BIOS760 Final Exam, Fall 2007

1. Suppose that we observe n independent pairs (Y_i, X_i), where

$$X_i = \frac{1}{2}\log i + U_i, \quad Y_i = \beta + e^{X_i}\epsilon_i$$

with $U_i, i = 1, ..., n$ i.i.d from Uniform(0,1), $\epsilon_i, i = 1, ..., n$ from $N(0, 1)$, and U_i independent of ϵ_i. By noting

$$Y_i e^{-X_i} = \beta e^{-X_i} + \epsilon_i,$$

we can obtain an alternative least-square estimator for β given as

$$\hat{\beta} = \frac{\sum_{i=1}^n Y_i e^{-2X_i}}{\sum_{i=1}^n e^{-2X_i}}.$$

   (a) (5 points) Write down the observed likelihood function for β.

   (b) (5 points) What is the Cramer-Rao lower bound for β?
   We want to derive the asymptotic distribution for $\hat{\beta}$ after a proper normalization. Particulary, we take the following steps.

   (c) (5 points) Derive the asymptotic distribution of $\sum_{i=1}^n i^{-1/2} e^{-U_i}\epsilon_i$, after a proper normalization;

   (d) (5 points) Show

$$\frac{\sum_{i=1}^n i^{-1} e^{-2U_i}}{(1 - e^{-2})/2 \sum_{i=1}^n i^{-1}} \rightarrow_p 1;$$

   (e) (5 points) Obtain the asymptotic distribution for $\hat{\beta}$ after a proper normalization.

2. Suppose $X_1, ..., X_n$ are i.i.d from density

$$f(x) = I(x \geq \theta)\exp\{-(x - \theta)\}.$$

   (a) (5 points) Find a complete and sufficient statistic for θ.

   (b) (5 points) Derive the UMVUE for θ and denote it as $\hat{\theta}_1$.

   (c) (5 points) Find the maximum likelihood estimate for θ and denote it as $\hat{\theta}_2$.

   (d) (5 points) Calculate the variance of $\hat{\theta}_2$ and show that $\hat{\theta}_2$ is consistent using the Chebyshev's inequality.

   (e) (5 points) Find some constant a_n such that $a_n(\hat{\theta}_2 - \theta_0)$ converges in distribution to a non-degenerate distribution. Identify the limiting distribution. Here, θ_0 is the true value for θ.

   (f) (5 points) Find some constant b_n such that $b_n(\hat{\theta}_1 - \theta_0)$ converges in distribution to a non-degenerate distribution. Identify the limiting distribution.

3. Let Z_1, ..., Z_n be the measurements of body mass index (BMI) and X_1,..., X_n be the daily fat intake from n independent adults. Assume

$$Z_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where ϵ_i follows distribution $N(0, \sigma^2)$ and is independent of X_i. Assume X_i has a known density $g(x)$. In a real study, we observe $X_1, ..., X_n$ but instead of Z_i, only the categories of BMI are observed: such categories are defined as

$$Y_i = \begin{cases} \text{normal}, & Z_i < 25, \\ \text{overweight}, & 25 \leq Z_i. \end{cases}$$

Therefore, the observed data are $(Y_1, X_1), ..., (Y_n, X_n)$. We may code $Y_i = 0$ for "normal" category and $Y_i = 1$ for "overweight" category.

(a) (5 points) Write down the observed likelihood function.

(b) (5 points) Write down the likelihood function of the complete data.

(c) (15 points) We want to use the EM algorithm to calculate the maximum likelihood estimates for β_0, β_1 and σ^2. Give the explicit calculations in both E-step and M-step. In the expressions, you may use $\Phi(x)$ and $\phi(x)$ to denote the respective cumulative distribution function and density function of standard normal distribution.