STAT 855. Lifetime Data Analysis (Spring 2022)

Homework 1 (due on Friday Jan 28 by 8:00pm PT)

Problem 1.1. Consider a lifetime $T \sim NE(\theta)$: $f(t;\theta) = \frac{1}{\theta}e^{-t/\theta}, t \geq 0$ with $\theta > 0$.

(i) Suppose a study recruits n independent subjects and the lifetime associated with subject i: $T_i \sim NE(\theta)$. Derive the MLE of θ with $\{T_i : i = 1, \ldots, n\}$, state its consistency and the asymptotical normality, and verify the required conditions for the asymptotics. Denote the MLE by $\hat{\theta}_A$.

(ii) Suppose the observation of the study in (i) 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. Derive the MLE of θ based on the right-censored lifetime data, and state its consistency and asymptotical normality. Denote this MLE by $\hat{\theta}_B$.

(iii) With the right-censored data in (ii), let $\hat{\theta}_C = \sum_{i=1}^{n} U_i/n$. If C_i are iid realizations of a r.v. $C \sim g(\cdot)$, what is the expectation of $\hat{\theta}_C$? When can be $\hat{\theta}_C$ an asymptotically unbiased estimator of θ?

Problem 1.2. Consider the model in Problem 1.1, and choose $\theta = 1$ and $n = 100$.

(i) Generate $t_i, i = 1, \ldots, n$ independently from $NE(1)$ as the observations from the population. Evaluate $\hat{\theta}_A$ given in Problem 1.1. Plot the following curves and comment on them:

- the true cumulative distribution function (cdf) of the population
- the emperical distribution with the first 10 observations, $\{t_1, \ldots, t_{10}\}$
- the emperical distribution with all the 100 observations, $\{t_1, \ldots, t_{100}\}$
- the estimated cdf of T using $\hat{\theta}_A$

(ii) Generate $c_i, i = 1, \ldots, n$ independently from the uniform distribution $U(0.5, 1.5)$ and obtain $u_i = \min(t_i, c_i)$ and $\delta_i = I(t_i \leq c_i)$ for $i = 1, \ldots, n$. Evaluate $\hat{\theta}_B$ and $\hat{\theta}_C$ with the generated right-censored lifetimes, as given in Problem 1.1.

(iii) Repeat (i)-(ii) $M = 200$ times, except for the plotting part in (i), plot histograms of the evaluations of the estimators $\hat{\theta}_A$, $\hat{\theta}_B$ and $\hat{\theta}_C$, and summarize the evaluations by their sample means and sample standard derivations. Comment on the three estimators based on the numerical results.

(iv) What do you envision the performance of $\hat{\theta}_B$ compared to $\hat{\theta}_A$, if c_i's are generated from (a) $U(0, 1)$, or (b) $U(2, 3)$? Why?

Problem 1.3. AIDS studies in late 1990s used HIV RNA copies by an assay to quantify HIV viral load. The assay had a lower detection limit, say, 1000 copies/ml. That is, any HIV RNA copies obtained by the assay were not reliable if they were below 1000 copies/ml. The appendix gives the set of HIV RNA copies collected from subjects at their study week 24 in an AIDS clinical trial with two treatment groups, where those HIV RNA copies originally lower than 1000 copies/ml are presented by NA.

There were three approaches used in analysis of such data. Approach 1 analyzed the data given in the appendix, using 1000 copies/ml as the true values for those presented as NA and were quantified by the assay as lower than 1000 copies/ml. Approach 2 replaced all those NA's by 500 copies/ml, a half of the detection limit, and analyzed the resulting data. Approach 3 viewed the observation of HIV RNA was subject to the left-censoring at the lower detection limit, and adapted the methods for right-censored lifetimes to analyze the data. (The observation of the negative HIV RNA could be then viewed as subject to a right-censoring.)

People believe that it is appropriate to assume a $\log_{10}$-transformed HIV RNA follows a normal distribution.

(i) Comment on Approaches 1-3.

(ii) Give 95% confidence intervals of the population mean of the $\log_{10}$-transformed HIV RNA copies for each of the two study treatments by the three approaches.

(iii) Conduct a hypothesis testing at the significance level of $\alpha = 0.05$ by Approach 3, to compare the two study treatments in the $\log_{10}$-transformed HIV RNA copies.