DS 102 Discussion 5

Monday, March 2, 2020

The past few lectures have looked at how to perform Bayesian inference using Markov Chain Monte Carlo sampling methods, such as Gibbs sampling. Today, we'll practice deriving one iteration of Gibbs sampling for the Gamma-Poisson model.

The goal of Bayesian inference is to get the posterior distribution of the parameters given the data, $\mathbb{P}(\theta \mid X)$. Often this is difficult to derive in closed form, so instead we'll try to sample from it. However, sampling from the posterior over *all* the parameters θ is also often difficult. The main insight behind Gibbs sampling is that it can be much easier to sample the posterior over just a *single* parameter, $\mathbb{P}(\theta_i \mid X, \theta_{-i})$ (where we use the index $-i$ to mean all indices except for i). Gibbs sampling then iterates through each parameter θ_i and samples from $\mathbb{P}(\theta_i \mid X, \theta_{-i})$. This loop is repeated, each time conditioning on the newly sampled values.

Iterating through each parameter θ_i and sampling from $\mathbb{P}(\theta_i \mid X, \theta_{-i})$ is not the same thing as sampling from $\mathbb{P}(\theta \mid X)$. However, the good news is that given enough iterations, the former converges to the latter.

1. **Gibbs sampling for Gamma-Poisson model.**

   Consider the hierarchical Bayes model where

$$\begin{aligned} \beta &\sim \text{Gamma}(m, \alpha) \\ \theta_i \mid \beta &\sim \text{Gamma}(k, \beta), \quad i = 1, \dots, n \\ X_i \mid \theta_i &\sim \text{Pois}(\theta_i), \quad i = 1, \dots, n, \end{aligned}$$

   where the θ_i are independent of each other and the X_i are independent of each other. The β and θ_i are unknown parameters, and m, α, and k are fixed and known.

   We'd like to infer the parameters β and the θ from the data X. That is, we'd like to sample from the posterior distribution $\mathbb{P}(\beta, \theta \mid X)$ using Gibbs sampling. This entails deriving the posterior of each parameter, conditioned on the data and all the other parameters.

   (a) We'll start with β. Derive $\mathbb{P}(\beta \mid \theta_1, \dots, \theta_n, X_1, \dots, X_n)$.

   Solution:

We have

$$
\begin{aligned}
\mathbb{P}(\beta \mid \theta_1, \ldots, \theta_n, X_1, \ldots, X_n) &= \mathbb{P}(\beta \mid \theta_1, \ldots, \theta_n) \\
&= \frac{\mathbb{P}(\beta) \prod_{i=1}^n \mathbb{P}(\theta_i \mid \beta)}{\int_0^\infty \mathbb{P}(b) \prod_{i=1}^n \mathbb{P}(\theta_i \mid b)\, db} \\
&\propto_\beta \beta^{m-1} e^{-\alpha\beta} \prod_{i=1}^n \beta^k e^{-\beta\theta_i} \\
&\propto_\beta \beta^{nk+m-1} e^{-(\alpha + \sum_{i=1}^n \theta_i)\beta} \\
&\propto_\beta \text{Gamma}(nk + m, \alpha + \sum_{i=1}^n \theta_i).
\end{aligned}
$$

(b) Next, we'll look at each θ_i. Derive $\mathbb{P}(\theta_i \mid \beta, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n, X_1, \ldots, X_n)$

Solution:

$$
\begin{aligned}
\mathbb{P}(\theta_i \mid \beta, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n, X_1, \ldots, X_n) &= \mathbb{P}(\theta_i \mid \beta, X_i) \\
&= \frac{\mathbb{P}(\theta_i, \beta, X_i)}{\mathbb{P}(\beta, X_i)} \\
&= \frac{\mathbb{P}(X_i \mid \theta_i, \beta)\mathbb{P}(\theta_i \mid \beta)\mathbb{P}(\beta)}{\mathbb{P}(\beta, X_i)} \\
&= \frac{\mathbb{P}(\beta)\mathbb{P}(\theta_i \mid \beta)\mathbb{P}(X_i|\beta, \theta_i)}{\mathbb{P}(\beta) \int_0^\infty \mathbb{P}(u|\beta)\mathbb{P}(X_i \mid \beta, u)\, du} \\
&= \frac{\mathbb{P}(\theta_i \mid \beta)\mathbb{P}(X_i|\beta, \theta_i)}{\int_0^\infty \mathbb{P}(u|\beta)\mathbb{P}(X_i \mid \beta, u)\, du} \\
&\propto_{\theta_i} \theta_i^{k-1} e^{-\beta\theta_i} \theta_i^{X_i} e^{-\theta_i} \\
&\propto_{\theta_i} \theta_i^{X_i+k-1} e^{-(\beta+1)\theta_i} \\
&\propto_{\theta_i} \text{Gamma}(X_i + k, \beta + 1).
\end{aligned}
$$

(c) Using the posteriors you derived in the last two parts, write out the algorithm for the Gibbs sampler.

Solution:
Initialize $\beta^{(0)} \sim \text{Gamma}(m, \alpha)$ and $\theta_i^{(0)} \mid \beta = \beta^{(0)} \sim \text{Gamma}(k, \beta^{(0)})$ for all i.
For $t = 1, \ldots, T$ for some large stopping time T:

1. Start with $(\beta^{(t-1)}, \theta_1^{(t-1)}, \ldots, \theta_n^{(t-1)})$ from the previous iteration.

2. Sample $\beta^{(t)}$ according to

$$\beta^{(t)} \sim \mathbb{P}(\beta \mid \theta_1 = \theta_1^{(t-1)}, \ldots, \theta_n = \theta_n^{(t-1)}) = \mathrm{Gamma}(nk + m, \alpha + \sum_{i=1}^{n} \theta_i^{(t-1)})$$

3. Sample the $\theta_i^{(t)}$ in parallel according to

$$\theta_i^{(t)} \sim \mathbb{P}(\theta_i \mid \beta = \beta^{(t)}, X_i = x_i) = \mathrm{Gamma}(x_i + k, \beta^{(t)} + 1)$$