Markov Chain Monte Carlo

1. Examples

Arnab Chakraborty

In Part 1 of this two-part article, we give two examples – one to illustrate the Markov chain, and the other to illustrate sampling of units in a Markovian dependent manner and explain how it can be used in a Monte Carlo sampling procedure.

Arnab Chakraborty did his BStat (Hons.) and MStat degrees from the Indian Statistical Institute (Calcutta). He is now a PhD student in the Department of Statistics at Stanford University, California. His research interest lies in areas of statistics and computational algebra. Apart from that he is interested in playing the piano and wishes he could really play well.

The Game of Ludo

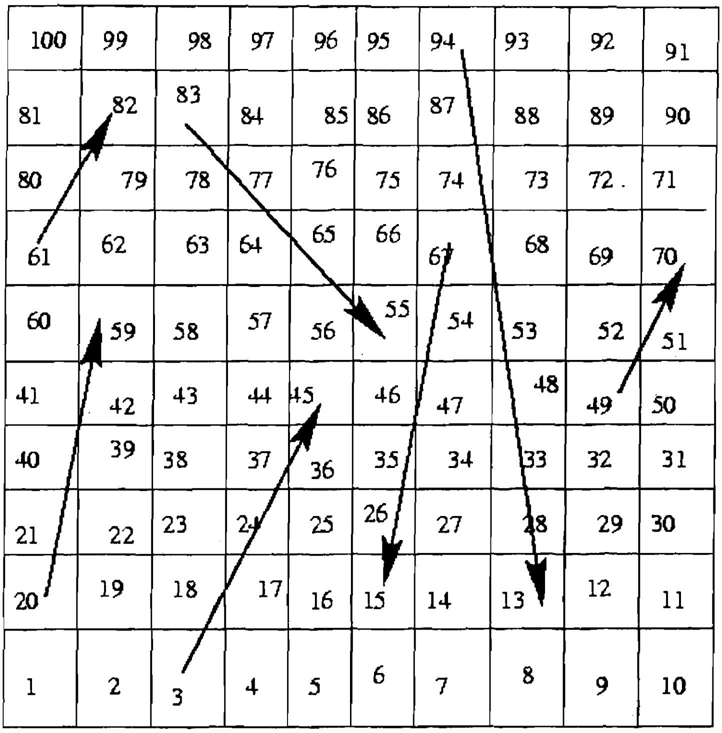
Have you ever played a game of Ludo? There are many different versions of the game – the one I liked playing may be called the Snake-n-Ladder Ludo. Some people call it as Chutes-n-Ladders, though. The game consists of a square board 10 by 10 in size, two differently coloured counters (one for each player) and a die. (*Figure* 1)

The 100 small squares are numbered from 1 to 100 as shown, and the game starts with both the counters on square 1. The players take turns at rolling the die, and advance their counters accordingly. In case the counter lands on a square with a snake's mouth in it, the counter has to move down to the tail of the snake. The reverse is the case for a ladder, which takes your counter up from the bottom to the top. The aim is to arrive at square 100 before your opponent. Obviously, this is not a game of strategy. You are simply a mute spectator to what chance does to your counter. Nevertheless, this simple game has a potential that can really do wonders to your mathematical mind and make you ask very interesting questions. We shall learn about this in this article.

Keywords.
Gibbs sampler, Markov chain Monte Carlo.

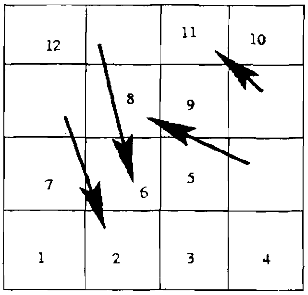
If you really like playing the game you will start asking yourself all sorts of questions like, "On an average how

Figure 1.



long does the game last?" or "How often do you get gobbled up by a particular snake?" or "Is it better to have longer ladders or shorter snakes?" In this mood I once tried to figure out what's the chance of reaching the end after, say, 30 steps. I explain it with the simple case of a 4 by 4 board and a single player (*Figure* 2).

Figure 2.



Note that I have numbered only those squares that are 'stable', *i.e.*, do not have arrows leading out from them. There are some simple things that you can compute rather easily. For instance, after the first roll you may be at 1, 2, 3, 4, 8 or 5, each with probability 1/6. Then the next roll will take you to the 6 next 'stable' squares after that point. One can draw a tree diagram for that (*Figure* 3).

So what is the chance that I ever arrive at 8 after the second roll? I may get there if my first roll shows 1 and the second roll shows 4, which has probability $1/6 \times 1/6 = 1/36$. Similarly, I can arrive at 8 after the second

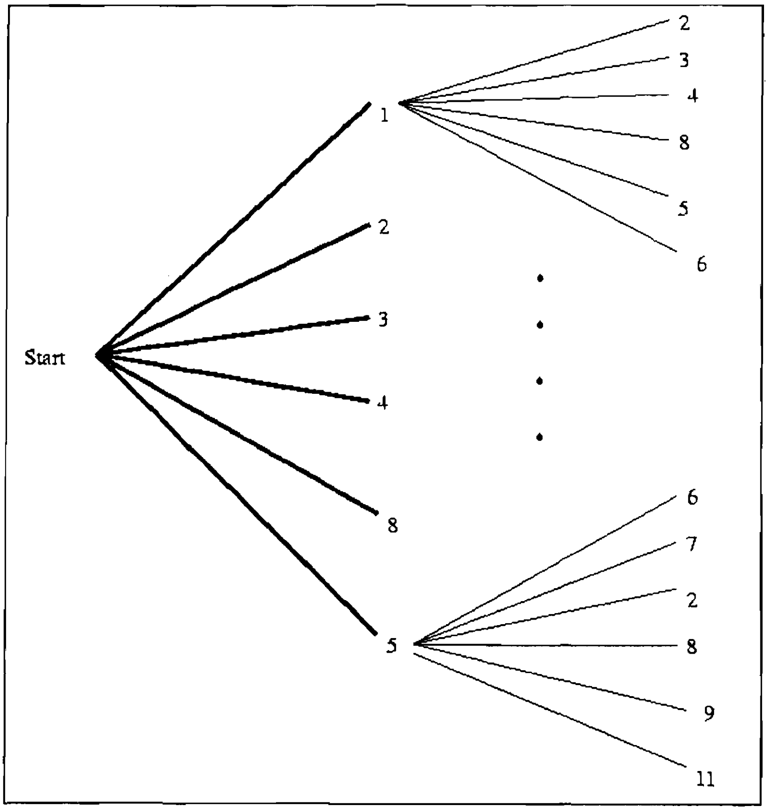


Figure 3.

roll if I get any of the pairs (2,3), (3,2), (4,1) in my first two rolls. Each such pair has a chance 1/36 to turn up. So the probability of hitting 8 after a couple of rolls is 4/36. However, such computations become mind-bogglingly complicated for more than three rolls. Fortunately, there is an elegant way of solving the problem that requires explicit computations for only single rolls. We shall discuss that method now.

Such computations become mind-bogglingly complicated for more than three rolls. Fortunately, there is an elegant way of solving the problem that requires explicit computations for only single rolls.

First make a big table for the game: a 12 by 12 table, each 'stable' square in the board getting one row and one column. We shall call each 'stable' square as a *state*. Thus we have 12 states. In the $(i, j)^{\text{th}}$ place we write the chance T_{ij} that a single roll takes me from state i to state j. This is easily computed. See the next page.

What is the chance that you are at some given state after the first roll of the die? We can be at any of the

$$
\begin{array}{cccccccccccc}
0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & 0 \\
0 & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\
0 & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{1}{6} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
$$

states 1,2,3,4,8 and 5, each with probability $\frac{1}{6}$. For a reason that will be obvious in a moment we shall write this information as a 12 dimensional vector (12 being the total number of states). This vector will be called $\mathbf{v_1}$. The subscript 1 tells us that it deals with the situation after 1 roll of the die. The i^{th} entry of $\mathbf{v_1}$ will be the chance that the counter is at square i. So $\mathbf{v_1}$ is the vector

$$\mathbf{v_1} = (\mathbf{1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 0, 0,} \qquad \mathbf{0}).$$

Similarly, we can compute $\mathbf{v_2}$, the probability vector after 2 rolls of the die.

It can be computed easily using the matrix $\mathbf{T}$ and the vector $\mathbf{v_1}$. You may be surprised to find that

$$\mathbf{v_2} = \mathbf{v_1} \times \mathbf{T}.$$

If you carefully remember the rule for matrix multiplication you will see that this is no coincidence, and indeed we have

$$\mathbf{v_t} = \mathbf{v_{t-1}} \times \mathbf{T} \quad \text{for any } \mathbf{t}.$$

Since matrix multiplication is a simple thing to do (at least with a computer) you can easily compute the probability vector $\mathbf{v_{30}}$ and read off it the chance that the

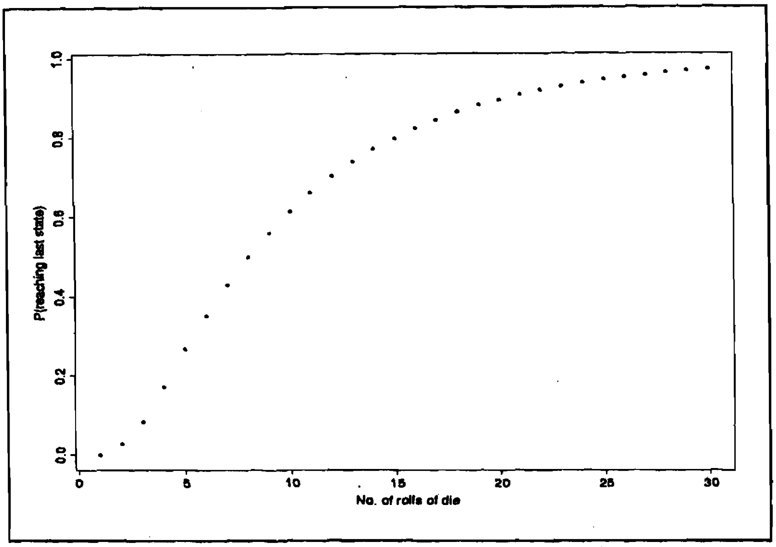


Figure 4.

counter is in the last square. The chance of reaching the last square just after the t^{th} roll is given by the last entry of $\mathbf{v_t}$. We have plotted it against t (*Figure* 4).

Note that it starts at zero, then slowly starts rising, and finally approaches one in the limit. Well, the fact that the probability approaches 1 in the limit as t tends to infinity was obvious even without all this mathematics, since it is a common experience that all games of Ludo eventually end since there is no snake to devour you at the final square. In probabilistic language, since there is a positive probability of reaching the last square, sooner or later (in a finite number of moves) the last square will be reached. Now this is a good observation since this lets us avoid all those matrix computations! Let us make a note of it.

The fact that the probability approaches 1 in the limit as *t* tends to infinity was obvious even without all this mathematics, since it is a common experience that all games of Ludo eventually end since there is no snake to devour you at the final square.

Theorem 1: Consider a game of Ludo as described above, where every stretch of 6 consecutive squares has at least one square that has no ladder or snake starting from it. If there is no snake's mouth at the last square, then whatever be the numbers of ladders and snakes, and whatever be their positions, the probability vector $\mathbf{v_t}$ approaches $(0, 0, 0, \quad 1)$ as t tends to infinity.

The method of matrix and vectors we have learned above is, however, applicable to many more things than just the plain Snake-n-Ladder Ludo.

"Can you see why the condition about 6 consecutive squares is required ?"

Of course, one needs to prove this theorem mathematically. But we shall not go into it here. However, the proof is not difficult and the interested reader is encouraged to give it a try.

The method of matrix and vectors we have learned above is, however, applicable to many more things than just the plain Snake-n-Ladder Ludo. Here is a more complicated version of a Ludo. Here at each step you toss a coin ($P(H) = 1/2$), and also roll a die. If the coin shows head, then you move forward as many squares as your die shows, else you move backwards by the same amount. If you are at square 1, and you want to move backwards, then you go to the last square. Thus the two ends of the board are tied together in a cycle. Of course, this is not very interesting as a game since it has no end. But this 'game' often occurs in practice under different disguises. We shall stop calling it a 'game' and shall consider it as a Markov chain.

A (finite) Markov chain is just like a Ludo. One main ingredient of a Markov chain is its *state space*, a finite set that serves the purpose of the board. A main point about a Markov chain is that the future sequence of states depends only on the current state and not on the past sequence of states.

A (finite) Markov chain is just like a Ludo. One main ingredient of a Markov chain is its *state space*, a finite set that serves the purpose of the board. A main point about a Markov chain is that the future sequence of states depends only on the current state and not on the past sequence of states. We need not necessarily have a die or a coin for making random moves. However, we shall have a table like the one shown above which will tell us the probability that we move from one point in the state space to another in a single step. We shall call it our *transition matrix*. To start the Markov chain we need an *initial probability vector*, which we call $\mathbf{v}_0$.

In this case what can you say about the limit of $\mathbf{v}_t$? Possibly you have guessed it already. $\mathbf{v}_t$ will tend to the (discrete) uniform distribution (on the state space) as t tends to infinity.

Theorem 2: If we have a Ludo such that we can move from any square to any square (maybe using multiple moves) and the chance of moving from i to j in a single move is the same as that of moving from j to i then $\mathbf{v_t}$ converges to uniform distribution as t tends to infinity.

Notice the difference between the Markov chains associated with the plain Ludo and this more complicated Ludo and the consequent limits of $\mathbf{v_t}$. The plain Ludo does not satisfy the conditions of this theorem because once the last square is reached, it stays there and has no chance of moving to any other square and the corresponding row of the transition matrix is $(0, 0, \quad 1)$. In the plain Ludo, the transition matrix is not symmetric either.

Martians on Earth

Once upon a time the Martians planned to attack the city of State Space, situated in the northern part of the Island of Statistica. To capture the city, the Martian Chief (MC) wanted to drop his paratroopers from a spacecraft onto the city and make them occupy various street crossings. For the mission to succeed, it was absolutely important that the Martian soldiers should be scattered throughout the city as uniformly as possible. Having too many soldiers in one part, and too few in another would simply disrupt the balance of the attack. This would have been as easy as pie, only if the MC had a good map of the city. Unfortunately, he did not. Also, as you all should know if you have read the Martian anatomy book, Martians are very short-sighted. So it was impossible for them to make a map after landing in the city at night. However, the MC had one very valuable piece of information about the city road map. All the roads in the city were either east-west (EW) or north-south (NS). The MC did not know how many roads there were, or how many crossings they formed, or what the shape of the city was. But he wanted to

For the mission to succeed, it was absolutely important that the Martian soldiers should be scattered throughout the city as uniformly as possible.

post his soldiers at the crossings as uniformly scattered throughout State Space as possible. If he had a map of the city, all that he needed to do was to label the crossings from 1 to say N and for each Martian i, select independently a crossing X_i at random (according to discrete uniform from 1 to N) and post him there. Incidentally, notice that, unlike us Earthies, Martians are gender-biased and all soldiers and the chief are males!

If he had a map of the city, all that he needed to do was to label the crossings from 1 to say N and for each Martian i, select independently a crossing X_i at random (according to discrete uniform from 1 to N) and post him there.

Furthermore, the MC was not at all sure about where the paratroopers will land after they leave the spacecraft. It depended on the cloud and wind conditions. As he was brooding about this in his tent, he suddenly recalled having once learnt about the game of Ludo that was so popular among the children of the Earth. He soon wrote some elaborate plans on a big piece of paper, and shouted "Ghuytu I faglfos!" (which in Martian means something like 'Eureka').

Once the paratroopers were ready, the MC gave them each two coins – a red one and a blue one. The coins were specially constructed so that they were perfectly balanced (*i.e.*, both sides were equally likely). The red coins were marked with NE and SW on their sides. The blue ones bore the marks NS and EW. Then the MC gave the soldiers the following instruction (which I translate into English for the benefit of those who do not know the Martian language well): "Soldiers, once you find yourselves on some road in the city of State Space, start walking down the road in any direction until you come to a crossing. Then toss the blue coin. If it shows EW, you are to take the EW road from that crossing, else the NS road. Now toss the red coin. If it shows NE, go North if you are to take the NS road, or East if you are to take the EW road. Do the opposite else. Once you decide on the road and direction, run until you come to the next crossing, and do the tossings again."

Notice that the probability is zero of landing exactly at a crossing and not having to move.

"But Sir" piped out a nervous-looking Martian soldier, who was somewhat weak in mathematics, "if there is no NS road leading from my crossing, and the blue coin shows NS, then what should I do?" – "Then you will repeat the tossings again, idiot!" was the curt reply of his chief. – "How long are we to continue this tossing and running business, MC?" asked a captain, who was holding dubious views about the sanity of his chief. "Ah, that's an important question, and the answer is: Run as fast as you can, and do as many tossings as you can until the day breaks."

At the end of which, they were indeed scattered rather uniformly over the city, despite the fact that they still did not know what the city really looked like!

Well, Martians can run and toss coins *really* fast, and that night the soldiers managed to perform one million tossings each! At the end of which, they were indeed scattered rather uniformly over the city, despite the fact that they still did not know what the city really looked like!

Well, what do you think about the MC's plan? Too complicated? Too obvious? Much ado about nothing? Let's see. What the soldiers did is called a 'random walk', for they did nothing but move around as directed by the random outcomes of the coin tosses. Consider the (finite) set of crossings as the state space. For each Martian, for each movement in one step from one crossing to another, the four 'adjacent' crossings are equally likely. This results in the conditions of Theorem 2 being satisfied and so we can conclude that after the one million crossings (t= 1 million = ∞!), the limiting distribution is reached and each Martian has equal probability of being in any of the crossings and so we could be satisfied that the innumerable Martians would be fairly uniformly distributed over the collection of crossings at day-break.

What the soldiers did is called a 'random walk', for they did nothing but move around as directed by the random outcomes of the coin tosses.

Here, lack of knowledge of the actual number of crossings and the consequent inability to carry out random sampling is overcome by the generation of a random sequence from a Markov chain with known transition probabilities (and known next states at any given state) and the limiting distribution of which is known to be the target distribution, albeit not knowing what this target distribution is (in the sense that its support is not known). Since the simulation (Monte Carlo procedure) is from a Markov chain, this procedure is called Markov chain Monte Carlo or more simply MCMC or more stylishly $(\mathrm{MC})^2$. We are not really interested in the intermediate states, but only in the 'final' state reached at infinite 'time'. We have here a large number of independent selections from the limiting uniform distribution as a result of our exercise. The Markov chain is not a part of the original problem, but is only a tool in the generation of random samples from the incompletely-known distribution.

In the next and concluding part, we investigate the mathematics behind the Markov chain Monte Carlo procedure.

Suggested Reading

[1] S Chib and E Greenberg, Understanding the Metropolis-Hastings algorithm, *The American Statistician*, Vol.49, pp.327-335, 1995.

Address for Correspondence
Arnab Chakraborty
Department of Statistics
Stanford University
California, USA.

Eigenvalues of AB and BA

Let A, B be $n \times n$ matrices with complex entries. In *Resonance*, Vol.7, p.88-93, several proofs of the fact that AB and BA have the same eigenvalues were discussed. Each proof brings out a different viewpoint and may be presented at the appropriate time in a linear algebra course. Here is another proof due to Roger Horn.

Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ be a block matrix. If A_{11} is nonsingular, then multiplying A on the right by $\begin{bmatrix} I & A_{11}^{-1}A_{12} \\ O & I \end{bmatrix}$ we get the matrix $\begin{bmatrix} A_{11} & O \\ A_{21} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$. Hence, $\det(A) = \det(A_{11})\det(A_{22} - A_{21}A_{11}^{-1}A_{12})$. [The matrix $A_{22} - A_{21}A_{11}^{-1}A_{12}$ is called the *Schur complement* of A_{11} in A. This determinant identity is one of the several places where it shows up.] In the same way, if A_{22} is invertible, then $\det(A) = \det(A_{22})\det(A_{11} - A_{12}A_{22}^{-1}A_{21})$. So, if A_{11} commutes with A_{21}, then $\det(A) = \det(A_{11}A_{22} - A_{21}A_{12})$; and if A_{22} commutes with A_{12}, then $\det(A) = \det(A_{22}A_{11} - A_{12}A_{21})$. Now let A, B be any two $n \times n$ matrices, and consider the block matrix $\begin{bmatrix} \lambda I & A \\ B & \lambda I \end{bmatrix}$. This is a very special kind of block matrix satisfying all conditions in the preceding lines. So $\det(\lambda^2 I - AB) = \det(\lambda^2 I - BA)$. This is true for all complex numbers λ. So, AB and BA have the same characteristic polynomial.

Rajendra Bhatia,
Indian Statistical Institute, New Delhi 110016.