STAT 740, Fall 2017: Homework 4

1. Let $\mathbf{x} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Find a $k \times k$ matrix $\mathbf{M}$ such that the elements of $\mathbf{y} = \mathbf{M}\mathbf{x}$ are independent. Hint: S.V.D.

2. In the finite mixture of normals MCMC example, argue that

$$P(z_i = j|\text{else}) = \frac{\pi_j \phi(x_i|\mu_j, \tau_j^{-1})}{\sum_{k=1}^J \pi_k \phi(x_i|\mu_k, \tau_k^{-1})}.$$

3. Read Prof. Charlie Geyer's post on "one long run" linked from the course website. Write a brief paragraph outlining Charlie's main arguments on why one long run is preferred over many short runs to obtain MCMC inference and diagnose "convergence."

4. **Beta regression with random effects**. Beta regression is useful for responses that live between zero and one $y_i \in (0,1)$, e.g. proportions. In general,

$$y_i \sim \text{beta}(\mu_i\phi, (1-\mu_i)\phi), \;\; g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}.$$

`betareg` is an R package that performs beta regression for various links $g(\cdot)$. Briefly read through

`https://cran.r-project.org/web/packages/betareg/vignettes/betareg.pdf`

   (a) In the beta regression model above, what is the mean and variance of y_i? Note: ϕ is called the 'precision.' The variance changes naturally with the mean like in binary and Poisson regression; variability decreases when μ_i is close to zero or one.

   (b) Install `betareg` and examine the dataset `GasolineYield`, e.g.

```
install.packages("betareg")
library(betareg)
data(GasolineYield)
attach(GasolineYield)
?GasolineYield
GasolineYield
f=betareg(yield~batch+temp)
summary(f)
```

You can see a MLE fit of these data in Section 4.1 of the PDF file referenced above. Let y_i be the ith `yield` from `batch` b_i at `temp` t_i. Let $\mathbf{x}_i = (1, t_i)'$. Use JAGS to fit the nonlinear mixed effects model

$$y_i|\boldsymbol{\beta}, \mathbf{u} \overset{ind.}{\sim} \text{beta}\left(\frac{e^{\mathbf{x}_i'\boldsymbol{\beta}+u_{b_i}}\phi}{1+e^{\mathbf{x}_i'\boldsymbol{\beta}+u_{b_i}}}, \frac{\phi}{1+e^{\mathbf{x}_i'\boldsymbol{\beta}+u_{b_i}}}\right), \;\; i = 1, \ldots, 32,$$

$$u_1, \ldots, u_{10}|\tau \overset{iid}{\sim} N(0, \tau^{-1}).$$

Assume $\beta_0, \beta_1 \overset{iid}{\sim} N(0, 10^5)$ (variance=10^5) independent of $\tau, \phi \overset{iid}{\sim} \Gamma(10^{-5}, 10^{-5})$. Remember: JAGS uses the precision, not the variance in its normal parameterization. Here, $g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta} + u_{b_i}$ where $g(\mu) = \log(\frac{\mu}{1-\mu})$. Use something like `yield[i] dbeta(alpha[i],beta[i])` in your model specification. Also note `ilogit` is the logistic function $f(x) = \frac{e^x}{1+e^x}$.

(c) Plot estimates of $\mu(t) = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{1+e^{\mathbf{x}'\boldsymbol{\beta}}}$ over the range of `temp` values, the mean for a typical batch with $u = 0$, i.e. the average yield for *the median batch effect*. There are two ways to do this:

- (Quick and dirty) Use posterior mean $\hat{\boldsymbol{\beta}}$ as "plug-in" value

$$\hat{\mu}(t) = \frac{e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}}}{1+e^{\mathbf{x}'\hat{\boldsymbol{\beta}}}},$$

- (Preferred Bayes' estimate w.r.t. squared-error loss) Find the posterior mean of the density over a grid of `temp` values t

$$\tilde{\mu}(t) = \tfrac{1}{M}\sum_{m=1}^{M} \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}^m}}{1+e^{\mathbf{x}'\boldsymbol{\beta}^m}}.$$

  You can define this over a grid using a loop in JAGS and monitor them. Then simply pull out posterior means after fitting. Obtain both estimates, plot them, and compare them, e.g. in the JAGS model have

```
for(i in 1:100){
 tg[i]<-200+(i-1)*250/99 # grid from 200 to 450
 mu[i]<-ilogit(b[1]+b[2]*tg[i]) # mean for median batch
}
```

Make sure you include "mu" in the parameters to save. If f is your fitted JAGS object, then do f2=print(f) and then

```
plot(seq(200,450,length=100),f2$mean$mu,type="l")
```

EXTRA CREDIT: from the same fitted JAGS object, pull out the $2.5th$ and $97.5th$ percentiles of each mu and add pointwise 95% CI's to the plot you just made as dashed-lines.

(d) Generalize the model to regress the precision on temperature as well, i.e. $\phi_i = e^{\gamma_0+\gamma_1 t_i}$. Which model has lower DIC? We will talk about DIC shortly, but it is essentially a "Bayesian AIC."

5. Recall Problem 2 in Homework 3. Run my code to fit a Bayesian version of this model on the V.A. data in JAGS. We didn't get to it in class yet, but it's in my R code examples under `JAGS: censored normal data`. Compare the posterior means (or medians) to the MLEs $\hat{\mu}$ and $\hat{\sigma}$. Also compare the posterior standard deviations to the standard errors. Recall you can get both MLEs and standard errors from `survreg` if weren't able to finish Homework 3 Problems 2 & 3.