Markov chain Monte Carlo sampling studies

Fu Amos <fs1984@msn.com>

Fri 2016-06-10 1:20 PM

To: nizar.bouguila@concordia.ca <nizar.bouguila@concordia.ca>;

1 attachments (4 KB)

MCMC.zip;

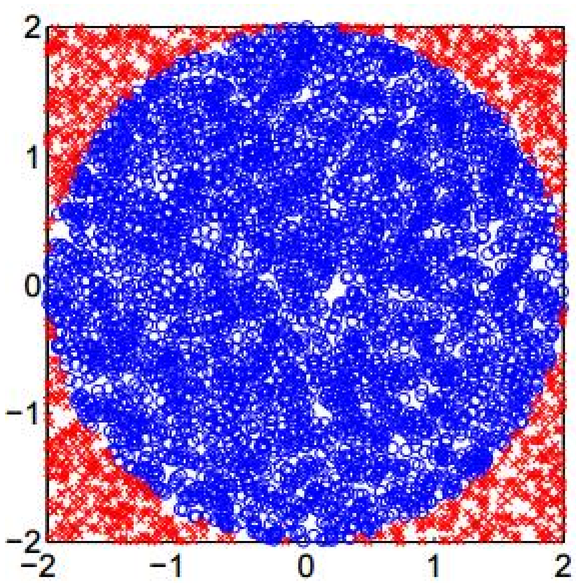
Dear Doctor Bouguila,

I've been studying MCMC sampling methods following your instructions.I started from the Chapter 11 of Bishop's textbook and read some related papers. For better understanding of the algorithms, I also implemented some small examples in Matlab and the source code is attached in this mail. The source code demonstrates basic Monte Carlo method (calculate PI), Metropolis-Hastings method(1-D mixed Gaussian sampling) and Gibbs sampling of 2-D Gaussian.

From best of my understanding, MCMC sampling consists of the following aspects:

- **Monte Carlo Method**: Algorithms using random sampling method to approximate target distribution can be seen as an implementation of Monte Carlo Method,specially when sampling from target distributions is hard.

  Example: Calculate PI using Monte Carlo method:



As we know the blue area is PI*r^2 and the red square are is 4*r^2, so the Blue/Red = PI/4. To approximate the PI,firstly, we simply generate samples (x,y) (-2<x<2, -2<y<2), then check them if they are inside of the blue circle (x^2 + y^2 < r^2). So the count of samples inside blue circle divided by the total count of samples is PI/4.

- **Metropolis-Hastings Algorithm:** If the target distribution is a probability density function and direct sampling from it is difficult.So a proposal distribution is introduced to approximate the target because sampling from proposal distribution is easy. Also the sampling procedure should be guided by an acceptance criteria since the sampling domain can be very large. From my understanding, every sampling step should be better than the previous one, otherwise this step should be discard. The acceptance probability is:

$$A_k(\mathbf{z}^\star, \mathbf{z}^{(\tau)}) = \min\left(1, \frac{\widetilde{p}(\mathbf{z}^\star) q_k(\mathbf{z}^{(\tau)}|\mathbf{z}^\star)}{\widetilde{p}(\mathbf{z}^{(\tau)}) q_k(\mathbf{z}^\star|\mathbf{z}^{(\tau)})}\right)$$

In the example (1-d mixed Gaussian), the proposal distribution is a normal distribution which is symmetric, so it can be also called random walk MCMC:

$$A(\mathbf{z}^\star, \mathbf{z}^{(\tau)}) = \min\left(1, \frac{\widetilde{p}(\mathbf{z}^\star)}{\widetilde{p}(\mathbf{z}^{(\tau)})}\right)$$

The procedure of MH algorithm includes the following steps:

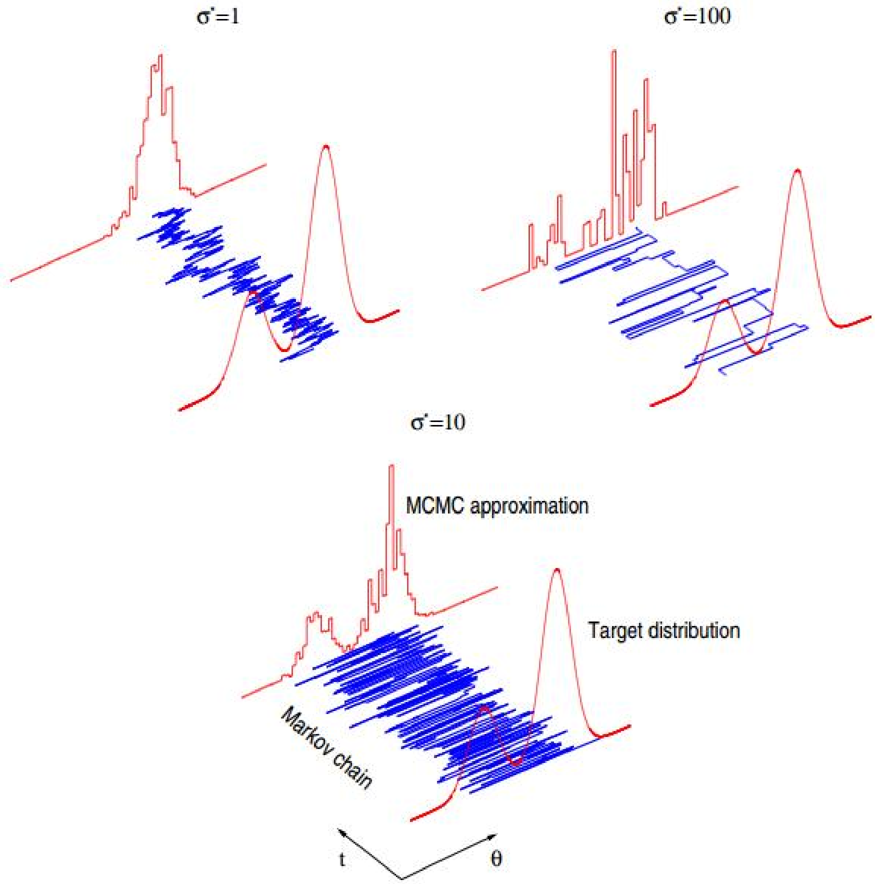
1. Initialize X_0 arbitrarily.
2. For $s = 0, 2, \ldots$
   (a) Generate a proposed state $x' \sim q(x'|x_s)$
   (b) Evaluate the acceptance propability

$$\begin{aligned} \alpha &= \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)} = \frac{\pi(x')/q(x'|x)}{\pi(x)/q(x|x')} \\ r(x'|x) &= \min\{1, \alpha\} \end{aligned}$$

   (c) Set

$$X_{s+1} = \begin{cases} x' & \text{with probability } r \\ x_s & \text{with probability } 1 - r \end{cases}$$

The main problem of MH algorithm is that finding a suitable proposal distribution sometimes is difficult. If proposal distribution does not fit, rejection rate could be very high and the result won't be accurate.



- **Gibbs sampling:** the main advantage of Gibbs sampling compared to MH algorithm is that proposal distribution is not needed for Gibbs sampling and the acceptance probability is always 1:

$$A(\mathbf{z}^\star, \mathbf{z}) = \frac{p(\mathbf{z}^\star) q_k(\mathbf{z}|\mathbf{z}^\star)}{p(\mathbf{z}) q_k(\mathbf{z}^\star|\mathbf{z})} = \frac{p(z_k^\star|\mathbf{z}_{\setminus k}^\star) p(\mathbf{z}_{\setminus k}^\star) p(z_k|\mathbf{z}_{\setminus k}^\star)}{p(z_k|\mathbf{z}_{\setminus k}) p(\mathbf{z}_{\setminus k}) p(z_k^\star|\mathbf{z}_{\setminus k})} = 1$$

Gibbs sampling procedure can be described as following:

1. Initialize $\{z_i : i = 1, \ldots, M\}$
2. For $\tau = 1, \ldots, T$:
   - Sample $z_1^{(\tau+1)} \sim p(z_1|z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.
   - Sample $z_2^{(\tau+1)} \sim p(z_2|z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$.

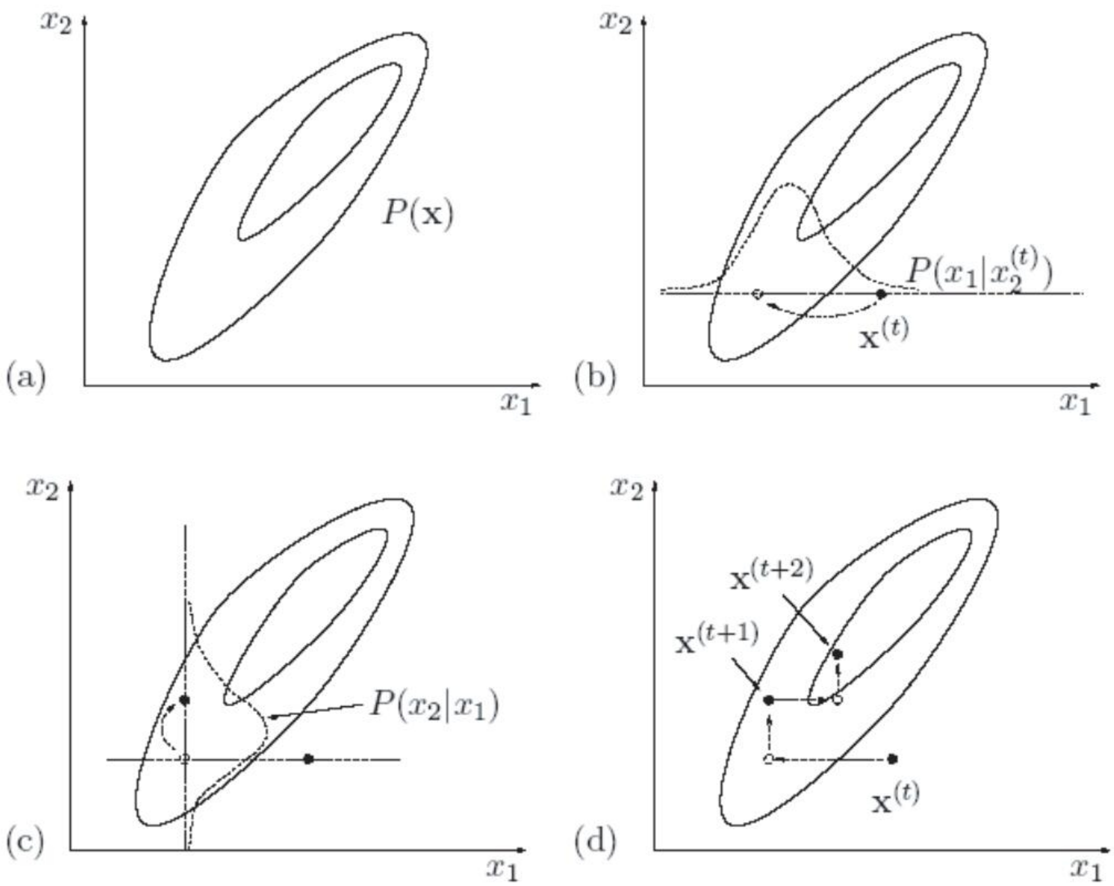
     $\vdots$
   - Sample $z_j^{(\tau+1)} \sim p(z_j|z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots, z_M^{(\tau)})$.

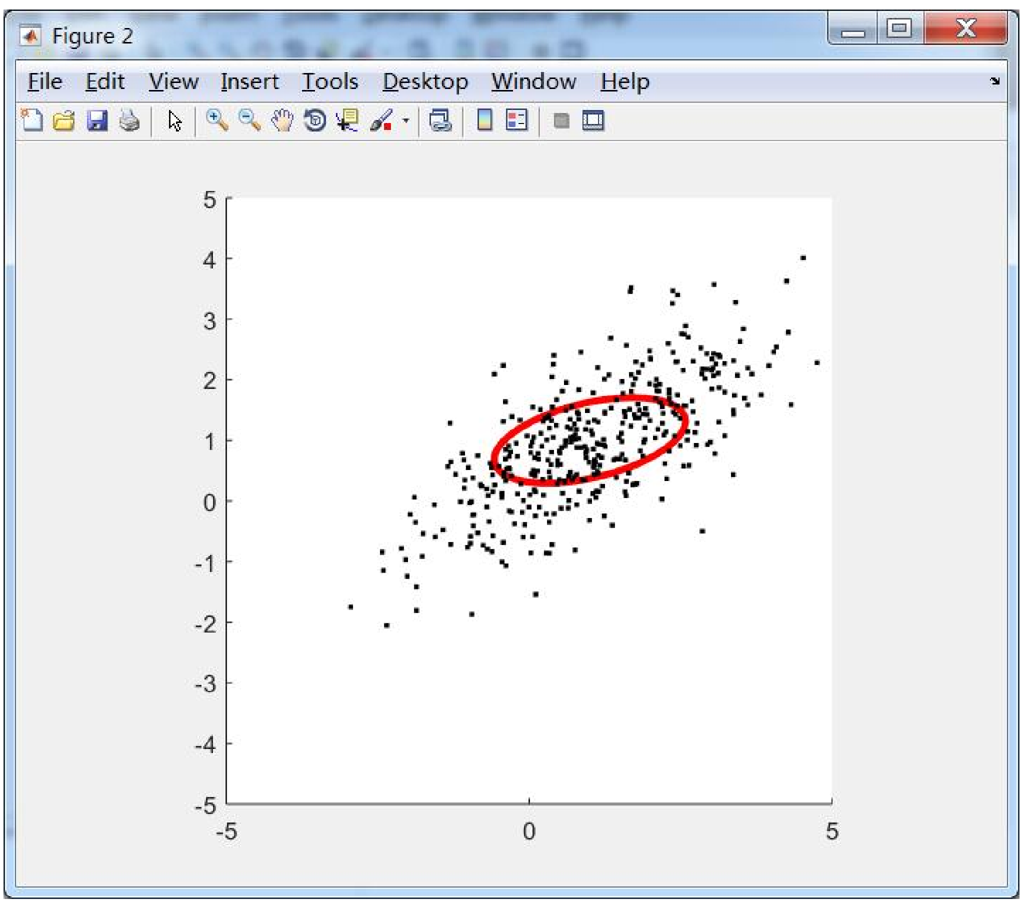
     $\vdots$
   - Sample $z_M^{(\tau+1)} \sim p(z_M|z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{M-1}^{(\tau+1)})$.

The key of Gibbs sampling is to find the conditional probability of all variables of target distribution, for example, the 2-D Gaussian distribution:

$$
\begin{aligned}
p(x_1|x_2) &= \mathcal{N}(x_1; \mu_{1|2}, \Sigma_{1|2}) \\
\mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\
\Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
\end{aligned}
$$

the sampling procedure and result:





reference: http://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall06/reading/mcmc.pdf

- Summary

Compared to clustering methods, sampling is a reversed procedure and related to many statistic concepts. When direct calculation is impossible, MCMC method can be very useful to approximate target distribution. Besides those discussed above, MCMC also have many improvements like Hybrid Monte Carlo and Gibbs with HM algorithm etc, that I need to work with in the future. I'll be very glad to have your corrections and suggestions regarding to my understanding of MCMC and looking forward to your further instructions.

Thank you,
Shuai Fu