Statistics 582, Problem Set 5

Wellner; 2/4/98

Reading: Chapter 5, sections 1-3; Ferguson, MS, Chapter 1, pages 1 - 43.
Due: Wednesday, February 11, 1998.
Reminder: Midterm Exam: Friday, February 13.

1. In class on 1/28 we showed that the nonparametric maximum likelihood estimator of F in the (right) censored data problem, possibly with ties, is the Kaplan-Meier (product limit) estimator $\widehat{\mathbb{F}}_n(t)$ given by

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

where $\widehat{\Lambda}_n(t)$ is the *Nelson-Aalen* estimator of

$$\Lambda(t) \equiv \Lambda_F(t) \equiv \int_0^t \frac{1}{1 - F_-} dF \,,$$

given by

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{1}{1 - \mathbb{H}_n(s-)} d\mathbb{H}_n^{uc}(s) \qquad 0 \le s \le t \,.$$

Here

$$\mathbb{H}_n^{uc}(t) = \frac{1}{n}\sum_{i=1}^n \delta_i 1_{[Z_i \le t]} \,, \qquad \mathbb{H}_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{[Z_i \le t]}$$

are the sub-empirical distribution function of the uncensored observations and the marginal empirical distribution of all the Z's uncensored or censored.
A. Compute $1 - \widehat{\mathbb{F}}_n$ 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):

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

here $+$ indicates censoring ($\delta = 0$).
B. In class on 1/30 I gave a heuristic derivation of

$$\sqrt{n}(\widehat{\mathbb{F}}_n(t) - F(t)) \Rightarrow (1 - F(t))B(C(t))$$

as a process uniformly in $t \in [0, \tau]$ for any $\tau < \tau_H$ (i.e. for any τ with $1 - H(\tau) = (1 - F(\tau))(1 - G(\tau)) > 0$, where B is a standard Brownian motion process and where

$$C(t) \equiv \int_0^t \frac{1}{(1 - H_-(s))^2} dH^{uc}(s) \,, \qquad 0 \le s \le t$$

Thus we have, for each fixed $t < \tau$,

$$\sqrt{n}(\widehat{\mathbb{F}}_n(t) - F(t)) \rightarrow_d N(0, (1 - F(t))^2 C(t))$$

Suggest an estimator of $C(t)$ and hence an estimator of $(1 - F(t))^2 C(t)$.

C. Show that your estimator of $(1 - F(t))^2 C(t)$ is consistent.

D. Use the estimator you suggest in B to obtain an approximate 90% confidence interval for $F(30)$ based for the data given in A.

2. Let $\Theta = \{1, 2\} = \mathcal{A}$ where $1 =$ a patient has tuberculosis, $2 =$ a patient does not have tuberculosis. Let X be the number of positive reactions to two different tuberculosis tests, so that $\mathcal{X} = \{0, 1, 2\}$, and suppose that X has the following distributions

x	0	1	2
$p_1(x)$	.04	.12	.84
$p_2(x)$	.64	.28	.08

If the losses are given by $L(1,1) = L(2,2) = 0$ $L(1,2) = 100$, $L(2,1) = 10$, and the prior $\lambda = (\lambda_1, \lambda_2) = (.3, .7)$, find the Bayes rule d_B and the minimax rule d_M. Plot the risk set and label the non-randomized decision rules.

3. Study for the midterm exam.