The role of Markov chain in Monte Carlo simulation

Sicheng Wang

School of Applied Statistics and Mathematics, Stony Brook Institute, Anhui University, Hefei, Anhui, 230039, China

r22214100@stu.ahu.edu.cn

Abstract. The Markov Chain Monte Carlo method has been applied in several fields and has thrived in academic research over the past few decades. The method is utilized to estimate the interest parameter's posterior distribution through random sampling in the probability space. Monte Carlo method alone can accomplish random simulations for some complicated continuous integrals or discrete summations. However, the method requires the sample set corresponding to the distribution, which demands the application of Markov chain. Nevertheless, problems occur when the probability function of the samples remain unknown, under which circumstance the Markov Chain Monte Carlo method is no longer available. Therefore, the article introduced Metropolis-Hastings Algorithm as a resolution. The paper is intended to provide specific insights about Metropolis-Hastings Algorithm for Markov Chain Monte Carlo and the corresponding interdisciplinary applications. Specifically, the paper will discuss about using the theory mentioned above to check the relationships among species in a common ecosystem. The paper will cover several prepositive theories, including the Law of Large Numbers for Markov Chains, binary sequences, and Metropolis-Hastings Algorithm. Overall, specific cases and the way to draw correct conclusions will be most concerned.

Keywords: Markov chain Monte Carlo; Metropolis-Hastings algorithm; Co-occurrence matrix; Stationary distribution.

1. Introduction

The Monte Carlo method was announced in the 1940s when the Manhattan Project took place. The brief idea of the method is stated as to collect many samples from the target distribution and analyze the properties of the distribution accordingly. The theory is supported by Law of Large Numbers, which suggests sample means converge to the expectation abiding by the distributions. To introduce it more specifically, assume that it is required to calculate the size of a specific graph of which the function is not accessible, one can put some points on the graph and make them randomly scattered. In this way, the points in the graph to the points out of the graph ratio is accessible. The theory needs an assumption that the point is equally distributed in the domain to hold. In the case the assumption fails, one can calculate the probability distribution function of the points in the domain, in order to draw a new function replacing the original distribution function, in which way the Monte Caro Method is applied [1].

The Markov chain serves to draw sample sets of distributions for Monte Carlo method. A sample set of a stationary distribution can be obtained through its corresponding state-transition matrix. For instance, firstly, people can generate sample based on an initial distribution such as Gaussian distribution $\pi_0(x)$ and acquire the status value x_0, then create sample according to conditional probability distribution $P(x|x_0)$ and get x_1. When the process is driven sufficient times, the set can be recognized as the correspondence sample set for the stationary distribution, which is qualified for Monte Carlo summation. Nevertheless, it is notable that Markov Chain Monte Carlo method fails when the probability density function is not accessible, which requires Metropolis-Hastings Algorithm.

The article is intended to briefly expound on the Markov chain Monte Carlo its related algorithms, and will further demonstrate the method using an interdisciplinary example “Darwin’s finch”. The rest of this work is organized as follows. In Section 2, the author introduces Markov chain Monte Carlo. In Section 3, the author will focus on the Metropolis-Hastings Algorithm. In Section 4, an example applying Markov Chain Monte Carlo method and Metropolis-Hastings algorithm is expounded. The last Section is devoted to the conclusion.

2. Markov chain Monte Carlo

Markov Chain Monte Carlo is an algorithm simulating baffling distributions through Markov sequences. The algorithm constructs a Markov Chain with limiting distribution in agreement with the expectation. Then, the process is repeated for enough times to converge to the limiting distribution mentioned above, from which the result of the Markov sequence is generated as a sample set in the end. The part intends to focus on Markov Chain Monte Carlo sampling.

2.1. Law of large numbers

If X_1, X_2... is an independent identically distributed sequence sharing the mean $\mu < \infty$, then the law suggests that with probability 1,

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu. \tag{1}$$

Similarly, one shall order that X is a random variable bearing identical distribution as above and assume that $f(X)$ is a real function with boundaries. Under this circumstance, $f(X_1), f(X_2)$... is also an independent identically distributed sequence with probability 1,

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right) = E(f(X)). \tag{2}$$

Order that X_0, X_1... is a Markov chain with stationary distribution $\boldsymbol{\pi}$ (where X is a random variable) and assume that assume that $f(X)$ is a real function with boundaries. In this case, with probability 1,

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} r(X_i)\right) = E\big(r(X)\big) = \sum_{i} r(i)\pi_i. \tag{3}$$

The significance of the theorem is that despite a random variable sequence $\{X_n\}$ does not follow the independent identical distribution, if it is a Markov chain, the Law of Large Numbers holds.

2.2. Rejection sampling method

When utilizing the Rejection Sampling Method to sample from a distribution whose probability density function is $f(x)$, it is required to obtain an available distribution $g(x)$ and a constant c which satisfies that $f(x)/g(x) \leq c$ for each X. The loop will be shown as follow. First, access a sample $X_i$from $g(x)$; Second, generate a random number u_i from the uniform distribution $U(0,1)$. Third, judge that whether $u_i \leq f(x_i)/cg(x_i)$ holds. If it holds, the sampling succeeds, otherwise fails, then repeat the loop [2].

To prove the feasibility of the sampling measure above, which is equivalent to the sample drawn by the method obeys $f(x)$, the author shall prove that $F(x) = P(X \leq x|U \leq f(x)/cg(x))$, where $F(x), G(x)$ refers to the cumulative distribution function of $f(X), g(X)$, given that $x{\sim}g(x)$. According to Bayes formula

$$p(A|B) = p(A)p(B|A)/p(B)\,, \tag{4}$$

the RHS equals to

$$RHS = p(X \le x)p\left(U \le \frac{f(X)}{cg(X)} \middle| X \le x\right)/p\left(U \le \frac{f(X)}{cg(X)}\right). \tag{5}$$

According to complete probability formula, $p\left(U \le \frac{f(X)}{cg(X)}\right) = \int p(X = x)p\left(U \le \frac{f(X)}{cg(X)} \middle| X = x\right)dx$. Since x is sampled from g, the author claims that $p(X = x) = g(x)$. In this case, the conditional probability for successful sampling is $p\left(U \le \frac{f(X)}{cg(X)} \middle| X = x\right) = \frac{f(X)}{cg(X)}$, from which people obtain

$$p\left(U \le \frac{f(X)}{cg(X)}\right) = \int p(X = x)p\left(U \le \frac{f(X)}{cg(X)} \middle| X = x\right)dx = \int \frac{f(x)}{cg(x)}g(x)dx = \frac{1}{c}. \tag{6}$$

For $p(X \le x)$, since it is clear that $x \sim g(x)$, so that one has $p(X \le x) = G(x)$. Eventually, utilizing the conditional probability formula, one has

$$p\left(U \le \frac{f(X)}{cg(X)} \middle| X \le x\right) = p\left(U \le \frac{f(X)}{cg(X)}, X \le x\right)/p(X \le x) = p\left(U \le \frac{f(X)}{cg(X)}, X \le x\right)/G(x)\,. \tag{7}$$

Proceed the calculation, it is found that

$$p = \int_{-\infty}^{x} \frac{p\left(U \le \frac{f(t)}{cg(t)} \middle| X = t \le x\right)}{G(x)} g(t)dt = \frac{1}{G(x)}\int_{-\infty}^{x} \frac{f(t)}{cg(t)} g(t)dt = \frac{1}{cG(t)}F(t). \tag{8}$$

Combining the results of Eq. (6), (7) and (8), the author acquires that $RHS = LHS$, such that the feasibility is proved.

2.3. An example demonstrating the advantage of Markov chain Monte Carlo

Given a sequence x consists of 0s and 1s with length l, where $f(x)$ is a function generated to indicate the number of 1s in the whole sequence. Grant the feasibility to obtain the uniformly random sequence x with no two 1s adjacent. Repeat the process and generate such independent sequences X_1, X_2... X_n. According to the Monte Carlo method, the estimation is as follow

$$\mu \approx \frac{\sum_1^n r(X_i)}{n}. \tag{9}$$

A way based on the rejection method alone, without Markov chain, is generating the sequence containing elements 0 and 1 randomly. If the resulting sequence satisfies the demand, it is accepted, otherwise denied and the process is repeated. However, this method doesn't function practically. The acceptance of the sequences is dreadfully low. By using the Markov chain Monte Carlo method, one can launch a Markov chain $\{X_i\}$ with the uniformly limiting distribution on the accepted sequences, of which the state space refers to the set of acceptance among all sequences. Under these circumstances, one will generate the Markov chain, whose estimate for the expectation is as follow

$$\mu \approx \frac{\sum_1^n r(\,Y_i\,)}{n}. \tag{10}$$

For further explanations, the Markov process can be made manifest on a graph, whose vertices refer to the sequences accepted and edges refer to one step of progress of the chain. The process is expounded below. Randomly pick one of the l elements from an accepted sequence. If the element is 1, then switch it into 0, such that the value of $f(x)$ declines, which indicates the acceptance of the resultant sample set. After that, proceed the process. Otherwise, if the element is 0, turn it into 1only for accepted consequences, after which the process continues. Nevertheless, if the new 1 result in adjacent 1s, which refer to the denied situation, pause the process [3].

To forge a uniform distribution, people can assign weights to the edges (including the loops of the sequences). Edges between the adjacent sequences, which need only one step to transfer to one another, are assigned with weight 1. In this case, all vertices weigh equivalently. In this case, the embryonic form of uniform distribution is forged. The graph for sequence x of which $l = 4$ is shown in Figure 1.

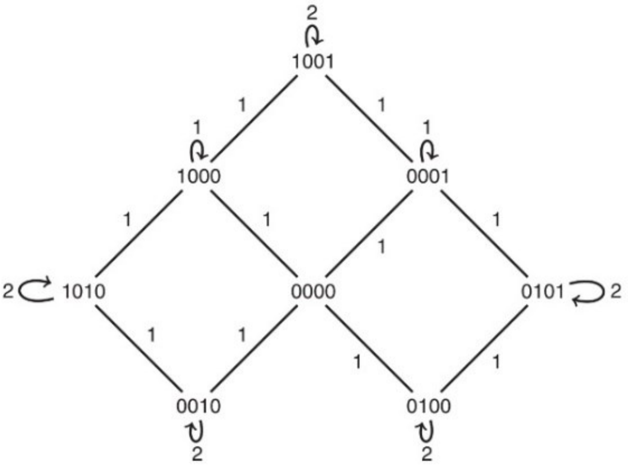

Figure 1. The weight graph of the walk [2].

3. Metropolis-Hasting algorithm

It is notable that the Markov Chain Monte Carlo method algorithm works only when the probability density function $p(x)$ is accessible. However, in most conjugate prior distributions, people cannot calculate the posterior distribution due to the barrier of solving the value of normalized constant c ($c = \int p(X|x)p(x)dx$). Therefore, Metropolis-Hasting Algorithm is necessary.

To give an example, set up a truncated normal distribution as the prior distribution with domain $\theta \in [0,1]$, average value $\mu = 1/2$, the author obtains the posterior distribution

$$p(\theta|x) = \frac{\theta^x(1-\theta)^x \cdot \frac{c}{\sqrt{2\pi}\sigma} e^{-\left(\frac{1-\theta}{2}\right)^2/2\sigma^2}}{\int_0^1 \theta^x(1-\theta)^x \cdot \frac{c}{\sqrt{2\pi}\sigma} e^{-\left(\frac{1-\theta}{2}\right)^2/2\sigma^2} d\theta}. \tag{11}$$

In this distribution, people are not able to solve the denominator of the fraction, however, the problem might be avoided through special methods. The author shall select two random values θ_i and θ_j, then calculate the relative density of the two probabilities in order to eliminate the c. The posterior distribution formula

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{c}, \tag{12}$$

and the relative density is given by

$$\gamma = \frac{p(\theta_i|x)}{p(\theta_j|x)} = \frac{p(x|\theta_i)p(\theta_i)}{p(x|\theta_j)p(\theta_j)}. \tag{13}$$

Under the circumstance of big samples, the value of γ is approximately equivalent to the ratio of θ_i to θ_j. Given a sample $\{\theta_1, \theta_2 \dots \theta_n\}$, among which i equals θ_i, j equals θ_j, and randomly pick θ', when

$\theta' = \theta_i$ holds, the author notices that the number of $\{\theta_i\}$ is so big that one tends to reject it when $i/j > \gamma$; conversely, if $i/j < \gamma$, include θ' and thus the value of i/j will approach γ. In Metropolis-Hastings algorithm, people assume that the sample is a sequent Markov chain, where one picks the new sample according to the latest sample selected.

Overall, people operate the following loop when sampling using the Metropolis-Hastings algorithm. First, sample θ_0 using θ_i based on a certain distribution, then determine the value of the new sample through $\gamma = \frac{p(\theta_0|X)}{p(\theta_i|X)}$. Notably, in step 1, people generate symmetric proposal distribution satisfying $J(\theta_i|\theta_j) = J(\theta_j|\theta_i)$. The symmetric proposal distribution is usually uniform distribution and normal distribution. The complete process is shown as follow. Initially, randomly pick the first sample θ_1. Afterwards, sample from $J(\theta|\theta_i)$ and obtain θ_0, then calculate $\gamma = \frac{p(\theta_0|X)}{p(\theta_i|X)} = \frac{p(X|\theta_0)p(\theta_0)}{p(X|\theta_i)p(\theta_i)}$. If $\gamma > 1$, assign $\theta_{i+1} \leftarrow \theta_0$,otherwise randomly sample $p \sim U(0,1)$. When $p < \gamma$, assign $\theta_{i+1} \leftarrow \theta_0$, otherwise assign $\theta_{i+1} \leftarrow \theta_i$. Finally, repeat step 2 until the sample meets the expectation.

4. An application using Metropolis-Hastings algorithm

4.1. Background information of the model

On Charles Darwin's journey to the Galapagos, he discovered various species of finches on the islands, either coexisting or competing. Seemingly, the research on the diversity of the birds has partially influenced the natural selection theory [4]. The biologist discovered 13 species of finches on the total 17 islands. The co-occurrence matrix below is used for studying the relationships among species [5]. It is found that the matrix is

$$T = \begin{Bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{Bmatrix}. \qquad (14)$$

4.2. Checkerboards and the expected measure of application

A checkerboard matrix consists of two 1s from different species on two islands independently, with all 1s on the main diagonal and 0s on the other, or vice versa. It indicates that two species are distributed respectively and independently on two different islands, implying their competitive relationship. The checkerboard matrices are shown as $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ [6]. Competition rate between species can be indicated through the number of the checkerboards in the whole matrix. Larger numbers stand for more intense relationships. For the co-occurrence matrix of Darwin's finches, there are 333 checkerboards in total [7]. Nevertheless, the author can barely judge whether the number 333 is large or small. Thus, one must solve the expectation of the number of such checkerboards.

It is reported that Doctor Susan Holmes estimated that the value of expectation is approximately 6.715×10^{16} [8]. Considering the astronomical number, it is not realistic to select all checkerboards from these matrices. Therefore, this work applies Metropolis-Hasting method to launch uniform co-occurrence matrices randomly, then calculate the probability that this matrix possesses equal to or more than 333 checkerboards. The process is the same as the one stated in section 3. It begins by generating an initial matrix A with not less than one single checkerboard, from which one can select one in the big matrix and exchange the 0s and 1s. The resultant matrix is B, possessing the same row and column sums as the previous matrix A. In this way, in a graph, these two matrices, which can be transferred to each other through one exchange operation of a checkerboard matrix, are adjacent vertices [9].

4.3. A simplified model for the problem

For example, given all 3 × 3 matrices

$$\begin{pmatrix}0&1&1\\1&1&0\\1&0&0\end{pmatrix},\begin{pmatrix}1&1&0\\0&1&1\\1&0&0\end{pmatrix},\begin{pmatrix}1&1&0\\1&1&0\\0&0&1\end{pmatrix},\begin{pmatrix}1&0&1\\1&1&0\\0&1&0\end{pmatrix},\begin{pmatrix}1&1&0\\1&0&1\\0&1&0\end{pmatrix}, \tag{15}$$

whose row sums and column sums are all 2, 2, and 1 as the co-occurrence matrices and collect the checkerboard matrices. After the transition graph of the five matrices is drawn, the stationary distribution of five vertices can be generated according to the number of edges converging on the same vertex. The graph referring to the five matrices is shown in Figure 2.

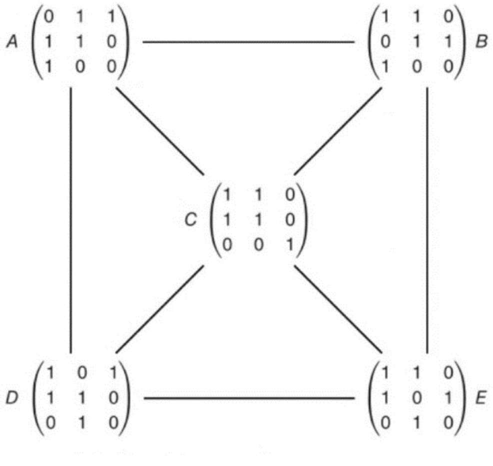

Figure 2. Illustration of transition graph [9].

The stationary distribution is shown below

$$s=(s_A,s_B,s_C,s_D,s_E)=\left(\frac{3}{16},\frac{3}{16},\frac{4}{16},\frac{3}{16},\frac{3}{16}\right). \tag{16}$$

The process of the algorithm is as follow. First, generate an initial matrix i with not less than one single checkerboard. Second, randomly select two rows and two columns, pause when the selected two forge exchangeable checkerboard matrix, then exchange the elements and form the new co-occurrence matrix j. After that, calculate the number of checkerboard matrices in both big matrices. In the end, generate the acceptance function to decide whether to include the next sample. The function is shown below:

$$a(i,j)=\frac{\pi_j T_{ji}}{\pi_i T_{ij}}=\frac{\deg(j)}{\deg(i)}=\frac{Numbers\ of\ checkerboards\ in\ j}{Numbers\ of\ checkerboards\ in\ i}. \tag{17}$$

The algorithm is processed for 5,000 steps [10]. Among all 5,000 matrices, only two had 333 or more checkerboards. Therefore, it can be announced that the analysis shows evidence of competition among finches in Galapagos.

5. Conclusion

The article has expounded two advanced algorithms based on the Monte Carlo method, which can enhance the efficiency of generating the sample sets and do further analyses. In Section 2, the article focused on how the combination of Markov chain with Monte Carlo method can improve the original rejection sampling method applied in the latter. The author generated detailed calculations to illustrate the feasibility of Markov chain application in Monte Carlo method. It is notable that Markov Chain Monte Carlo is the core method utilized in Section 4 to solve the "Darwin's finches" issue. Section 3 contains a detailed account of Metropolis-Hastings Algorithm in order to prepare for the data analysis process in Section 4. In Section 4, the author demonstrated the methods and algorithms mentioned above by "Darwin's finches" problem. Owing to the complexity of the original problem, the article stated the procedures through a simplified model sharing the same properties with the original problem. Significantly, the section showed the advantages of Metropolis-Hasting Algorithm over the ordinary Markov Chain Monte Carlo method to process enormous data. Overall, from the example provided by the paper, it is without doubt that Markov chain Monte Carlo method possesses a bright predisposition to give resolutions to a variety of issues among different subjects. For example, Markov chain applied to defend against cyberattacks, Markov chain-based IoT techniques, etc. The author will devote further studies to the application of Markov chain Monte Carlo in modern lifestyles.

References

[1] Ching, W.-K., Huang, X., Ng, M. K., & Siu, T.-K. (2013). Markov chains. International Series in Operations Research and Management Science, Springer.
[2] P. Dobrow, Robert. (2016). Introduction to Stochastic Processes with R, John Wiley & Sons, Inc.
[3] Geweke, J., and Tanizaki, H. (2003). Note on the Sampling Distribution for the Metropolis-Hastings Algorithm. Communications in Statistics - Theory and Methods, 32(4), 775–789.
[4] Zhang, Guoqi, Hou, Yue, Wang, Kangbo. (2023). Vulnerability Analysis of Ship in Preliminary Design Stage Based on Markov Chain, Ship Engineering, 45(3), 67-72.
[5] Abzhanov, Arhat. (2010). Darwin's Galapagos finches in modern biology. Philosophical transactions of the Royal Society of London, 365(1543), 1001-7.
[6] Arita, Héctor T. (2016). Species co-occurrence analysis: pairwise versus matrix-level approaches. Global Ecology and Biogeography, 25(11), 1397-1400.
[7] Presley, S. J. et al. (2019). Checkerboard metacommunity structure: an incoherent concept. Oecologia, 190, 323–331.
[8] Holmes, Susan et al. (2005). Sequential Monte Carlo Methods for Statistical Analysis of Tables. Journal of the American Statistical Association, 100(469), 109-120.
[9] Sami, Abu-El-Haija, Bryan, Perozzi, Rami, Al-Rfou, and Alex, Alemi. (2018). Watch your step: learning node embeddings via graph attention. Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). 18, 9198–9208.
[10] Zink, Robert, Hernán, Vázquez-Miranda. (2019). Species Limits and Phylogenomic Relationships of Darwin's Finches Remain Unresolved: Potential Consequences of a Volatile Ecological Setting", Systematic Biology, 68(2), 347–357.