

Seeking Application of Markov Chain Monte Carlo in Different Fields

Bairun Chen

North Broward Preparatory School, Florida, USA
Bairun_chen@mynbps.org

Abstract: The Markov Chain Monte Carlo (MCMC) methods build the Markov chain which the stable distribution matches the expected distribution. It can approximate sampling and estimate to the complex probability distributions by generating dependent samples iteratively. The MCMC methods are algorithms used to generate samples in the complex probability distributions. This article introduces the basic principles, theory background and the applications of MCMC in various fields such as Bayesian inference, statistical physics, and machine learning. By discussing the particular challenge that MCMC is facing now and the future research directions of MCMC, this paper is focused on providing a comprehensive overview about MCMC methods. And this discussion also emphasizes the historical development of MCMC, its mathematical foundations, and its practical implications in modern computational science. Thus, MCMC methods make the field of computational statistics different and continue to play a key role in the modern data science. As the development of the increasing of computing power and new algorithm, the potential of MCMC will expand further.

Keywords: Markov Chain, Monte Carlo, Probability distributions.

1. Introduction

It is well-known that the Markov Chain Monte Carlo (MCMC) methods provide an available structure for sampling in the complex probability distributions. These methods use the Markov Chains to generate approximate samples so that it can make the MCMC very useful in high-dimensional statistical modeling. Traditional Monte Carlo methods are often difficult in direct sampling from the complex distributions, especially in Bayesian inference and stochastic optimization, because computing exact integrals is computationally not work. MCMC solve this issue through building a Markov Chain, the stationarity of its distribution corresponds to the probability distribution of the target.

The original of MCMC can travel back to the middle of 20th century, the seminal work from the Metropolis (1953) brings Metropolis algorithm [1]. Then Hastings (1970) generalized the algorithms into the Metropolis-Hastings algorithm. This algorithm remains one of the most widely used MCMC methods today. Over decades, MCMC has developed very well. There are many advancements such as in the perspective of Gibbs sampling, Hamiltonian Monte Carlo, and adaptive MCMC methods. They further expand its applicability.

MCMC has become a indispensable part in various subjects such as Bayesian statistics, artificial intelligence, computational physics, and genomics. It allows the parameter estimating in probabilistic

models and significantly bolster the feasibility of analyzing large-scale data. For instance, in the Bayesian inference, MCMC can gauge the rear distributions in complex hierarchical models, but if one uses the traditional methods, it will be more difficult to achieve the goals.

2. Methods and Theory

2.1. Markov Chains

A Markov chain is a stochastic process where the changing to the next state is determined only on the current status. Mathematically, it is expressed as:

$$P(X_{t+1}|X_t, X_{t-1}, \dots, X_0) = P(X_{t+1}|X_t) \quad (1)$$

This property is called Markov property. It makes sure the future status only be determined by the present state but it will not influence the sequence of events that happen before. Markov chain consists of a state space S , a transition probability matrix P , and an initial state distribution π_0 . If the Markov chain has a unique stationary distribution π , it satisfies the balance equation:

$$\pi(X) = \sum_{X'} P(X|X') \pi(X') \quad (2)$$

where $P(X|X')$ represents the probability of transitioning from state X' to state X .

The key properties of Markov chains relevant to MCMC include the following properties. Ergodicity: If a Markov Chain can get to any state from any other state in a finite number of steps. The Markov chain is ergodicity. This makes sure that the Markov chain can travel through the entire state space. Reversibility: If Markov Chain satisfies the detailed balance condition, the chain will be reversible. It is a necessary condition for the existence of the stable distribution. Mixing Time: The mixing time of the Markov chain means the time that the chain spends in converging to its stationary distribution. The faster of the mixing times is, the more efficient of sampling is.

2.2. Monte Carlo Methods

Methods of Monte Carlo approximate numerical results by repeating random sampling. Given a function $f(x)$ and a probability distribution $p(X)$, the expectation is defined as:

$$E[f(X)] = \int f(X)p(X)dX \quad (3)$$

When direct computation is not work, Monte Carlo estimation approximates the integral using samples [2]:

$$E[f(X)] \approx \frac{1}{n} \sum_{i=1}^n f(X_i) \quad (4)$$

However, it is often difficult to direct sample from $p(X)$. Especially, in high-dimensional spaces or when the distribution is complex, the situation will be tricky. This limitation leads to the development of importance sampling. This is a technique of variance reduction. The principle is using the proposal distribution to generate samples more efficiently. Although one made these advancements, Monte Carlo methods are still facing many challenges in processing highly correlated or multimodal distributions (common in real-world applications).

The efficiency of Markov Chain Monte Carlo (MCMC) is determined by the mixing time which refers to how fast it converges to the stationary distribution and convergence properties. One of the key advantages of MCMC is that it can process high-dimensional and multimodal distributions. It is very common in Bayesian inference and machine learning. However, the MCMC methods are not without any limitations. For instance, the “curse of dimensionality” can lead to bad status of convergence(slow) and mixing(poor). To solve these challenges, researchers developed an advanced MCMC techniques such as Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS) [3]. They improve sampling efficiency by using gradient [4].

3. Results and Applications

3.1. Metropolis-Hastings Algorithm and Gibbs Sampling

The Metropolis-Hastings algorithm is a common MCMC method. It is used for sampling from a target distribution $\pi(X)$. It employs a proposal distribution $Q(X'|X)$ and an acceptance probability. The algorithm proceeds as follows. The first is to set up state X_0 . At each iteration t , the second is to propose a new state $X' \sim Q(X'|X_t)$. The third is to compute the acceptance ratio:

$$P = \min \left(1, \frac{\pi(X') Q(X_t|X')}{\pi(X_t) Q(X'|X_t)} \right) \quad (5)$$

One can accept X' with probability A , otherwise retain X_t . The fourth is to repeat until convergence.

The Metropolis-Hastings algorithm is widely applied in Bayesian inference. And it can be used in estimating the posterior distributions in complex models. For example, in hierarchical Bayesian models, the Metropolis-Hastings algorithm can be applied for sampling from a joint rear distribution of parameters and hyperparameters.

The Gibbs sampling is an unique method of Markov Chain Monte Carlo (MCMC). Every time it updates a new variable, the other variables will still be kept the same at the same time. This algorithm will be especially useful for the Bayesian inference in high-dimensional distributions. The first is to initialize state $X_0 = (\chi_1, \chi_2, \dots, \chi_n)$. The second is to sequentially update each variable according to

$$\chi_i^{(t+1)} \sim P \left(\chi_i \mid \chi_1^{(t+1)}, \dots, \chi_{i-1}^{(t+1)}, \chi_{i+1}^{(t)}, \dots, \chi_n^{(t)} \right), \quad (6)$$

and the procedure can iterate until convergence.

Gibbs sampling is widely used in the Latent Dirichlet Allocation (LDA) of topic modeling and Bayesian hierarchical models. In the Latent Dirichlet Allocation (LDA), Gibbs sampling is used for the position of words in the topics and inferring the position of topics in documents.

MCMC has widespread applications. The first is Bayesian inference. Estimating posterior distribution in the hierarchical model. For example, MCMC is applying in estimating the transmission rates of infectious diseases in the epidemiology. The second is Statistical Physics. MCMC methods such as the Metropolis algorithm are used for the study of phase transitions in materials. Simulation thermodynamic states. The third is Machine Learning. MCMC is also used in bolstering learning for policy optimization. Training probabilistic models like Bayesian neural networks. The fourth is Computational Biology. In the genomics, MCMC is used in analyzing DNA sequences and identifying genetic variants. The fifth is Financial Modeling. MCMC is used in modeling the behavior of financial markets and estimating the value of complex financial instruments. Risk assessment and Monte Carlo pricing of derivatives.

3.2. Applications of Markov Chains in Finance

Because Markov chains can model stochastic processes and predict future status based on current information, it has gotten many extensive applications in the field of finance. There are three examples below which show that the application of Markov chains in finance. Each one example includes details to explain.

3.2.1. Risk Management and Extreme Value Theory

Markov chains are widely used in the perspective of risk management which is especially in the background of EVT. Extreme value theory (EVT), a branch of statistics, is used for processing the extreme deviations from the median of probability distributions. In the field of finance, EVT is used for simulating tail behavior of asset returns. It is vital for the estimation of risk of extreme losses.

One important application is about the estimation of Value at Risk (VaR) and Expected Shortfall (ES), they are the key metrics to the management of financial risk. VaR can estimate the potential loss in the value of a portfolio given period for a given confidence interval [5], and the ES will provide the estimate of the average loss in VaR threshold. MCMC methods are often applied in estimating the parameters of the GPD. It is used for modeling the tail of the loss distribution.

For instance, in a study by Li (2017), MCMC was used in comparing the extreme risks of shadow banking and the stock market in China [6]. This research shows that compared to the shadow banking, the risk of the stock market will be higher. The application of MCMC can help estimate tail risks more accurately which is vital for management of their exposure to extreme market events. Markov chains play a key role in the background of the application of the risk of finance. It is necessary to maintain financial stability [7].

Markov chains are also used in the stress testing and the analysis of scenario [8]. It is important to estimate the resilience of financial agencies in the extreme conditions. By modeling the transitions of economic states (recession, recovery, and boom), financial agencies can stimulate different impacts of adverse scenarios on their portfolios [9]. For example, in the period of financial crisis in 2008, the banks used the model which is based on the Markov chain to evaluate the effects of serious downturns in economic on loan portfolios. These models can recognize where the weak part is and introduce how to ease the risks. In addition, the Markov chain can combine with the machine learning so that it can increase the accuracy in the stress testing. And then it can archive the goal which means the more dynamic management in the risk in the background of complex financial environments.

3.2.2. Credit Rating Transitions

Another important application in finance of Markov chains is building the model of credit rating transitions. Credit agencies will rate the debt issuers with their creditworthiness. These ratings will change with the factors such as the economic conditions, company performance, and market sentiment.

Markov chains are used in estimating the transition probabilities in different credit ratings. For example, the company which got “A” rating will probably change to “AA” rating or “BBB” rating in each period. These transition probabilities are usually represented in the form of transition matrix. Here, each element in the matrix represents the probability of one credit rating being changed to another one.

The application of Markov chains in the credit rating transition is not only useful in the pricing credit derivatives such as the credit default swaps (CDS) but also helpful in the estimation of the credit risk of portfolios. Financial not only can estimate the probability of default but also the potential losses associated with their credit portfolios through simulating the future credit rating paths of issuers.

For example, a study from Jarrow, Lando, and Turnbull (1997) used a model of Markov chain to estimate the probability of transition between two different credit ratings and applied this model in pricing the corporate bonds [10]. The study shows that the Markov chains provide a strong structure for the modeling of credit risk which is especial in the background of dynamic and uncertain financial markets.

3.2.3. Algorithmic Trading and Market Regime Detection

Markov chains is widely used in the algorithms trading. They are especially used in the detection of market regimes. Financial markets usually have different forms or status, such as bull markets, bear markets, and periods of high volatility. It is essential that can recognize these regimes and make some trading tactics to adapt the changing market conditions.

Markov chains can be used in stimulating the transitions between different market regimes. For example, the Hidden Markov Model (HMM) can recognize the basic status of the market through the pricing movement that it observes. The Hidden Markov Model assumes that the market is in one of several hidden status, these status are changed with the following of the Markov process.

In fact, the HMM can be used by traders in detecting the shifts in market conditions and then adjust their strategies accordingly. For example, the traders may manage risk by reducing their position size or increasing their use of hedging instruments, during the high-volatility regime. In contrast, in the period of low-volatility regime, the trader may take advantage of stable market conditions by increasing their exposure.

A study from the Hassan and Nath (2005) will apply the HMM in the stock market data to prove the HMM can recognize different market regimes effectively [11]. The study shows that the trading tactics which is based on the HMM is better than the traditional tactics which is not considered the transitions of regime [12].

4. Conclusion

The MCMC methods provide a strong approach to sample from the distributions of complex probability. This approach makes it become a part of various scientific and engineering fields which can't be replaced. Although the MCMC is widely used, it still faces some challenges such as the problem of slow convergence in high-dimensional spaces and the difficulty of tuning algorithm parameters. The future research should focus on increasing the convergence rates, developing the adaptive Markov chain MCMC strategies, and decreasing the computational cost in high-dimensional applications. A promising direction is combining the MCMC with the deep learning techniques. For example, the MCMC can improve sampling process efficiency and explore complex distributions by combining it with VAEs and GANs. In addition, the analysis of massive datasets in real-time can be achieved by the development of parallel and distributed MCMC algorithms.

References

- [1] Metropolis, N., et al. (1953). "Equation of State Calculations by Fast Computing Machines." *Journal of Chemical Physics*.
- [2] Betancourt, M. (2017). "A Conceptual Introduction to Hamiltonian Monte Carlo." *arXiv preprint arXiv:1701.02434*.
- [3] Neal, R. M. (2011). "MCMC Using Hamiltonian Dynamics." *Handbook of Markov Chain Monte Carlo*.
- [4] Robert, C. P., & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.
- [5] Hull, J. C. (2015). *Risk management and financial institutions (4th ed.)*. John Wiley & Sons.
- [6] Li, J. (2017). *Financial Extreme Risk Measurement Based on Markov Chain Monte Carlo Simulation*. *Wuhan Finance*, 8, 35-39.
- [7] Rebonato, R. (2010). *Plight of the fortune tellers: Why we need to manage financial risk differently*. Princeton University Press.

- [8] *Basel Committee on Banking Supervision (BCBS). (2009). Principles for sound stress testing practices and supervision. Bank for International Settlements (BIS).*
- [9] *Glasserman, P., & Xu, X. (2014). Stress testing and risk integration. Journal of Banking & Finance, 47, 1-14.*
- [10] *Lando, D. (1998). On Cox processes and credit risky securities. Review of Derivatives Research, 2(2-3), 99-120.*
- [11] *Hassan, M. R., & Nath, B. (2005). Stock market forecasting using hidden Markov model: A new approach. Intelligent Systems Design and Applications, 2005. ISDA'05. Proceedings. 5th International Conference on, 192-196.*
- [12] *Guidolin, M., & Timmermann, A. (2007). Asset allocation under multivariate regime switching. Journal of Economic Dynamics and Control, 31(11), 3503-3544.*