Homework 1 - Bayesian Statistics - Due February 16th

1. In this first question, we will consider a very simple problem for practice. Suppose that $x_n \sim N(\mu,\sigma^2)$ are i.i.d. and N is the sample size.

(a) Write down the likelihood of (μ,σ^2)

(b) Write down an expression for the posterior distribution of (μ,σ^2), assuming a uniform (improper) prior for (μ,σ^2)

(c) Write down an expression for the posterior distribution of (μ,σ^2), assuming that μ and σ^2 have independent priors $\mu \sim N(\mu_{pr},\tau_{pr}^2)$ and $\sigma^2 \sim IG(\alpha_{pr},\beta_{pr})$ (here, IG refers to the inverse gamma distribution, which is characterized by the parameters α_{pr} and β_{pr} and has a density function of $\frac{\beta_{pr}{}^{\alpha_{pr}}}{\Gamma(\alpha_{pr})}(\sigma^2)^{-\alpha_{pr}-1}e^{-\beta_{pr}/(\sigma^2)}$).

(d) Demonstrate that the full conditionals for the posterior distribution are given by,

$$f(\mu\,|\,\sigma^2,x) = N(\left(\tfrac{1}{\sigma^2/N}+\tfrac{1}{\tau^2_{prior}}\right)^{-1}\left[\tfrac{\bar{x}}{\sigma^2/N}+\tfrac{\mu_{prior}}{\tau^2_{prior}}\right],\left(\tfrac{1}{\sigma^2/N}+\tfrac{1}{\tau^2_{prior}}\right)^{-1})$$

$$f(\sigma^2\,|\,\mu,x) = IG(N/2+\alpha_{pr},\tfrac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2+\beta_{pr})$$

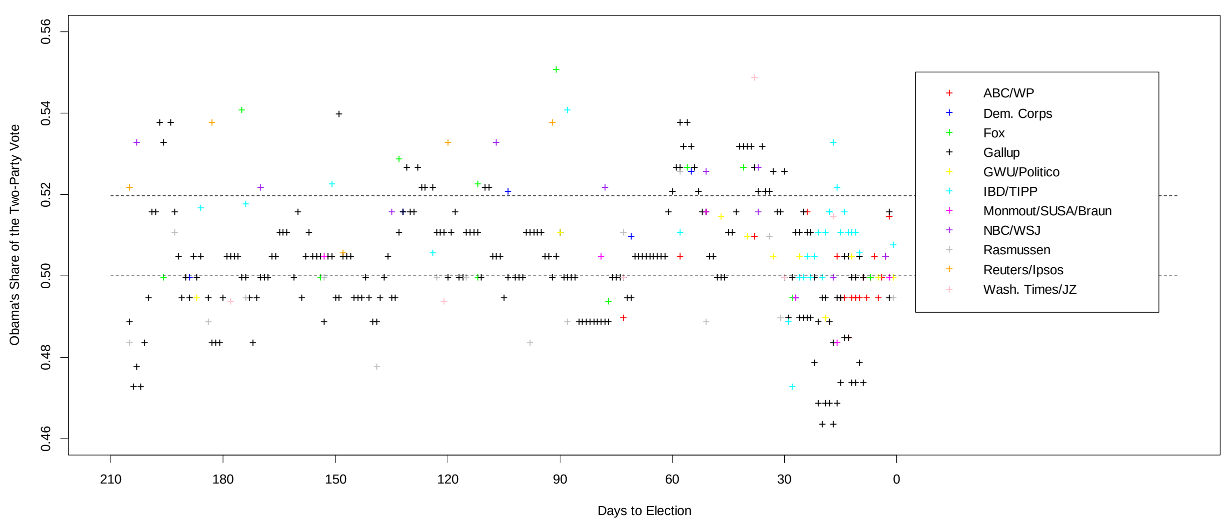
(e) Use the expressions from (d) to generate a Gibbs sampler for (μ,σ^2) using direct sampling to sample from $\mu\,|\,\sigma^2,x$ and $\sigma^2\,|\,\mu,x$. Note that the "pscl" library has a function that can be used to sample from the Inverse Gamma distribution.

(f) Program up a sampler for (μ,σ^2) using either (1) Gibbs-Metropolis Hastings (2) Gibbs-slice sampling, (3) JAGS, or (4) STAN.

(g) Generate a sample of $N = 10000$ data points from the data generating process assuming that $(\mu, \sigma^2) = (-0.5, 0.8)$. Use this to test the sampler your programmed in parts (e) and (f). Provide some graphical evidence that samplers worked correctly.

2. Consider the data file polls.dat on the course website. The data file has polling data from the 2012 Presidential race from various polling firms. The variable *DateEnd* is the last day of the poll. For simplicity, you can assume that all the interviews were conducted on the last day. Note that you can convert the data from a string to a date that r can recognize using the command *as.Date(DateEnd,"%m/%d/%Y")*. The variable *DemSh* has Obama's share of the two party vote. The variable *Survey.House* has a unique string identifier for each polling company. *N* is the sample size of the poll.

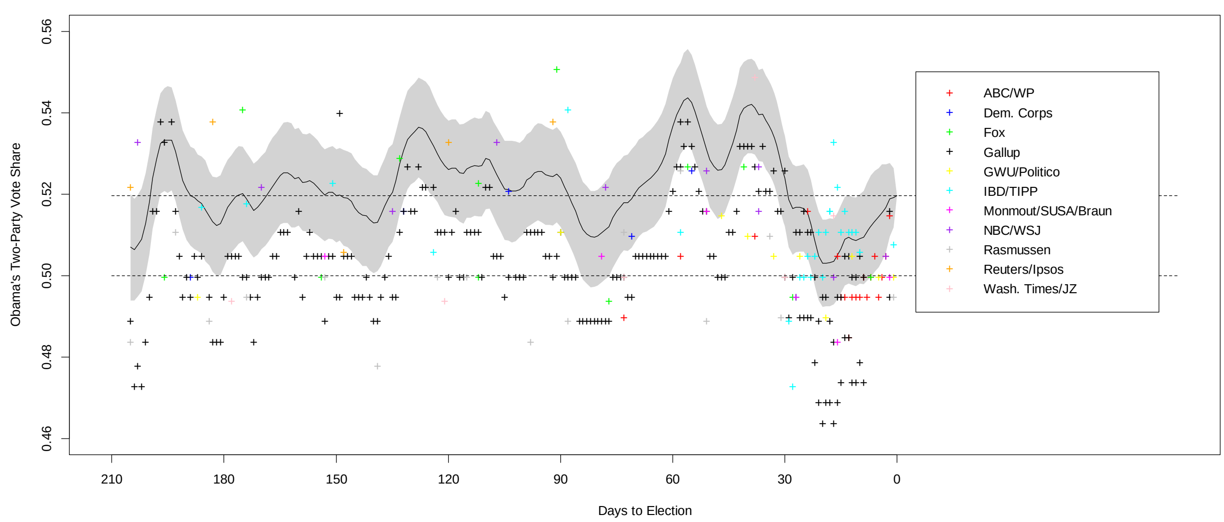
(a) Create a plot of the polling results over time (using different colors to denote different survey houses). Below is what my version looks like.



(b) Estimate a model similar to the one Simon Jackman estimated in his Australian Journal of Political Science paper. You may do this however you like, but the easiest way will be to adapt the JAGS code in the appendix of Jackman's paper, which is mostly correct but has an error in it. You will have replace the loop for alpha and the prior for alpha with,

```
for(i in 1:(NPERIODS-1)) { alpha[i] ~ dnorm(alpha[i+1],tau) }
alpha[NPERIODS] = 0.519638643 ## initialization based on election outcome
```

Once you estimate this model, report the estimates of alpha along with confidence intervals, superimposed on the plot you created in (a). Below is what my version looks like.



(c) Report the estimates of the house effects along with confidence intervals. Interpret the house effects. Which survey houses were the most accurate?