ST 740: Markov Chain Monte Carlo

Alyson Wilson

Department of Statistics
North Carolina State University

October 2, 2013

Quick Review

- Model
  1. Likelihood
  2. Prior
     - Conjugate
     - Diffuse
     - Informative/noninformative
     - Proper/improper
- Bayes' Theorem: Posterior is proportional to prior times likelihood

Quick Review

- Summarizing the posterior distribution: usually one dimensional summaries of marginal posterior distributions, posterior distributions of functions of parameters, or predictive distributions

  1. Analytical
     - Conjugate prior distribution: posterior distribution has "nice" form
     - Marginal distributions
     - Conditional distributions
     - $$\pi(\theta_1 \mid \mathbf{y}) = \int \pi(\theta_1 \mid \theta_2, y)\pi(\theta_2 \mid \mathbf{y})d\theta_2$$
  2. Simulation
     - Nonconjugate prior distributions
     - Functions of parameters
     - Brute force, rejection, SIR/weighted bootstrap, Metropolis, Metropolis-Hastings
     - MCMC: Gibbs sampling

Short Stochastic Process Tutorial

A stochastic process is a sequence of random variables $\{X(t), t \in T\}$ where

- $X(t)$ is the *state* of the process at time t
- T is the set of time points at which we observe $X(t)$
- The *state space* is the set of possible values of $X(t)$

Short Stochastic Process Tutorial

- Here we consider discrete time stochastic processes, with either discrete or continuous state space
- Colloquially, a stochastic process has the *Markov property* if, given the present, the future does not depend on the past.

  For a discrete-time Markov chain, we can write:

$$\begin{aligned} P(X(t) &\in A \,|\, X(t-1) = x_{t-1}, ..., X(1) = x_1) \\ &= P(X(t) \in A \,|\, X(t-1) = x_{t-1}) \end{aligned}$$

- A process in discrete time with the Markov property is called a *Markov chain*

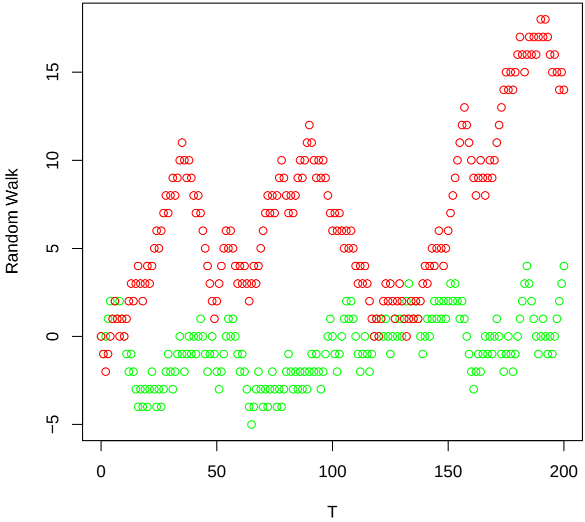
Discrete Time, Discrete Space Markov Chain

A simple example of a Markov chain is the random walk. At each time point, move right one step with probability p or move left one step with probability $1 - p$.

$$
\begin{aligned}
p(j+1 \,|\, j) &= p \\
p(j-1 \,|\, j) &= 1-p
\end{aligned}
$$

Starting at $X(0) = 0$ move left or right by 1 with probability $p = 0.5$ over $T = 200$ steps.

Random Walk



Another Example: Gary's Mood

On any given day, Gary is either cheerful (C), so-so (S), or grumpy (G). If he is cheerful today, then he will be C, S, or G tomorrow with respective probabilities 0.5, 0.4, 0.1. If he is feeling so-so today, then he will be C, S, or G tomorrow with probabilities 0.3, 0.4, 0.3. If he is grumpy today, then he will be C, S, or G tomorrow with probabilities 0.2, 0.3, 0.5.

Let $X(n)$ denote Gary's mood on the nth day. We have a discrete time, discrete state space (C, S, G) Markov chain.

Another Example: Gary's Mood

Let $p(j \mid i)$ be the probability that the process, when in state i, next make a transition to state j. We can collect these probabilities in a matrix P that we call the *one-step transition probability* matrix.

$$P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}$$

Define $p^n(j \mid i)$ as probability of getting to state j from state i in n steps. You can calculate the n-step transition matrix by matrix multiplying P by itself n times.

Properties of Markov Chains

Theorem: Irreducible, ergodic Markov chains have a limiting distribution:

$$\lim_{n \to \infty} p^n(j \mid i) = \pi_j$$

with $p^n(j \mid i)$ the probability that we reach j from i after n steps or transitions.

We can think of π_j as the long-run proportion of the time that the stochastic process will be in state j.

Markov Chain Simulation

- **Idea:** Suppose that sampling from $\pi(\theta \mid \mathbf{y})$ is hard, but that we can generate (somehow) a Markov chain $\{\theta(t), t \in T\}$ with stationary distribution $\pi(\theta \mid \mathbf{y})$.
- Situation is different from the usual stochastic process case:
  - Here we know the stationary distribution.
  - We seek transitions $\pi(\theta^{(t+1)} \mid \theta^{(t)})$ that will take us to the stationary distribution.
- Idea: start from some initial guess $\theta^{(0)}$ and let the chain run for n steps (n large), so that it reaches its stationary distribution.

Markov Chain Simulation

- After convergence, all additional steps in the chain are draws from the stationary distribution $\pi(\theta \mid \mathbf{y})$.
- MCMC methods all based on the same idea; difference is just in how the transitions in the MC are created.

Gibbs Sampling

- A MCMC iterative algorithm that produces Markov chains with joint stationary distribution $\pi(\theta \,|\, \mathbf{y})$ by cycling through all possible full conditional (posterior) distributions.
- Example: suppose that $\theta = (\theta_1, \theta_2, \theta_3)$, and that the target distribution is $\pi(\theta_1, \theta_2, \theta_3 \,|\, \mathbf{y})$. Steps in the Gibbs sampler are:
  1. Start with a guess $(\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)})$
  2. Draw $\theta_1^{(1)}$ from $\pi(\theta_1 \,|\, \theta_2 = \theta_2^{(0)}, \theta_3 = \theta_3^{(0)}, \mathbf{y})$
  3. Draw $\theta_2^{(1)}$ from $\pi(\theta_2|\theta_1 = \theta_1^{(1)}, \theta_3 = \theta_3^{(0)}, \mathbf{y})$
  4. Draw $\theta_3^{(1)}$ from $\pi(\theta_3|\theta_1 = \theta_1^{(1)}, \theta_2 = \theta_2^{(1)}, \mathbf{y})$
- Steps above complete **one iteration** of the Gibbs sampler
- Repeat the steps above *n* times, and after convergence (more later), draws $(\theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)})$ are sample from stationary distribution $\pi(\theta \,|\, \mathbf{y})$.

Two-Parameter Case

Consider the two parameter case, where we are interested in sampling from $\pi(\theta_1, \theta_2 \mid \mathbf{y})$.

- The full conditional distributions are $\pi(\theta_1 \mid \theta_2, \mathbf{y})$ and $\pi(\theta_2 \mid \theta_1, \mathbf{y})$.
- If $\pi(\theta_1 \mid \theta_2, \mathbf{y})$ and $\pi(\theta_2 \mid \theta_1, \mathbf{y})$ are easier to sample from than $\pi(\theta_1, \theta_2 \mid \mathbf{y})$, then we should consider using Gibbs sampling to get a sample from $\pi(\theta_1, \theta_2 \mid \mathbf{y})$.

The Non-Conjugate Normal Example

Let $Y \mid \mu, \sigma^2 \sim \text{Normal}(\mu, \sigma^2)$, with (μ, σ^2) unknown.

Consider the prior:

$$\begin{aligned} \mu &\sim \text{Normal}(\theta_0, \tau_0^2) \\ \sigma^2 &\sim \text{InverseGamma}(\nu_0, \eta_0). \end{aligned}$$

The joint posterior distribution is

$$\begin{aligned} \pi(\mu, \sigma^2 \mid \mathbf{y}) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2} \sum (y_i - \mu)^2\} \\ &\times \exp\{-\frac{1}{2\tau_0^2}(\mu - \theta_0)^2\} \\ &\times \left(\frac{1}{\sigma^2}\right)^{\nu_0 + 1} \exp\left(-\frac{\eta_0}{\sigma^2}\right) \end{aligned}$$

Full Conditional Distributions

For μ

Collect terms with μ in them

$$
\begin{aligned}
\pi(\mu \,|\, \sigma^2, \mathbf{y}) &\propto \exp\left(-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2\right)\exp\left(-\frac{1}{2\tau_0^2}(\mu-\theta_0)^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(n\mu^2-2n\bar{y}\mu)-\frac{1}{2\tau_0^2}(\mu^2-2\theta_0\mu)\right) \\
&\propto \exp\left(-\frac{1}{2}\frac{n\tau_0^2+\sigma^2}{\sigma^2\tau_0^2}[\mu^2-2\left(\frac{n\tau_0^2\bar{y}+\sigma^2\theta_0}{n\tau_0^2+\sigma^2}\right)\mu]\right) \\
&\propto \text{Normal}(\theta_n, \tau_n^2),
\end{aligned}
$$

Full Conditional Distributions

$$
\begin{aligned}
\theta_n &= \frac{n\tau_0^2\bar{y} + \sigma^2\theta_0}{n\tau_0^2 + \sigma^2} \\
&= \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\tau_0^2}\theta_0}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}} \\
\tau_n^2 &= \frac{\sigma^2\tau_0^2}{n\tau_0^2 + \sigma^2} \\
&= \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}.
\end{aligned}
$$

Full Conditional Distributions

For σ^2

Collect terms with σ^2 in them

$$\begin{aligned}\pi(\sigma^2 \mid \mu, \mathbf{y}) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\nu_0+1} \exp\left(-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2 - \frac{\eta_0}{\sigma^2}\right) \\ &\propto \text{InverseGamma}(\frac{n}{2}+\nu_0, \eta_0 + \frac{1}{2}\sum(y_i-\mu)^2)\end{aligned}$$

How do we implement the Gibbs sampler?

Gamma Example

Suppose that our sampling distribution is Gamma(α, β) and we choose independent marginal prior distributions $\pi(\alpha) \sim$ Gamma$(2, 1)$ and $\pi(\beta) \sim$ Gamma$(5, 1)$.

$$
\begin{aligned}
\pi(\alpha, \beta \mid \mathbf{y}) &\propto \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \exp(-\beta \sum y_i)(\prod y_i)^{\alpha-1} \exp(-(\alpha+\beta))\alpha\beta^4 \\
&\propto \frac{\beta^{n\alpha+4}}{\Gamma(\alpha)^n} \exp(-\beta(\sum y_i + 1))\alpha \exp(-\alpha)(\prod y_i)^{\alpha-1}
\end{aligned}
$$

How do we implement the Gibbs sampler?