The Gibbs Sampler

Yingbo Li

Clemson University

MATH 9810

Markov Chain Monte Carlo Sampling

- We have seen that Monte Carlo sampling is a useful tool for simulating prior and posterior distributions
- By limiting to conjugate prior distributions, all models have tractable posterior distributions so sampling easy
- What if we want to use a non-conjugate prior distribution?
  - If posterior distribution is not "recognizable" but cdf is tractable, we can use inverse cdf method to sample.
- What if we cannot sample from the joint posterior distribution?

We can use Markov Chain Monte Carlo sampling!

Semi-Conjugate Prior for Normal

Revisit the normal model:

$$Y_i \mid \mu, \phi \overset{\text{iid}}{\sim} \mathsf{N}(\mu, \frac{1}{\phi})$$

But now assume that μ is independent of ϕ a priori:

$$\begin{aligned} \mu &\sim \mathsf{N}(\mu_0, 1/\omega_0) \\ \phi &\sim G(\nu_0/2, \mathsf{SS}_0/2) \end{aligned}$$

Note: Hoff uses τ_0^2 in the variance for μ.
I use $\omega_0 = 1/\tau_0^2$, but results are identical.

Posterior Distribution of μ, ϕ

Posterior Distribution:

$$p(\mu, \phi \mid Y) \propto \prod_i p(y_i \mid \mu, \phi) p(\mu) p(\phi)$$

$$\begin{aligned}
p(\mu, \phi \mid Y) &\propto \phi^{n/2} \exp\left(-\frac{\phi}{2} \sum_i (y_i - \mu)^2\right) \\
&\times \omega_0^{1/2} \exp(-\frac{\omega_0}{2} (\mu - \mu_0)^2) \;\; \phi^{\nu_0/2-1} \exp(-\phi \mathsf{SS}_0/2) \\
&= p(\mu \mid \phi, Y)\; p(\phi \mid Y) \\
&= p(\mu \mid Y)\; p(\phi \mid \mu, Y)
\end{aligned}$$

First Factorization: $p(\mu \mid \phi, Y)$

For $\mu \mid \phi, Y$ complete the square to show that

$$\mu \mid \phi, Y \sim \mathsf{N}\left(\frac{n\phi\bar{y} + \omega_0\mu_0}{n\phi + \omega_0}, (n\phi + \omega_0)^{-1}\right)$$

Can we recognize the marginal distribution for ϕ? No!

Second Factorization: $p(\phi \mid \mu, Y)$

$$\begin{aligned} p(\mu, \phi \mid Y) &\propto \phi^{n/2} \exp\left(-\frac{\phi}{2}\sum_i (y_i - \mu)^2\right) \\ &\times \omega_0^{1/2} \exp(-\frac{\omega_0}{2}(\mu - \mu_0)^2) \;\; \phi^{\nu_0/2-1} \exp(-\phi \mathsf{SS}_0/2) \end{aligned}$$

Hence,

$$p(\phi \mid \mu, Y) \propto \phi^{(n+v_0)/2-1} \exp\left(-\frac{\phi}{2}[\sum_i (y_i - \mu)^2 + \mathsf{SS}_0]\right)$$

Posterior Distribution of ϕ Given μ

Can recognize

$$\phi \mid \mu, Y \sim \mathsf{G}\left(\frac{n+\nu_0}{2}, \frac{\sum_i (y_i-\mu)^2 + \mathsf{SS}_0}{2}\right)$$

An equivalent expression useful for coding is

$$\phi \mid \mu, Y \sim \mathsf{G}\left(\frac{n+\nu_0}{2}, \frac{\mathsf{SS} + n(\bar{y}-\mu)^2 + \mathsf{SS}_0}{2}\right)$$

where $\mathsf{SS} = \sum_i (y_i - \bar{y})^2$

But, cannot recognize form of marginal distribution for μ

Approach for Posterior Sampling?

- Suppose we were given a value of μ that comes from the marginal posterior distribution (say $\mu^{(1)}$)
- We could draw a value of ϕ from the conditional Gamma distribution given $\mu = \mu^{(1)}$, which would give us a draw from the joint distribution $(\mu^{(1)}, \phi^{(1)})$
- $\phi^{(1)}$ could be viewed as a draw from the marginal distribution of ϕ (based on the first factorization), so if we now use the conditional distribution of $\mu \mid \phi, Y$ to draw a new $\mu^{(2)}$, we have another sample from the joint distribution $(\mu^{(2)}, \phi^{(1)})$
- Taking $\mu^{(2)}$, draw $\phi^{(2)}$ from the conditional distribution for $\phi \mid \mu, Y$, and so on....

Gibbs Sampler

Start with initial values for parameters, $(\mu^{(0)}, \phi^{(0)})$
For $t = 1, \ldots, T$, generate from the following sequence of full conditional distributions:

- $\mu^{(t)} \sim p(\mu \mid \phi^{(t-1)}, Y)$
- $\phi^{(t)} \sim p(\phi \mid \mu^{(t)}, Y)$
- Set $\theta^{(t)} = (\mu^{(t)}, \phi^{(t)})$

The sequence $\{\theta^{(t)}\colon\ t = 1, \ldots, T\}$ may be viewed (but is not necessarily...yet) as a dependent sample from the joint posterior distribution of (μ, ϕ)

MCMC Algorithms

The Gibbs sampler is an example of a Markov Chain Monte Carlo (MCMC) algorithm

- $\phi^{(t)}$ depends on the past sequence of draws $\phi^{(0)}, \ldots, \phi^{(t-1)}$ only through $\phi^{(t-1)}$ (Markov property)
- The empirical distribution of $(\mu^{(t)}, \phi^{(t)})$ approaches the posterior distribution as $t \to \infty$ no matter what the starting point
- As $T \to \infty$, the sample or ergodic average approaches the posterior mean, i.e., $\frac{1}{T} \sum_t g(\theta^{(t)}) \to \mathsf{E}[g(\theta) \mid Y]$
- Extends to more than 2 parameters

Extends to More than 2 Parameters

Suppose the full conditionals of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ are all tractable.

- Start the Gibbs sampler at the initial value $(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_m^{(0)})$.
- For the $(t+1)$ th iteration: generate from the following full conditionals sequentially:
  - $\theta_1^{(t+1)} \sim p(\theta_1 \mid \theta_2^{(t)}, \theta_3^{(t)}, \theta_4^{(t)}, \ldots, \theta_m^{(t)}, Y)$
  - $\theta_2^{(t+1)} \sim p(\theta_2 \mid \theta_1^{(t+1)}, \theta_3^{(t)}, \theta_4^{(t)}, \ldots, \theta_m^{(t)}, Y)$
  - $\theta_3^{(t+1)} \sim p(\theta_3 \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_4^{(t)}, \ldots, \theta_m^{(t)}, Y)$
  - $\vdots$
  - $\theta_m^{(t+1)} \sim p(\theta_m \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, \ldots, \theta_{m-1}^{(t+1)}, Y)$