

## Statistics 347, Homework 7, due Feb 28

Discussion of homework problems among students is encouraged. However, all material handed in for credit must be your own work.

Please hand in each problem in a separate file with your name on it.

1. Faraway, Chapter 11, Problem 1, page 251
2. Consider the following model:  $Y_n \sim \text{Pois}(\exp(\alpha X_n + u_n))$ , where  $X_n = 2 + .01n$ ,  $n = 0, \dots, 100$ , with  $\alpha = 1$  and  $u_n \sim N(0, \sigma^2)$ ,  $\sigma = 1.2$ . exclude intercept

- (a) Simulate data from this model. Apply the `glm` and `glmer` fits to the data. Show what formulas you use in each case. Discuss the difference of the two fits in terms of estimates of the parameters, the likelihood and residuals etc.
- (b) Assume  $\sigma$  is known. Write the Laplace approximation to the joint distribution of Y,U in  $u$  and  $\alpha$  (second order approximation of the log-density around the maximal values  $u^*, \alpha^*$ ). The Newton-Raphson iterations to compute  $u^*$  and  $\alpha^*$  involve an iteration

$$J \begin{pmatrix} \delta u \\ \delta \alpha \end{pmatrix} = V,$$

$$\alpha^{new} = \alpha^{old} + \delta \alpha, u^{new} = u^{old} + \delta u.$$

What are  $J$  and  $V$ .

3. The National Youth Survey collected a sample of 11-17 year olds, 117 boys and 120 girls, asking questions about marijuana usage. The data is presented in `potuse` in library `faraway`.
  - (a) Plot the total number of people falling into each usage category as it varies over time separately for each sex.
  - (b) Condense the levels of the response into whether the person did or did not use marijuana that year. Turn the year into a numerical variable. Fit a `GLMM` for the now binary response with an interaction between sex and year as a predictor. Comment on the effect of sex.
  - (c) Fit a reduced model without sex and use it to test for the significance of sex in the larger model.
  - (d) Fit a model with year as a factor. Should this model be preferred to the model with year as just a linear term? Interpret the estimated effects in the year as a factor version of the model.

add "year" as  
random slope

remove the  
random slope just  
for simple

qualitatively answer the question  
plot shows linear effect

use the model from (c)