Masters Comprehensive Exam
Stat 505-506 August 2013
100 points

1. Biologists observed 27 owl nests and randomly applied a treatment or a control during each of several observation times on each nest. The treatment involved feeding the owlets so that they were not hungry (satiated). The response measured was the number of calls made by the owlets in the 30 seconds before the parent arrived. The sex of the parent was also recorded. Note that the number of calls (y) naturally depends on the size of the brood in the nest. The following was fit using the R package lme4.

```
> g1 <- glmer(calls~FoodTrt*SexP+offset(log(BroodSize)) + (1|Nest),
              family=poisson,data=owlets)
> summary(g1)
Generalized linear mixed model fit by the Laplace approximation
Formula: Calls ~ FoodTrt * SexP + offset(log(BroodSize)) + (1|Nest)

Data: owlets
AIC BIC logLik deviance
3532 3554 -1761   3522

Random effects:
Groups Name        Variance Std.Dev.
Nest   (Intercept) 0.20631  0.45421
Number of obs: 599, groups: Nest, 27(Intercept)

Fixed Effects:
                        Estimate StdErr    z value   Pr(>z)
(Intercept)              0.65584  0.09564   6.857     7.03e-12
FoodTrtSatiated         -0.65612  0.05612  -11.691   < 2e-16
SexPMale                -0.03705  0.04506  -0.822     0.4110
FoodTrtSatiated:SexPMale 0.13130  0.07047   1.863     0.0624
```

(a) What does the `family = poisson` argument do? (3 pts)

(b) In what scale is this linear model fit? (3 pts)

(c) What assumption about variability is implicit in the family specification? (3 pts)

(d) Write out the model as fit with the R code using Greek letters for each parameter. Describe all distributions used. (10 pts)

(e) Interpret the last line in the coefficient table. What does it mean, and what strength of evidence does the model provide for this term? (10 pts)

(f) Similarly, interpret the `SexPMale` coefficient estimate. (10 pts)

(g) Interpret the random effects as provided in the output. (5 pts)

(h) Overdispersion for this model is estimated as 5.6. What effect does that have on our inference? How can you modify the code to fix the problem? (5 pts)

2. Suppose we have five subjects and measure the "math aptitude" of each using 3 different tests. Model the vector of 3 responses for the ith subject as:

$$\boldsymbol{y}_i = \boldsymbol{\beta} + \mathbf{1}b_i + \boldsymbol{\epsilon}_i, i = 1, \ldots, 5$$

where $\boldsymbol{y}_i$, $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}_i$ are 3 by 1 vectors, $\boldsymbol{\epsilon}_i \sim iid(\mathbf{0}, \sigma^2\boldsymbol{I})$ and $b_i \sim iid(0, \sigma_b^2)$ are random subject effects.

   (a) What is the variance-covariance matrix for the three measurements on one subject? (Use Greek letters.) (6 pts)

   (b) What do the three coefficients in $\boldsymbol{\beta}$ represent? (6 pts)

   (c) Next, suppose that the tests have three different variances. How does that change your above answers? (3 pts)

   (d) Now stack the five $\boldsymbol{y}_i$'s up into a single response vector of length 15 and call it $\boldsymbol{y}$. Describe the variance-covariance of the entire response vector. (6 pts)

3. Causal inference.

   (a) Explain what a counterfactual is and how the process of randomization allows us to make causal inference in experiments where treatments are assigned randomly. (10 pts)

   (b) Social scientists like Andrew Gelman, cannot randomly assign treatments to people (like race or ideology), yet they still want to make causal inferences. What assumptions are needed to allow them to make such inference? (6 pts)

4. Relating to running MCMC iterations in BUGs or JAGS:

   (a) Why is it important to check convergence when using MCMC methods to fit models? (4 pts)

   (b) Describe one way to do so. (4 pts)

   (c) Suppose we obtain a 90% interval estimate for parameter θ of (1.37, 2.54). What is the Bayesian interpretation of this interval? (6 pts)