Stat 513
Emerson, Winter 04
Homework #2
January 15, 2004

**Supplemental written problems due Friday, January 23, 2004 at the beginning of class.**

1. Let $Y_i$, $i = 1, \ldots, n$ be independent Bernoulli random variables with $Y_i \sim \mathcal{B}(1, p_i)$ where

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$$

and $x_i$, $i = 1, \ldots, n$ are known covariates. The file 513hw2data.txt contains 100 simulated values for $Y$ and $x$. Find maximum likelihood estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ using a Newton-Raphson algorithm in the statistical program of your choice (S+, R, and Excel come to my mind), and starting with initial estimates of $\hat{\beta}_{0(0)} = \text{logit}(\overline{Y})$ and $\hat{\beta}_{1(0)} = 0$. Provide the value of the loglikelihood, score function, Fisher's information matrix, and updated estimates at each iteration. Iterate until the absolute difference between successive estimates is no more than 0.000001 for either parameter.

2. Let $Y_i$, $i = 1, \ldots, n$ be independent Normal random variables with $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, with $\mu_i = \beta_0 + \beta_1 x_i$, where $x_i$, $i = 1, \ldots, n$ are known covariates.

   a. Find formulas for the maximum likelihood estimates and the covariance matrix of those estimates. What is the distribution of the estimates? Are they unbiased? Consistent? Efficient?

   b. Suppose now that it is merely known that $E[Y_i] = \mu_i = \beta_0 + \beta_1 x_i$ and $Var(Y_i) = \sigma^2$, but that the distribution of the $Y_i$s is otherwise unknown. Discuss the issues in finding the asymptotic distribution of $\hat{\vec{\beta}}$.

3. Let $T_i$, $i = 1, \ldots, n$ be independent Weibull random variables measuring time (in years) until failure with $T_i \sim Weib(p, \lambda)$ where $Pr(T_i > t) = \exp\{-(\lambda t)^p\}$ is the survivor function for $t > 0$, $p > 0$, and $\lambda > 0$. Let $C_i$, $i = 1, \ldots, n$ be independent (of each other and the $T_i$s) censoring times having distribution $C_i \sim \mathcal{U}(a, b)$ with $a < b$ known constants. Define $Y_i = \min(T_i, C_i)$ as the possibly censored observation time for the $i$th individual and $\delta_i = \mathbf{1}_{[Y_i = T_i]}$ as the indicator that $Y_i$ corresponds to an observed failure. Derive formulas for the maximum likelihood estimate of $\vec{\theta} = (p, \lambda)$, and find its asymptotic distribution. Also find the MLE of the probability of a subject surviving 5 years, and find its asymptotic distribution.

4. Let $\hat{\theta}$ be the maximum likelihood estimate of $\theta$ derived from i.i.d. random variables $X_1, \ldots, X_n$ from a statistical estimation problem satisfying the regularity conditions. Let $I_1(\theta)$ be the contribution to Fisher's information from a single observation. Thus we have asymptotic distribution

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d \mathcal{N}\left(0, I_1^{-1}(\theta)\right),$$

and our MLE has an asymptotic distribution that achieves the Cramér-Rao lower bound for variance of unbiased estimators of $\theta$. Show that the MLE for $g(\theta)$ has an asymptotic distribution which achieves the Cramér-Rao lower bound for variance of unbiased estimators of $g(\theta)$.

5. Suppose we are interested in determining the factors which affect the number of offspring in a herd which are sired by a single male when more than one male is present. Using fish, we perform an experiment in which two two males are placed with a female in varying conditions. We might, for instance, be interested in whether order of mating, size of males, age of males, etc. affects the fertilization of eggs. In order to easily distinguish which male sired which fry, we use our knowledge of the inheritability of the color of the fishs. Suppose that it is known that fish color follows simple Mendelian genetics in which black is dominant over red. That is, each fish possesses two genes for color. If both of those genes are for black or if the fish possesses one black and one red gene, the

fish is black. Only if both genes are red will the fish be red. Since each parent gives one of its genes (at random) to its offspring, the offspring from the breeding of a homozygous (both genes the same) black male and a homozygous red female will always be black (since the offspring will have one black gene from the male and one red gene from the female). If we breed two homozygous red fishs, all the offspring will be red. However, if we breed a heterozygous black male (with one black gene and one red gene) with a homozygous red female, we expect half the offspring to be black (when the male gave a black gene to the offspring) and half the offspring to be red (when the male gave a red gene to the offspring).

a. Suppose we place a homozygous black male and a homozygous red male with a homozygous red female. Let $N$ be the number of fry which are produced, let $X$ count the number of fry sired by the black male, and let $Y$ count the number of black fry. In this case, $X = Y$. Using the easily observed value of $X$, find an estimate for the probability $p$ that the black male sires a fry. Derive the asymptotic (as $N$ gets large) distribution for your estimate.

b. Now suppose the black male is heterozygous. In this case, all the black fry will have been sired by the black male, but some of the red fry may also have been sired by the black male. In fact, conditional upon the value of $X$, we expect $Y|X \sim \mathcal{B}(X, 0.5)$. Since $X$ is not readily observable in this case, find an estimate for the probability $p$ that the black male sires a fry using only the information in $Y$. Derive the asymptotic distribution for your estimate.

c. Find the asymptotic relative efficiency of the estimate derived in (a) compared to the estimate derived in (b). Interpret your results with regard to the importance of knowing the genotype of the males.