STAT 855. Lifetime Data Analysis (Spring 2022)

Homework 2 (due on Feb 11 by 8:00pm)

Problem 1. Consider that a study recruits n independent subjects from a population, the lifetime of which is r.v. $T \sim F(\cdot)$. Let the lifetimes of the subjects in the study be T_i, $i = 1, \ldots, n$. 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 $\delta_i = I(T_i \le C_i)$, 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) (a) Give the empirical distribution $F_n(\cdot)$ with the iid observations $\{T_i : i = 1, \ldots, n\}$, and state its pointwise consistency and asymptotic normality. (b) Give the Kaplan-Meier estimator with the right-censored data.

(ii) (a) Taking $n = 20$, generate 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. (b) Evaluate the following estimators of T's survivor function and based on them to construct approximate 95% pointwise CIs for the survivor function:

(b1) $S_n(\cdot) = 1 - F_n(\cdot)$ with $F_n(\cdot)$ the empirical distribution;

(b2) $S_{KM}(\cdot)$, the Kaplan-Meier estimator;

(b3) $S(t; \hat{\alpha}, \hat{\theta})$ with $S(t; \alpha, \theta)$ the surivor function of the Weibull distribution $Weibull(\alpha, \theta)$ and $\hat{\alpha}, \hat{\theta}$ the MLE of α, θ with the right-censored data under the Weibull distribution assumption;

(b4) $S(t; \hat{\theta})$ 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.

(c) Display the four estimates, and comment. (d) Obtain an estimate for the median of the lifetime from each of the four survivor function estimates.

(iii) Repeat (a) and (b) of (ii) $M = 100$ times, and summarize the 100 sets of the evaluations of the four estimators by the sample means and the sample standard errors. Comment on the four estimators based on the summary statistics.

(iv) Recall that the true median of $Weibull(\alpha, \theta)$ is $\theta(\ln 2)^{1/\alpha}$. Repeat (d) in (ii) $M = 100$ times, and obtain an interval estimate for the median using the 5% and 95% percentiles of the 100 point median estimates associated with each of the four estimators. Tabulate the four interval estimates and comment on the simulation results.

(v) (a) Re-do (ii) with $n = 200$ and comment on the difference of the simulation outcomes with $n = 200$ and the ones with $n = 20$. (b) Conduct a model checking regarding each of the two model assumptions: $T \sim NE(\theta)$, or $T \sim Weibull(\alpha, \theta)$.

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 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) (a) Generate $n_k = 100$ for $k = 1, 0$ iid samples (T_{ki}, C_{ki})'s with T_{ki} from the exponential distirbuton $NE(\theta_k)$ with mean $\theta_k = \exp(k)$ and C_{ki} from $NE(1)$ independently. Form two sets of right-censored lifetimes. (b) Implement the following testing procedures for comparing the two study groups and comment on three testing procedures, using the significance level of 5%:

(b1) the Wald-type test using the MLE $\hat{\theta}_k$, $k = 0, 1$ derived under the exponential distn assumption.

(b2) the logrank (Mantel) test.

(b3) the Wilcoxon-Gehan test.

(c) Display the KM estimates together with the true survivor functions of the two groups in the same plot. Are your findings from (b) consistent with what is shown in the plot regarding the difference between the two distributions?

(ii) Repeat (a),(b), and (c) in (i) with T_{0i} generated from the lognormal distribution: $\log(T_{0i}) \sim N(0, 1)$, instead. Comment based on your findings from the simulation results.

(iii) Repeat (a), (b), and (c) in (i) with each C_{ki} generated from the uniform distribution $U(3T_{ki}/4, 5T_{ki}/4)$ given each generated T_{ki}, instead. Comment based on your findings from the simulation results.

(iv) List at least two general statements about the three testing proceures described by (b1), (b2) and (b3) in (i) that can be verified using the outcomes of (i), (ii), and (iii).