Statistics 322, Statistical Computing, Home Work 7

Posted: Oct. 10, 1996 Due: Oct. 24, 1996

The goal of this is a simulation investigation of the sample properties of the Gibbs sampling examples studied in the script file `s322eg9.m`:

J1. 'Good Circulation' joint distribution.
J2. 'Sticky States' joint distribution.
J3. 'Jump Away Fast' joint distribution.
J4. 'Nearly Cyclic' joint distribution.

1. Write a MATLAB script, which makes a plot (one for each joint distribution, but put all four on the same page) of a single realization of the "sample paths" $\left\{X_1^{(n)} : n = 1, ..., 100\right\}$ and $\left\{X_2^{(n)} : n = 1, ..., 100\right\}$ as a function of n. Comment on the results, and how they compare to the intuitive ideas discussed in class.

2. Extend your MATLAB script, to study simulated versions of the sampling properties of the statistics:
   S1. $\overline{X}_1 = \frac{1}{n} \Sigma_{i=m}^{n+m} X_1^{(i)}$
   S2. $\hat{cov}\left(X_1, X_2\right) = \frac{1}{n} \Sigma_{i=m}^{n+m} X_1^{(i)} X_2^{(i)} - \left(\frac{1}{n} \Sigma_{i=m}^{n+m} X_1^{(i)}\right)\left(\frac{1}{n} \Sigma_{i=m}^{n+m} X_2^{(i)}\right)$,

   for $m = 40$ and $n = 400$. Calculate the values of S1 and S2 for 200 pseudo data sets, and print out a table that shows:
   a. the sample standard deviations, for easy comparisons across distributions.
   b. the fact that the bias (as estimates of the corresponding theoretical versions) is negligible (compared to the standard deviation) in each case.

   Discuss how the results appear to be driven by:
   i. the sample path characteristics observed in part 1.
   ii. the convergence properties observed in `s322eg9.m`.
   iii. the eigenvalues of the $X_1 \to X_1$ transition matrices given in `s322eg9.m`.
   iv. the eigenvalues of the $X_1 \to X_2$ and $X_2 \to X_1$ transition matrices given in `s322eg9.m`.

Notes:

1. Copy code from `s322eg9.m` to cut down on the amount of needed rewriting.
2. It would be good to account for the Monte Carlo variability, but the time series methods needed to handle the dependence would need too much additional effort, so we will skip that this time.
3. The Gibbs sampling run requires a loop for each step, but the simulated data sets can be done simultaneously, by updating a vector of values (for computational speed).