Markov chains and random walks

1 Choosing at random

Suppose I have a fair coin. How can I choose a random Latin square of order 99?

A fair coin is a device which can in one step (or toss) produce one bit of information (a 0 or a 1, or informally, "heads" or "tails"), in such a way that the results of different tosses are independent – this means that each of the 2^n sequences of results produced by n tosses occurs with the same probability, namely $1/2^n$.

Given a fair coin, there is a simple algorithm for choosing a random integer x in the range $[0, 2^n - 1]$. We just toss the coin n times and interpret the sequence of bits as the expansion of x in base 2.

What about choosing an integer in the range $[0, N-1]$, where N is arbitrary? We cannot do this with a bounded number of coin tosses if N is not a power of 2, since the probability of any event defined by n coin tosses is a rational number with denominator 2^n. So we have to make a small compromise, as follows. Choose n to be the least integer such that $2^n \geq N$. Choose an integer in the range $[0, 2^n - 1]$ as before. If it is smaller than N, we accept this result; otherwise we try again, and continue until a result is obtained. It is not hard to show that, if $p = N/2^n$, then

- the resulting integer is uniformly distributed in the range $[0, N-1]$;
- the expected number of attempts is $1/p$;
- the probability that more than m attempts are required is q^m, where $q = 1 - p$.

(The last two statements follow because the number of attempts is a geometric random variable). Since $p > 1/2$, the expected number of attempts is less than 2 and the probability of needing a long series of tries is exponentially small.

Now we can choose a random structure of a certain type in some situations. If we can count the structures, then we may suppose there are N altogether; choose a random number x from $[0, N-1]$, skip over the first x structures in the count, and take the next one. If each structure is determined by a sequence of choices, and making these choices uniformly gives the uniform distribution, then we can make the choices at random as long as we know how many choices there are at each stage.

For example, how can we choose a random permutation σ of the set $\{1, \ldots, n\}$? The image 1σ of 1 can be any of $1, \ldots, n$; choose this image at random. Then 2σ

can be any of $1,\ldots,n$ except 1σ; choose a random number x from $1,\ldots,n-1$ and add one if $x \geq 1\sigma$, then set $2\sigma = x$. Continuing in this way gives the required random permutation.

Choosing a random graph on n vertices is even easier: simply decide with a single coin toss whether each pair of vertices is joined by an edge or not. (Indeed, the number of graphs is $2^{n(n-1)/2}$, and counting them is equivalent to choosing one at random.)

In other cases, e.g. Latin squares, we don't even know how many structures there are, so choosing one at random cannot be done by these methods; we need a new idea.

2 Markov chains and random walks

We consider only Markov chains with a finite number of states. Let $S = \{s_1,\ldots,s_m\}$ be a finite set of *states*. Suppose we are given a matrix $P = (p_{ij})$ of order m, whose entries are non-negative real numbers satisfying

$$\sum_{j=1}^{m} p_{ij} = 1.$$

The interpretation is that we have a marker on one of the states; at a certain moment it moves to a new state, where the probability of moving from state s_i to state s_j is p_{ij}. The displayed equation just reflects the fact that the marker is bound to move to some state!

We can now iterate this procedure. The marker starts out on some state, possibly chosen at random from an arbitrary probability distribution. At each positive integer time it makes a transition according to the specification above. We are interested in how it behaves in the long term. There are two extremes of behaviour:

- Suppose that $p_{ii+1} = 1$ for $i = 1,\ldots,m-1$ and $p_{m1} = 1$, all other probabilities being zero. Then the marker simply marches around the m-cycle in a mechanical way; if it starts at state s_i then after n steps it is certainly at state s_j, where $j \equiv i+n \pmod{m}$.

- Suppose that $p_{ij} = 1/m$ for all i, j. Then, no matter where the marker starts, after one transition it is in a random state chosen uniformly from $s_1,\ldots,s_m$, and this remains true after any number of transitions.

For most interesting chains, we don't have either of these extremes, but instead, under certain hypotheses the marker's position approaches a limiting distribution as the number of transitions increases.

The displayed equation above can be rewritten as $Pj^\top = j^\top$, where j is the all-one vector, $j = (1, 1, \ldots, 1)$. So 1 is a right eigenvalue of P. Since the left and right eigenvalues of a matrix are the same, there is a vector $q = (q_1, \ldots, q_m)$ such that $qP = q$.

It can be proved that we can choose q to have all its entries non-negative, so we can normalise the entries so that $\sum_{i=1}^m q_i = 1$. Then we can interpret q as a probability distribution on the states. Suppose that the marker starts in this distribution. Then after one transition, its probability of being in state s_j is

$$\sum_{i=1}^{m} q_i p_{ij} = q_j,$$

that is, the same as before the transition! So if the marker starts in the distribution q, then it remains in this distribution. So q is certainly a candidate for a limiting distribution.

We need a couple of conditions on the chain to guarantee good limiting behaviour and rule out cases like the first example above. Let p_{ij}^n be the probability of moving from state s_i to state s_j after n transitions. (Exercise: this is just the (i, j) entry of the matrix P^n.) The chain is said to be *irreducible* if, for any two states s_i and s_j, there exists n such that $p_{ij}^n > 0$, that is, it is possible to move from s_i to s_j. The chain is said to be *aperiodic* if, for any state s_i, the greatest common divisor of the set

$$\{n : p_{ii}^n > 0\}$$

is equal to 1.

Theorem 1 *Let P be an irreducible and aperiodic Markov chain, and q the normalised left eigenvector of P with eigenvalue* 1*. Then, starting from an arbitrary initial distribution, the distribution after n steps approaches q as $n \to \infty$.*

The particular type of Markov chain we consider is the random walk on an undirected graph. The states are the vertices of the graph, and a transition consists of choosing an edge through the vertex on which the marker sits (all edges being equally likely) and moving to the other end of this edge. In other words, p_{ij} is the reciprocal of the valency of the ith vertex v_i if v_i and v_j are adjacent, and is zero if they are non-adjacent. It is not hard to see that the random walk is irreducible if

and only if the graph is connected, and is aperiodic if and only if the graph is not bipartite. (Since the graph is undirected, we can always return to the start vertex after 2 steps with non-zero probability.)

With our fair coin we can do a random walk on a graph, since we have to choose among a number of edges at each step, giving each edge the same probability.

It is also simple to compute the limiting state. We claim that the vector whose ith component is the valency of v_i is a left eigenvector with eigenvalue 1. This is an easy exercise; here is a heuristic argument. If the probability of starting at v_i is ck_i, where k_i is the valency of v_i and c is a constant, then the probability of passing along any given edge is c, and so the probability of arriving at v_j is ck_j.

In other words, if a graph is connected and non-bipartite, then the random walk on that graph has the property that, in the limit, the probability of being at any vertex is proportional to its valency. In particular, if the graph is regular, then the limiting distribution is uniform.

3 Random Latin squares

A *Latin square* of order n is an $n \times n$ array, each cell containing a symbol from the set $\{1,\ldots,n\}$, such that each symbol occurs once in each row and once in each column of the array.

In order to use this method to choose a random Latin square, we need to find a set of "moves" connecting up the set of all Latin squares of given order. This was done by Jacobson and Matthews [1]; we outline their method.

First we re-formulate the definition of a Latin square. We regard it as a function from N^3 to $\{0,1\}$, where $N = \{1,\ldots,n\}$, having the property that, for any given $x,y \in N$, we have

$$\sum_{z \in N} f(x,y,z) = 1, \tag{1}$$

with similar equations for the sum over y (for given x,z) and the sum over x (for given y,z). The interpretation is that $f(x,y,z) = 1$ if and only if the entry in row x and column y of the array is z.

We also have to enlarge the space in which we walk, by admitting also "improper" Latin squares. Such a square is a function from N^3 to $\{-1,0,1\}$ having the properties that it takes the value -1 exactly once, and that Equation (1) and the two similar equations hold. Effectively, we allow one "negative" entry which must be compensated by additional "positive" entries.

To take one step in the Markov chain starting at a function f, do the following:

(a) If f is proper, choose any (x,y,z) with $f(x,y,z) = 0$; if f is improper, use the unique triple with $f(x,y,z) = -1$.

(b) Let $x',y',z' \in N$ satisfy

$$f(x',y,z) = f(x,y',z) = f(x,y,z') = 1.$$

(If f is proper, these points are unique; if f is improper, there are two choices for each of them.)

(c) Now increase the value of f by one on the triples (x,y,z), (x,y',z'), (x',y,z') and (x',y',z), and decrease it by one on the triples (x',y,z), (x,y',z), (x,y,z') and (x',y',z'). We obtain another proper or improper Latin square, according as $f(x',y',z') = 1$ or 0.

Jacobson and Matthews showed that the graph which has the set of all proper and improper Latin squares as vertices and the moves of the above type as edges is connected and non-bipartite, and so the random walk on this graph has a unique limiting distribution. Moreover, in this distribution, all (proper) Latin squares have the same probability.

4 Random isomorphism classes

Another application of Markov chains was given by Jerrum [2]. Suppose that we can choose a random element of a large set (for example, a random graph on a given set of vertices). Now the number of times that an isomorphism class of objects appears in the list is inversely proportional to the number of automorphisms of an object in the class. So we are relatively unlikely to find highly symmetric objects. Jerrum gave a Markov chain whose limiting distribution makes all isomorphism types equally likely.

More generally, let G be a group which operates on a set Ω. Starting from an element $\omega \in \Omega$, a move consists of the following two operations:

- move to a random element g in the stabiliser of ω;
- move to a random element ω' fixed by g.

In other words, we form a bipartite graph on $\Omega \cup G$, in which ω is joined to g if $\omega^g = \omega$; a step in Jerrum's Markov chain consists of two steps in the random walk on this graph.

For example, suppose we want to choose a random graph on n vertices, with each isomorphism class equally likely. Here Ω is the set of all $2^{n(n-1)/2}$ graphs on the given vertex set V, and G is the symmetric group S_n consisting of all permutations of V. Starting with a graph Γ, we calculate the automorphism group G of Γ (for example, using a program such as `nauty` [3]), choose a random element g of G, and then choose a random graph Γ' preserved by g.

References

[1] M. T. Jacobson and P. Matthews, Generating uniformly distributed random Latin squares, *J. Combinatorial Design* **4** (1996), 405–437.

[2] M. R. Jerrum, Computational Pólya theory, pp. 103–118 in *Surveys in Combinatorics, 1995* (Peter Rowlinson, ed.), London Math. Soc. Lecture Notes **218**, Cambridge University Press, Cambridge, 1995.

[3] B. D. McKay, `nauty` user's guide (version 1.5), Technical report TR-CS-90-02, Computer Science Department, Australian National University, 1990.

Peter J. Cameron

May 30, 2003