STAT 740, Fall 2017: Homework 5

1. Consider the Challenger data fit via logistic regression.

   (a) Fit the Bayesian model under Jeffreys' prior (p. 9 in Tim's bootstrap notes) using adaptive Metropolis via the `adaptMCMC` package. You will need to write a function that gives the log-posterior as the log-likelihood plus the log of Jeffreys' prior. Use a target acceptance rate of 30% and the asymptotic covariance matrix as a starting `scale`, e.g.

   ```
   f=logistf(y~t,family="binomial")
   vcov(f) # asymptotic posterior covariance under Jeffreys' prior
   ```

   (b) Compare the bootstrapped sampling distributions (the default histogram from `plot`ing a `boot` object is fine) of $\hat{\beta}_0$ and $\hat{\beta}_1$ (see my R code) under Firth's method to the marginal posteriors $[\beta_0|\mathbf{y}]$ and $[\beta_1|\mathbf{y}]$ from a Bayesian fit using Jeffreys' prior. Are they skewed? Are the similar?

   (c) Compare the bootstrapped CIs to the Bayesian CIs.

   (d) Compare the Bayesian estimate (favor the posterior median) and CI to the penalized MLE and bootstrapped CI for the relative risk in Problem 2, Homework 1, e.g.

   $$\theta = \frac{\frac{e^{\beta_0+\beta_1 36}}{1+e^{\beta_0+\beta_1 36}}}{\frac{e^{\beta_0+\beta_1 70}}{1+e^{\beta_0+\beta_1 70}}}.$$

   How do these compare to your large sample results in the unpenalized case from Problem 2, Homework 1?

2. For the beta regression Problem 4 of Homework 4, compare four Bayesian models via LPML and WAIC: (a) random $u_1, \ldots, u_{10}$ and precision ϕ, (b) fixed $u_1, \ldots, u_{10}$ and precision ϕ (maybe each u_j could have a $N(0, 10^4)$ prior?), (c) random $u_1, \ldots, u_{10}$ and precision $\phi_i = e^{\gamma_0+\gamma_1 t_i}$, (d) fixed $u_1, \ldots, u_{10}$ and precision $\phi_i = e^{\gamma_0+\gamma_1 t_i}$. Just use the `loo` package. Which model is best? Rank the models from best to worst, giving a pairwise pseudo Bayes factor for each adjacent pair (three pseudo Bayes factors total).

3. Look through the paper on LOO and WAIC I posted on the course webpage. Why does the simple "trick" (8) for estimating the conditional predictive ordinate $p(y_i|\mathbf{y}_{-i})$ fail? How does PSIS-LOO over come this? Which do the authors recommend overall between PSIS-LOO and WAIC? Why?

4. Look at the BMA slide, first formula. For $J = 2$ and $P(M = 1) = P(M = 2) = \frac{1}{2}$, show that the odds of $[M = 1|\mathbf{x}]$ vs. $[M = 2|\mathbf{x}]$ is the Bayes factor BF_{12}, i.e. the Bayes factor is the posterior odds of the two models.

5. For the Ache hunting data, fit several models via `glmer` in the `lme4` package, obtaining both the marginal AIC and the conditional AIC (`cAIC4` package). Fit models with both linear and quadratic terms, and with and without random effects. Which of the four models is preferred via the usual marginal AIC? Which is preferred by conditional AIC?

6. Compare bootstrapped CIs for the population $90th$ percentile to the large sample estimate as in the notes for (a) $exp(1)$ data, (b) $N(0,1)$ data, (c) $U(0,1)$ data, and (d) χ_1^2 data. For sample sizes of $n = 100$ and $n = 500$ replicate $M = 500$ data sets and compute coverage probabilities of the two intervales and average interval length. Construct a table summarizing your results. Which is better?

7. Consider gamma data,

$$x_1, \ldots, x_n \overset{iid}{\sim} f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

Using Jeffreys' prior

$$\pi(\alpha, \beta) \propto \frac{\sqrt{\alpha\psi'(\alpha) - 1}}{\beta}$$

where $\psi'(\cdot)$ is the trigamma function, obtain the full conditional distribution for $\beta|\alpha, \mathbf{x}$. What is this distribution? Also find the full conditional distribution for $\alpha|\beta, \mathbf{x}$. Outline a component-at-a-time MCMC scheme for updating α and β where $\beta|\alpha, \mathbf{x}$ should be closed-form and $\alpha|\beta, \mathbf{x}$ updated with an adaptive M-H step (see Tim's MCMC notes, p. 25). Suggest starting values (α^0, β^0) obtained from method-of-moments estimators.

Simulate gamma data of size $n = 200$ for values $\alpha = 20$ and $\beta = 10$ and implement your MCMC scheme. Superimpose a density estimate over the range $(0, 5)$ on top of a histogram of the data and the true density.