Gibbs sampling

Setup: Suppose we have a set of parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_d$

and some data $\mathcal{D}$.

Goal: Generate samples from the posterior distribution $p(\theta \mid \mathcal{D})$

Gibbs sampling:

1.) Pick some initial $\theta^{(i)} = (\theta_1^{(i)}, \ldots, \theta_d^{(i)})$.

2.)
$$\theta_1^{(i+1)} \sim \underbrace{p(\theta_1 \mid \theta_2^{(i)}, \theta_3^{(i)}, \ldots, \theta_d^{(i)}, \mathcal{D})}_{\text{conditional posterior of } \theta_1}$$

$$\theta_2^{(i+1)} \sim p(\theta_2 \mid \theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_d^{(i)}, \mathcal{D})$$

$$\theta_3^{(i+1)} \sim p(\theta_3 \mid \theta_1^{(i+1)}, \theta_2^{(i+1)}, \theta_4^{(i)}, \ldots, \theta_d^{(i)}, \mathcal{D})$$

$$\vdots$$

$$\theta_d^{(i+1)} \sim p(\theta_d \mid \theta_1^{(i+1)}, \theta_2^{(i+1)}, \ldots, \theta_{d-1}^{(i+1)}, \mathcal{D})$$

i.e. sample n values for a g parameter from its conditiona posterior, while hol the othe variable constan

3.) Increment $i = i+1$, and repeat N times to generate N sampl

Pros: Does not require tuning any parameters (e.g. most MCMC samplers require choosing a proposal / tuning free parameters e.g. step-size)

Cons: Assumes knowledge of the conditional posterior densiti which may be hard to derive in practice.

Example : Bayesian linear regression.

$$y_i \overset{i.i.d.}{\sim} \mathcal{N}(y_i \mid w_0 + w_1 x, \gamma^{-1}) \iff y_i = w_0 + w_1 x_i + \varepsilon, \ \varepsilon \sim \mathcal{N}($$

Likelihood : $\mathcal{L}(y_1, y_2, \ldots, y_n \mid x_1, x_2, \ldots, x_n, w_0, w_1, \gamma) = \prod_{i=1}^{n} \mathcal{N}(y_i \mid w_0 + w_1$

Priors : $w = (w_0, w_1) \sim \mathcal{N}\left(\begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}, \begin{bmatrix} \lambda_0^{-1} & 0 \\ 0 & \lambda_1^{-1} \end{bmatrix}\right)$

$\gamma \sim \text{Gamma}(\gamma \mid \alpha, \beta)$

Goal : Design a Gibbs sampler for $\theta \sim p(\theta \mid \mathcal{D})$, $\theta := \{w_0, w_1, \gamma\}$

General approach :

(i) Write down the posterior conditional density in log-form.

ii) Throw away all terms that do not depend on the current sampling variable.

iii) Pretend that this is the pdf for the current sampling variable, while all other variables are kept fixed.

- Gibbs update for w_0 :

$$p(w_0 \mid w_1, \gamma, \mathcal{D}) \propto \underbrace{p(\mathcal{D} \mid w_0, w_1, \gamma)}_{\text{likelihood}} \underbrace{p(w_0)}_{\mathcal{N}(\mu_0, \lambda_0^{-1})}$$

$$\Rightarrow \log p(w_0 \mid w_1, \gamma, \mathcal{D}) \propto -\frac{\gamma}{2} \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2 - \frac{\lambda_0}{2} (w_0 - \mu$$

$$\propto \gamma \sum_{i=1}^{n} (y_i - w_1 x_i) w_0 - \frac{\lambda_0}{2} w_0^2 + \lambda_0 \mu_0 w_0 - \frac{\gamma}{2} n w_0$$

$$\Longrightarrow \quad p(w_0 \mid w_1, \gamma, \mathcal{D}) \sim \mathcal{N}\left(w_0 \,\middle|\, \frac{\lambda_0 \mu_0 + \gamma \sum_{i=1}^{n} (y_i - w_1 x_i)}{\lambda_0 + n \cdot \gamma}, (\lambda_0 + n.\right.$$

- Gibbs update for w_1 (same as above)

- Gibbs update for γ:

Recall $\quad \mathrm{Gamma}(\gamma \mid \alpha, b) \propto b^{\alpha} x^{\alpha - 1} e^{-b\gamma}$

- $\log \mathrm{Gamma}(\gamma \mid \alpha, b) \propto (\alpha - 1) \log \gamma - b\gamma$

$$p(\gamma \mid w_0, w_1, \mathcal{D}) \propto p(\mathcal{D} \mid w_0, w_1, \gamma)\, p(\gamma)$$

$\Longrightarrow \quad \log p(\gamma \mid w_0, w_1, \mathcal{D}) \propto \quad \ldots\ldots$ (keep terms that only depend on γ

$\ldots \quad \Longrightarrow \quad p(\gamma \mid w_0, w_1, \mathcal{D}) \sim \mathrm{Gamma}(\ ? \ , \ ? \ )$