Lecture 15

- Assignment 3 well done. 2nd lowest score 44/50. Solutions posted.
  Also posted musbayes.do musbayes2.do
  Soon post assignment 4 due Tuesday Bayesian.

- Maximum simulated likelihood (MSL)
  Idea is simple. Implementation is tricky.
Density \( f(y|x, \beta) = \int g(y|x, \beta, \alpha) h(\alpha) \, d\alpha \)

\[
\ln L = \sum_{i=1}^{n} \ln f(y_i|x_i, \beta) \quad \text{gives independence over } i
\]

Use \( \ln L = \frac{1}{s} \sum_{i=1}^{s} \ln g(y_i|x_i, \beta, \alpha^{(s)}) \)

where \( \alpha^{(s)} \) are draws from \( h(\alpha) \)

Problems \( \hat{f}(y_i|x_i, \beta, \alpha) = \frac{1}{s} \sum_{s} g(y_i|x_i, \beta, \alpha^{(s)}) \)

is unbiased for \( f(y_i|x_i, \beta) \)

But \( \ln \hat{f} \) is not unbiased for \( \ln f \)

so need \( s \to \infty \) need to avoid chatter; need
Need the simulator used (to estimate $f()$) to be continuous in $\beta$.

Often used in latent variable models - multinomial probit - generalizations

Use the GHK simulator.

State uses this in asmprobit for multinomial probit.
Bayesian methods

Recall \[ p(\theta | y, X) = \frac{L(y | \theta, X) \cdot \pi(\theta)}{f(y | X)} \]

\[ C = \int L(y | \theta, X) \pi(\theta) d\theta \]

Use MCMC methods.

- Make sequential random draws \( \theta^{(1)}, \theta^{(2)} \), ...
  where \( \theta^{(s)} \) depends on \( \theta^{(s-1)} \) but not on \( \theta^{(s-2)} \) once we condition on \( \theta^{(s-1)} \). So a Markov chain.
- When chain converges, \( \theta^{(s)} \) are correlated draws from \( p(\theta | y, X) \).
Metropolis Algorithm

At the $s^{th}$ round

- draw $\Theta^a$ from candidate distribution $g(.)$
  where $g(.)$ is symmetric i.e. $g(\Theta^* | \Theta^{(s-1)})$ is symmetric
  so $g(\Theta^a | \Theta^b) = g(\Theta^b | \Theta^a)$

- set $\Theta^{(s)} = \Theta^a$ if $u < \frac{p(\Theta^*)}{p(\Theta^{(s-1)})}$ where $u \sim \text{Uniform}(0,1)$
  i.e. if $u < \frac{L(y | \Theta^a, x) \pi(\Theta^a)}{L(y | \Theta^{(s-1)}, x) \pi(\Theta^{(s-1)})}$

Take logs. Then easy to implement.
We need a symmetric distribution.

Random walk Metropolis use

$$\theta^* \sim \mathcal{N}(\theta^{(s-1)}, V) \quad \exp \left( -\frac{1}{2} (\theta^2 - \theta^{(s-1)})' V^{-1} (\theta^2 - \theta^{(s-1)}) \right)$$

$V$ does not depend on $\theta^{(s-1)}$.

In practice $V$ may be updated every so often until chain converges.
Metropolis-Hastings Algorithm

- Generalizes to allow the candidate distribution to be asymmetric in $\Theta^*$ and $\Theta^{(s-1)}$.

- The acceptance rule is then

$$u \leq \frac{p(\Theta^*) \times g(\Theta^* | \Theta^{(s-1)})}{p(\Theta^{(s-1)}) \times g(\Theta^{(s-1)} | \Theta^*)}$$

- For explanation see CT (2005).

- Independent chain MH uses $g(\Theta^* | \Theta^{(s-1)})$ does not depend on $\Theta^{(s-1)}$ where $g(\cdot)$ is a good approx. to $p(\cdot)$

  e.g. Do MCMC for $p(\Theta)$ then $g(\Theta)$ is multivariate $\mathcal{T}(\hat{\Theta}, \tilde{V}(\hat{\Theta}))$. 
Gibbs Sampler

- Partition \( p(\theta) = p(\theta_1, \theta_2) \)

Make alternating draws from \( p(\theta_1|\theta_2) \) and \( p(\theta_2|\theta_1) \), assuming \( p(\theta_1,\theta_2) \) & \( p(\theta_2|\theta_1) \) are known.

- Special case of MHT
  Usually quicker than the usual MHT.
  If need MHT to draw from \( p(\theta_1|\theta_2) \) or \( p(\theta_2|\theta_1) \)
  called MHT within Gibbs.

- Extend to eg. \( p(\theta_1, \theta_2, \theta_3) \)
  Need to know \( p(\theta_1|\theta_2, \theta_3), p(\theta_2|\theta_1, \theta_3), p(\theta_3|\theta_1, \theta_2) \).
Bayes2-clean.do does Metropolis random walk for a probit example.
I went through the code.
It also does data augmentation and Gibbs sampler.
Next lecture.
Specification of prior

- As $N \to \infty$ data dominates the prior
  \[ p(\theta | y) \sim N(\hat{\theta}_m, \sigma^2 I(\hat{\theta}_m)^{-1}) \]

- Noninformative and possibly improper prior is one that
  - has little effect on posterior
  - uniform prior (all values equally likely) is obvious choice
  improper usually causes no problem
  - not invariant to transformation $\theta \to e^\theta$
Jeffreys prior set \( p(\theta) \propto \det | I(\theta) |^{-1/2} \)

where \( I(\theta) = \frac{\partial^2 \ln}{\partial \theta \partial \theta} \), invariant to transformation.

For linear regression under normality this is a uniform prior on \( \beta \).

- Proper priors (informative or uninformative).
- Becomes uninformative as variance gets big.
- Use conjugate prior if possible.
- Hierarchical priors are used a lot.