Midterm 2: Project. Stat 157, Fall 2012

Due – November 6, 2012 at 3:30 PM Pacific Time

In this take-home exam, your objective is to **predict the result of the 2012 U.S. Presidential Election** with Bayesian statistical modeling. *You may work in teams of two or individually.* If you work in teams of two, please turn in one copy of your final report, with both team members' names on it. *You can ask me questions, but do not discuss the project with any classmates that are not on your team.*

Your task for the exam is to **determine the posterior probability that President Obama will win the 2012 Election given the data of your choosing**. For those unfamiliar with the U.S. Presidential Voting System, each state (and the District of Columbia) is given a pre-determined number of electoral votes based on its number of legislators. Each state allocates its votes to the candidate receiving the largest proportion of votes from that state's electorate. The winner of the election is the candidate receiving the most total electoral votes [1].

A complete project will:

- characterize the (joint) posterior distribution of Obama's margin of victory, $\theta_1, ..., \theta_{51}$, in each state (and the District of Columbia), and
- use that joint distribution to find the posterior distribution of the number of electoral votes won by Obama and his posterior probability of winning the election. The current allocation of electoral votes by state is provided in the file `electoral_college_2012.dat` on bspace.

Since the report is due on Election Day, you will be able to predict the outcome of the election and then validate the performance of your predictions!

The exam should be treated like a research project, meaning that **the final report should contain separate sections for Introduction, Data, Methodology, and Results, and a bibliography with (at least two) references**. The following will be essential for a satisfactory completion of the exam:

1. **Literature Review:** Read about what previous researchers have done in the field of Bayesian election forecasting. I have placed three papers on bspace that should be very helpful in guiding you towards relevant data and models. In addition, you can search the academic literature using Google Scholar and for extra inspiration there are dozens of websites that do (mostly non-Bayesian) election prediction, such as `http://fivethirtyeight.blogs.nytimes.com`, `PollyVote.com`, and `http://www.realclearpolitics.com`.

[1] For more information about this unique system see `http://en.wikipedia.org/wiki/Electoral_College_(United_States)`

2. **Data Collection:** I have compiled an extensive data set of state-wise polling data, state-wise results of previous elections (2004 and 2008), state-wise and national unemployment data (from 2004–2012), and 2012 election betting data from Intrade `www.intrade.com`. All of this data is bundled up in the file `project_data.tar` on bspace complete with a `README` file. You may use all, some, or none of this data. You may use any other data sources you wish. If you have an online data set that you want to use but are having problems downloading and organizing data from the web, please see me.

3. **Model Choice:** A significant part of this project is to choose a likelihood that appropriately describes the data and to pick good prior distributions on all of the model parameters. You can use any of the models that we have talked about in class, or any other model that you want. Feel free to use any of the papers in your literature review for inspiration. As always, start simple and add more complexity if it is needed. You should describe all of your modeling choices in the Methodology section.

4. **Model Fitting:** Use `JAGS` to fit your model and sample from the joint posterior. Try to get a simple model to fit in the first week of the project so you can be sure that your code is working properly. Use diagnostics to test convergence of the MCMC.

5. **Examining the Posterior:** Use plots and numerical summaries to answer the questions of interest. All data from the 2008 election is provided for you to validate your methods on the results of that election.

All resources (data sets, papers, etc.) are in bspace under the Resources page under the Midterm 2: Projects directory.

Please arrange to see me early on (within the first week of the project) during office hours so we can make sure that you are on the right track to a successful project.

Most importantly, have fun trying to predict the election!