STAT 855. Lifetime Data Analysis (Spring 2022)

Homework 3 (due on Feb 25, 2022 by 8:00pm)

Problem 1. 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 a right-censoring. That is, the available data are right-censored and presented as $\{(U_{ki}, \delta_{ki}) : i = 1, \ldots, n_k\}$ with $U_{ki} = \min(T_{ki}, C_{ki})$ and $\delta_{ki} = I(T_{ki} \leq C_{ki})$ with $k = 1, 0$ for the two groups, respectively. Assume that the censoring times C_{ki}'s are independent of the lifetimes T_{ki}'s.

(i) Introduce Z as the indicator for the treatment group, and assume T_k follow Cox proportional hazards model $h_0(t)\exp(\beta Z)$ with $Z = k$, for $k = 1, 0$. Write down (a) the likelihood function of $h_0(\cdot)$ and β with the righ-censored data, and (b) the partial likelihood function of β and the partial score funciton.

(ii) (a) Generate $n_k = 100$ for $k = 1, 0$ iid samples (T_{ki}, C_{ki})'s with T_{ki} from the exponential distirbuton $NE(\exp(k))$ with mean $\exp(k)$ and C_{ki} from $NE(1)$ independently. Form two sets of right-censored lifetimes. (b) Using the generated data, give the partial score evaluated at $\beta = 0$ under the Cox model in (i), and the numerator of the log-rank test statistic (the Mantel test). (c) Evaluate the MPLE $\hat{\beta}$ of β with the right-censored data under the Cox model in (i), and its asymptotic variance estimate. (d) Implement the following testing procedures for comparing the two study groups:

(d1) the Wald-type test using $\hat{\beta}$;

(d2) the partial score test (or the log-rank test).

(e) Estimate the baseline cumulative hazard function, and in the same plot with the true survivor functions, display the two estimates of the survivor function for each of the two groups, the KM estimate and the estimate based on the Cox model.

(iii) Repeat (ii)(a,c,d2,e) but with T_{0i} generated from the lognormal distribution: $\log(T_{0i}) \sim N(0, 1)$. Comment.

(iv) Repeat (iii) $M = 200$ times, and summarize the 200 sets of MPLE estimates and their asymptotic variance estimates, and the test outcomes. Comment.

(v) Repeat (ii) (a,c,d2,e) but simulate each C_{ki} from the uniform distribution $U(3T_{ki}/4, 5T_{ki}/4)$ given each generated T_{ki}. Comment.

Problem 2. The appendix presents the dataset from a hematological study. The study objectives include (a). to estimate the survivor function of time to disease progression since the diagnosis; (b). to compare the survival rates to progression in females vs males and in white people vs others; (c) to examine the effects of potential covariates to the disease progression.

Apply the inference procedures studied in Parts II and III to analyze the data, aiming to achieve the study objectives. An ad hoc requirement: your analysis needs to use at least once each of the parametric, nonparametric, and semiparametric approaches.