Homework 2 for STA 841

(Due in class October 1)

Problem 1

First, derive the deviance for the following familar families:

1. Binomial(n, π): $D(\pi_1, \pi_2)$.
2. Gamma(N, λ): $D(\lambda_1, \lambda_2)$, where N the shape parameter is known and λ is the rate parameter.
3. Negative-Binomial(k, p): $D(p_1, p_2)$.

Show that for $Y \sim$ Gamma(N, λ), find the deviance residual. Moreover, show that the deviance residual has the same distribution for all λ.

Problem 2

(To be done individually) This problem is taken from McCullagh and Neldar (1991). The data, taken from Snedecor and Cochran (1967, p.354), were obtained as part of an experiment to determine the effects of temperature and storage time on the loss of ascorbic acid in snap-beans. The beans were all harvested under uniform conditions at the Iowa Agricultural Experiment Station before eight o'clock one morning. They were prepared and quick-frozen before noon the same day. Three packages were assigned at random to each temperature and storage-time combination. The sum of the three ascorbic acid determinations is shown in the Table.

	Weeks of storage				
Temp.^oF	2	4	6	8	Total
0	45	47	46	46	184
10	45	43	41	37	166
20	34	28	21	16	99
Total	124	118	108	99	449

Suppose for the purpose of model construction that the ascorbic acid concentration decays exponentially fast, with a decay rate that is temperature-dependent. In other words, for a given storage temperature T, the expected concentration after time t (measured in weeks) is $\mu = E(Y) = e^{\alpha - \beta_T t}$. The initial concentration e^{α} is assumed in this model to be independent of the storage temperature.

1. Express the above theory as a generalized linear model—clearly specify the random component, systematic component and the link function, treating temperature as a factor and storage time as a variate.

   [The above model is unusual in that it contains an interaction between time and temperature, but no main effect of temperature. By design, the concentrations are equal at time zero.]

2. Estimate the times taken at each of the three temperatures for the ascorbic acid concentration to be reduced to 50% of its original value. Construct confidence intervals for this half-life. (Be careful with the bounds.)

3. The mean squared error for individual packages, obtained from the replicates, was 0.706 on 24 degrees of freedom. Is this value consistent with the above analyses? Why or why not?

Problem 3

(To be done individually) Estimating the LD_{50}. Experiments are often carried out at a sequence of dose levels, $x_0, x_1, \ldots$ each dose being twice the preceding dose. The model most commonly used in toxicology is linear in log dose. Suppose that the following results have been obtained in an experiment at various multiples of the baseline dose.

Dose $\log_2(x)$	0	1	2	3	4	5
Mortality y/m	0/7	2/9	3/8	5/7	7/9	10/11

Here y/m is the number of deaths occurring in a sample of m individuals.

1. Plot the data, i.e. the mortality fraction against log dose.

2. Fit a probit linear model in which the probit of the mortality rate is linear in log dose using the `glm` function in `R`. Superimpose the fitted probabilities on the plot.

3. Obtain the estimate of the $\log_2 LD_{50}$, and use Fieller's method to generate a 95% confidence interval.

4. One could consider computing a bootstrap CI for $\log_2 LD_{50}$. What may be a potential concern for doing that in the current problem? Compute a 95% bootstrap CI and comment on how well it matches the CI you computed in the previous part.

5. Fit a Bayesian probit linear model using the data augmentation technique by Albert and Chib introduced in class. Please specify the priors you adopt as well as any other assumptions you impose. Based on a posterior sample, construct a 95% Bayesian credible interval for $\log_2 LD_{50}$.