

# SFI5704 – Mecânica Estatística A

## Projeto Computacional 2: Modelo de Ising

### Instruções

Elabore o presente projeto na forma de um artigo (com 8 páginas no máximo), utilizando preferencialmente o programa `latex`. Os códigos podem ser escritos na linguagem computacional de sua preferência. Seu artigo deve conter gráficos e tabelas, resumo, introdução e conclusão, como descrito abaixo. Data de entrega: 02/12/2011.

### O Método de Monte Carlo

Em mecânica estatística, a probabilidade de uma configuração  $\psi$  para um sistema físico em equilíbrio a temperatura  $T$  é dada (no ensemble canônico) em termos da hamiltoniana do sistema  $\mathcal{H}(\psi)$  pela distribuição de Boltzmann

$$P(\psi) = \frac{e^{-\beta\mathcal{H}(\psi)}}{Z}; \quad Z = \int d\psi e^{-\beta\mathcal{H}(\psi)}; \quad \beta = 1/KT. \quad (1)$$

Médias termodinâmicas, e.g.  $E = \langle \mathcal{H}(\psi) \rangle$ , são dadas por

$$\langle A \rangle = \int d\psi A(\psi) P(\psi). \quad (2)$$

Note que esta é uma integral em muitas dimensões, extremamente difícil de calcular!

Um dos métodos para cálculo numérico de integrais consiste (como vimos no primeiro projeto) em estimar a integral por uma soma de variáveis aleatórias independentes (igualmente distribuídas) com distribuição uniforme no intervalo de integração. Por exemplo, para a integral de uma função  $f(x)$  em  $[0,1]$  temos

$$I \equiv \langle f \rangle = \int_0^1 f(x) dx \quad \rightarrow \quad \bar{I} \equiv \frac{\sum_i f(x_i)}{N}, \quad (3)$$

com  $x_i$  distribuída uniformemente em  $[0,1]$ . Na verdade, para  $N$  finito, a estimativa  $\bar{I}$  também é uma variável aleatória, que converge para o seu valor médio  $I$  com um erro (i.e. largura da distribuição) proporcional a  $1/\sqrt{N}$ , pelo teorema central do limite. De fato, pode-se demonstrar que

$$\sigma_{\bar{I}}^2 = \frac{\sigma_f^2}{N} = \frac{\langle f^2 \rangle - \langle f \rangle^2}{N}. \quad (4)$$

Resumindo, para um número grande de “sorteios”  $N$ , a soma  $\bar{I}$  calculada tende para a integral desejada, à medida que a largura da distribuição  $\sigma_{\bar{I}}$  tende a zero. Essa é a essência do método de Monte Carlo. Mas, será que o método é eficiente na prática?

O comportamento do erro com  $1/\sqrt{N}$  significa que para obter um resultado com precisão 10 vezes maior (ou seja, para ganhar um dígito de precisão) é preciso realizar um cálculo com custo computacional 100 vezes maior. Podemos argumentar que os métodos de integração determinísticos para cálculo de integrais têm erros tipicamente muito menores em função do número de operações  $N$ , como  $O(N^{-2})$  para a regra do trapezóide e  $O(N^{-4})$  para a regra de Simpson. Isto quer dizer apenas que o método de Monte Carlo não é vantajoso no cálculo de integrais *unidimensionais*. A situação é claramente diversa, porém, para integrais em muitas dimensões como na Eq. (2) acima.

De fato, para uma integral em  $d$  dimensões, o erro no caso da regra trapezoidal é de  $N^{-2/d}$ , para a regra de Simpson é de  $N^{-4/d}$ , enquanto que o erro de Monte Carlo é sempre proporcional a  $N^{-1/2}$ . O método torna-se atraente, portanto, a partir de  $d = 8$ . Qual a dimensão típica do espaço de integração para sistemas de mecânica estatística? tomando o caso muito simples de um modelo de Ising em 3 dimensões espaciais com 10 sítios por direção, temos a soma de  $10^3$  variáveis, que neste caso tomam apenas 2 valores. (Isto equivale a uma discretização grosseira da “integral” unidimensional, para comparação com os métodos determinísticos.) Pode ser frustrante calcular o tempo necessário para integração determinística desta soma: em um computador com capacidade de 1 Teraflops (i.e.,  $10^{12}$  operações de ponto flutuante por segundo, centenas de vezes mais rápido do que um PC) a conta, que envolve  $2^{1000}$  operações, levaria um tempo  $t \approx 10^{288} s$ , correspondente a aproximadamente  $10^{270}$  vezes a idade do universo! Vemos portanto que o cálculo por método de Monte Carlo não é a melhor escolha, é a *única* escolha...

Mesmo assim, calcular

$$I = \int_0^1 f(x) w(x) dx = \frac{\sum_i f(x_i) w(x_i)}{N} \quad (5)$$

[onde  $\int_0^1 w(x) dx = 1$ ] com  $x_i$  uniformemente distribuída em  $[0,1]$  [ou seja, considerar a média de  $f(x) w(x)$  na distribuição uniforme] é muito ineficiente se  $w(x)$  estiver concentrada em uma pequena região do espaço de integração. Nesse caso, é essencial poder calcular  $\sum_i f(x_i)/N$  com  $x_i$  distribuída de acordo com  $w(x_i)$ , o que é conhecido como *amostragem por importância*.

Há diversos métodos para a amostragem de variáveis aleatórias com distribuição não-uniforme a partir da distribuição uniforme em  $[0,1]$ , que é fornecida por geradores de números aleatórios para cálculo computacional. Infelizmente, porém, não há esperança de calcular a integral da Eq. (2) por amostragem direta de (grupos de) variáveis independentes para a distribuição de Boltzmann, já que isto envolveria a

amostragem de uma distribuição conjunta de um número grande demais de graus de liberdade. Para o cálculo da integral

$$\langle A \rangle = \int A(x) w(x) dx, \quad w(x) = \frac{e^{-\beta \mathcal{H}(x)}}{Z} \quad (6)$$

devemos, portanto, encontrar uma maneira de amostrar as variáveis  $x$  segundo  $w(x)$  que seja viável no caso da distribuição de Boltzmann. A solução para este problema, como descrevemos a seguir, é o chamado método de Monte Carlo dinâmico, baseado na evolução temporal de cadeias de Markov.

Já que não é possível amostrar a distribuição de Boltzmann diretamente, “inventar-se” um processo estocástico com uma evolução temporal tal que as configurações de variáveis aleatórias geradas para *simulação* do processo obedeçam à distribuição  $w(x)$  após um tempo longo. Isto pode ser feito tomando-se uma cadeia de Markov, i.e. um processo estocástico  $X_0, X_1, \dots, X_t$  com história determinada pela distribuição de probabilidade inicial  $P(X_0)$  e por probabilidades de transição  $P(X_{t+1}|x_0 \dots x_t) = P(X_{t+1}|x_t)$ , de forma que só é preciso conhecer as probabilidades de ir de um estado (e.g.  $x$ ) para outro (e.g.  $y$ ) em um passo temporal, o que é dado pelos elementos  $p_{xy}$  da *matriz de transição*  $p$ . [Note que devemos ter  $\sum_y p_{xy} = 1$  para todo  $x$  e que  $(p^n)_{xy}$  dá a probabilidade de ir de  $x$  para  $y$  em  $n$  passos temporais.] Em outras palavras, para cada tempo  $t$  só importa o estado atual do sistema, não a sua história para chegar até ele. Nessas condições, pode-se demonstrar que: se existe uma distribuição  $w(x)$  tal que  $\sum_x w(x) p_{xy} = w(y)$  para todo  $y$ , então o processo converge para a *distribuição estacionária*  $w(x)$ , independentemente de  $P(X_0)$ . Em relação ao estudo usual de processos estocásticos markovianos temos, portanto, um problema inverso: dada uma distribuição de probabilidades  $w(x)$ , encontrar  $p_{xy}$  que tenha  $w(x)$  como sua distribuição estacionária.

Lembremos [1] as condições sobre a dinâmica  $p_{xy}$  para que a cadeia de Markov convirja para a distribuição  $w(x)$  a tempos longos:

(A) irreducibilidade (ergodicidade): para quaisquer  $x, y$  há  $n$  tal que  $(p^n)_{xy} \neq 0$ .

(B) que  $w(x)$  seja estacionária, i.e. que  $p_{xy}$  satisfaça  $\sum_x w(x) p_{xy} = w(y)$ .

Note que ao invés da condição (B) acima podemos impor a condição *suficiente*

(B') balanço detalhado:  $w(x) p_{xy} = w(y) p_{yx}$ .

**Exercício:** mostre que (B')  $\Rightarrow$  (B).

Em geral, a evolução temporal das configurações do sistema (as variáveis  $x$  ou  $\psi$  nas equações acima) é feita utilizando-se um método *local*, que consiste em manter

as variáveis fixas em todos os pontos do sistema exceto em um dado ponto escolhido, que é atualizado com a distribuição de probabilidade (condicional) resultante. Uma iteração do método, i.e. um passo da cadeia de Markov, é obtida percorrendo-se todos os pontos do sistema desta forma. Esta é claramente uma maneira válida de amostrar a distribuição conjunta das variáveis em todos os pontos. Há duas formas principais de atualização da distribuição condicional para o ponto escolhido:

- **método de banho térmico:** amostragem exata da distribuição de probabilidade local, na qual os pontos vizinhos fazem o papel de um “campo” externo (banho térmico);
- **método de Metropolis,** detalhado abaixo.

Embora a amostragem local exata pelo método de banho térmico possa apresentar dificuldades dependendo da complexidade da distribuição a ser amostrada, o método de Metropolis pode ser aplicado sempre, de maneira muito simples. Desta forma, apesar de ser menos eficiente do que uma amostragem local exata, este é o principal algoritmo usado em simulações de Monte Carlo, há mais de 50 anos! [2]. A amostragem pelo método de Metropolis baseia-se em propor a mudança da variável atual  $x$  para uma nova variável  $y$ , que é aceita ou não de acordo com a seguinte regra:

- aceita se  $w(y)/w(x) \geq 1$ ;
- do contrário, aceita com probabilidade  $w(y)/w(x)$ .

A probabilidade de aceitação resultante é  $A_{xy} = \min \{1, w(y)/w(x)\}$ . A matriz de transição é então tomada como:  $p_{xy} = T_{xy} A_{xy}$ , onde  $T_{xy} = T_{yx}$  é a probabilidade de propor  $y$  como o novo valor da variável a ser atualizada.

**Exercício:** mostre que a escolha acima satisfaz o balanço detalhado.

Para a distribuição de Boltzmann, temos

$$w(x) = \frac{e^{-\beta E(x)}}{Z}, \quad (7)$$

o que implica

$$\frac{w(y)}{w(x)} = e^{-\beta \Delta E}; \quad \Delta E \equiv E(y) - E(x). \quad (8)$$

Portanto, o valor proposto  $y$  é aceito se  $\Delta E \leq 0$ ; do contrário, ele será aceito com probabilidade  $e^{-\beta \Delta E}$ . Nota: se o passo proposto for rejeitado, mantém-se o valor antigo e passa-se a um novo ponto do sistema. Quando possível, é conveniente escolher  $T_{xy}$  tal que a taxa de aceitação das mudanças propostas seja de 50%.

De acordo com o procedimento descrito até aqui, nosso objetivo será seguir a dinâmica  $X(t) = x_i$  e calcular as médias

$$\bar{A} = \frac{\sum_i A(x_i)}{N} \equiv \frac{\sum_i A_i}{N} \rightarrow \langle A \rangle = \int A(x) w(x) dx, \quad (9)$$

que serão valores esperados da distribuição desejada, i.e. a distribuição de Boltzmann. Note que o transiente inicial deve ser descartado. As médias (e respectivos erros) resultantes são os resultados da simulação de Monte Carlo.

Ainda há, porém, um problema: as amostras obtidas não serão independentes. Isto significa que não podemos garantir que  $\bar{A}$  vai convergir para  $\langle A \rangle$  com largura  $\sim 1/\sqrt{N}$ , como tínhamos demonstrado na Eq. (4). O procedimento acima está sujeito a erros sistemáticos. Devemos portanto levar em conta a correlação entre diferentes amostras (diferentes passos da cadeia de Markov), dada por

$$C(k) = \frac{\langle A_i A_{i+k} \rangle - \langle A_i \rangle^2}{\langle A_i^2 \rangle - \langle A_i \rangle^2}, \quad (10)$$

de forma que as amostras serão independentes após passado um tempo  $k$  tal que  $C(k) \approx 0$ . Pode-se concluir que a simulação deve ser feita tomando-se a média apenas sobre valores  $A_i$  efetivamente independentes (ou seja, mantendo apenas uma em cada  $k$  medidas).

Alternativamente, podemos considerar todas as medidas  $A_i$  e calcular o efeito disso para a variância de  $\bar{A}$ . Temos

$$\sigma_{\bar{A}}^2 = \frac{\sigma_A^2}{N} \left[ 1 + 2 \sum_{k=1}^{N-1} C(k) \right] = \frac{\sigma_A^2}{N} (2\tau), \quad (11)$$

com a correlação temporal  $C(k)$  dada acima. Neste caso, consideramos  $N$  amostras, mas é como se contássemos apenas  $N/(2\tau)$  delas. (Correspondendo ao número de amostras efetivamente independentes.) Pode-se demonstrar que  $\tau$  é o tempo efetivo de auto-correlação para o observável  $A$ . De fato, supondo um decaimento  $C(k) = e^{-k/\tau}$  para  $\tau$  grande (mas  $\tau \ll N$ ) obtemos

$$\begin{aligned} 1 + 2 \sum_{k=1}^{N-1} C(k) &\approx 2 \sum_{k=0}^{\infty} e^{-k/\tau} - 1 \\ &\approx 2\tau \int_0^{\infty} e^{-u} du - 1 \approx 2\tau, \end{aligned} \quad (12)$$

o que justifica a definição  $\tau \equiv 1/2 + \sum_{k=1}^{N-1} C(k)$  para o tempo de auto-correlação  $\tau$ . (Vemos que o tempo  $k$  para obter  $C(k) \approx 0$  deve ser da ordem de  $\tau$ .)

De qualquer forma, o “preço” a pagar por considerar amostras dependentes entre si é o fator  $2\tau$  a mais no número de amostras necessário para obter uma certa precisão nos resultados. (Ou seja, o custo computacional é proporcional a  $\tau$ .) Isto representa um problema para a simulação da região crítica do sistema pois, da mesma forma que há divergência da correlação espacial (o comprimento de correlação  $\xi$  tende a infinito quando a temperatura  $T$  tende a seu valor crítico  $T_c$ ), há divergência da correlação temporal  $\tau$  quando  $T \rightarrow T_c$ . Este fenômeno é conhecido como *critical*

*slowing-down*. Na prática, como os sistemas simulados são finitos, não há uma verdadeira divergência, mas um comportamento crescente com o tamanho linear do sistema (e.g. lado da rede)  $L$ , da forma  $\tau \sim L^z$ , com  $z \approx 2$  para os métodos de Monte Carlo usuais. (Note que o valor 2 para o expoente  $z$  pode ser entendido em termos de um fenômeno de difusão.)

Em **resumo**, vimos que no cálculo de médias termodinâmicas pelo método de Monte Carlo a integral pode ser estimada por uma soma de variáveis aleatórias

$$\int f(x) d\mu, \quad d\mu = \frac{e^{-\beta\mathcal{H}(x)}}{Z} dx \quad \Rightarrow \quad \frac{1}{N} \sum_{i=1}^N f(x_i)$$

onde  $x_i$  tem distribuição estatística  $\mu$ . A amostragem direta (método de Monte Carlo estático) corresponderia a um erro  $\sim 1/\sqrt{N}$ , mas não é viável. A alternativa é a simulação de uma cadeia de Markov com distribuição de equilíbrio  $\mu$  (erro  $\sim \sqrt{\tau/N}$ ), onde o tempo de auto-correlação  $\tau$  está relacionado ao fenômeno de critical slowing-down.

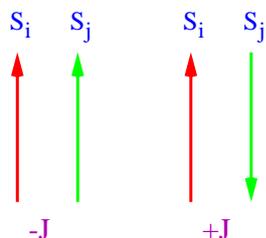
Note que a *simulação* do modelo estudado, seguindo a sequência de configurações dada pelo algoritmo, pode parecer semelhante a um experimento “teórico” — como a evolução de um processo estocástico, feita no primeiro projeto — sendo frequentemente usado o termo “medir” um observável ao invés de “calcular”. É necessário lembrar, porém, que o modelo original (i.e. a hamiltoniana que define o peso de Boltzmann) representa um sistema em equilíbrio, sem dependência temporal, e a dinâmica considerada aqui foi introduzida artificialmente, por conveniência, para o cálculo de médias na distribuição desejada. (Essa dinâmica não corresponderá necessariamente à dinâmica física do sistema fora do equilíbrio.)

A correlação entre amostras (introduzida pela dinâmica markoviana) deve ser levada em conta no cálculo do erro associado aos resultados. Isto pode ser feito de duas maneiras:

- podem-se “descorrelacionar” as amostras, tomando-se um espaçamento entre elas de forma que as mesmas sejam independentes, através de uma análise prévia da correlação temporal  $C(k)$ . O erro será então calculado como anteriormente, pelo desvio padrão para as amostras escolhidas. (São também usados métodos mais sofisticados, como *jack-knife* e *bootstrap* para observáveis mais complicados do que as médias discutidas aqui.)
- ou, consideram-se todas as amostras, mas levando em conta as correlações. Isto é feito com métodos específicos, como *binning* [3] e *self-consistent windowing* [1]. Note que o cálculo direto de  $\tau$  a partir da soma de correlações como definido acima não será confiável, devido a flutuações estatísticas para os valores grandes de  $k$ .

## Aplicação ao Modelo de Ising

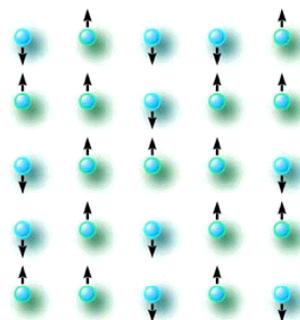
Considere o modelo de Ising [4], em que spins  $S_i$  em uma rede tomam valores  $\pm 1$  e interagem entre si e com o campo magnético externo de acordo com a hamiltoniana dada abaixo, onde a soma em  $i, j$  é tomada sobre pares de primeiros vizinhos, contando cada par uma vez só.



$$\mathcal{H}(S) = -J \sum_{\langle i,j \rangle} S_i S_j - H \sum_i S_i.$$

Vemos que para o modelo ferromagnético, i.e. para  $J > 0$ , os spins preferem estar alinhados com seus vizinhos. Observáveis de interesse para o modelo podem ser calculados a partir da configuração de spins, como abaixo.

- Energia:  $E = \langle \mathcal{H}(S) \rangle$
- Calor Específico:  $C_V = \partial E / \partial T$
- Magnetização:  $M = \langle \sum_i S_i \rangle$
- Suscetibilidade:  $\chi = \partial M / \partial H$



**Projeto:** escreva programas simples para simular o modelo de Ising bidimensional usando os métodos de Metropolis e banho térmico. Calcule a magnetização do sistema, e seu erro estatístico. (Escolha uma condição inicial para a rede de spins, por exemplo todos os spins “congelados” em um dado valor, ou spins escolhidos aleatoriamente.) Compare a eficiência dos métodos, fazendo gráficos da dependência da magnetização com o campo magnético e a temperatura. Compare o resultado da magnetização a campo zero com a solução exata do modelo. Você consegue extrair o expoente crítico  $\beta$  a partir dos seus dados?

Discuta os seguintes fatores que podem influenciar sua simulação:

- cálculo dos erros (dependência com a temperatura, tamanho da rede, algoritmo utilizado)
- dependência em relação à condição inicial
- dependência com o volume
- será que a definição utilizada para calcular a magnetização é confiável? (note

que a distribuição das magnetizações medidas na simulação pode ser obtida de um histograma)

## Apêndice

Metropolis method for the Ising model: sweep over the lattice, at each site propose to flip the spin, i.e.  $S_i \rightarrow -S_i$ . Acceptance probability

$$\frac{e^{-\beta E(y)}}{e^{-\beta E(x)}} = \frac{e^{+\beta J S_i \sum_{j \text{ n.n. } i} S_j}}{e^{-\beta J S_i \sum_{j \text{ n.n. } i} S_j}} = e^{2\beta J S_i h_i}$$

an iteration consists of a complete “sweep” over the lattice. At the end of the iteration compute  $A(S)$  for the generated configuration, and restart the process of generating configurations

Heat-bath method for the Ising model: exact sampling of the conditional probability at site  $i$ . Sweep over the lattice, at each site pick a new value for  $S_i$  independently of the old one, keeping all other spins fixed. Unnormalized probability  $P(S_i) = e^{\beta J S_i \sum_{j \text{ n.n. } i} S_j} \times \text{const.}$

Thus

- $P(S_i = +1) = e^{\beta J h_i} / (e^{\beta J h_i} + e^{-\beta J h_i}) \equiv p$
- $P(S_i = -1) = 1 - p$

## Referências

- [1] Sokal, A. (1997). “Monte Carlo methods in statistical mechanics: foundations and new algorithms”. In: Functional integration (Cargèse, 1996), Plenum, New York, pp. 131–192.
- [2] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. (1953). “Equations of State Calculations by Fast Computing Machines”. Journal of Chemical Physics **21** (6), pp. 1087–1092.
- [3] Wolff, U. (2004). “Monte Carlo errors with less errors”. Computer Physics Communications **156** (2), pp. 143–153. Erratum: Computer Physics Communications **176** (5), p. 383 (2007).
- [4] Huang, K. (1987). “Statistical Mechanics” (John Wiley & Sons, New York).