Survival Data Analysis (BIOS 7210)
Breheny

Assignment 1
Due Thursday, August 31

1. Suppose T is a continuous, nonnegative random variable and that $\mathbb{E}T$ exists. Show that $\mathbb{E}T = \int_0^\infty S(u)du$.

2. Suppose $T_1, T_2, \ldots, T_n$ are independent, continuous, nonnegative random variables with hazard functions $\lambda_1(t), \lambda_2(t), \ldots, \lambda_n(t)$. Now suppose we observe $T = \min(T_1, \ldots, T_n)$; for example, perhaps an object has multiple components and the object as a whole fails when any one of its components fails. Show that T has hazard function $\sum_j \lambda_j(t)$.

3. As we showed in class, for a discrete distribution

$$S(t) = \prod_{t_j \le t} \{1 - \lambda_j\}$$

while for a continuous distribution

$$S(t) = \exp\left\{-\int_0^t \lambda(s)\, ds\right\}.$$

Reconcile these two results in the case of the exponential distribution. In particular, consider discretizing the exponential distribution by dividing the interval $[0, t]$ into n intervals of equal length. In this case, the conditional probability[1] of failing in any given interval (i.e., the hazard) is $\lambda_j = t\lambda/n$. Show that as $n \to \infty$, taking the limit of the first expression above yields the second expression.

4. This example involves data from the Human Mortality Database. As we discussed in class, please assume for the sake of simplicity that that the size of the U.S. population is constant over time, even though this is clearly not entirely accurate. Please note that in the data, all ages over 109 are collapsed into a single category, "110+". You must deal with this properly (Hint: simply throwing these lines out does **not** work).

   (a) In class we constructed hazard plots based on 1960 U.S. deaths. Construct hazard plots like those we constructed in class, but for the years 1933 and 2010 (the earliest and latest years in the data set). Comment on the differences between the hazards and what they represent in terms of historical changes.

   (b) Construct separate hazard plots for males and females (you can choose whatever year you'd like for this). Again, comment on differences between the hazards and what they mean. In particular, some people think that young men live more recklessly and dangerously than young women; do you see any evidence of this?

[1] Technically, $\lambda_j = t\lambda/n$ plus a small remainder term that goes away as $n \to \infty$, but for the purposes of this problem, you don't need to worry about this.