

Application of Markov chain Monte carlo method in Bayesian statistics

Qi Zhao

Department of Foundation, Shandong Yingcai University, Jinan 250104, China

Abstract. In statistical inference methods, bayesian method is a method of great influence . This paper introduces the basic idea of the bayesian method . However, the widespread popularity of MCMC samplers is largely due to their impact on solving statistical computation problems related to Bayesian inference. Markov chain Monte Carlo method is essentially a Monte Carlo synthesis procedure. The random sample of it is related to a Markov chain. It is a widely used stochastic simulation method. This paper mainly introduces Gibbs sampling, and its application in Bayesian statistics. We use a simple example to illustrate how to tackle a typical Bayesian problem via the MCMC method

1 Introduction

Bayesian statistical originated in the eighteenth century British scholar Bayesian. His method was later developed into a system of statistical inference method. By the 1930 s has formed a bayesian school of thought, To 50 ~ 60 s has developed into an influential school of statistics, the effects are still growing. An important unfavorable factors hampering the bayesian method widely used in the further, that is obtaining the posterior distribution is often complex, high dimension, no standard form, the function on the integral calculation is often very difficult. Recently, several approaches short of direct integration have been proposed . we focus here on Markov Chain Monte Carlo methods. Markov chain monte carlo method, the basic idea is through the establishment of a stationary distribution $\pi(x)$ for markov chains to get the samples, And then based on these samples can do all kinds of statistical inference.

2 Bayesian statistical inference method and bayesian formula

The basic point of bayesian school is: Any unknown quantity θ can be regarded as random variables, A probability distribution is available to describe it, this distribution as prior distribution, denoted by $\pi(\theta)$.suppose $p(x|\theta)$ is conditional density function depend on the parameters θ . Our mission is to make statistical decision for θ ,According to the bayesian school point of view that should be divided into the following steps

1. Produce the sample $X = (X_1, \dots, X_n)$ needs to be done in two steps: First of all a observed value θ generated by the prior distribution $\pi(\theta)$. Then from the conditions of density function $p(x|\theta)$ to produce sample observations $x = (x_1, \dots, x_n)$. then the joint conditions density function of the sample as follows

$$p(x|\theta) = \prod_{i=1}^n p(x_i|\theta) \quad (1)$$

E-mail addresses: zaho231200@163.com

the joint distribution comprehensive sample information, referred to as the likelihood function.

2. $\pi(\theta)$ is prior distribution of the θ . Synthetic a priori information and samples after get sample joint distribution of X and θ .

$$h(x, \theta) = p(x|\theta)\pi(\theta) \quad (2)$$

3. In order to make statistical decision to θ , decompose the joint distribution

$$h(x, \theta) = \pi(\theta|x)m(x) \quad (3)$$

here $m(x)$ is the marginal density function of x

$$m(x) = \int_{\Theta} h(x, \theta) d\theta = \int_{\Theta} p(x|\theta)\pi(\theta) d\theta \quad (4)$$

The parameter space of θ is Θ

the $m(x)$ does not contain θ any priori information. Therefore can be used to make statistical decision to θ is only conditional distribution of $\pi(\theta|x)$. Its computation formula is

$$\pi(\theta|x) = \frac{h(x, \theta)}{m(x)} = \frac{p(x|\theta)\pi(\theta)}{\int_{\Theta} p(x|\theta)\pi(\theta) d\theta} \quad (5)$$

This is a density function of the bayesian formula. This is under the condition of the given samples, the conditional distribution of θ , Called the posterior distribution of θ .It focus on the sample and all information related to the θ in the prior distribution. It is more close to the actual situation than prior distribution $\pi(\theta)$.So using posterior distribution $\pi(\theta|x)$ to make statistical decision may be improved.

Example 1 Gelfand and Smith, consider the following example. Set a test may have five as a result, the probability of their occurrence, respectively

$$\frac{\theta}{4} + \frac{1}{8}, \frac{\theta}{4}, \frac{\eta}{4}, \frac{\eta}{4} + \frac{3}{8}, \frac{1}{2}(1 - \theta - \eta)$$

It contains two unknown parameters θ and η , and between 0 and 1. The existing 22 test results observed value as follows. $y = (y_1, \dots, y_5) = (14, 1, 1, 1, 5)$, y_i denotes the number of occurrences of the ith a result. $i = 1, \dots, 5$. Now, put the x, y is decomposed into two

parts, respectively. let

$$y_1 = z_1 + (y_1 - z_1), y_4 = z_2 + (y_4 - z_2), \text{ such that}$$

$$(Z_1, Y_1 - Z_1, Y_2, Y_3, Z_2, Y_4 - Z_2, Y_5) \sim M[22; \frac{\theta}{4}, \frac{1}{8}, \frac{\theta}{4}, \frac{\eta}{4}, \frac{\eta}{4}, \frac{3}{8}, \frac{1}{2}(1-\theta-\eta)]$$

M denotes the multinomial distribution. z_1 and z_2 are not observed, Can be seen as defect data, (θ, η) are assumed to be evenly distributed. i.e.

$$\pi(\theta, \eta) \propto 1$$

Let $z = (z_1, z_2)$ is the value of $Z = (Z_1, Z_2)$, then, (θ, η) likelihood function and posterior distribution, respectively

$$L(y, z | \theta, \eta) \propto \left(\frac{\theta}{4}\right)^{z_1+y_2} \left(\frac{1}{8}\right)^{y_1-z_1} \times \left(\frac{\eta}{4}\right)^{y_3+z_2} \left(\frac{3}{8}\right)^{y_4-z_2} \left(\frac{1-\theta-\eta}{2}\right)^{y_5}$$

$$\pi(\theta, \eta | y, z) \propto \theta^{z_1+y_2} \eta^{y_3+z_2} (1-\theta-\eta)^{y_5}$$

3 Markov Chain Monte Carlo method

Markov chain monte carlo method is essentially a monte carlo comprehensive program. The generation of random sample is associated with a markov chain. The basic idea of this method is to create a stationary distribution is $\pi(x)$ markov chain, Samples of $\pi(x)$ is obtained by the markov chain, Then based on these samples can do all kinds of statistical inference, In summary is divided into the following three steps

1. To construct a "suitable" markov chains, its transition kernel is $P(\cdot, \cdot)$, "suitable" is refers to the meaning of $\pi(x)$ is its corresponding stationary distribution
2. Starting from a point $X^{(0)}$, creating a point sequence $X^{(1)}, \dots, X^{(n)}$ using the markov chains
3. for a certain number of m and a big enough number n, the expectations of any function $f(x)$ is estimated as follows

$$\hat{E}_\pi f = \frac{1}{n-m} \sum_{t=m+1}^n f(X^{(t)})$$

Using MCMC methods, construct transition kernel is crucial, different MCMC method often is the transition kernel structure is different

Based conditional distribution of iteration sampling is a kind of important markov chain monte carlo method. A special case of the most famous is the Gibbs sampling, Now has become the standard of statistical computing tools, Its potential of markov chains is established by decomposition a series of conditional distribution, It's the most appealing features. Here to use conditional distribution, Especially with full conditional distribution. The so-called full conditional distribution is shaped as the conditional distribution of $\pi(x_T | x_{-T})$ among them $x_T = \{x_i, i \in T\}, x_{-T} = \{x_i, i \notin T\}, T \subset N = \{1, \dots, n\}$. take notice, In the conditional distribution of the above, All variables appears (either appear in conditions or in the variable). Because the Gibbs sampling involves only a

single variable sampling, it is the most attractive. The following will introduce the Gibbs sampling method

Let $\pi(x_1, \dots, x_m)$ denotes m-dimensional joint distribution, constructor the transition nuclear in Gibbs sampling as follows

$$P_{x,y} \triangleq P(x, y) = \prod_{k=1}^m \pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m)$$

Among them

$$x = (x_1, \dots, x_m), y = (y_1, \dots, y_m), x_i \in D, y_i \in D$$

$\pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m)$ is the conditional distribution. Verifiable $(P_{x,y})$ is indeed a transfer matrix

$$\text{i.e. } \sum_y P_{x,y} = 1$$

$$\sum_{y_1} \pi(y_1 | x_2, \dots, x_m) = \sum_{y_1} \frac{\pi(y_1, x_2, \dots, x_m)}{\pi(x_2, \dots, x_m)} = \frac{\pi(x_2, \dots, x_m)}{\pi(x_2, \dots, x_m)} = 1$$

$$\sum_y P_{x,y} =$$

$$\sum_{y_m} \sum_{y_{m-1}} \dots [\sum_{y_1} \pi(y_1 | x_2, \dots, x_m)] \pi(y_2 | y_1, x_3, \dots, x_m) \dots \pi(y_m | y_1, \dots, y_{m-1}) = 1$$

Verifiable, $\pi(x_1, \dots, x_m)$ is stationary distribution, it's probability transition matrix is $p(x, y)$

$$\sum_{(x_1, \dots, x_m)} \pi(x_1, \dots, x_m) p((x_1, \dots, x_m), (y_1, \dots, y_m)) =$$

$$\sum_{x_m} \sum_{x_{m-1}} \dots \sum_{x_1} \pi(x_1, \dots, x_m) \pi(y_1 | x_2, \dots, x_m) \pi(y_2 | y_1, x_3, \dots, x_m) \dots \pi(y_m | y_1, \dots, y_{m-1})$$

$$= \sum_{x_m} \sum_{x_{m-1}} \dots \sum_{x_1} \pi(x_1, \dots, x_m) \cdot \frac{\pi(y_1, x_2, \dots, x_m)}{\sum_{x_1} \pi(x_1, \dots, x_m)} \pi(y_2 | y_1, x_3, \dots, x_m) \dots \pi(y_m | y_1, \dots, y_{m-1})$$

$$=$$

$$\sum_{x_m} \sum_{x_{m-1}} \dots \sum_{x_2} \pi(y_1, x_2, \dots, x_m) \pi(y_2 | y_1, x_3, \dots, x_m) \dots \pi(y_m | y_1, \dots, y_{m-1})$$

$$= \dots$$

$$=$$

$$\sum_{x_m} \pi(y_1, \dots, y_{m-1}, x_m) \frac{\pi(y_1, \dots, y_m)}{\sum_{x_m} \pi(y_1, \dots, y_{m-1}, x_m)} = \pi(y_1, \dots, y_m)$$

Gibbs sampling specific steps are as follows

By markov chain $X_n(\omega)$ sample, a sample (y_1, \dots, y_m) of $X_{n+1}(\omega)$ can be obtained according to the following procedure

1 first, y_1 is obtained by a random variable $X_{n+1,1}(\omega), X_{n+1,1}(\omega)$ obey the distribution of $\{\pi(y_1 | x_2, \dots, x_m), y_1 \in D_1\}$ (x_2, \dots, x_m from $X_n(\omega)$)

2 secondly, y_2 is obtained by a random variable $X_{n+1,2}(\omega), X_{n+1,2}(\omega)$ obey the distribution of $\{\pi(y_2 | y_1, x_3, \dots, x_m), y_2 \in D_2\}$, In this down, $y_k (k=1, \dots, m-1)$ is obtained by a random variable $X_{n+1,k}(\omega), X_{n+1,k}(\omega)$ obey the distribution of $\{\pi(y_k | y_1, \dots, y_{k-1}, x_{k+1}, \dots, x_m), y_k \in D_k\}$

In the end , y_m is obtained by a random variable

$X_{n+1,m}(\omega), X_{n+1,m}(\omega)$ obey the distribution of

$$\{\pi(y_m | y_1, \dots, y_{m-1}), y_m \in D\}$$

Definition $y \triangleq (y_1, \dots, y_m)$ is a sample of $X_{n+1}(\omega)$

Now, Take an initial value $X_0(\omega) = y^{(0)}$, $y^{(1)}$ is obtained by the random variable $X_1(\omega)$ According to the above methods. To get the sample $y^{(2)}, \dots, y^{(n)}$ from $X_2(\omega), \dots, X_n(\omega)$. When n is large, the distribution of $X_n(\omega)$ approximation of $\pi(x_1, \dots, x_m)$, Can approximate thought, $y^{(n)}$ is a sample obey the distribution of $\pi(x_1, \dots, x_m)$

Example 2 from example 1, can obtained full conditional distribution of (θ, η) from its Posterior distribution

$$\pi(\theta | y, z, \eta) \propto \theta^{z_1+y_2} [(1-\eta)-\theta]^{y_3} \propto \left(\frac{\theta}{1-\eta}\right)^{z_1+y_2} \left(1-\frac{\theta}{1-\eta}\right)^{y_3}$$

$$= (1-\eta)Be(z_1 + y_2 + 1, y_3 + 1)$$

$$\pi(\eta | y, z, \theta) = (1-\theta)Be(y_3 + z_2 + 1, y_3 + 1)$$

$$\pi(z_1 | y, \theta, \eta) = \pi(z_1 | y_1, \theta) = b(y_1, \frac{2\theta}{2\theta+1})$$

$$\pi(z_2 | y, \theta, \eta) = \pi(z_2 | y_4, \eta) = b(y_4, \frac{2\eta}{2\eta+3})$$

the Gibbs sampling can be carried out on the full conditions of distribution.

Let The initial value $\theta = 0.5, \eta = 0.5$, Iterative sampling program compiled with MATLAB, Below is the first 1000 θ and η values draw the image

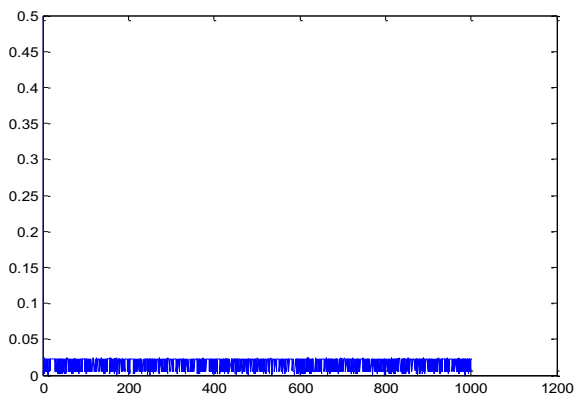


Figure 1

Figure of θ

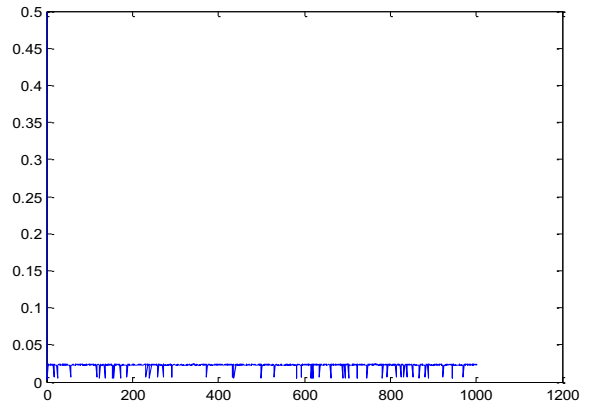


Figure 2

Figure of η

4 On the convergence of Markov chains

To make the markov chain monte carlo method is applied to practical problems, to determine Markov chains when to converge to the stationary distribution. convergence of markov chains is a property in theory, i.e. There is always a time t ,

Markov chains convergence to a stationary distribution when $t' > t$. We generally through a special markov chains to judge whether the convergence, but state is usually high dimension, and so it is difficult to estimate. Most judge convergence of the method is based on the monitoring of one or some state the number of functions, The most common way is as time increases, simply draw these functions, and then on intuitive determine when to stationary distribution

Example 3 Gelfand and Smith to Gibbs sampling data in case 1. they judge convergence adopt the mean value of ergodic, Every 10 data in the markov chain, take a sample, Calculate a traverse the mean every 20 sample. After 5000 times of the calculation, the Gibbs sampling has converged, Get the following estimate

$$E(\theta | y) = 0.520$$

$$E(\eta | y) = 0.123$$

$$Var(\theta | y) = 0.018$$

$$Var(\eta | y) = 0.0065$$

Gelfand and Smith still calculated $E(\theta | y) = 0.520$ and $E(\eta | y) = 0.123$ use quadratic gaussian integral method, Both consistent

Conclusion

to integrate complex and high dimensional functions is extremely difficulty in Bayesian statistics, however, the Markov chain Monte Carlo method provides a good way, we sample from the posterior directly, and obtain sample estimates of the quantities of interest, thereby performing the integration implicitly. This method is also applicable to other fields, has the widespread application prospect

References

- [1] G. L. Gong , M. P. Qian. The application of stochastic process . Tsinghua University press, Beijing, 2004
- [2].Sh.S.Mao,J.L.Wang,X.L.Pu. Advanced Mathematical Statistics, Higher education press, Beijing, 1998
- [3] Y.Q. Li, Ch. Liu, Q. Wang. stochastic process. National Defense Industry Press, Beijing, 2009
- [4] Gelfand A E, Smith A F M. samling based approaches to calculating marginal densities. J Amer Statist Assoc, 1990, 85: 398-409
- [5] Gelfand A E, Smith A F M. Gibbs sampling for marginal posterior expectations. Communications in Statistics A, 1991, 20: 1747-1766

- [4] Walsh, B. (2004). Markov Chain Monte Carlo and Gibbs Sampling. Lecture Notes for EEB 581, version 26 April 2004.
- [5] Gareth O. Roberts , Jeffrey S . Rosenthal., General state space Markov chains and MCMC algorithms, Probability surveys Vol.1 (2004) 20-71