Sample Test Problems, Stat 798L Spring 2008

The problems from the 2005 sample test are still appropriate except for #4 since we have not covered anything about kernel (smoothed) hazard intensitiy estimators this Spring. Here are some others.

1. Suppose that a subject population consists of two population subgroups, respectively correcponding to 40% and 60% of the total. In the first group, death X and censoring C times are independent exponential random variables with respective hazards 3 and 2; in the second group, X and C are independent exponential random variables with respective hazards 2 and 4. Assume that a random sample of subjects is drawn and their right-censored survival data $T = \min(X, C)$ and $\Delta = I_{[X \le C]}$ are observed, but that one cannot observe which of the two population subgroups each patient comes from. *Find the net and crude hazard rates for the mortality random variable* X.

2. Consider the following life-table constructed from a right-censored survival dataset of size 18:

time	0.1	0.3	0.6	1.2	1.8	3.0	3.3	4.2	5
deaths	1	1	1	1	1	0	1	1	0
at-risk	18	17	16	15	14	13	12	11	10

and that all patient observations are administratively censored at $t = 5.0$. (Assume that right-censorship is independent of survival.)

(a). Find the Kaplan-Meier estimator and the Greenwood-formula standard error for $S_X(5)$.

(b). Suppose that observations with $T < 0.5$ are dropped because they are deemed unreliable, i.e., the observations at times 0.1, 0.3 are simply omitted, so that only 16 observations are used. Write the Likelihood for this 16-element left-truncated and right-censored dataset based on hazard-parameter λ under the assumption that the failure time random variables X_i are $Weibull(\lambda, 2)$ distributed.

(c) Re-calculate the likelihood in (b) if the observations from the two subjects with event-times $0.1, 0.3$ are not left-truncated but rather left-censored at time $t = 0.5$. (That is, the dataset now has 18 elements, and the observations $0.1 and 0.3$ now become left-censored observations $\mathbf{.5}-$).

3. Give an estimator and the delta-method based standard error for the median survival time in the dataset of problem 2 based on the $Weibull(\lambda, 2)$ survival distribution.

4. The **R** dataset **rats** contains *time, status, rx* data for 150 rats, all of which were injected with a carcinogen and 50 (randomly chosen to be the $rx = 1$ group) were given a drug. The observed times are waiting times until either a tumor developed or the followup on the rat was right-censored. Selected output from a survival analysis of these data is as follows:

	N	Observed	Expected	S^(80)	SE(S^(80))
rx=0	100	19	27.5	.8733	0.0343
rx=1	50	21	12.5	.8102	0.0572

The $\hat{S}$ values are Kaplan-Meier estimators, and the standard errors are based on the Greenwood formula. The 'expected' number of failures in each treatment group is based on the formula $\int (Y_z/Y)\, dN$ associated with the centering for the logrank statistic.

Perform two different (approximate, large-sample-based) hypothesis tests at significance level $\alpha = .05$ of difference between survival for the rx=1 and rx=0 groups based on these data. Explain clearly what each of the two null hypotheses is and (roughly) how strongly, if at all, you can reject it. Use the facts that $z_{.025} = 1.96$, $\chi^2_{1,.05} = 3.84$, $\chi^2_{1,.02} = 5.41$. Would your answer to either test change if you were given the additional information that $\int (Y_1(t)\, Y_0(t)/Y^2(t))\, dN(t) = 8.491$?

Other problem types that might be asked: Wald and Likelihood Ratio tests of significance of coefficients from survival regression (based on *parametric or Cox-model* **`survreg` outputs).**