BIOS760 Final Exam, Fall 2006

1. Suppose that $\epsilon_1, ..., \epsilon_n$ are i.i.d from $N(0, \sigma^2)$. Let $Y_i = \mu + \epsilon_i/\sqrt{i}$.

   (a) (5 points) What are the complete sufficient statistics for parameters (μ, σ^2)?

   (b) (5 points) Find the UMVU estimator for μ.

   (c) (5 points) Does the UMVU estimator attain the Crame-Rao lower bound? Justify.

2. Suppose that $\epsilon_1, ..., \epsilon_n$ are i.i.d from a distribution with mean zero and variance σ^2. Let $Y_i = \mu + \epsilon_i/\sqrt{i}$.

   (a) (5 points) Clearly, for any non-negative constants $a_1, ..., a_n$ satisfying $\sum_{i=1}^n a_i = 1$, $\sum_{i=1}^n a_i Y_i$ is an unbiased estimator for μ. From this class of estimators, find the one which has the smallest variance, denoted as $\bar{Y}_n^*$.

   (b) (10 points) Derive and justify the asymptotic distribution of $\bar{Y}_n^*$ (after proper normalization).

3. Suppose that $Y_1, ..., Y_n$ are i.i.d from $N(\mu, \sigma^2)$ where $\mu \geq 1$ and $\sigma > 0$ are two parameters. One important quantity measuring the dispersion of the distribution is called the coefficient of variation (CV), defined as σ/μ.

   (a) (10 points) Suppose that the true mean is larger than 1. What are the efficient influence function and information bound for CV? Hint: the efficient influence function and information bound are defined based on only one observation.

   (b) (10 points) Clearly, we can estimate CV by

   $$\widehat{CV} = \frac{\sqrt{n^{-1}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}{\bar{Y}_n}$$

   where $\bar{Y}_n$ is the sample mean. Find the asymptotic distribution of $\widehat{CV}$ (after proper normalization).

   (c) (10 points) Another procedure to estimate CV is: we first calculate the maximum likelihood estimates for μ and σ and denote them as $\hat{\mu}_n$ and $\hat{\sigma}_n$; then the estimate for CV is $\hat{\sigma}_n/\hat{\mu}_n$. Give the explicit expression for $(\hat{\mu}_n, \hat{\sigma}_n)$ and show that if the true mean is larger than 1, then $\hat{\sigma}_n/\hat{\mu}_n$ has the same asymptotic distribution as $\widehat{CV}$. Hint: in calculating the maximum likelihood estimates, pay attention to the condition $\mu \geq 1$ and $\sigma > 0$.

4. Let $Y_1, ..., Y_n$ denote the log(CD4 cell count) from n HIV patients. Suppose that $Y_1, ..., Y_n$ are i.i.d from a normal distribution with mean μ and variance 1. However, due to the limitation of techniques measuring the CD4 cell count, we can only observe those Y_i's such that $Y_i > y_0$, where y_0 is a constant denoting the threshold value. We thus introduce indicators $R_1, ..., R_n$ where $R_i = I(Y_i > y_0)$ so $R_i = 1$ indicates that Y_i is observable and vice versa. Note the observed statistics for patient i are R_iY_i and R_i.

   (a) (5 points) Write out the observed likelihood function in terms of Y's and R's.

   (b) (5 points) Show that the parameter μ is identifiable from the observed data. Hint: to verify the identifiability, you need only consider $n = 1$ case.

   (c) (5 points) Give the likelihood equation for calculating the maximum likelihood estimator for μ and describe the Newton-Raphson algorithm for solving this equation.

   (d) (15 points) Provide an EM algorithm for calculating the maximum likelihood estimator. Make the final expressions as simple as possible.

   (e) (5 points) What is the asymptotic distribution of the maximum likelihood estimator?

   (f) (5 points) Construct an asymptotic confidence interval for μ with coverage 95%.