Statistics 582, Problem Set 1

Wellner; 1/7/2015

Reading:

- Chapter 6, Sections 4-6 and 10; Lehmann and Casella, TPE, pages 456-487 and 515-519.
- Course Notes, sections 4.5 and 4.6.
- Groeneboom's notes on the EM algorithm.

Due: Wednesday, January 14, 2015.

1. Lehmann and Casella, TPE, Problem 4.9, page 504.

2. Compare the explanation of the EM algorithm in Lehmann and Casella TPE, pages 458-459 with the explanation given in Groeneboom's notes, pages 1 - 3 and 10 - 12. Correct the expressions given in (4.24) of TPE, page 459.

3. Lehmann and Casella, TPE, Problem 4.16, page 506, modified as follows: First, it seems to me that ζ_i in the third line of the problem statement should be just ζ. (For alternative formulations involving different ζ_i's, see TPE section 3.6.)
   We observe independent Bernoulli variables $X_1, \ldots, X_n$ which depend on unobservable variables Z_i distributed independently as $N(\zeta, \sigma^2)$ where
   $$X_i = \begin{cases} 0, & \text{if } Z_i \le u_i, \\ 1, & \text{if } Z_i > u_i. \end{cases}$$
   Assuming that $u_1, \ldots, u_n$ are known, we are interested in obtaining maximum likelihood estimates of ζ and σ^2 based on $X_1, \ldots, X_n$.
   (aa) Show that if $u_1 = u_2 = \cdots = u_n \equiv u$ and $\theta = (\zeta, \sigma)$, then the distribution P_θ of X_1 is *not identifiable*.
   (a) Show that the likelihood function is $\prod_{i=1}^n p_i^{X_i}(1-p_i)^{1-X_i}$ where $p_i = P(Z_i > u_i) = \Phi((\zeta - u_i)/\sigma)$, $i = 1, \ldots, n$. You will need to make further appropriate changes in Lehmann and Casella parts (c)-(e) as well.

4. Suppose that $X \sim F_\theta =$ exponential(θ) with density $f_\theta(x) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$ and $Y \sim G_\eta$ independent of X with densities $\{g_\eta : \eta \in R^+\}$, a regular parametric model on $(0, \infty)$. In (optional) problem xx of problem set 8, Statistics 581, we considered the following three scenarios for observation of X or functions of X:
   (a) Uncensored: we observe X and Y.
   (b) Right-censored: we observe $T(X,Y) = (X \wedge Y, 1\{X \le Y\}) \equiv (\min\{X,Y\}, 1\{X \le Y\}) \equiv (Z, \Delta)$.
   (c) Interval-censored (case 1): we observe $S(X,Y) = (Y, 1\{X \le Y\}) \equiv (Y, \Delta)$.
   In that problem, in each of the three scenarios (a), (b), (c), we computed: (i) The joint density of (X,Y) and joint distributions of $T(X,Y)$ and $S(X,Y)$.
   (ii) The scores for θ and η. (Let $(\partial/\partial\eta)\log g_\eta(y) \equiv a(y)$ with $a \in L_2^0(G_\eta)$.)
   (iii) And we compared $I_{X,Y}(\theta)$, $I_{T(X,Y)}(\theta)$, and $I_{S(X,Y)}(\theta)$.
   Here we will additionally assume that $Y \sim$ exponential(η) (independent of X). In scenarios (b) and (c), find EM algorithms for computation of the MLE's $\hat{\theta}_n$ of θ based on $T_1, \ldots, T_n$ and $S_1, \ldots, S_n$ i.i.d. as $T(X,Y)$ and $S(X,Y)$ respectively. How would you estimate the variance of $\hat{\theta}$ in each case?

5. **Optional bonus problem 1:** (Profile likelihood) [For nice plots to accompany this exercise, see pages 41 - 43 of Cox, D. R. and Oakes, D. (1984); *Analysis of Survival Data*, Chapman and Hall.] Consider the Weibull family of example 3.2.5 (581 Course Notes) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^{+2}$ given by the (Lebesgue) densities
$$p_\theta(x) = \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1}\exp(-(\frac{x}{\alpha})^\beta)1_{[0,\infty)}(x)$$
where $\theta \equiv (\alpha, \beta) \in (0,\infty) \times (0,\infty) \subset \mathbb{R}^2$.
(a) For a sample of n observations from p_θ, we know that, for each fixed value of β the value of α which maximizes the likelihood as a function of α is
$$\widehat{\alpha}(\beta) = \{\frac{1}{n}\sum_{i=1}^{n} X_i^\beta\}^{1/\beta}\,.$$
Use this to compute the *profile likelihood* $l_{\text{profile}}(\beta) = l_{\text{profile}}(\beta|\underline{X})$ defined by
$$l_{\text{profile}}(\beta) = l(\widehat{\alpha}(\beta), \beta) = l(\widehat{\alpha}(\beta), \beta|\underline{X})\,.$$
(b) Use what we know from Statistics 581 problem 9.2 to show that the profile likelihood is strictly concave and hence has a unique maximum. Show that maximizing the profile likelihood as a function of β yields the maximum likelihood estimate: i.e. that $(\widehat{\alpha}, \widehat{\beta}) = (\widehat{\alpha}(\widehat{\beta}_{\text{profile}}), \widehat{\beta}_{\text{profile}})$.

6. **Optional bonus problem 2.** This is a continuation of problem 5 above.
(a) What is the relationship of the score function for β from the profile likelihood, $\dot{l}_{\beta,\text{profile}}$ to the (efficient) score for β from the full likelihood? Prove or disprove my claim: the profile score for β (based on n observations) is asymptotically equivalent to the sum of efficient scores for β over the sample in the sense that their difference divided by $\sqrt{n}$ converges to 0 in probability.
(b) What is the relationship of the observed information from the profile likelihood $-\ddot{l}_{\beta\beta,\text{profile}}$ to information quantities from the full likelihood?

7. **Optional bonus problem 3.** Lehmann and Casella, problem 3.15, page 502: note that part (a) is false in the modern understanding of "log-concave" since all uniform distributions $f(x) = (b-a)^{-1}1_{[a,b]}(x)$ are log-concave, but the mode is not unique. (See Dharmadhikari and Jogdeo (1988), *Unimodality, Convexity, and Applications*, Section 1.4 and page 23.)