Statistics 581, Problem Set 10

Wellner; 12/04/2002

Reading: Chapter 4, Sections 1-4 and 6;
Ferguson, ACLST, Chapter 20, pages 133-139; Chapter 22, pages 144-150;
Lehmann and Casella, Chapter 6, especially section 6.5, pages 461-468.
Due: Wednesday, December 11, 2002.

1. A. Ferguson, ACLST, page 139, problem 3.
   B. What if Ferguson's density $f(x|\theta)$ with $\theta \in (0,1)$ is replaced by

$$f(x|\gamma,\eta) = \{(1-\gamma)e^{-x} + \gamma\eta^2 x \exp(-\eta x)\}1_{[0,\infty)}(x)$$

   with $\gamma \in (0,1)$ and $\eta > 0$?

2. Ferguson, ACLST, page 149, problem 2 modified as follows:
   (a) Find the LR test statistic of the null hypothesis $H_0 : \mu = c\theta$ for any fixed number $c > 0$, and find the asymptotic distribution of the LR statistic under H_0.
   (b) Does the theory of our chapter 4 (or Ferguson's chapter 22) apply directly?
   (c) Does the local asymptotic power of your test depend on c?

3. Suppose that $X \sim F$ on $R^+ \equiv [0,\infty)$, $Y \sim G$ on R^+, and X and Y are independent random variables. Let $Z = \min\{X,Y\} = X \wedge Y$ and $\Delta = 1\{X \le Y\}$. (This is *right-censored data*: if we view X as a survival time, and Y as a censoring time, then $Z = X$ when $X \le Y$, but $Z = Y$ when $X > Y$.)
   (a) Find the joint distribution of (Z,Δ).
   (b) If $X \sim$ Exponential(λ) and $Y \sim$ Exponential(μ), show that Z and Δ are independent.
   [Hint: for (a), compute $P(Z \le z, \Delta = 1)$ and $P(Z \le z, \Delta = 0)$.]

4. (Right censoring – again.) Consider nonparametric maximum likelihood estimation of F in the censored data problem considered in section 4.6 but extend the argument to include ties as follows:
   A. When there are ties, let the distinct Z's be denoted by $T_1 < \ldots < T_k$. Let $m_1,\ldots,m_k$ and $n_1,\ldots,n_k$ be defined by $m_j \equiv \#$ of $Z_i\delta_i = T_j$, $n_j \equiv \#$ of $Z_i(1-\delta_i) = T_j$, and let $p_j \equiv \Delta F(T_j) \equiv F(T_j) - F(T_j-)$, $= 1,\ldots,k$, $p_{k+1} = 1 - F(T_k)$. Show that the likelihood (for F) is

$$L(F|\underline{Z},\underline{\delta}) = \prod_{i=1}^{k} p_i^{m_i}(\sum_{j=i+1}^{k+1} p_j)^{n_i}.$$

   B. By defining $a_i \equiv p_i/\sum_{j=i}^{k+1} p_j$ for $i = 1,\ldots,k$ and $a_{k+1} = 1$, and rewriting the likelihood in terms of the a_i's, show that the likelihood is maximized by

$$\hat{a}_i = m_i/\sum_{j=i}^{k}(m_j+n_j) = n\Delta\mathbb{H}_n^{uc}(T_i)/n(1-\mathbb{H}_n(T_i-)),$$

and hence that the nonparametric MLE of F is (again) the Kaplan - Meier estimator

$$1 - \hat{\mathbb{F}}_n(t) = \prod_{0 \le s \le t} (1 - \Delta \hat{\Lambda}_n(s)).$$

C. Compute $1 - \hat{\mathbb{F}}_n$ for the following data (length of time until complete remission in weeks for the "maintained group") from a study of the efficacy of chemotherapy for acute Myelogenous leukemia (AML):

$$9, 13, 13+, 18, 23, 28+, 31, 31, 34, 45+, 48, 161+;$$

here "$+$" indicates censoring ($\delta = 0$).

5. **Optional bonus problem 1.**
Ferguson, ACLST, page 150, problem 3. Does the theory in our chapter 4 (or Ferguson's chapter 22) apply directly?

6. **Optional bonus problem 2:** Suppose that X and Y are as in the preceding problem, but that we now observe (Y, Δ); this is called "interval censored" or "current-status" data.
(a) Find the joint distribution of (Y, Δ).
(b) Specialize the result in (a) when $X \sim Exp(\lambda)$ and $Y \sim Exp(\mu)$ as in (b) of problem 3.