Monte Carlo Methods — Why Do They Work? — Part I

Physicists came up with Monte Carlo methods because they do what nature does! Suppose all the air in a room is in one corner. Maybe someone just opened a compressed air tank in that corner. In a fraction of a second, or perhaps seconds if the air tank open is constricted, simply through random motion, the air molecules will spread through the room, almost perfectly uniformly.

If the room is a very tall room, the result will not be quite uniform. It will turn out that the density of the air at the top of the room is just a bit thinner than the density of the air at the bottom of the room, due to the force of gravity. So nature can even produce non-uniform probability distributions through random motion! When nature does this through random motion, and you analyze how it is happening, it is called the Principle of Detailed Balance. In my mind this principle: (a) tells you about the behavior at equilibrium, and (b) also tells you about the mechanism of the approach to equilibrium.

So nature is doing exactly what we want Monte Carlo to do: we want to produce representative samples of non-uniform probability distributions using randomness. In Chapters 13 to 16, Donovan and Mickey give us three specific Monte Carlo methods to consider:

* The Metropolis algorithm, published 1953, but developed during the Manhattan project
* The Metropolis-Hastings algorithm, published 1970, where Hastings made a critical improvement to the algorithm
* The "Gibbs Sampling" (GS) algorithm, published 1984 by Geman and Geman, who were smoothing images

Furthermore, you even have two versions of the original Metropolis algorithm:

* the version with Brian's "appropriate ratio"
* and the rest of the world's version with a slightly different appropriate ratio

The goal of this writeup is to just convincingly explain how the Metropolis algorithm works, with my version of the appropriate ratio, on something simple, like our 4-bin example of quarterly iPhone sales, and you can take it from there to demonstrate that the other, more complicated Monte Carlo methods also work. Here we go....

The Simplest Monte Carlo Algorithm

You have studied the version of the Metropolis algorithm with Brian's "appropriate ratio" and played with it enough to both:

(a) totally get what it does, and
(b) see "experimentally" that it works.

Let us summarize the Metropolis algorithm. First the situation:

* We have n bins and a set of desired probabilities p_i, where i runs from 1 to n.

* In our quarterly iPhone sales example, n was 4 and the four p_i's were 0.1, 0.2, 0.4, and 0.3.

Then the core of the algorithm (repeated *ad nauseam)* was:

Step 1. You are in bin i. You flip a coin. This determines with equal probability, $\frac{1}{2}$, whether you will propose to go to bin $i+1$ (if it is heads), or bin $i-1$ (if it is tails).

Step 2. Compute Brian's "appropriate ratio." If it was heads, and the proposed bin is $i+1$, then the appropriate ratio is $\frac{p_{i+1}}{p_i+p_{i+1}}$.

Step 3. If it was tails, the proposed bin is $i-1$ and the appropriate ratio is $\frac{p_{i-1}}{p_i+p_{i-1}}$.

Step 4. Generate a random number between 0 and 1. If the number is less than the appropriate ratio, move to the proposed bin, and make a tally there. Otherwise stay in the current bin and make another tally in the current bin.

Thinking Probabilistically Instead of Algorithmically

Suppose after running the algorithm for a while, the chance that the algorithm is currently in bin i is q_i where $i = 1, \ldots, n$. Note that the q_i do not have to be related to the p_i. That relationship is what we are trying to prove! In fact, at the beginning, we start the algorithm off somewhere, and of course the q_i are initially concentrated wherever we chose to start. We want to derive something about the q_i.

Let's think about the $j+1 \to j$ transition and the $j \to j+1$ transition. From bin $j+1$, the chance of going left and contributing on the next step to bin j is

$$q_{j+1} * \frac{1}{2} * \frac{p_j}{p_{j+1}+p_j}$$

While the chance of being in bin j and going right and contributing on the next step to bin $j+1$ is:

$$q_j * \frac{1}{2} * \frac{p_{j+1}}{p_{j+1}+p_j}$$

The difference of the flow to the right minus the flow to the left represents the ***net flow*** to the right between these two neighbors. Let's write down the net flow:

$$q_j * \frac{1}{2} * \frac{p_{j+1}}{p_{j+1}+p_j} - q_{j+1} * \frac{1}{2} * \frac{p_j}{p_{j+1}+p_j}$$

The Principle of Detailed Balance

It is in principle possible that the q_j will never settle down to anything. There are weird cases where that can happen. However, we are going to ignore the weird cases and assume that the q_j eventually settle down to some values. In that case, the net flow from $j \to j+1$ must be zero, at least on average, otherwise the q_j are still settling down! So let's see what setting the net flow to zero gets us:

$$q_j * \frac{1}{2} * \frac{p_{j+1}}{p_{j+1}+p_j} - q_{j+1} * \frac{1}{2} * \frac{p_j}{p_{j+1}+p_j} = 0$$

or

$$q_j * \frac{1}{2} * \frac{p_{j+1}}{p_{j+1}+p_j} = q_{j+1} * \frac{1}{2} * \frac{p_j}{p_{j+1}+p_j}$$

Multiply through by $2\left(p_{j+1} + p_j\right)$ and get

$$q_j * p_{j+1} = q_{j+1} * p_j$$

or

$$\frac{q_j}{p_j} = \frac{q_{j+1}}{p_{j+1}}$$

Now once things have settled down ***this is true for every adjacent pair***. So that means this ratio is the same ratio for all j. Let's call the constant ratio ρ. We have shown that

$$\frac{q_j}{p_j} = \rho$$

or

$$q_j = \rho p_j$$

The very last thing to observe is that the sum of all the p_j is 1, because the p_j are a probability distribution. And the sum of all the q_j is 1, because the q_j are also a probability distribution. The only way the p_j and the q_j can both be normalized is if ρ is 1.

We have shown that Monte Carlo works! Every q_j settles down to the corresponding p_j.

Comments / Looking Ahead

We have only shown that Monte Carlo works on average, once equilibrium is reached. Of course there are random fluctuations about the average. And of course while the algorithm is trying to settle down after being started in some possibly very non-representative initial position, during the settling down the principle of detailed balance does not yet apply.

So you will see practitioners of Monte Carlo methods do things like run their algorithm a 1,000,000 times, but dump the first 100,000 tallies under the assumption that they are less representative. The exact numbers are a matter of art, and provided N is sufficiently large, the initial tallies slowly get drowned out in the average anyway. To come all the way back to our example of a compressed air tank opened in the corner of the room, the air tank needs time to empty and the molecules need time to bounce throughout the room. Maybe a person entering the room would prefer to wait a bit, rather than entering immediately, which they could do if they were comforted in the knowledge that the air will eventually, and on average, become uniform.

Finally, note that the “appropriate ratio” was critical to the proof. What the Metropolis-Hastings algorithm did in 1970 was discover new and more sophisticated “appropriate ratios.” And what Geman and Geman did with Gibbs sampling in 1984 was get even more sophisticated. With every increase in sophistication of Monte Carlo methods, you need to select appropriate ratios that make a proof that uses the Principle of Detailed Balance work.

Metropolis-Hastings and Chapter 15

Just to be a bit less vague and explain a bit about Metropolis-Hastings, what Hastings did was to relax the assumption that the chance of moving left and the chance of moving right at the coin flip step of the algorithm was equal and $\frac{1}{2}$. Now you would think this would make an awful mess out of the Principle of Detailed Balance, but Hastings exactly compensated for the unevenness of the coin flip by adding a “correction factor” to the “appropriate ratio.”

You might also think that the Metropolis-Hastings generalization is useless. Why not just have an even chance of going left and right? You need uneven chances of going left and right when a distribution has an edge! We avoided this in the iPhone sales example by having the bin “to the right” of Q4 be Q1, and the bin “to the left” of Q1 be Q4. So we avoided having any edges. But you can’t have a 50% chance of going right from bin Q4 if you don’t do something hokey like defining the bin “to the right” of Q4 to be Q1. So if in some interesting probability distribution, there are bins near the edge from which you can’t blindly go right, that means you must have a larger chance of going left from those bins. The writeup of Metropolis-Hastings is in Chapter 15 of Donovan and Mickey. Unequal coin flips are forced upon us there because the interesting probability distribution is the binomial distribution, which has edges at $p = 0$ and at $p = 1$. Donovan and Mickey introduce the “correction factor” on p. 229.