Spring 1997 John Rust
Economics 551b 37 Hillhouse, Rm. 27

PROBLEM SET 2

Bayesian Basics (II)
(Due: February 10, 1997)

QUESTION 1 Derive the convergence of posterior distribution when support of prior does not include true parameter θ^*. Show that

$$p(\theta|x_1, \cdots, x_N) \to \delta_{\overline{\theta}} \qquad \text{with probability} \quad 1$$

where $\delta_{\overline{\theta}}$ is a point mass at $\overline{\theta}$ and $\overline{\theta}$ is defined by:

$$\overline{\theta} = \arg\min_{\theta \in \text{support}(p)} E_{\theta^*} \left\{ -\log \frac{f(x|\theta)}{f(x|\theta^*)} \right\}.$$

QUESTION 2 Derive the convergence of posterior distribution when there exists θ^{**} observationally equivalent to θ^*, i.e. where $f(x|\theta^{**}) = f(x|\theta^*)$ for almost all x (i.e. except for x in a set of Lebesgue measure zero).

QUESTION 3 Do extra credit problem on an example of computing an invariant density for a Markov process. This is *optional* but *recommended.*

QUESTION 4 This question asks you to employ the Gibbs-sampling algorithm in a simple example taken from the book *Bayesian Data Analysis* by Gelman *et. al.* Here the data consist of a single observation (y_1, y_2) from a bivariate normal distribution with unknown mean (θ_1, θ_2) and known covariance matrix Ω given by

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

With a uniform prior distribution over θ the posterior distribution of θ is bivariate normal with mean $(y_1, y_2)'$ and covariance matrix Ω. Although it is trivial to sample directly from a bivariate normal distribution of (θ_1, θ_2), the purpose of this question is to have you use the Gibbs sampler to draw from the posterior and compare how well the random samples from the Gibbs sampler approximate draws from the bivariate normal posterior.

1. If $(y_1, y_2) = (2, 1)$ and $\rho = -.3$, use the normal random number generator in Gauss (or some other language you are comfortable with) to draw a random sample of 500 observations from the posterior "directly". Report the sample mean and covariance matrix from this random sample. (**Hint:** you can draw from a bivariate normal with covariance matrix Ω by drawing a bivariate normal with covariance matrix I from Gauss's random normal generator (i.e. two independent univariate normal draws $\tilde{z} = (\tilde{z}_1, \tilde{z}_2)'$) and then compute the Cholesky decomposition of Ω, $\sqrt{\Omega}$ and multiply this "square root" matrix by $\tilde{z}$ to get a bivariate random draw with mean $(0, 0)'$ and covariance matrix Ω).

2. Now use the Gibbs sampler to generate a random sample from the posterior. Starting from 500 points randomly, uniformly drawn over the square of width and length 1 with center $(2, 1)$ run the Gibbs sampler for 50 iterations. The Gibbs sampler is constructed by generating a draw from the normal conditional density of θ_1 given θ_2 given by:

$$\theta_1|\theta_2, y \sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$

Given the realized value of θ_1 draw a value of θ_2 from the conditional density of θ_2 given θ_1 given by:

$$\theta_2|\theta_1, y \sim N(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2).$$

Starting from 500 different randomly drawn initial values of (θ_1^0, θ_2^0) perform $T = 50$ loops of the above Gibbs sampling algorithm and save the 500 final draws of (θ_1^T, θ_2^T) for each initial condition. Use these random draws to compute the sample mean and covariance matrix of the posterior. How well does it compare to the true mean and covariance matrix of the posterior? Is $T = 50$ a sufficient number of iterations for the Gibbs sampler to converge to the invariant density?

3. Repeat part 2, but now using $T = 1000$ draws of the Gibbs sampler for each of the 500 randomly drawn initial conditions (save the initial conditions in part 2 in a file so you can use the same initial conditions for comparison of the the results in parts 2 and parts 3). Do you think $T = 1000$ iterations is sufficient for the Gibbs sampler to have converged to the invariant density, i.e. the bivariate normal posterior distribution for θ?

4. Instead of using 500 different randomly chosen initial conditions, run the Gibbs sampler only once from one initial condition $(\theta_1^0, \theta_2^0) = (0, 0)$ for $T = 2500$ iterations. Discard the first 2000 values of θ and save the last 500 values of θ^t, $t = 2001, \ldots, 2500$. Compute the sample mean and covariance matrix of these θ^t's. How well do they approximate the true mean and covariance matrix of the posterior density? (Here we are illustrating the principle of *ergodicity* i.e. if the Gibbs sampler has converged to its invariant density, then time-series averages of the random draws of the Markov process converge to their expectations, i.e. their "long run" expectations with regard to the invariant density. Thus, even thought the draws of θ^t are not independent for successive values of t, sample averages of these values should be close to the corresponding expectations when T is large, i.e.

$$\frac{1}{T}\sum_{t=1}^{T} h(\theta^t) \to \int h(\theta)p(\theta), \quad \text{as } T \to \infty$$

where h is some function that has finite expectation with respect to the invariant density p of the Markov process $\{\theta^t\}$ generated by the Gibbs sampling algorithm.)