ST740 – Assignment 3 – Due 11/17

This homework is due via email to the TA by 5PM on Friday, 11/17. Include code used to produce the results. You may work in groups, but write up separate solutions.

The "dishonest casino" data are from the `R` package `aphid`.

```
library(aphid)
Y          <- as.numeric(casino)
true_state <- names(casino)
```

The dealer is rolling a die and the data consist of a sample of $n = 300$ rolls. Let $Y_t \in \{1, ..., 6\}$ be the result of roll $t \in \{1, ..., n\}$. The assumption is that there are two latent states: one is a fair die with equal probability on all six outcomes and the other is a rigged die with unequal (and unknown) probabilities. The dealer switches in the rigged die at times during in the game, and the goal is to compute the posterior probability that the rigged die is being used at a given time.

Your assignment is to fit a Hidden Markov Model (HMM) to these data and verify that the model fits well. The HMM is defined through the latent state at time t that takes values $X_t = 1$ if the fair die is in play and $X_t = 2$ if the rigged die is in play. The distribution of Y_t is $\text{Prob}(Y_t = y|X_t = 1) = \frac{1}{6}$ for all y under the fair die and $\text{Prob}(Y_t = y|X_t = 2) = p_y$ under the rigged die. The distribution of the latent states is determined by the transition probability matrix with $P_{ij} = \text{Prob}(X_t = j|X_{t-1} = i)$. Uninformative Dirichlet priors are given to the rigged die probabilities $(p_1, ..., p_6) \sim \text{Dirichlet}(1/6, ..., 1/6)$ and the rows of the transition probability matrix $(P_{11}, P_{12}) \sim \text{Dirichlet}(1/2, 1/2)$ and $(P_{21}, P_{22}) \sim \text{Dirichlet}(1/2, 1/2)$. Call this the "full model".

1. Derive the full conditional distributions of (a) X_t, (b) $(p_1, ..., p_6)$ and (c) (P_{11}, P_{12}).

2. Write an MCMC algorithm to approximate the posterior distribution of the latent states $X = (X_1, ..., X_n)$ and verify that the MCMC algorithm has convered.

3. (a) Plot the posterior distribution of the latent states over roll number. (b) Compare the posterior with the data $Y = (Y_1, ..., Y_n)$ and describe the types of data sequences that lead to the HMM flagging the die as rigged. (c) Compare the results with the true state and summarize your method's performance (of course, this step cannot usually be done in a real data analysis).

4. Compare the fit of three models: (1) the full model, (2) the model that assumes independence across rolls with $\text{Prob}(X_t = 2) = P$ independent across t and (3) the model without a rigged die, i.e., the latent state fixed at $X_t = 1$ for all t.

5. Verify that the full model fits the data well.

Turn in a single PDF document with the results summarized in the first two pages and the derivations and code in the remaining pages. Email this document to Chenyin Gao (cgao6@ncsu.edu) by 5PM on Friday, Nov 17.