Advanced Markov Chain Monte Carlo

Yee Whye Teh

Gatsby Computational Neuroscience Unit
University College London

April 3, 2009

Set-up

- Models: m belong to a class of models $\mathcal{M}$.
- Data: $X = \{x_1, x_2, \ldots, x_n\}$.
- Latent variables: $Y = \{y_1, y_2, \ldots, y_n\}$.
- Parameters: θ_m for model $m \in \mathcal{M}$.
- Likelihood:

$$p(X|Y, \theta_m) = \prod_{i=1}^{n} p(x_i|y_i, \theta_m)$$
$$p(Y|\theta_m) = \prod_{i=1}^{n} p(y_i|\theta_m)$$

- Prior:

$$p(m)$$
$$p(\theta_m|m)$$

Bayesian Inference

- Posterior distribution:

$$\pi(\theta_m, Y|m) = p(\theta_m, Y|m, X) = \frac{p(\theta_m|m)p(Y|\theta_m)p(X|Y, \theta_m)}{p(X|m)}$$

$$\pi(m) = p(m|X)$$

- Evidence:

$$p(X|m)$$

- Bayes Factors:

$$\frac{p(X|m)}{p(X|m')}$$

Metropolis-Hastings

- Metropolis-Hastings is a standard and flexible Markov chain Monte Carlo transition kernel.
- Distribution we wish to sample from:

$$\pi(\theta)$$

- Proposal kernel:

$$q(\theta \to \tilde{\theta})$$

- Acceptance probability:

$$\alpha(\theta \to \tilde{\theta}) = \min\left(1, \frac{\pi(\tilde{\theta})q(\tilde{\theta} \to \theta)}{\pi(\theta)q(\theta \to \tilde{\theta})}\right)$$

Reversible-Jump Markov Chain Monte Carlo

- To do Bayesian model averaging, we are interested in computing

$$\pi(m, \theta_m)$$

  where for different m's θ_m's come from different spaces with different dimensionalities.
- Most MCMC techniques can only deal with sampling from distributions over a single space, e.g. the posterior distribution over parameters of a *single* model.
- Reversible-Jump MCMC allows us to sample from the full posterior $\pi(m, \theta_m)$ over both models and parameters properly.

Simple Example

- Imagine mixtures of either 1 or 2 Gaussians in 1D:

$$
\begin{aligned}
m = 1: &\qquad \mathcal{N}(\theta_0, \sigma^2) \\
m = 2: &\qquad .5\mathcal{N}(\theta_1, \sigma^2) + .5\mathcal{N}(\theta_2, \sigma^2)
\end{aligned}
$$

  where σ^2 is known and parameters consist of the mean θ_0 for model 1 and the means (θ_1, θ_2) for model 2.
- Conditioned on m we can easily sample θ_0 or (θ_1, θ_2) depending on value of m.
- How do we sample m? The θ spaces are different for different m's, so we have to change the θ's if we propose changing m.

Simple Example

- Say $m = 1$ and we propose changing $m = 1 \rightarrow m = 2$. We would also need to propose (θ_1, θ_2) as well. Since this proposal is a "split" of one Gaussian into two, a reasonable proposal would be:

$$\delta \sim \mathcal{N}(0, v)$$
$$\theta_1 \leftarrow \theta_0 + \delta$$
$$\theta_2 \leftarrow \theta_0 - \delta$$

  This preserves the overall mean of the distribution over the data.
- To compute the MH acceptance probability we need the reverse proposal probabilities for $m = 2 \rightarrow m = 1$. Given a particular (θ_1, θ_2), notice that *the only* θ_0 *that could have be used to propose* (θ_1, θ_2) *is their average*:

$$\theta_0 = (\theta_1 + \theta_2)/2$$

Simple Example

- Recap of the proposal and the reverse proposal:

$$\begin{aligned} m = 1 &\to m = 2 & m = 2 &\to m = 1 \\ \theta_0 &\to (\theta_1, \theta_2) = (\theta_0 + \delta, \theta_0 - \delta) & (\theta_1, \theta_2) &\to \theta_0 = (\theta_1 + \theta_2)/2 \\ \delta &\sim \mathcal{N}(0, v) & & \end{aligned}$$

- The acceptance probability is:

$$\min\left(1, \frac{\pi(m = 2, \theta_0 + \delta, \theta_0 - \delta)1}{\pi(1, \theta_0)\mathcal{N}(\delta; 0, v)}\right)$$

Notice that the dimensions match up.

Another Example

- If the $\theta_0 \to (\theta_1, \theta_2)$ proposal is simply:

$$\theta_1 = \theta_0$$
$$\theta_2 \sim \mathcal{N}(0, v)$$

  i.e. the first Gaussian stays the same, create a second Gaussian.
- Then the reverse proposal $(\theta_1, \theta_2) \to \theta_0$ would have needed to be:

$$\theta_0 = \theta_1$$

  i.e. drop the second Gaussian and keep the first one.

General Case

- We want a proposal $q(m \to w)$ where $m, w \in \mathcal{M}$.
- We need to design proposals for the parameters $q(\theta_m \to \theta_w)$ and reverse proposals $q(\theta_w \to \theta_m)$.
- The acceptance probability of RJMCMC,

$$\min\left(1, \frac{\pi(w)\pi(\theta_w|w)q(w \to m)q(\theta_w \to \theta_m)}{\pi(m)\pi(\theta_m|m)q(m \to w)q(\theta_m \to \theta_w)}\right)$$

only makes sense if the θ densities are both wrt the same underlying measure. Simple case: measure lies on a 2D plane in $(\theta_0, \theta_1, \theta_2)$ space satisfying $\theta_0 = (\theta_1 + \theta_2)/2$.