STA250
Monday 10/21/13

Metropolis-within-Gibbs

Allows you to sample from high-dimensional distributions using a sequence of lower dimensional distributions. Generalizes Gibbs sampler in two ways:

1. If we can't sample directly/exactly then can use MH.
2. Sample sub-blocks of parameters, not necessarily full conditionals, i.e.
   $p(\theta_1, \theta_2, \phi|\mathbf{y})$
   $\rightarrow$ sample $p(\theta_1, \theta_2|\phi, \mathbf{y})$ exactly
   $\rightarrow$ sample $p(\phi|\theta_1, \theta_2, \mathbf{y})$ using MH.

Idea for hw
2 strategies:

1. Sample from $p(\boldsymbol{\beta}|\mathbf{y})$ using MH & a multivariate proposal for $\boldsymbol{\beta}$.
2. For $j = 1, ..., p$, sample β_j from $p(\beta_j|\beta_{[-j]}, \mathbf{y})$ [no closed form $\Rightarrow$ use MH].

When to use non-symmetric proposals? θ has compact support, then proposal can respect the boundaries.

Example $X_i \stackrel{\text{iid}}{\sim} \text{Pois}(\lambda)$, $i = 1, ..., n$, prior $\lambda \sim t_\nu \mathbb{1}\{(0, \infty)\}$
Use MH, proposal $\theta^* \sim \text{TN}(\theta^{(t)}, v^2, [0, \infty)) \Rightarrow$ no longer symmetric.

Checking your MCMC code
For one dataset, how to know if we reach convergence?
$\rightarrow$ "by eye" using traceplots
$\rightarrow$ use effective sample size to gauge roughly "how well" converged
$\rightarrow$ other diagnostics: Gelman-Rubin (multiple chains), Heidelberger, Geveke(?)

We can use simulation studies to check everything is working.
Idea:

1. Simulate $\theta_{(j)}$ from the prior $p(\theta)$
2. Simulate a dataset $\mathbf{y}_{(j)}$ from the model $p(y|\theta_{(j)})$
3. Sample from posterior for dataset $\mathbf{y}_{(j)}$
4. Find 100(1-α)% central credible interval for θ from dataset $\mathbf{y}_{(j)}$
5. Record Y/N whether the interval contained $\theta_{(j)}$
6. Check roughly 100(1-α)% of intervals contained their specific $\theta_{(j)}$

Why does this work?

$$\begin{aligned}
\int p(\theta) \int & \mathbb{1}\{\theta \in S^{1-\alpha}(\mathbf{y})\} p(\mathbf{y}|\theta) dy d\theta \\
&= \int \int \mathbb{1}\{\theta \in S^{1-\alpha}(\mathbf{y})\} p(\theta) p(\mathbf{y}|\theta) dy d\theta \\
&= \int \int \mathbb{1}\{\theta \in S^{1-\alpha}(\mathbf{y})\} p(\theta|\mathbf{y}) p(\mathbf{y}) d\theta dy \\
&= \int p(\mathbf{y}) \left[\int \mathbb{1}\{\theta \in S^{1-\alpha}(\mathbf{y})\} p(\theta|\mathbf{y}) d\theta \right] dy \\
&= 1 - \alpha
\end{aligned}$$

Bayes: $\int p(\theta) \int \mathbb{1}\{\theta \in S^{1-\alpha}(y)\} p(y|\theta) dy d\theta = 1 - \alpha$
Frequentist: $\int \mathbb{1}\{\theta \in C^{1-\alpha}(y)\} p(y|\theta) dy = 1 - \alpha \quad \forall\, \theta$

Posterior Predictive Checking

Validation simulation checks you can sample from the posterior under your model. It doesn't tell you if your model is a good fit to the data.
Idea: Having fit your model to the data, you know roughly what θ is, so if you simulate from $p(y|\theta)$ the simulated data should look "similar" to the real data.
Formally: $p(\tilde{\mathbf{y}}|\mathbf{y}) = \int p(\tilde{\mathbf{y}}, \theta|\mathbf{y}) d\theta = \int p(\tilde{\mathbf{y}}|\theta) p(\theta|\mathbf{y}) d\theta$

Recipe:

- Sample $\theta^{(t)}$ from the posterior $p(\theta|\mathbf{y})$
- For each $\theta^{(t)}$, sample a new dataset from $p(y|\theta^{(t)})$
- We now have M predictive datasets & one real dataset
- We can take univariate summary statistic of each dataset & the real dataset & compare

Could use min, mean, median, max.

Posterior Predictive p-value

$2\min\{\mathbb{P}(T(y^*) > T(\mathbf{y})), \mathbb{P}(T(y^*) < T(\mathbf{y}))\}$

In practice, $p = 2\min(\text{fraction to left, fraction to right})$. Heuristically, p small ($< .05$) $\Rightarrow$ model is NOT a good fit to the data.