MCMC Methods

- In many cases the posterior distribution does not have a simple **recognizable** form, and so we cannot sample from it using built-in `R` functions like "`rgamma`"
- In this case, **Markov chain Monte Carlo** (MCMC) sampling methods are used.
- A **Markov chain** is an ordered, indexed set of random variables (a stochastic process) in which the value of each quantity depends probabilistically **only** on the previous quantity.

MCMC Methods

- Specifically, if $\{\theta^{[0]}, \theta^{[1]}, \theta^{[2]}, \ldots\}$ is a Markov chain, then it has the **Markovian** property:
- For any set $\mathcal{A}$,

$$P\{\theta^{[t]} \in \mathcal{A} | \theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[t-1]}\} = P\{\theta^{[t]} \in \mathcal{A} | \theta^{[t-1]}\}$$

- So $\theta^{[t]}$ is **conditionally independent** of all earlier values **except** the previous one.
- So the values in a Markov chain are not independent, but are "almost independent."

Gibbs Sampling

- The **Gibbs Sampler** is a MCMC algorithm that approximates the **joint distribution** of k random quantities by sampling from each **full conditional** distribution in turn.
- **Example**: We are interested in the distribution of $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. The Gibbs algorithm is:

1. Choose initial values $\boldsymbol{\theta}^{[0]} = (\theta_1^{[0]}, \theta_2^{[0]}, \ldots, \theta_k^{[0]})$.
2. Cycle through each **full** conditional distribution, sampling, for $t = 1, 2, \ldots$

$$\begin{aligned}
\theta_1^{[t]} &\sim \pi(\theta_1 | \theta_2^{[t-1]}, \ldots, \theta_k^{[t-1]}) \\
\theta_2^{[t]} &\sim \pi(\theta_2 | \theta_1^{[t]}, \theta_3^{[t-1]}, \ldots, \theta_k^{[t-1]}) \\
&\vdots \\
\theta_k^{[t]} &\sim \pi(\theta_k | \theta_1^{[t]}, \theta_2^{[t]}, \ldots, \theta_{k-1}^{[t]})
\end{aligned}$$

3. Repeat steps in (2) until convergence.

Gibbs Sampling

- We must be able to sample from each of the full conditional distributions to use the Gibbs Sampler.
- Note that in each step, the **most recent** value of **each** θ_j is conditioned on.
- After many cycles, the sampled values of $(\theta_1, \dots, \theta_k)$ will approximate random draws from the joint distribution of $(\theta_1, \dots, \theta_k)$.
- Then we can summarize, say, a posterior distribution of interest as before.

A Simple Gibbs Example

- **Example 2**: Testing the effectiveness of a seasonal flu shot.
- 20 individuals are given a flu shot at the start of winter.
- At the end of winter, follow up to see whether they contracted flu.
  Let
  $$X_i = \begin{cases} 1 & \text{if shot effective (no flu)} \\ 0 & \text{if ineffective (contracted flu)} \end{cases}$$
- Suppose the 20th individual was unavailable for followup.
- Define $Y = \sum_{i=1}^{19} X_i$.

A Simple Gibbs Example

- If θ is the probability the shot is effective, then

$$p(y|\theta) = \binom{19}{y} \theta^y (1-\theta)^{19-y}$$

- If we had the complete data (for Y **and** X_{20}), then

$$p(\theta|y, x_{20}) = \binom{20}{y + x_{20}} \theta^{y+x_{20}} (1-\theta)^{20-y-x_{20}}$$

- If we put in "temporary" values θ^* and x_{20}^* for the unknown quantities, then

$$\theta|X_{20}^*, Y \sim \text{beta}(Y + X_{20}^* + 1, 20 - Y - X_{20}^* + 1)$$

$$\text{and } X_{20}|Y, \theta^* \sim \text{Bernoulli}(\theta^*)$$

A Simple Gibbs Example

- We can repeatedly sample from these "full conditional" distributions and eventually get a sample from the joint distribution of (θ, X_{20}).
- See `R` example with data.

A More Complicated Gibbs Example (Changepoint)

Example 3: (Coal Mining Disasters)

- Gill gives yearly counts of British coal mine disasters, 1851-1962.
- Relatively large counts in the early era, small counts in the later years.
- **Question**: When did the mean of the process change?
- We model the data using two Poisson distributions:
- "Early" data: $X_1, \ldots, X_k | \lambda \overset{\text{iid}}{\sim} \text{Pois}(\lambda), \quad i = 1, \ldots, k$
- "Later" data: $X_{k+1}, \ldots, X_n | \phi \overset{\text{iid}}{\sim} \text{Pois}(\phi), \quad i = k+1, \ldots, n$
- We must estimate each Poisson mean, λ and ϕ, and **also** the "changepoint" k.

A More Complicated Gibbs Example (Changepoint)

Consider the priors:

$$\lambda \sim \text{gamma}(\alpha, \beta)$$
$$\phi \sim \text{gamma}(\gamma, \delta)$$
$$k \sim \text{discrete uniform on}\{1, 2, \ldots, n\}$$

- If we believe the mean annual disaster count for early years is ≈ 4 and for later years is ≈ 0.5, let $\alpha = 4$, $\beta = 1$, $\gamma = 1$, $\delta = 2$ be the hyperparameters.

A More Complicated Gibbs Example (Changepoint)

Then the posterior is $\pi(\lambda, \phi, k|\mathbf{x})$

$$\propto L(\lambda, \phi, k|\mathbf{x}) p(\lambda) p(\phi) p(k)$$

$$= \left[\prod_{i=1}^{k} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] \left[\prod_{i=k+1}^{n} \frac{e^{-\phi}\phi^{x_i}}{x_i!}\right] \left[\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\right] \left[\frac{\delta^\gamma}{\Gamma(\gamma)}\phi^{\gamma-1}e^{-\delta\phi}\right] \left[\frac{1}{n}\right]$$

$$\propto e^{-k\lambda}\lambda^{\sum_{i=1}^{k} x_i} e^{-(n-k)\phi}\phi^{\sum_{k+1}^{n} x_i}\lambda^{\alpha-1}e^{-\beta\lambda}\phi^{\gamma-1}e^{-\delta\phi}$$

$$= \lambda^{\alpha+\sum_{i=1}^{k} x_i - 1} e^{-(\beta+k)\lambda}\phi^{\gamma+\sum_{k+1}^{n} x_i - 1} e^{-(\delta+n-k)\phi}$$

So full conditionals are:

$$\lambda|\phi, k \sim \text{gamma}(\alpha + \sum_{i=1}^{k} x_i, \beta + k)$$

$$\phi|\lambda, k \sim \text{gamma}(\gamma + \sum_{i=k+1}^{n} x_i, \delta + n - k)$$

A More Complicated Gibbs Example (Changepoint)

To get the full conditional for k, note the joint density of the data is:

$$
\begin{aligned}
p(\mathbf{x}|k,\lambda,\phi) &= \left[\prod_{i=1}^{k} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right]\left[\prod_{i=k+1}^{n} \frac{e^{-\phi}\phi^{x_i}}{x_i!}\right] \\
&= \left[\prod_{i=1}^{n} \frac{1}{x_i!}\right] e^{k(\phi-\lambda)} e^{-n\phi} \lambda^{\sum\limits_{i=1}^{k} x_i} \left[\prod_{i=k+1}^{n} \phi^{x_i}\right]\left[\frac{\prod\limits_{i=1}^{k} \phi^{x_i}}{\phi^{\sum_{i=1}^{k} x_i}}\right] \\
&= \left[\prod_{i=1}^{n} \frac{e^{-\phi}\phi^{x_i}}{x_i!}\right]\left[e^{k(\phi-\lambda)}\left(\frac{\lambda}{\phi}\right)^{\sum\limits_{i=1}^{k} x_i}\right] \\
&= f(\mathbf{x},\phi)g(\mathbf{x}|k)
\end{aligned}
$$

A More Complicated Gibbs Example (Changepoint)

By Bayes' Law, for any particular value k^* of k,

$$p(k^*|\mathbf{x}) = \frac{f(\mathbf{x},\phi)g(\mathbf{x}|k^*)p(k^*)}{\sum_{k=1}^{n} f(\mathbf{x},\phi)g(\mathbf{x}|k)p(k)}$$

Since $p(k) = 1/n$ (constant), we have

$$p(k^*|\mathbf{x}) = p(k^*|\mathbf{x},\lambda,\phi) \propto \frac{g(\mathbf{x}|k^*)}{\sum_{k=1}^{n} g(\mathbf{x}|k)}$$

(full conditional for k)

- This ratio defines a probability vector for k that we use at each iteration to sample a value of k from $\{1, 2, \ldots, n\}$.
- see R example (Coal mining data)