STAT 855. Lifetime Data Analysis (Spring 2022)

Homework 4 (due on March 28, 2022 by 8:00pm)

Problem 1. Consider that a study recruits n independent subjects from a population, the lifetime of which is r.v. $T \sim F(\cdot)$. Denote T's hazard function, cumulative hazard function, and survivor function by $\lambda(\cdot)$, $\Lambda(\cdot)$, and $S(\cdot)$, respectively. Suppose the study observation is subject to a right-censoring. That is, the available data are right-censored and presented as $\{(U_i, \delta_i) : i = 1, \ldots, n\}$ with $U_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$, where C_i is the censoring time associated with subject i. Assume that the censoring times C_i's are independent of the lifetimes T_i's.

(i) Let $N_i(t) = I(U_i \leq t, \delta_i = 1)$, $Y_i(t) = I(U_i \geq t)$ for $t \geq 0$, and $A_i(t) = \int_0^t Y_i(s)\lambda(s)ds$. Show that $\{M_i(t) = N_i(t) - A_i(t), t \geq 0\}$ is a zero-mean martingale wrt $\{\mathcal{H}_i(t), t \geq 0\}$ with $\mathcal{H}_i(t) = \sigma\{(N_i(s), Y_i(s)), 0 \leq s < t\}$. What is its quadratic predictable variation process $< M_i, M_i > (t)$, $t \geq 0$?

(ii) Present the Nelson-Aalen estimator for $\Lambda(\cdot)$, denoted by $\hat{\Lambda}_{NA}(\cdot)$, using the notation introduced above. Present $\sqrt{n}(\hat{\Lambda}_{NA}(t) - \Lambda(t))$ as a summation in the form of $\sum_{i=1}^n \int_0^t H_{ni}(s) dM_i(s)$ with $H_{ni}(\cdot)$ a predictable process.

(iii) Generate $n = 300$ iid samples T_i's from the Weibull distribution with the shape parameter $\alpha = 3$ and the scale parameter $\theta = 1$, and iid samples C_i's from the exponential distirbuton $NE(1)$ and then form a set of right-censored lifetimes. Evaluate the following estimators of T's survivor function and display them in a plot along with the ture survivor function:

(a) $S_{KM}(\cdot)$, the Kaplan-Meier estimator;

(b) $S_{FH}(\cdot)$, the Fleming-Harrington estimator;

(c) $S(t; \hat{\theta})$ with $S(t; \theta)$ the survivor function of the exponential distribution and $\hat{\theta}$ the MLE of θ with the right-censored data under the exponential distribution.

Problem 2. Consider that a study recruits independent subjects from a population, and randomly allocates them to a treatment arm and a placebo arm. Denote the lifetimes associated with the treatment arm and control arm by $T_1 \sim F_1(\cdot)$ and $T_0 \sim F_0(\cdot)$, respectively. Suppose the study observation is subject to an independent right-censoring. Generate $n_k = 200$ samples for groups $k = 1, 0$: $T_{1i} \sim F_1(\cdot) = NE(e)$ iid and $T_{0i} \sim F_0(\cdot) = lognormal(0, 1)$ iid; C_{ki} from $NE(1)$ independently for $k = 1, 0$. Form two sets of right-censored lifetimes.

(i) Evaluate the Fleming-Harrington estimator with each of the two sets of data. Display the two Fleming-Harrington estimates along with the two true survivor functions in a plot.

(ii) Introduce Z as the indicator for the treatment group, and assume T_k follow Cox proportional hazards model $\lambda_0(t) \exp(\beta Z)$ with $Z = k$, for $k = 1, 0$. Evaluate the MPLE of β and the Breslow estimator of $\Lambda_0(\cdot)$. Using them computes an estimate for the survivor function of each group. Add these two estimates to the plot obtained in (i). Comment.

(iii) Conduct a model checking regarding the assumption of $T_{0i} \sim NE(\theta)$ using the generated right-censored data for the placebo arm as follows.

(a) Obtain the MLE of θ and form an estimate for the placebo's survivor function based on the parametric model assumption.

(b) Obtain the Nelson-Aalen estimate for the cumulative hazard function, and then the Fleming-Harrington estimate for the survivor function.

(c) Adapting the idea of Lin, Wei and Ying (Biometrika, 1993), generate $L = 300$ approximate realizations of the Nelson-Aalen estimator and then form them into approximate realizations of the Fleming-Harrington estimator.

(d) Construct an approximate 95% confidence band based on the approximate realizations from (c).

(e) Present the estimates of the placebo's survivor function from (a) and (b) along with the 300 simulated approximate realizations from (c), and the approximate 95% CB from (d) in a plot with the true survivor function. Comment.