STAT 425: Introduction to Bayesian Analysis

Marina Vannucci

Rice University, USA

Fall 2018

Part 2: Markov chain Monte Carlo (MCMC) methods

- Markov chains
- MCMC methods - Metropolis-Hastings and Gibbs Sampler
- The multivariate normal model

Introduction

- We have seen that Monte Carlo sampling is a useful tool for sampling from distributions
- In simple models, especially with conjugate prior distributions, it is often easy to derive the posterior distribution in closed form. MC sampling is convenient (though not really necessary).
- When the posterior density does not have a recognizable form, it might be possible to factor the distribution analytically and simulate in parts, as we did for the Normal model.
- For more complicated problems, it may not possible to directly generate samples from the target distribution (e.g., non-conjugate settings and, in general, situations where we cannot sample from the joint posterior distribution)
- We can however build a sequence (Markov chain) whose distribution converges to the target distribution and use Markov chain Monte Carlo (MCMC) methods.

Markov Chains (discrete case - brief review)

A Markov chain is a sequence of random variables $X_1, X_2, X_3, \ldots$ with the Markov property, namely that, given the present state, the future and past states are independent. The possible values of X_i form a countable set S called the **state space** of the chain.

$$Pr(X_{n+1} = x_j | X_n = x_n, \ldots, X_1 = x_1) = Pr(X_{n+1} = x_j | X_n = x_n).$$

Properties of a Markov Chain:

1. Reducibility
2. Periodicity
3. Recurrence
4. Ergodicity

Markov Chains properties:

- **Reducibility** A state j is said to be accessible from a state i if a system started in state i has a non-zero probability of transitioning into state j at some point.

$$Pr(X_n = j | X_0 = i) = p_{ij}^{(n)} > 0$$

- **Irreducibility** A chain is irreducible if all states communicate (there's a positive probability to visit all states in a finite number of steps).
- **Periodicity** A state i has period k if any return to state i must occur in multiples of k time steps. If $k = 1$, then the state is said to be aperiodic i.e. returns to state i can occur at irregular times.

Markov Chains properties:

- **Recurrence** A state i is said to be **transient** if, given that we start in state i, there is a non-zero probability that we will never return to i. A state i is called **absorbing** if it is impossible to leave this state. A state i is **recurrent** if the expected number of visit to i is equal to infinity.
- **Ergodicity: to what is the chain converging?**

$$\lim_{n \to \infty} \|P^{(n)} - \pi\| = 0$$

that is the distribution $P^{(n)}$ of X_n converges to a target invariant distribution π irrespective of the initial conditions.
A chain is said to be ergodic if it is aperiodic and positive recurrent.
(positive = the chain has an invariant probability measure)

Additional property

- A chain is **reversible** if and only if $\pi_j p_{ji} = \pi_i p_{ij}$ for all $i \neq j$

$$p_{ij} = Pr(X_n = j | X_{n-1} = i) \text{ transition prob}$$

$$\pi_i \text{ equilibrium probability of being in state } i$$

  This implies that the joint probability $Pr(X_{n-1} = i, X_n = j)$ is symmetric in i and j.

- **Important result:** A Markov chain with transition matrix P will have an equilibrium distribution π if and only if it is reversible.

- In this way, to sample from the limiting distribution π, we run a Markov Chain with transition matrix P satisfying the detailed balance condition until the chain appears to have settle down to equilibrium.

Monte Carlo methods for Markov Chains

Classical LLN's and CLT's not directly applicable due to:

- Markovian dependence structure between the obs X_i,
- Non-stationarity of the sequence.

Theorem (Ergodic Theorem)

If the Markov chain (X_n) is Harris recurrent (irreducible, aperiodic and positive recurrent) then for any function h with $E|h| < \infty$,

$$\lim_{n \longrightarrow \infty} \frac{1}{n} \sum h(X_i) = \int h(x)\pi(x)dx$$

If the chain is also reversible $(X_{n+1}|X_{n+2} = x \sim X_{n+1}|X_n = x)$ then a "modified" version of the CLT holds.

The MCMC principle - general idea

- A *Markov Chain Monte Carlo (MCMC) method* for simulation from a distribution π is any method producing an ergodic Markov chain (X_n) whose stationary distribution is π.
- For an arbitrary starting value $x^{(0)}$, a chain (X_n) is generated using a transition kernel with stationary distribution π, which ensure the convergence in distribution of X_n to a random variable from π.
- In order for the Markov chain to converge to the target (stationary or equilibrium) distribution π, it must be irreducible, aperiodic and positive recurrent
- The main problem is how to construct a suitable chain; we will see some of the most used algorithms (namely, Metropolis-Hastings and Gibbs sampling) and also look into methods to assess whether the chain has reached its stationary distribution.
- MCMC was invented in 1953 but was not widely applied until statisticians became aware of it around 1990. Then problems that had previously been intractable suddenly became tractable.