BIOS 760 Final 2008

1. (10 Points + 3 Extra Credit Points). Let $X_1, \ldots, X_n$ be an i.i.d. sample from density $f_\alpha(x) = I\{0 \leq x \leq 1\}(\alpha+1)x^\alpha$, where $-1 < \alpha < \infty$. Our goal is to develop and study estimators for α.

   (a) We first obtain various moments for X_1 and $\log X_1$. Define functions

   $$g_0(u) = \frac{1}{u} - 1 \quad \text{and} \quad g_1(u) = \frac{u}{1-u} - 1.$$

      i. (2 Points) Let $m_0(\alpha) = -E(\log X_1)$ and calculate the form of the function m_0. Show that $g_0(m_0(\alpha)) = \alpha$ and that $\text{var}(\log X_1) = (\alpha+1)^{-2}$.

      ii. (2 Points) Let $m_1(\alpha) = E(X_1)$ and calculate the form of the function m_1. Show that $g_1(m_1(\alpha)) = \alpha$ and that $\text{var}(X_1) = (\alpha+1)(\alpha+2)^{-2}(\alpha+3)^{-1}$.

   (b) (2 Points) We now derive consistent estimators. Let $M_{0n} = -n^{-1}\sum_{i=1}^n \log X_i$ and $M_{1n} = n^{-1}\sum_{i=1}^n X_i$. Show that $g_0(M_{0n}) \to_p \alpha$ and $g_1(M_{1n}) \to_p \alpha$.

   (c) We now establish asymptotic normality.

      i. (3 Points) Show that $\sqrt{n}(g_0(M_{0n}) - \alpha) \to_d N(0, \sigma_0^2(\alpha))$, where $\sigma_0^2(\alpha) = (\alpha+1)^2$.

      ii. (Extra Credit: 2 Points) Show that $\sqrt{n}(g_1(M_{1n}) - \alpha) \to_d N(0, \sigma_1^2(\alpha))$, where $\sigma_1^2(\alpha) = (\alpha+1)(\alpha+2)^2(\alpha+3)^{-1}$.

   (d) Let $r(\alpha) = \sigma_1^2(\alpha)/\sigma_0^2(\alpha)$.

      i. (Extra Credit: 1 Point) Show that $r(\alpha) > 1$ for all $-1 < \alpha < \infty$ and that $\lim_{\alpha \to -1} r(\alpha) = \infty$, where the limit is taken from the right.

      ii. (1 Point) Interpret in words what the above statement means for the relative performance of $g_0(M_{0n})$ and $g_1(M_{1n})$ for estimating α.

2. (10 Points) Continuing with the above setting, let $T_n = g_0(M_{0n})$. The goal of this problem is to show that T_n is a regular and efficient estimator of α.

   (a) (3 Points) Show that T_n is asymptotically linear with influence function $H_\alpha(X) = (\alpha+1)^2(\log X + (\alpha+1)^{-1})$.

   (b) (3 Points) Show that $H_\alpha(X)$ is the efficient influence function for estimating α.

   (c) (1 Point) Argue briefly why this now implies that T_n is regular and efficient.

   (d) (3 Points) Let $\tilde{\alpha}_n = g_1(M_{1n})$. Derive a one-step estimator based on $\tilde{\alpha}_n$ that is also regular and fully efficient.

3. (10 Points) Assume that $Y_1, \ldots, Y_n$ are an i.i.d. sample from density $h_\alpha(y) = I\{0 \le y \le 1\}\,(1/2 + f_\alpha(y)/2)$, where f_α is as defined in Problem 1. We are going to develop an EM algorithm for calculating the maximum likelihood estimate for α based on $Y_1, \ldots, Y_n$.

   (a) (2 Points) Let Δ be a Bernoulli random deviate with $\Pr(\Delta = 1) = 1/2$ and let Y given $\Delta = \delta$ have density $I\{0 \le y \le 1\} f_\alpha^{1-\delta}(y)$. Show that the marginal density of Y is $h_\alpha(y)$.

   (b) (2 Points) Show that $E(\Delta | Y = y, \alpha) = (1 + f_\alpha(y))^{-1}$.

   (c) (1 Points) Show that the full log-likelihood $\ell_n(\alpha)$ for the sample $(\Delta_1, Y_1), \ldots, (\Delta_n, Y_n)$ is $-n\log(2) + \sum_{i=1}^n (1 - \Delta_i)\,(\alpha \log(Y_i) + \log(1 + \alpha))$.

   (d) (2 Points) Show that

$$E\left[\ell_n(\alpha)|Y_1, \ldots, Y_n, \alpha^{(k)}\right] = -n\log(2) + \sum_{i=1}^n w_i(\alpha^{(k)})\left[\alpha \log(Y_i) + \log(\alpha + 1)\right],$$

   where $w_i(\alpha) = f_\alpha(Y_i)\left[1 + f_\alpha(Y_i)\right]^{-1}$.

   (e) (3 Points) Now maximize over α and prove that

$$\alpha^{(k+1)} = \frac{-\sum_{i=1}^n w_i(\alpha^{(k)})}{\sum_{i=1}^n w_i(\alpha^{(k)}) \log(Y_i)} - 1.$$

4. (2 Extra Credit Points) Define mathematically the following concepts:

   (a) (1 Extra Credit Point) Complete Statistic.

   (b) (1 Extra Credit Point) Martingale.