

Notes for Bayesian

Jinze LI

June 13, 2024

Abstract

Notes Bayesian

1 Intro

conjugate prior

Drawing a dice: the draw follows the binomial dist, the prior follows the Beta dist. By computation, the posterior follows the Beta Dist also.

Books: 1. [知乎](#) 2. 统计学习方法, 李航 3. Geyer's Web [web1](#) 4. Gibbs sampling 和分布式[王益](#)

2 rejection method

simple rejection sampling [知乎](#)

3 Frequentist MSE of Bayesian estimator

Chapter6 page38 比较 bayesian estimation 的效果和 freq 的 estimator 的效果。因为 bayesian 估计出来是一个分布, 所以我们要给这个后验分布一个点估计 (eg. posterior mean)。其实这个后验的点估计和 freq 的比如 mle 方法估计出来的 estimator 都是 X 和 parameters 的函数而已。

因此，我们可以比较它们。只是 bayesian 的多了 prior 的参数，你甚至可以选择 prior 使得这个 posterior 的性质 (eg. MSE) 更好。

4 MCMC and Metropolis Hastings

这里面最基本而最重要的思想就是 MCMC。当 markov chain 收敛稳定的时候，我们从不同的状态之间的跳转会成为一个稳定的概率分布 $\pi P = \pi$ ，也就是相当于我们按这个 chain 的转移规律来抽样的话，最后长时间得到的不同 state 的概率分布就是 $\pi = [\pi_{s1}, \pi_{s2}, \dots]$ 那么我们想要的分布采样都将是平稳分布 π 的样本介绍 [mcmc 统计之都](#) want

$$p(i)q(i, j)\alpha(i, j) = p(j)q(j, i)\alpha(j, i) \quad (*) \quad (1)$$

The Metropolis-Hastings algorithm [Short Video from Youtube](#)

Metropolis-Hastings

Goal: Sample from $p(x)$

MCMC Idea: $p(x) = \frac{f(x)}{N}$

Algorithm Steps:

- ① Sample from $q(x_{t+1}|x_t)$ (Easier)
- ② Accept w/ Prob $A(x_t \rightarrow x_{t+1})$

Detailed Balance: $f_{a,b}$

$$p(a)T(a \rightarrow b) = p(b)T(b \rightarrow a)$$

$$\frac{f(a)}{N} g(b|a) A(a \rightarrow b) = \frac{f(b)}{N} g(a|b) A(b \rightarrow a)$$

$$A(a \rightarrow b) = \frac{f(b)}{f(a)} \times \frac{g(a|b)}{g(b|a)}$$

Acceptance Probability:

$$A(a \rightarrow b) = \min(1, r_f r_g)$$

Intuition: $A(a \rightarrow b) = \min(1, \frac{p(b)}{p(a)})$

Final Formula:

$$A(a \rightarrow b) = \min(1, \frac{p(b)}{p(a)})$$

Figure 1: Metropolis - Hastings

这里 rejection 如下，如果 ratio 大于 1，肯定 move 到新的，uniform[0,1] 怎么取也是小于 1。如果是 ratio 小于 1，我们想倾向于保留原来的 θ_{t-1} 作为下一个 θ_t 。但是！一般也不是 100% 这样，我们还是要给新的 θ_{new} 一个机会。(当 ratio 越大，新的机会越大) 所以我们还是把 $u < ratio$

作为一个接受新 θ_{new} 的一个 accepted probability.

$$Accepted \text{ Prob} = \min[ratio, 1]$$

where

$$ratio = \frac{\text{Posterior prob. of } \theta_{New}}{\text{Posterior prob. of } \theta_{t-1}}$$

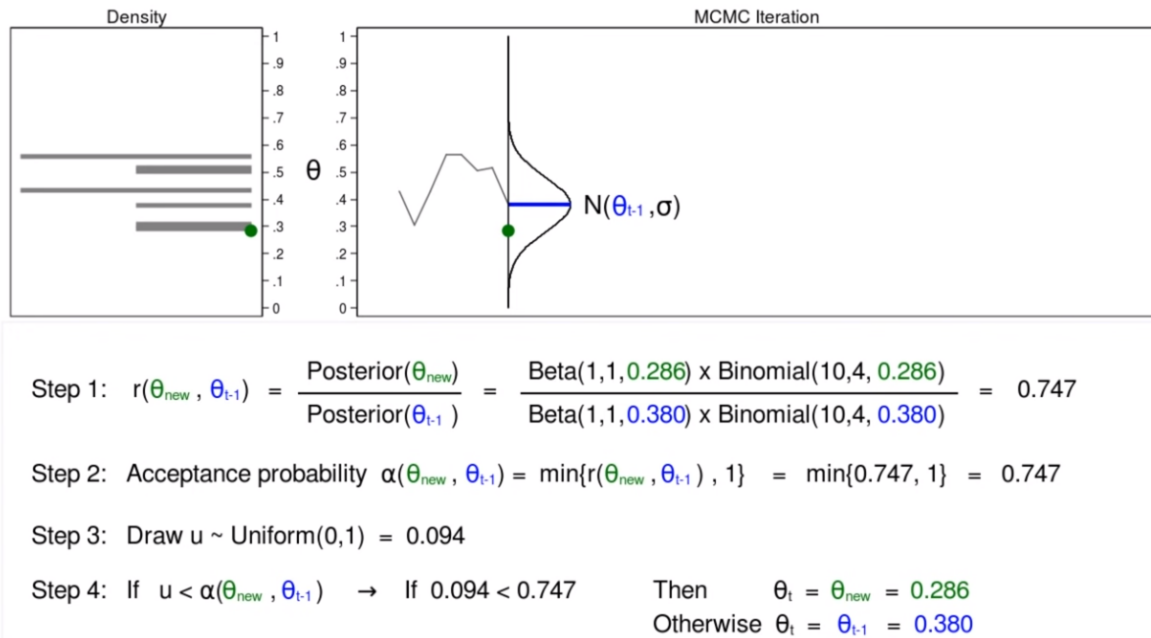


Figure 2: Enter Caption

其实，转移机会 $q(\theta^t|\theta')$ 并不一定是对称的 $q(\theta^t|\theta') \neq q(\theta'|\theta^t)$ 所以

$$ratio = \frac{(\text{Posterior prob. of } \theta_{New}) \cdot q(\theta^t|\theta')}{(\text{Posterior prob. of } \theta_{t-1}) \cdot q(\theta'|\theta^t)}$$

see Fig.3

issues: 1. dependent on the first value. 由此可以看出，初始值的选取其实很大程度影响了一段时间的采样，因此我们可以去除这一部分。which is called "burning in" period. 那多少的时间需要被去除？

2. auto-correlation due to the MC chain

<https://arxiv.org/pdf/1504.01896.pdf>

1. Given $\theta^{(t)}$, $t = 0, 1, 2, \dots$, generate a sample θ' from $q(\theta|\theta^{(t)})$, where q is called the **proposal distribution**.
2. Calculate the acceptance probability:

$$\alpha(\theta^{(t)}, \theta') = \min \left\{ \frac{L(\theta')f(\theta')q(\theta^{(t)}|\theta')}{L(\theta^{(t)})f(\theta^{(t)})q(\theta'|\theta^{(t)})}, 1 \right\}$$
3. Set $\theta^{(t+1)} = \theta'$ with probability $\alpha(\theta^{(t)}, \theta')$ and set $\theta^{(t+1)} = \theta^{(t)}$ with probability $1 - \alpha(\theta^{(t)}, \theta')$.

Figure 3: Enter Caption

5 Advanced Method

The beginners guide to Hamiltonian Monte Carlo [hmc](#)

<https://bobondemon.github.io/2022/05/07/Hamiltonian-Monte-Carlo/> ([Childers et al., 2022](#))

6 Exercise

SDSC8004

References

- Childers, David, Jesús Fernández-Villaverde, Jesse Perla, Christopher Rackauckas, and Peifan Wu. 2022. “Differentiable State-Space Models and Hamiltonian Monte Carlo Estimation.” Tech. rep., National Bureau of Economic Research.