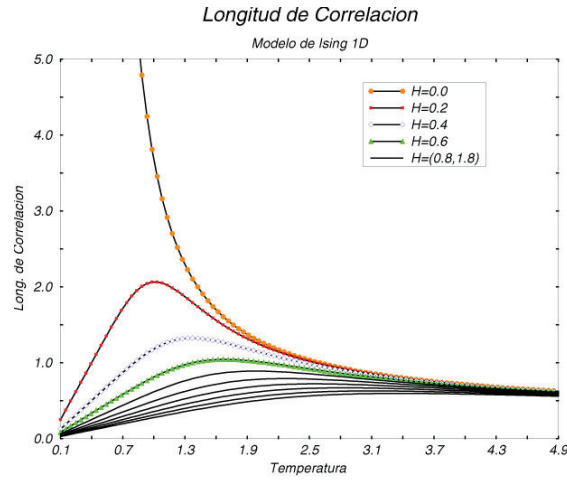


Figure 6.4:

mostrando un decaimiento exponencial para para $T > 0$. La **longitud de correlación** ξ se define como

$$G(i, i + j) = e^{-j/\xi}$$

donde ξ está medida en unidades de espaciamento de red. A partir de las dos últimas ecuaciones podemos escribir

$$\xi = \frac{1}{\ln(\coth(K))}$$

lo cual mide la distancia sobre la cual los espines están correlacionados con probabilidad ~ 1 . Cuando T tiende a cero ξ diverge exponencialmente mientras que cuando T tiende a infinito ξ tiende a cero, como puede apreciarse en la figura 6.4.