

Sample for Stat798S Test

The in-class test on Monday will consist of two or three problems something like these in level and difficulty and coverage. You may use calculators and a double-sided notebook sheet of notes, but no other books or notes. Wherever you are asked to ‘explain’ or show something, you may give either an analytical proof or a clear and detailed intuitive explanation.

(1). *Properties of Hazard functions.* Suppose that a survival r.v. T has density a mixture of two exponentials

$$f_T(t) = p\lambda e^{-\lambda t} + (1-p)\mu e^{-\mu t} \quad , \quad t \geq 0$$

where $0 < p < 1$, $0 < \lambda < \mu < \infty$ are parameters. Find the hazard (intensity) function for T and explain either analytically or intuitively why it is decreasing as a function of its positive argument t .

(2). *Dependent Times, Hazards & Cum. Incidence.* Suppose that two failure-time random variables are dependent, with joint density

$$f_{X,C}(s,t) = \frac{30}{7} \exp(-2s - t - 2tI_{[t \geq s]}) \quad , \quad s, t > 0$$

(a) Find the cumulative incidence function for X failures, i.e., $F_1(t) = P(X \leq \min(C, t))$.

(b) If a large sample of right-censored survival data $(\min(X_i, C_i), I_{[X_i \leq C_i]})$, $i = 1, \dots, n$, were observed (where (X_i, C_i) are *iid* r.v. pairs with joint density $f_{X,C}$), then what nonrandom function of t will the Nelson-Aalen estimator $\hat{H}_X(t)$ be close to with high probability?

(3). *Behavior of Estimators in Uncensored Settings.* Give a formula for the Nelson-Aalen estimator at the k 'th smallest observation $X_{(k)}$ (the k 'th order-statistic) in a sample of n positive *uncensored* random continuously distributed survival times X_i , $i = 1, \dots, n$. Explain why the estimator at the k 'th order statistic for $k = [n(1-r)]$ (where $[z] =$ largest integer $\leq z$) will be close to $\ln(1/r)$, for each fixed $r \in (0, 1)$.

(4). *Kernel estimators.* (a) Give a formula for the kernel-smoothed estimator of cumulative hazard based on a right-censored survival data sample (T_i, Δ_i) , $i = 1, \dots, n$ using a uniform kernel with bandwidth b (and no edge correction for small or large t).

(b) Use the formula in (a) to give the cumulative hazard function on the interval $[4, 5]$ using $b = 1$ based on the right-censored data-sample $1.7, 2.2, 3.5, 3.8+, 4.6, 5.5, 5.8+, 8.8$ where ‘+’ denotes a censoring time (and the absence of ‘+’ a failure time).

(c) Explain, in the general case with fixed n , what the limiting value of the estimator in (a) is when the bandwidth parameter $b \searrow 0$.

(5). Test Statistics & Confidence Intervals. The following is an excerpt from an Splus data output including Nelson-Aalen cumulative hazard estimators and Standard Errors at three ordered times, for each of two treatment Groups. Three columns are given, corresponding to the first event time, last death-time, and one intermediate time ($t = 10$). The notations t_{1i} and $t_{(j)}$ in the last two lines respectively denote ordered **Group 1** event times and combined-group death times. **Assume that the underlying survival distributions are continuous, that there are no tied death-times, that death and censoring are independent in each group, and that the censoring distribution is the same within each group.**

t	0.23	...	10	...	96.2
$N_1(t)$	0	...	70	...	150
$N_2(t)$	1	...	130	...	180
$Y_1(t)$	200	...	120	...	2
$Y_2(t)$	300	...	150	...	1
$\hat{H}_1(t)$	0	...	0.70	...	5.0
$\hat{H}_2(t)$	*	...	0.64	...	5.5
$SE_{\hat{H}_1}(t)$	0	...	0.05	...	*
$SE_{\hat{H}_2}(t)$	*	...	0.04	...	*
$\sum_{i=1}^{200} \min(t_{1i}, t)$	0	...	259.2	...	556
$\sum_{j:t_{(j)} \leq t} \frac{Y_1(t_{(j)})}{Y(t_{(j)})}$	0	...	64.7	...	135.0

(a) Give the approximate values of the Kaplan-Meier estimators *and their standard errors* for the two groups at time $t = 10$.

(b) Give a 95% confidence interval for the difference between the cumulative hazards at time 10 for the two Groups.

(c) Give an approximate value for the logrank test statistic of equality between the marginal survival functions for the two Groups.

(d) Test at the 5% significance level the hypothesis that survival in Group 1 is Expon(0.25) against alternatives with uniformly higher hazards.

(e) Estimate the cumulative excess hazards in Group 1 at time $t = 10$ with respect to the Expon(0.25) reference distribution.