Markov chain Monte Carlo (MCMC)

Thursday, November 8, 2018 00:19

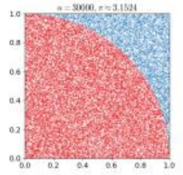
For the glory of God

Introduction

- A Markov chain Monte Carlo is one of algorithms for sampling from a probability distribution.
- Before we dive into the MCMC algorithm, let us summarize a few prerequisite concepts.

Monte Carlo Method

- Monte Carlo method is a computational algorithm that rely on repeated random sampling to obtain numerical results.
- The idea is using randomness to solve problems that might be deterministic in principle.
- Let us take an example ;
  - Let's say that we want to know the value π (Suppose that we don't know the value)
  - Steps are as following ;
    1) Draw a square and then inscribe a quadrant within it
    2) Uniformly scatter a given number of points over the square
    3) Count the number of points inside the quadrant



; Approximately 30,000 random points

4) Calculate π

$$\frac{\text{Area of quadrant}}{\text{Area of square}} = \frac{\text{Total number of points inside circle}}{\text{Total number of points inside square}} = \frac{\text{Inside count}}{\text{Total sample count}}$$

5) Note that ;
   - If the points are not uniformly distributed, then the approximation will be poor.
   - There should be a large number of points.

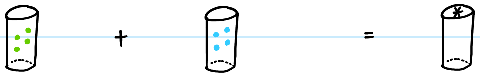
Bernoulli's trial

- In the theory of probability and statistics, a Bernoulli trial is a random experiment with exactly two possible outcomes.
- He actually tried to refine the idea of expectation ;
  - He was focused on a method of accurately estimating the unknown probability of some event based on

the number of times the event occurs in independent trials.

- He conducted a simple experiment to figure out it.

e.g. Let's say



30 green balls — 20 blue balls — mixed with green and blue balls ; Now, totally we have 50 balls

Then, suppose that you are going to take out one of balls from the mixed one.

↳ Let's say that you got the green one as the first try. Let's keep doing it!

- Finally, you will see that it will converge on the actual ratio as the number of trials increases.

$$\therefore \frac{\text{Count number of blue balls}}{\text{Count number of green balls}} = \frac{2}{3}$$; known as law of large numbers

· He concluded that everything in the world is governed by precise ratios.

· The idea was quickly spread but some people started to argue with his idea.

↳ They claimed that his theory is only valid for toy examples because the theory is based on independence.

(Most thing in reality are dependent on prior outcomes)

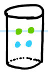
Markov Chain

· Markov extends Bernoulli's results ;

- Markov demonstrated that the law of large numbers can apply to dependent variables.

· In statistics and probability, a Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

· For example, let's say that we have the following states ;



State 1 : 50-50 mix with green and blue

State 2 : we have more green versus blue

- First, we simply start in a random state. e.g. either 1 or 2

- Second, we make a selection ; let's say that starting from state 2 and randomly green ball is chosen.

- Third, we move to either state 1 or 2 depending on that event.

· So, we can identify four possible transitions ;

1) If we are in state 1 and a green ball occurs, we loop back to the same state and select again.

2) If we are in state 1 and a blue ball occurs, we jump over to state 2 and select.

3) If we are in state 2 and a green ball occurs, ⋯

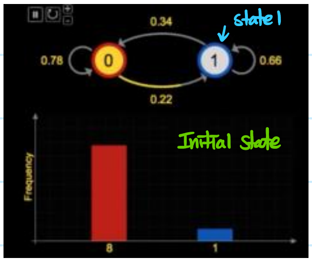
4) If we are in state 2 and a blue ball occurs, ⋯

· Actually, the probability of a green versus blue selection is clearly not independent.

↳ obviously, it depends on the previous outcome !

· However, Markov proved that as long as every state in the machine is reachable, when you run these machines in a sequence, they reach equilibrium. (called as stationary distribution)

↳ That is, no matter where you start, once you begin the sequence, the number of times you visit each state converges to some specific ratio or probability.



→

Here, $a_{ij} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

$= \begin{bmatrix} 0.5 & 0.5 \\ 0.35 & 0.65 \end{bmatrix}$

· The concept of modeling sequences of random events using states and transition between states became known as a Markov chain → It was emerged when Markov tried to extend Bernoulli's trial.

characteristic : $X_1 \to X_2 \to X_3 \to \cdots$ $P(X_k \mid X_{k-1}, X_{k-2}, \cdots, X_1, X_0) = P(X_k \mid X_{k-1})$

↳ In other words, they are all connected but only dependent on the previous one.

Markov chain Monte Carlo (MCMC)

a) Introduction

· It's time to talk about MCMC with the concepts that we discussed.

e.g. Sampling on high dimensional probabilities

· As always, there is a target probability in which it is hard to sample points from the distribution.

· The MCMC is used to sample points from the target probability.

b) How does MCMC work ?

· Basically, the MCMC is a mixed word between Monte Carlo simulation and Markov chain.

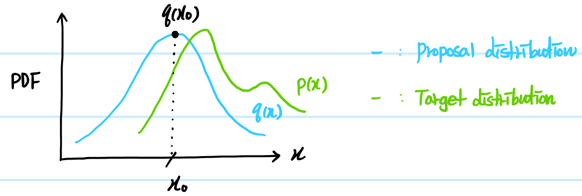
- Therefore, we may be able to say that the MCMC is a sort of Monte Carlo simulation where the simulation uses the characteristics of Markov chain to sample points from a target probability.

· The procedure is as following ;

1) Choose randomly an initial value x_0

2) Calculate the Probability Density Function of x_0 from the proposal distribution

e.g.



; Here, Note that { A commonly used proposal distribution is a symmetric Gaussian distribution.
In Metropolis-Hasting algorithm, the $q(x)$ must be symmetric. ⇒ It means $q(x|y) = q(y|x)$

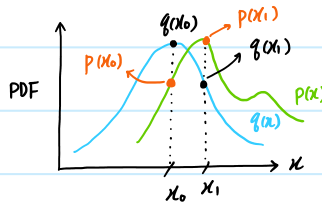
3) Perform Metropolis - Hastings algorithm → This is needed to use the characteristic of Markov chain.

- First, we are going to sample another point from the proposal distribution (It's possible because it's easy to sample)

- Second, we need to calculate the acceptance probability (α) ;

$$\alpha = \min\left(1, \frac{p(x_1)\, q(x_0|x_1)}{p(x_0)\, q(x_1|x_0)}\right)$$

e.g.



; Since we have a symmetric distribution for $q(x)$,

$$\alpha = \min\left(1, \frac{p(x_1)}{p(x_0)}\right)$$

- Third, generate a uniform random number u on $[0, 1]$

- Fourth, check the criteria to accept or reject as following ;

① If $u < \alpha$, accept the candidate by setting $x_{t+1} = x'$

e.g. if $u < \alpha$, then we will accept the x_1 as one of sample points on the target probability.

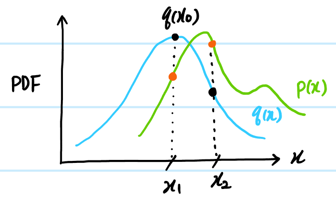
② If $u > \alpha$, reject the candidate and set $x_{t+1} = x_t$ instead

e.g. Let's say that we accepted the x_1 as a sample point from the target probability.

In addition, we sample another point x_2 from the proposal probability distribution.

→ It would be randomly selected.

(It's called as random walk)



Let's say we obtained $u > \alpha$; then, we're not choosing x_2 as one of sample points on the target distribution.

Instead, we are going to set x_1 as a current point ; and then sample another point on the proposal distribution.

4) Repeat this process enough such that we will be able to hit every point in the space.

· From the procedure, we may notice that ;

- This is obviously considered as Monte Carlo method because we repeat the process randomly with a lot of iterations.

- This is also considered as Markov chain method because choosing random point is only dependent on the previous step.

· As the simulation goes to very large number of iteration,

- We may want to cut off the beginning events (Burn out) because it's not stabilized.

- Once the events reach to the stable status (e.g. stationary distribution in Markov chain), we may be able to use the values from the simulation.

Can we throw it away ?

: Yes, because of Markov chain characteristic (only dependent on the previous one),

we know that the effect of beginning points disappears as the number of iteration increases.

This stationary distribution is not affected by the beginning status.