Eric Slud 10/23/2021

STAT 702 Sample Problems for In-Class Test.

The test will consist of 3 problems, each no longer than the ones given here.

(1). Suppose that the death and censoring variables for each subject in a randomly right-censored survival study have joint probability density

$$f_{X,C}(x,y) \;=\; \begin{cases} (21/2)\,(1+x)^{-5}\,(1+y)^{-4} & \text{if} \quad x < y \\ 14\,(1+x)^{-5}\,(1+y)^{-4} & \text{if} \quad x \ge y \end{cases}$$

(a) Find $S_X(t)$.

(b) Find $P(X < C)$.

(c) Find $f_1(t)$ and $S_T(t)$.

(d) Find the limit of the Kaplan-Meier survival function estimator for large n in a survival study based on right-censored data with underlying death and censoring joint density $f_{X,C}(x,y)$.

(2). Consider the following data derived from a cohort life table constructed from right-censored survival data $(T_i, \Delta_i, \quad i = 1, \ldots, 110)$. In this table, t_j are the sorted increasing distinct event (death or censoring) times, l_j the number at risk (alive and uncensored) at time t_j, and d_j the number of observed failures at time t_j.

j	t_j	l_j	d_j	$\sum_{k=1}^{j} d_k/l_k$	$\sum_{k=1}^{j} d_k/(l_k(l_k-d_k))$	$\prod_{j=1}^{k}(1-d_k/l_k)$
...						
52	1.737	59	0	0.488	0.077	0.612
53	1.743	58	1	0.506	0.079	0.601
54	1.747	57	0	0.506	0.079	0.601
55	1.749	56	1	0.524	0.081	0.590
56	1.791	55	1	0.542	0.083	0.580
57	1.796	54	1	0.560	0.085	0.569
58	1.798	53	1	0.579	0.087	0.558
59	1.809	52	0	0.579	0.087	0.558
...						

(a) Find $\hat{H}(1.78)$ (Nelson-Aalen Estimator) and $\hat{S}^{KM}(1.78)$.

(b) Test the null hypothesis H_0 that the median survival time is 1.80 (i.e., that $S_X(1.80) = 0.5$) at signficance level 0.05, against the two-sided alternative.

(c) Find a 95% two-sided confidence interval for $S_X(1.75)$ based on transforming the one for $\hat{H}(1.75)$ or for $-\log(\hat{S}^{KM}(1.75))$, and compare it with the symnmetric CI around $\hat{S}^{KM}(1.75)$.

(d) Estimate the standard error of $\hat{H}(1.8) - \hat{H}(1.737)$.

(3)*. Based on large samples of *iid* right-censored survival data $\{(T_i, \Delta_i)\}_{i=1}^n$, **without** the assumption of independence of death and censoring,

(a) Is it possible to estimate $F_1(t) = P(X_i \le \min(t, C_i))$ consistently ? If so, what estimator has this property ?

(b) What information if any can be obtained from identifiable functions of the right-censored survival data about the conditional density $f_{X|X \ge C, C=y}(x|y)$? What does that suggest to you about the possibility of identifying $S_X(t)$ from right-censored survival data, even if $f_C(y)$ is known in advance ?

(4). Suppose right-censored survival data, with independent death and censoring variables, are given in the form

t_j	11	13	15	16	18	21	24	27
d_j	1	0	1	1	1	0	0	1
c_j	0	1	0	0	0	1	1	0

(a) Find the likelihood and MLE for the parameter λ if the failure-time random variables X_i are Weibull$(2, \lambda)$ with density $2\lambda x\, e^{-2\lambda\, x^2}$.

(b) Find the likelihood (but do not solve for the MLE) with these same data if you are told that the data were left-truncated at time 10, i.e. that the *iid* data sample omits any data for subjects with $X_i < 10$.

See next page for a list of the most important topics to study for the test.

List of Most Important Topics for STAT 702 in-class Test

(I) Format of right-censored survival data, and definitions and relationships of functons S_X, S_T, F_1, f_X, h_X, S_C, f_C , in general and under the assumption of independent random right-censoring.

(II) Functions identifiable (without assumption of independent death and censoring) from right-censored survival data

(III) Definitions of Cohort Life Tables

(IV) Definitions of Right-Censoring, Left Truncation, Double Censoring

(V) Kaplan-Meier Estimator of S_X, Definition and Characterization as "Product Limit" estimator, as nonparametric Maximum Likelihood estimator, and as self-constistent estimator

(VI) Parametric Model Likelihoods for Right-censored survival data with and without left-truncation under assumption of independent death and censoring variables

(VII) Greenwood Formula and variance Estimates for Kaplan-Meier Estimator, Confidence Intervals for $S_X(t)$ and $H_X(t)$, Large-sample limit (under independent death and censoring) for Kaplan-Maier Estimator and Greenwood formula

(VIII) Martingale property of random functions $N(t) - \int_0^t Y(x) h_X(x)\, dx$, $\hat{H} - H_X$ and $\hat{S}(t \wedge t_r)/S(t \wedge t_r) - 1$, and large-sample CLT and random-function distributional limits, with particular attention to independent increments property. Definition of Confidence Bands