Homework 1 Solutions: Bayesian Inference Module

STA 250 Fall 2013, Prof. Baines (10/26/13)

Q1 Under the assumption that the resulting Markov Chain is ergodic (i.e., aperiodic and positive recurrent) with continuous state space, we know that the chain converges to a stationary distribution π satisfying:

$$\pi(y) = \int \pi(x)\mathcal{P}(x,y)dx \tag{1}$$

Therefore, we must show that for $\mathcal{P}$ corresponding to a Gibbs sampler, the target density π satisfies equation (1). If we can show this, then we have shown that the target density is the stationary distibution for the Gibbs sampler. Note that the stationary distribution can be seen to be unique under certain regularity conditions. For a two-dimensional Gibbs sampler, the states are $x = (x_1, x_2)$ and $y = (y_1, y_2)$ with the transition density given by:

$$\mathcal{P}(x,y) = \pi_{1|2}(y_1|x_2)\pi_{2|1}(y_2|y_1),$$

where $\pi_{1|2}$ and $\pi_{2|1}$ are the conditional distributions defined by the joint target density π. For brevity we drop the subscripts on π as the conditional densities are clear from the context. Therefore, we see that:

$$\begin{aligned}
\int \pi(x)\mathcal{P}(x,y)dx &= \int \pi(x_1,x_2)\pi(y_1|x_2)\pi(y_2|y_1)dx_1dx_2 \\
&= \int \pi(x_2)\pi(y_1|x_2)\pi(y_2|y_1)dx_2 \\
&= \int \pi(y_1,x_2)\pi(y_2|y_1)dx_2 \\
&= \pi(y_1)\pi(y_2|y_1) = \pi(y_1,y_2).
\end{aligned}$$

Therefore the target density π is seen to be the stationary distribution of the chain.

For a $p-$dimensional (sequential) Gibbs sampler we just extend the result slightly. The transition density becomes:

$$\mathcal{P}(x,y) = \prod_{j=1}^{p} \pi(y_j|y_{[1:j-1]}, x_{[(j+1):p]}),$$

where $y_{[1:0]} := \varnothing$. Therefore, we see that:

$$\begin{aligned}
\int \pi(x)\mathcal{P}(x,y)dx &= \int \pi(x_1,\ldots,x_p)\prod_{j=1}^{p}\pi(y_j|y_{[1:j-1]},x_{[(j+1):p]})dx_1\cdots dx_p \\
&= \int \pi(x_2,\ldots,x_p)\prod_{j=1}^{p}\pi(y_j|y_{[1:j-1]},x_{[(j+1):p]})dx_2\cdots dx_p \\
&= \int \pi(y_1,x_2,\ldots,x_p)\prod_{j=2}^{p}\pi(y_j|y_{[1:j-1]},x_{[(j+1):p]})dx_2\cdots dx_p \\
&= \int \pi(y_1,x_3,\ldots,x_p)\prod_{j=2}^{p}\pi(y_j|y_{[1:j-1]},x_{[(j+1):p]})dx_3\cdots dx_p \\
&= \int \pi(y_1,y_2,x_3,\ldots,x_p)\prod_{j=3}^{p}\pi(y_j|y_{[1:j-1]},x_{[(j+1):p]})dx_3\cdots dx_p \\
&= \cdots \\
&= \pi(y_1,y_2,\ldots,y_p),
\end{aligned}$$

as needed. A little bit more elegantly, let:

$$Q_k = \pi(y_{[1:(k-1)]},x_{[k:p]})\prod_{j=k}^{p}\pi(y_j|y_{[1:(j-1)]},x_{[(j+1):p]})$$

Then it can be seen that:

$$\begin{aligned}
&\int Q_k dx_k = \pi(y_{[1:k]},x_{[(k+1):p]})\prod_{j=k+1}^{p}\pi(y_j|y_{[1:(j-1)]},x_{[(j+1):p]}) = Q_{k+1}, \qquad k=1,\ldots,p-1, \\
&\int Q_p dx_p = \pi(y_{[1:p]}).
\end{aligned}$$

This gives us:

$$\int \pi(x)\mathcal{P}(x,y)dx = \int Q_1 dx_1\cdots dx_p = \int Q_2 dx_2\cdots dx_p = \cdots = \int Q_p dx_p = \pi(y).$$

Q2 The two main algorithms most people implemented here were a multivariate Metropolis algorithm (with a Normal random walk proposal) and a Metropolis-within-Gibbs algorithm (again, using Normal random walk proposals). Most coverage plots looked something like Figure 1.

Q3 There were differing degrees of success in tackling this problem. The course GitHub repo has solution code for those who couldn't get things to run. A couple of notes:

- Run more iterations! Unless computing time is a big issue, run for more iterations than you need. For example, set the number of iterations to 500k and leave things to run on Gauss (unless your code was very inefficient this shouldn't take too long. For example, my code takes 9 minutes to run 500k iterations on an old laptop).
- Using `mcmc` objects from `library(coda)` (when using `R`) makes life much simpler.

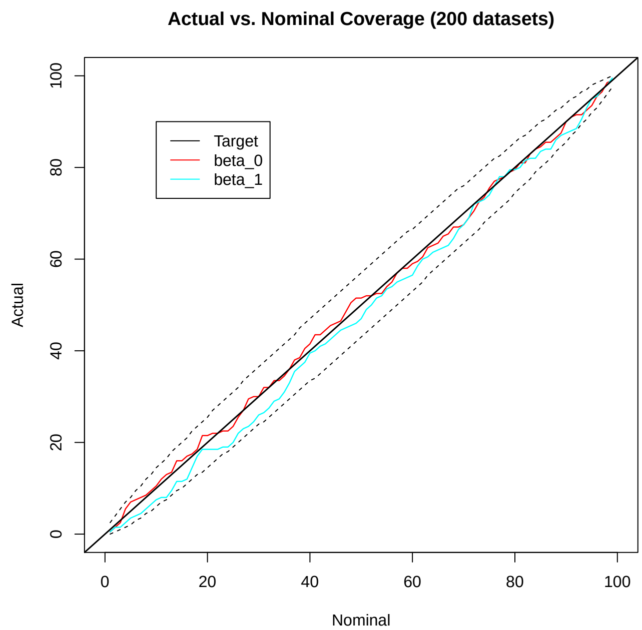

Figure 1: Coverage plot for Q2

- Covariance matrices for the proposal distribution needed to be selected carefully. Most successful choices were based on the covariance matrix from a regular `lm` or `glm` fit to the data.

Example traceplots are given in Figure 2.

Posterior estimates for the standardized and non-standardized cases are given below.

```
  Unstandardized Results:
  =====================
                   2.5%       25%       50%      75%     97.5%
Intercept   -33.646603 -19.97332 -13.37311 -6.97462  4.56086
area          0.002933   0.02297   0.03339  0.04341  0.06282
compactness -31.961992 -13.29931  -3.53277  6.12656 24.34174
concavepts   22.091979  46.37750  59.29670 72.32769 97.64969
concavity    -5.033865   3.65120   8.17016 12.83947 22.23377
fracdim     -67.036530 -29.34780  -9.68547  9.80690 47.50416
perimeter    -0.949401  -0.36876  -0.06827  0.22715  0.81614
radius       -8.038762  -3.58885  -1.38465  0.86469  5.11523
smoothness   13.793833  39.65263  53.44324 67.38485 94.48697
symmetry     -2.035344  11.20961  17.95906 24.70283 37.49552
texture       0.255499   0.32747   0.36717  0.40900  0.49198


  Standardized Results:
```

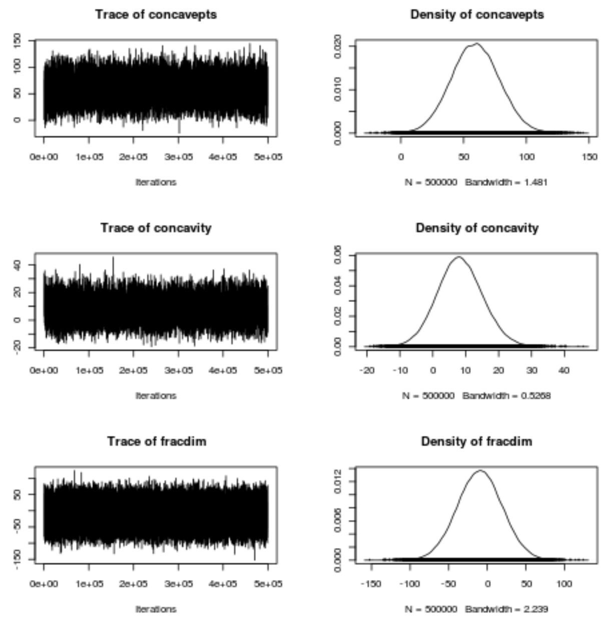

Figure 2: Traceplots for Q3 (unstandardized)

```
  =====================
                 2.5%      25%     50%       75%   97.5%
Intercept     -0.7355   0.0267  0.4389   0.81986  1.5809
area           1.9959   9.6230 13.8054  18.13752 26.2322
compactness   -2.2540  -0.8715 -0.1518   0.58897  2.0037
concavepts     0.5985   2.0095  2.7577   3.55913  5.2764
concavity     -0.5620   0.3531  0.8288   1.28917  2.1910
fracdim       -1.7799  -0.9534 -0.5175  -0.09353  0.7399
perimeter    -24.7027  -9.6716 -2.0238   5.48899 19.7902
radius       -29.5539 -14.5519 -6.5121   1.50326 17.3198
smoothness     0.2732   0.8634  1.1713   1.48762  2.1227
symmetry      -0.1017   0.2856  0.4901   0.69491  1.1048
texture        1.2646   1.6107  1.8069   2.01321  2.4510
```

The posterior predictive distribution for the mean (i.e., proportion of 1's) are shown in Figure 3 for the unstandardized setting. Note that other predictive statistics that aggregate over observations without accounting for the x_i's do not provide additional information beyond the mean.

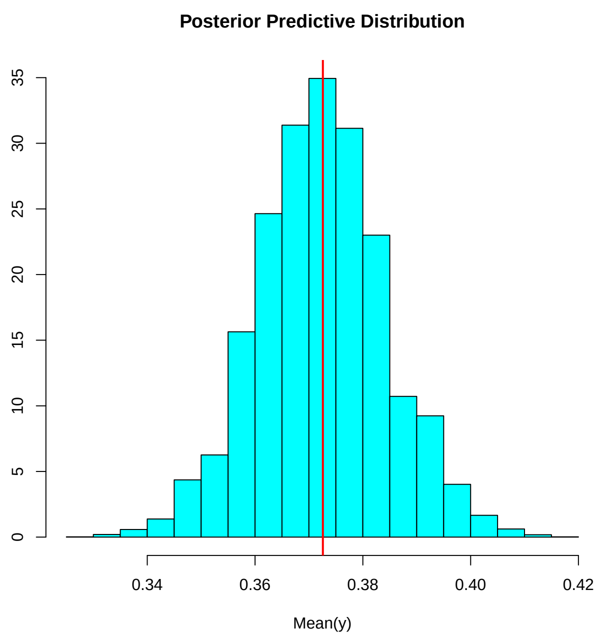

Figure 3: Posterior predictive distribution of proportion of 1's in the dataset (unstandardized)