ORIE 671: Intermediate Applied Statistics

Homework Assignment 2: Fall 2003
Due date: Friday October 17

1. Suppose that $X \sim B(m, \mu)$ and that m is large. By expanding in a Taylor series, show that the random variable

$$Z = \arcsin\left\{(X/m)^{1/2}\right\}$$

has approximate first two moments:

$$\begin{aligned} E(Z) &\approx \arcsin(\mu^{1/2}) - \frac{1-2\mu}{8m\sqrt{\mu(1-\mu)}} \\ \mathrm{Var}(Z) &\approx (4m)^{-1} \end{aligned}$$

[McCullagh and Nelder (1989,4.8)]

2. Suppose that $X \sim B(m, \mu)$, and let $\lambda = \mathrm{logit}(\mu)$.

(a) Show that $m - X$ also has the binomial distribution and that the induced parameter is $\lambda' = -\lambda$.

(b) Consider

$$\tilde{\lambda} = \log\left(\frac{X + c_1}{m - X + c_2}\right)$$

as an estimator of λ. Show that, in order to achieve consistency under the transformation $X \to m - X$, we must have $c_1 = c_2$. (Consistency in the sense that the corresponding estimate of λ based on $m - X$ is the same.)

[McCullagh and Nelder (1989,4.14)]

3. Using the notation of the previous exercise, write

$$X = m\mu + \sqrt{m\mu(1-\mu)}Z\,,$$

where $Z = O_p(1)$ for large m.

(a) Show that

$$E\left\{\log\frac{X+c}{m\mu}\right\} = \frac{c}{m\mu} - \frac{1-\mu}{2m\mu} + O(m^{-3/2})\,.$$

(b) Find the corresponding expansion for

$$E\left\{\log \frac{m - X + c}{m(1-\mu)}\right\}.$$

(c) Hence, if $c_1 = c_2 = c$, deduce that

$$E(\tilde{\lambda}) = \lambda + \frac{(1-2\mu)(c-\frac{1}{2})}{m\mu(1-\mu)} + O(m^{-3/2}).$$

[McCullagh and Nelder (1989,4.15), Cox (1970,§3.2)]

4. Derive the score test for testing equality of two binomial probabilities based on independent binomial counts. Show that the score test is equivalent to the Pearson chisquared test.

5. Consider the logistic regression model for the space shuttle data described in class.

   (a) Show that $(T_1, T_2) = (\sum Y_i, \sum x_i Y_i)$ is sufficient for (β_0, β_1) and that the conditional distribution of T_2 given T_1 does not depend on β_0.

   (b) Write an R program to obtain a Monte Carlo approximation to the P-value for the conditional test of $H_o : \beta_1 = 0$. Give an error bound on your approximation and explain how you got it. Construct a plot of the P-value and 2 standard error bounds against iteration number.

   (c) Write an R program to obtain the exact P-value for the conditional test of $H_o : \beta_1 = 0$.

   Tabulate the P-values obtained using the Wald statistic, the signed likelihood root, MC approximation and exact methods.

6. The following data come from a dose-response experiment on beetles, where ldose = log(dose), nbeetle = number of beetles exposed, and nkilled = number killed at the given dose level.

ldose	1.6907	1.7242	1.7552	1.7842	1.8113	1.8369	1.8610	1.8839
nkilled	6	13	18	28	52	53	61	60
nbeetle	59	60	62	56	63	59	62	60

(a) Fit binomial GLM's to this data using probit, logit and cumulative log-log links. Tabulate the deviance values from the three model fits.

(b) Construct a plot of the fitted tolerance densities implied by the three choices of link.

INSTRUCTIONS: Call your programs "yourloginname1-5b.R" and 1-5c.R, etc. Email them to me as attachments that I can click on and save directly. The files should be plain text. Do not send me an MS Word file. You should write the R code in a form that I can source and see the output without having to edit your files. Test your programs on our system before you send them to me.