THE MONTE CARLO METHOD IN QUANTUM STATISTICAL MECHANICS

By D. C. HANDSCOMB

Received 4 April 1962

ABSTRACT. This paper describes, in general terms, a Monte Carlo method for estimating statistical parameters of quantum-mechanical systems. In this method, we construct a Markov chain of transitions between finite sequences of indices, and obtain these parameters in terms of parameters of the limit distribution. This is an extension of the method of Metropolis for classical systems, and may have equally wide application.

1. The Monte Carlo method has been applied in the past to several problems of classical statistical mechanics (see, for example, **1, 2, 4**), and has succeeded in producing quite good estimates of the parameters of such systems as the hard-sphere gas model and the Ising model of ferromagnetism. These applications all are based on the idea, first put forward in 1953 by Metropolis and his collaborators (**4**), of generating a required probability distribution in configuration space as the limit distribution of a Markov chain. A system of reversible transitions between configurations is set up such that, for any two configurations C and C',

$$\pi(C)\,\mathrm{pr}\,(C \to C') = \pi(C')\,\mathrm{pr}\,(C' \to C). \tag{1}$$

Provided that the system is irreducible and aperiodic, it has a limit distribution whose probabilities are proportional to $\pi(C)$. The mean value of any observable is then estimated by calculating its average value over the configurations occurring in a realization of the Markov chain, which estimate will improve as sampling progresses. The attractive feature of this method is that it requires one to know only the ratios, not the absolute values, of the limit probabilities $\pi(C)$.

It is natural to ask whether a similar method can be found for a quantum-mechanical system. The obvious approach is to replace the configurations C in the classical method by the eigenstates of the Hamiltonian, but in general these eigenstates will not be readily obtainable. Instead of this, we shall describe a method in which we take samples from a space of finite sequences of indices, transforming the observables accordingly. The manner in which the distribution is generated bears a strong resemblance to the method of Metropolis, except that the transitions are no longer required to be reversible.

2. We start by supposing that the Hamiltonian operator of the system may be expressed in the form

$$H = H_0 + \sum_{i=1}^{N} H_i, \tag{2}$$

where H_0 is a scalar constant and N is a large but finite number. The mean value of any operator O is then given by the standard expression

$$\langle O \rangle = \frac{\mathrm{trace}\,\{O \exp(-\beta H)\}}{\mathrm{trace}\,\{\exp(-\beta H)\}}, \tag{3}$$

where $\beta = (kT)^{-1}$,

$$= \frac{\sum_{r=0}^{\infty} \frac{(-\beta)^r}{r!} \sum_{C_r} \text{trace}\,(OH_{i_1} \dots H_{i_r})}{\sum_{r=0}^{\infty} \frac{(-\beta)^r}{r!} \sum_{C_r} \text{trace}\,(H_{i_1} \dots H_{i_r})}, \tag{4}$$

where C_r denotes any sequence $i_1, i_2, \dots, i_r$ of r indices in the range $1 \leqslant i \leqslant N$.

We define the sample space to consist of the set of all such sequences C_r for all non-negative values of r. Let

$$\Omega(C_r) = \frac{\text{trace}\,(OH_{i_1} \dots H_{i_r})}{\text{trace}\,(H_{i_1} \dots H_{i_r})}. \tag{5}$$

Then

$$\langle O \rangle = \mathscr{E}\{\Omega(C_r)\}, \tag{6}$$

the classical expectation value, when the probability of selection of each C_r is proportional to

$$\pi(C_r) = \frac{(-\beta)^r}{r!} \text{trace}\,(H_{i_1} \dots H_{i_r}), \tag{7}$$

which must be non-zero if $\Omega(C_r)$ is to be well defined. For (7) to define a proper distribution, we must have

$$\left.\begin{aligned} &\pi(C_r) > 0, \quad \text{for all } C_r \\ \text{and} \quad &\sum_r \sum_{C_r} \pi(C_r) < \infty. \end{aligned}\right\} \tag{8}$$

A sufficient condition for (8) to hold is that

$$\frac{\text{trace}\,(H_{i_1} \dots H_{i_{r+1}})}{\text{trace}\,(H_{i_1} \dots H_{i_r})} \tag{9}$$

should be strictly negative and uniformly bounded.

Our method of estimating the mean value of O is, then, to set up a Markov chain in C_r-space, having a limit distribution whose probabilities are proportional to $\pi(C_r)$, and to calculate the average value of $\Omega(C_r)$ over the sequences occurring in a realization of this chain.

In the particular case where $O = H$, we have also

$$\langle H \rangle = \mathscr{E}\left\{H_0 - \frac{r}{\beta}\right\}. \tag{10}$$

This does not necessarily lead to the best way of estimating the mean energy, but it does also give us the information that the mean length of the sequences C_r is

$$\mathscr{E}\{r\} = \beta(H_0 - \langle H \rangle). \tag{11}$$

3. Let $p(i)$ be any N numbers which satisfy

$$\left.\begin{aligned} &p(i) > 0 \quad \text{for} \quad 1 \leqslant i \leqslant N, \\ \text{and} \quad &\sum_{i=1}^{N} p(i) = 1, \end{aligned}\right\} \tag{12}$$

let λ be any positive constant, and define

$$\tau(C_r) = \frac{(-\beta\lambda)^r \operatorname{trace}(H_{i_1} \dots H_{i_r})}{p(i_1) \dots p(i_r)}, \tag{13}$$

which will then be positive. (The condition (9) is then equivalent to the condition that

$$\tau(C_r i)\, p(i)/\tau(C_r) \tag{14}$$

should be uniformly bounded.) Also define

$$f_r = e^{-1/\lambda} \lambda^r \int_0^{1/\lambda} e^t t^r dt, \tag{15}$$

so that

$$f_r = 1 - \lambda r f_{r-1}, \quad \text{for all } r > 0. \tag{16}$$

We now set up a Markov chain on the sample space, defining the probability of transition from the present state C_r to each possible new state as follows.

(i) At each step we first select the 'forward' direction with probability f_r, or the 'backward' direction with probability $(1 - f_r)$.

(ii) If the forward direction is selected, we next choose an index i according to the probability distribution $p(i)$. If $\tau(C_r i) \geqslant \tau(C_r)$, the new state is $C_r i$, which denotes the sequence $i_1, i_2, \dots, i_r, i$. If $\tau(C_r i) < \tau(C_r)$, then the new state is taken to be $C_r i$ with probability $\tau(C_r i)/\tau(C_r)$ only; otherwise it is taken to be C_r itself again.

(iii) If the backward direction is selected, and if $r > 0$, let $C_r = iC'_{r-1}$. Then, if $\tau(C'_{r-1}) > \tau(C_r)$, the new state is C'_{r-1}. If $\tau(C'_{r-1}) \leqslant \tau(C_r)$, the new state is taken to be C'_{r-1} with probability $\tau(C'_{r-1})/\tau(C_r)$ only, but otherwise it is taken to be $C'_{r-1} i$, so that C_r is cyclically permuted. If the backward direction is selected when $r = 0$, the new state is always taken to be C_0.

This completes the definition of the Markov chain.

4. We now prove that the Markov chain defined in § 3 has a limit distribution whose probabilities are proportional to $\pi(C_r)$.

The only possible one-step transitions to the state $C_r i$ are the following:

(i) from $C_r i$ itself, with probability

$$\sum_{j:\, \tau(C_r i) > \tau(C_r ij)} \left\{1 - \frac{\tau(C_r ij)}{\tau(C_r i)}\right\} p(j) f_{r+1}; \tag{17}$$

(ii) from C_r, with probability

$$\left.\begin{aligned} p(i) f_r \quad &\text{if} \quad \tau(C_r i) \geqslant \tau(C_r), \\ p(i) \frac{\tau(C_r i)}{\tau(C_r)} f_r \quad &\text{if} \quad \tau(C_r i) < \tau(C_r); \end{aligned}\right\} \tag{18}$$

(iii) from $jC_r i$, with probability

$$\left.\begin{aligned} (1 - f_{r+2}) \quad &\text{if} \quad \tau(C_r i) > \tau(jC_r i), \\ \frac{\tau(C_r i)}{\tau(jC_r i)} (1 - f_{r+2}) \quad &\text{if} \quad \tau(C_r i) \leqslant \tau(jC_r i); \end{aligned}\right\} \tag{19}$$

(iv) from iC_r, if $\tau(iC_r) \geqslant \tau(C_r)$, with probability

$$\left\{1 - \frac{\tau(C_r)}{\tau(iC_r)}\right\}(1 - f_{r+1}). \tag{20}$$

We form the sum

$$\sum_s \sum_{C_s'} \mathrm{pr}\,(C_s' \to C_r i)\,\pi(C_s'), \tag{21}$$

making use of the identities

$$\pi(C_r i) = \pi(iC_r) \quad \text{and} \quad \tau(C_r i) = \tau(iC_r). \tag{22}$$

Combining terms which satisfy equivalent inequalities, we have

$$\begin{aligned}
&\sum_s \sum_{C'} \mathrm{pr}\,(C_s' \to C_r i)\,\pi(C_s') \\
&\qquad = \sum_{j:\,\tau(C_r i) > \tau(C_r ij)} \left[\pi(C_r i)\left\{1 - \frac{\tau(C_r ij)}{\tau(C_r i)}\right\} p(j) f_{r+1} + \pi(C_r ij)\,(1 - f_{r+2})\right] \\
&\qquad\quad + \sum_{j:\,\tau(C_r i) \leqslant \tau(C_r ij)} \left[\pi(C_r ij)\,\frac{\tau(C_r i)}{\tau(C_r ij)}\,(1 - f_{r+2})\right] \\
&\qquad\quad + \left[\pi(C_r)\,p(i) f_r + \pi(C_r i)\left\{1 - \frac{\tau(C_r)}{\tau(C_r i)}\right\}(1 - f_{r+1})\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad \tau(C_r i) \geqslant \tau(C_r)
\end{aligned}$$

or

$$+ \left[\pi(C_r)\,p(i)\,\frac{\tau(C_r i)}{\tau(C_r)}\,f_r\right] \quad \text{if} \quad \tau(C_r i) < \tau(C_r). \tag{23}$$

But, from (7), (13), and (16), we have

$$\frac{\tau(C_r i)}{\tau(C_r)} = \frac{\lambda(r+1)}{p(i)}\,\frac{\pi(C_r i)}{\pi(C_r)} = \frac{(1 - f_{r+1})}{p(i) f_r}\,\frac{\pi(C_r i)}{\pi(C_r)}. \tag{24}$$

By means of this substitution, we may combine the terms of the expression (23) still further, reducing it to the form

$$\sum_{j=1}^{N} [\pi(C_r i)\,p(j) f_{r+1}] + \pi(C_r i)\,(1 - f_{r+1}) = \pi(C_r i), \quad \text{since} \quad \Sigma p(j) = 1. \tag{25}$$

We have thus established that

$$\sum_s \sum_{C'} \mathrm{pr}\,(C_s' \to C_r i)\,\pi(C_s') = \pi(C_r i), \tag{26}$$

and we may similarly verify the special case

$$\sum_s \sum_{C_s'} \mathrm{pr}\,(C_s' \to C_0)\,\pi(C_s') = \pi(C_0). \tag{27}$$

The Markov chain defined by these transition probabilities is clearly irreducible and aperiodic, and has only an enumerable number of states, and we have assumed (8) that $\pi(C_r) > 0$ and $\sum_r \sum_{C_r} \pi(C_r) < \infty$. It follows (Foster (3), Theorem 1) that the chain is ergodic, having the limit distribution required.

5. As we remarked in § 1, the transitions of this Markov chain are not reversible. We could have made them so, and somewhat simplified the arguments of § 3 and §4, by making both forward and backward steps affect one end only of the sequence C_r. Practical experience, however, has convinced us that only a method such as we describe will give sufficiently representative coverage of the sample space in a short time. On account of the low connectivity of its possible transitions, the simpler method is inclined to generate many steps confined to one small region of the sample space, before moving on to another region.

The practical usefulness of the proposed method depends upon the ease with which the trace-ratios $\Omega(C_r)$ and $\tau(C_r i)/\tau(C_r)$ may be determined, and the requirement that $\pi(C_r) > 0$. The work of determining the latter ratio may be reduced, or the convergence of the Markov chain speeded up, by a wise choice of λ and $p(i)$, which are entirely at our disposal. At the present time, the method has been applied to one problem only, the simple Heisenberg ferromagnet, where it has met with reasonable success. We intend to discuss this application in a future paper.

The author wishes to thank the United Kingdom Atomic Energy Authority for the award of a research studentship under which this work was carried out, and also to acknowledge his debt to Dr J. M. Hammersley for inspiration and advice.

REFERENCES

(1) EHRMAN, J. R., FOSDICK, L. D. and HANDSCOMB, D. C. Computation of order parameters in an Ising lattice by the Monte Carlo method. *J. Mathematical Phys.* 1 (1960), 547–558.

(2) FISHER, I. Z. Applications of the Monte Carlo method in statistical physics (translated from the Russian). *Soviet Physics, Uspekhi*, 2 (1960), 783–796.

(3) FOSTER, F. G. On the stochastic matrices associated with certain queuing processes. *Ann. Math. Statist.* 24 (1953), 355–360.

(4) METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21 (1953), 1087–1092.

OXFORD UNIVERSITY COMPUTING LABORATORY
9 SOUTH PARKS ROAD
OXFORD