EXERCISE DUE 10/5

(1) Consider the model we discussed in class as an example of where the SLLN ought not to be relied on to justify estimating an expectation as a sample mean of i.i.d. draws. That is,

$$X_t = \begin{cases} 0 & \text{w.p. } 1-\varepsilon \\ \varepsilon^{-2} & \text{w.p. } \varepsilon \end{cases}$$

Discuss how the posterior mean of the X_t's (i.e., $E[1/\varepsilon \,|\, \{X_1,\ldots,X_T\}]$) behaves as a function of T when T successive observations of $X_t = 0$ are drawn. Consider the following two cases:

(a) The prior pdf for ε is uniform on (0,1).

(b) The prior pdf for ε is Beta(3,2), i.e. proportional to $\varepsilon^2(1-\varepsilon)$.

The Beta distribution

The Beta(p,q) distribution has pdf

$$\text{Beta}(x \,|\, p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{(p-1)}(1-x)^{(q-1)} .$$

The normalizing constant for this distribution is what is known as the Beta function:

$$\text{Beta}(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} .$$

The Beta distribution is well defined for any $p > 0$ and $q > 0$. When p or q is less than 1, the pdf is unbounded at zero or 1, respectively, though of course still integrable. The expectation of a Beta(p,q) variable is $p/(p+q)$.

(2) Forecasting the election. The course website has three versions of the data set underlying Ray Fair's presidential election forecasting regression. `FairElectionData.txt` has the data as a text file, with variable names across the top and election years down the left. The `fairData` file has the same data arranged in an R `ts` object. `FairElectionData.xls` has the data in a spreadsheet format. Fair's web site `http://fairmodel.econ.yale.edu/vote2004/` shows his forecast as of July 2004 of the right-hand-side variables for his estimated equation for this election. For PARTY, PERSON, and DURATION the values are -1, 1, and 0. (These variables are certain now.) For GROWTH, INFLATION, and GOODNEWS, they were forecast by Fair in July as 2.7, 2.1, and 2.

(a) See if you can reproduce Fair's own estimates of his equation. He uses only the data from 1916 onward. His estimates are displayed in the November 2004 update paper `http://fairmodel.econ.yale.edu/RAYFAIR/PDF/2002DHTM.HTM`.

(b) Use the full data set, back to 1880, and compare the results informally. (Look at differences in estimated coefficients to see if they imply substantively important differences in effects. Look at whether probability bands around estimates overlap.)

(c) For both sets of estimates, treating Fair's forecasts of November 2004 values for his right-hand-side variables as exactly known values, assuming the SNLM applies and using a σ^{-1} improper prior on σ, construct 95% probability intervals for the Bush vote percentage. Take account of both parameter uncertainty and disturbance uncertainty, including uncertainty about σ^2. Also calculate the probability, given the data, that Bush gets more than 50%.

(d) Redo this analysis using a proper prior. Use a prior in the conjugate class and choose it so that it does not imply unreasonable beliefs about how much variation in the vote percentage can be explained by the regression equation. Also keep it centered at a model that simply predicts 50%, regardless of the values of the right-hand-side variables. By this I mean that at the prior mean values of β, the constant term is 50% and every other coefficient is zero.

Setting up the election forecast regression: The R dataset is in an R "ts" (for "time series") object. You could recast it to an R data frame by writing `fairData <- as.data.frame(fairData)` right at the start. Then you can specify your regression using variable names, along the lines of

```
faireg <- lm(VOTES~PARTY+PERSON+ etc., data=fairData)
```

If you stick with the ts object version you have to set up the regression with columns of the matrix, e.g.

```
faireg <- lm(fairData[,"VOTES"]~fairData[,2:8])
```

The advantage of sticking with the time series object is that then residuals are automatically properly associated with dates on plots and listings of data. If you use the data.frame object, you have to do some work yourself to get correct year labels on plots.

In R you'll want to use the `window()` function to use the shorter data set.

For Matlab, you can strip away the labeling in the text file, using an editor, and then read in the data matrix as an ascii file, or try matlab's commands that read spreadsheet formats. You'll need up to set up the regression algebra and labeling of output yourself, though that's not too hard.

For RATS, you can read in the xls data file, using the for=xls option in DATA. It looks like RATS can't automatically keep track of the gaps in the dates in these series, so its best to treat them as undated.

CONJUGATE PRIORS FROM DUMMY OBSERVATIONS

Supposse we want to implement a prior that has a marginal distribution on $\nu = 1/\sigma^2$ that is Gamma(p,α) and conditional distribution for $\beta \mid \sigma^2$ that is $N(\mu, \sigma^2\Omega)$. Begin by finding W such that $W'W = \Omega^{-1}$. There is always such a W, indeed many of them. One is the inverse of the cholesky decomposition. In R, this would be `W<-t(solve(chol(Omega)))`, in Matlab `W=inv(chol(Omega)')`, in Rats `comp W = inv(decomp(Omega))`. Then choose as a first set of dummy observations

$$Y^* = W\mu$$
$$X^* = W$$

This will take care of the conditional distribution of $\beta \mid \sigma^2$.

To add a factor reflecting the proper prior on ν, add $2p$ additional dummy observations, each with $Y = \sqrt{\alpha/p}$ and $X = 0$. **Note that this is different from what I said at the end of the 9/30 lecture! Then I said you set both Y and X to zero in these additional dummy observations. That will not implement a proper prior.** Note that since the number of dummy observations is discrete, this approach can only implement Gamma priors on ν that have p an integer multiple of $\frac{1}{2}$.

While this formally takes care of the problem, one more often works more directly with the dummy observations. One can think of reasonable X_t^* vector values (where X_t^* is a single hypothetical row of the X matrix), and for each one what a reasonable corresponding Y^* value would be. Then one can ask oneself what would a reasonable "standard error" on this correspondence between X_t^* and Y_t^* be, and multiply both Y_t^* and X_t^* by the ratio of what you think the equation residual standard error should be to your subjective standard error on this dummy observation. Proceeding this way, one need not be limited to a specific number of dummy observations. So long as there are at least $k+1$ dummy observations, with a full column rank X^* matrix, the implied prior pdf will be proper.

One reasonable standardized proper prior might set $\Omega = (X'X/T)^{-1}\kappa/k$, where k is the number of columns in X and $\kappa/(1+\kappa)$ is the fracation of pre-sample uncertainty about $Y'Y$ one wants to attribute to uncertainty about β, as opposed to uncertainty about the error terms. This of course still leaves open the choice of the mean of β and the values of p and α. Choosing $p = \frac{1}{2}$ or $p = 1$ is reasonable, as these priors do not imply a local peak in the prior pdf on ν anywhere except at zero.