Multistage Gibbs Sampler

The Gibbs Sampler extends easily to higher dimensions. Suppose that for some $p > 1$, the random variable X (often parameters) can be written as $X = (X_1, \ldots, X_p)$ and we can simulate from the corresponding conditional densities $f_1, \ldots, f_p$.

$$X_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p \sim f_i(x_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$$

called the full conditional distributions.

Multistage Gibbs Sampler

The Multistage Gibbs Sampling algorithm is then:
at iteration $t = 1, 2, \ldots$, given $x_{(t)} = (x_1^{(t)}, \ldots, x_p^{(t)})$, generate

1. $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \ldots, x_p^{(t)})$;
2. $X_2^{(t+1)} \sim f_1(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)})$
3. $\vdots$
4. $X_p^{(t+1)} \sim f_1(x_p | x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})$

Note that it is easy to pick out a full conditional distribution from looking at the joint distribution. Pull out all pieces invovlving X_k (called the kernel), and try to match it to a known distribution.

Example:Changepoint Analysis in a Poisson Model

Carlin, Gelfand, and Smith (1992) analyzed data on the yearly counts of coal mining disasters in the United Kingdom from 1851 to 1962. Counts are higher in the earlier years and lower in the later years.
We will model the counts as Poisson random variables, but allow for a different parameter in the later years than in the earlier years.

Example:Changepoint Analysis in a Poisson Model

$$\begin{aligned} y_i|\lambda &\sim Poisson(\lambda) \quad i = 1, \ldots, k \\ y_i|\phi &\sim Poisson(\phi) \quad i = k+1, \ldots, n \end{aligned}$$

The conjugate prior for the Poisson distribution is the gamma distribution.

$$\begin{aligned} \lambda &\sim gamma(a, b) \\ \phi &\sim gamma(c, d) \end{aligned}$$

Example:Changepoint Analysis in a Poisson Model

A noninfomative prior on the discrete changepoint variable k is the discrete uniform distribution.

$$k \sim uniform(1, \ldots, n)$$

Using b and d as rate parameters in the gamma, the joint posterior distribution of λ, ϕ, and k is expressed as

$$\begin{aligned}
\pi(\lambda, \phi, k|y) &\propto f(y|\lambda, \phi, k)\pi(\lambda|a, b)\pi(\phi|c, d)\pi(k) \\
&= \left(\prod_{i=1}^{k} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}\right)\left(\prod_{i=k+1}^{n} \frac{e^{-\phi}\phi^{y_i}}{y_i!}\right)\left(\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}\right)\left(\frac{d^c}{\Gamma(c)}\phi^{c-1}e^{-d\phi}\right)\frac{1}{n} \\
&\propto \lambda^{a-1+\sum_{i=1}^{k} y_i}\phi^{c-1+\sum_{i=k+1}^{n} y_i}e^{-\lambda(k+b)-\phi(n-k+d)}
\end{aligned}$$

Example:Changepoint Analysis in a Poisson Model

We then obtain the following posterior distributions:

$$\lambda|\phi, k, y = \lambda|k, y \sim \qquad gamma(a + \sum_{i=1}^{k} y_i, b + k)$$

$$\phi|\lambda, k, y = \phi|k, y \sim \qquad gamma(c + \sum_{i=k+1}^{n}, d_n - k)$$

$$\pi(k|\lambda, \phi, y) = \frac{\lambda^{a-1+\sum_{i=1}^{k} y_i} \phi^{c-1+\sum_{i=k+1}^{n} y_i} e^{-\lambda(k+b)-\phi(n-k+d)}}{\sum_{j=1}^{n} \lambda^{a-1+\sum_{i=1}^{j} y_i} \phi^{c-1+\sum_{i=j+1}^{n} y_i} e^{-\lambda(j+b)-\phi(n-j+d)}}$$

Example:Changepoint Analysis in a Poisson Model

- The posterior distribution for k is a discrete distribution from which we can draw using the sample command in R.
- We call these *full conditional* distributions when we condition a parameter on all of the other pieces of the model.
- By sampling from each of the full conditional distributions in turn, we obtain a Markov chain that converges to the joint posterior distribution of λ, ϕ, and k. This is the essence of the Gibbs Sampler.

Implementation

For the parameters of the prior distribution we will use $a = 4, b = 1, c = 1$, and $d = 2$. Code to implement this example in R is adapted from Gill.

Missing data and latent variables

In Chapter 5, we saw a few examples of missing data.

- Mixture of normals - component unknown
- Censored exponential data
- Genetic multinomial data

We represented these models in terms of observed data x, and missing data, z.

$$g(x|\theta) = \int_Z f(x, z|\theta) dz$$

Which can also be expressed as

$$g(x|\theta) = \int_Z f(x|z, \theta) h(z|\theta) dz,$$

which puts us in great position for Gibbs Sampling.

Example: Censored Exponential

Consider data that is $exp(rate = \theta)$ (so that the mean is $1/\theta$) and censored at a. m values less than a are observed and $n - m$ values are truncated at a.

$$g(x|\theta) = L(\theta|x) \propto \theta^m \prod_{i=1}^{m} e^{-x_i\theta} = \theta^m e^{-\theta \sum_{i=1}^{m} x_i}$$

and

$$f(x, z|\theta)(\theta|x, z) \propto \theta^m e^{-\theta \sum_{i=1}^{m} x_i} + \theta^{(n-m)} e^{-\theta \sum_{i=m+1}^{n} z_i}$$

If we put a flat prior distribution on θ, $f(\theta) \propto 1$, then we can take draws from the posterior distribution of θ by alternating draws between $f(\theta|x, z)$ and $f(z|x, \theta)$ in a Gibbs sampler.

$$\theta|x, z \sim Gamma(n, m\overline{x} + (n-m)\overline{z})$$

$$z|\theta, x \sim Exp(rate = \theta, a(truncated))$$