B5440 – Exercise 3

Exercises

1. Construct an example of a pair of random variables (T, U) representing failure times and censoring times that are dependent but the "independent censoring" assumption holds.

2. (Fleming and Harrington Exercise 1.9) Suppose T, U have joint distribution given by $H(t, s) = P(T > t, U > s) = \exp(-\lambda t - \mu s - \theta ts)$ where $0 \leq \theta \leq \lambda\mu$.

a. Find the marginal survival functions and corresponding hazard functions for T and U.
b. Calculate
$$\alpha_1^{\#}(t) = \frac{-\frac{\partial}{\partial u}P(T \geq u, U \geq t)|_{u=t}}{P(T \geq t, U \geq t)}$$
and
$$\alpha_2^{\#}(t) = \frac{-\frac{\partial}{\partial u}P(T \geq t, U \geq u)|_{u=t}}{P(T \geq t, U \geq t)}.$$
c. Suppose $\lambda = 1, \mu = 2$ and $\theta = 2$. Plot and compare i) the marginal survival function for T to $H_1^{\#}(t) = \exp(-\int_0^t \alpha_1^{\#}(s)\,ds)$ and ii) the marginal survival function for U to $H_2^{\#}(t) = \exp(-\int_0^t \alpha_2^{\#}(s)\,ds)$. What do you notice and why is this comparison meaningful?
d. Suppose one erroneously assumes T and U are independent, and takes their joint distribution to be
$$H^{\#}(t, s) = H_1^{\#}(t)H_2^{\#}(s).$$
Find $H^{\#}(t, s)$ and observe that $H^{\#}(t, t) = H(t, t)$.
e. Suppose the observed data consist of $\{(\min(T_i, U_i), I(T_i \leq U_i)), i = 1, \ldots, n\}$. The crude survival functions are defined by $Q_1(t) = P(T > t, U > T)$ and $Q_2(t) = P(U > t, T > U)$. Show that whether the joint distribution of (T, U) is given by $H(t, s)$ or $H^{\#}(t, s)$ as calculated in part d, one obtains the same $Q_1(t)$ and $Q_2(t)$.

3. For a fixed t, a $1 - \alpha\%$ confidence interval for the $S(t)$ the survival probability at t years based on an estimate $\hat{S}(t)$ with estimated standard error $\hat{se}(t)$ is
$$\hat{S}(t) \pm z_{1-\alpha/2}\hat{se}(t),$$
where $z_{1-\alpha/2}$ is the upper $\alpha/2$ quantile of a standard normal distribution. This manner of constructing confidence intervals provides the guarantee that for the year t, the true survival will between the lower and upper limits, over approximately $(1 - \alpha)\%$ of repeated samples.

If we consider the collection of pointwise confidence intervals in the range $[a, b]$

$$\{\hat{S}(t) \pm z_{1-\alpha/2}\hat{se}(t) : a \leq t \leq b\}$$

we have defined a *confidence band* for the survival curve, which is the process or collection $\{S(t) : a \leq t \leq b\}$. We have gone from defining an interval on a one-dimensional quantity to defining a continuous banded region for a function of time. We would like a confidence band $\{[l(t), u(t)] : a \leq t \leq b\}$ that has simultaneous coverage properties, i.e.,

$$P\{l(t) \leq S(t) \leq u(t) : a \leq t \leq b\} \approx 1 - \alpha.$$

Demonstrate or argue why this naive construction of a confidence band based on the collection of pointwise confidence intervals does not satisfy the simultaneous coverage property.

Reading

ABG Chapter 3, sections 3.1.1-2, 3.1.5-6, 3.2.1, 3.2.4-6, 3.3.1-3,5.

Hall and Wellner 1980 (on canvas).