Econ. 556b Spring 1997 C. Sims

Take-Home Final Exam

The exam is due at 2PM Tuesday, 5/6/97.

1. We know that in the standard univariate Gaussian AR model

$$y(t) = \rho y(t-1) + \varepsilon(t)\,, \tag{1}$$

with ε i.i.d. as $N(0,\sigma^2)$ and a prior flat on (ρ,σ) or flat on $(\rho,\log\sigma)$, use of the unconditional likelihood leads to a posterior density that drops to zero as $\rho \to 1$ from below. Suppose that we use a different parameterization, with ε distributed as $N\left(0,\nu^2 \cdot \left(1-\rho^2\right)\right)$ and the prior flat on $(\rho,\log\nu)$. Is it still true that the posterior based on the unconditional likelihood has density dropping to zero as $\rho \to 1$? Consider both the joint posterior, with ν held fixed as $\rho \to 1$, and the marginal posterior for ρ. Is the posterior p.d.f. based on the conditional likelihood Gaussian for ρ under this parameterization?

2. Consider an identified VAR model, parameterized as

$$A_0(\theta)y(t) = \sum_{s=1}^{k} A_s y(t-s) + \varepsilon(t) \tag{2}$$

with ε i.i.d. $N(0,I)$. We assume that the elements of the A_s matrices are unrestricted for $s>0$, but that $A_0(\theta)$ depends on the parameter vector θ that is of lower dimension than $\left(n^2-n\right)/2$. We suppose there is a Gaussian prior p.d.f., conditional on A_0, for the A_s matrices and that inference is to be based on the p.d.f. for $\{y(t)\}_{t=1}^{T}$ conditional on the intitial conditions $y(0),\dots,y(-k+1)$. There is also a prior, possibly non-Gaussian, on θ. There is a redundancy in the parameterization because of the fact that, for any row of $[A_0\ A_s]$, we can multiply by -1, changing the signs of all the coefficients in the equation, and the likelihood value will be unaffected. We have to avoid this redundancy by restricting the parameter space. One standard method would be to consider only A's such that the diagonal of A_0 is all positive.

Discuss the difficulties such a restriction raises for a Metropolis sampling scheme, because of the requirement that jump distributions be symmetric. Distinguish the case where A_0 depends on θ linearly (as when θ is simply the nonzero elements of A and identifying restrictions are all zero-restrictions) from the case where the dependence of A on θ is nonlinear. Suggest approaches that would work for Markov chain Monte Carlo sampling from the posterior in both cases.

3. In a 2 by 2 identified VAR model with one lag

$$A_0 y(t) = A_1 y(t-1) + \varepsilon(t), \tag{3}$$

where ε is i.i.d. $N(0, I)$. We have a prior distribution based on data observed through time t=5 on the vector of 8 elements of A_0 and A_1, stacked by columns, with A_0 on top, that is given by

$$N\left(\begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, .25I\right). \tag{4}$$

We observe $y(6) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and we had already observed $y(5) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ (before we formed (4)). Find the posterior p.d.f. for the elements of A_0 and A_1.

4. a) Is the posterior variance of a value of the state vector based on full-sample Kalman smoothing always lower than the variance of the same value of the state vector based on the Kalman filter? Explain why or why not.

b) Is ordinary rejection sampling, based on an approximating p.d.f. that dominates the true p.d.f., a special case of Metropolis-Hastings? Explain why or why not.

c) Suppose we want to predict $X(T+1)^2$ and we have available observations on $X(1), \ldots, X(T)$, which are i.i.d. with the same distribution as $X(T)$. Consider the following two strategies: i) predict $X(T+1)^2$ as the sample mean of the squared values of $X(1), \ldots, X(T)$; ii) predict $X(T+1)^2$ as twice the square of the sample mean of the levels of $X(1), \ldots, X(T)$ themselves. Are there assumptions under which each of these would be an optimal way to form predictions? What can you say about the range of assumptions on the distribution of X under which each approach will give good results?