Survival Analysis
Biostatistics 140.641

Problem Set 1
Due date: Monday, 9/12/2022

1. (a) Given the hazard function $\lambda(t) = c$, where $c > 0$, derive the survival function and the density function. Derive the median failure time for $c = 2, 5$ and 11.
   (b) Given the survival function $S(t) = \exp(-\theta t^{\beta})$, where $\theta, \beta > 0$, derive the density function and the hazard function.

2. Assume the hazard function starts at 0 and increases linearly with time. Suppose the slope of the line is a positive "b".

   (a) Comment in a couple of sentences on the meaning of the above assumption.
   (b) Derive a formula for the survival function based on this assumption.
   (c) Derive a formula for the density function based on this assumption.
   (d) Sketch three graphs: the hazard vs time; the survival function vs time; and the density vs time. To do this, assume b=.08.

3. Suppose T follows Uniform$(0, \theta)$ distribution, $\theta > 0$, and $C = c_0$ is a constant variable, $c_0 > 0$. Define

$$Y = \begin{cases} T & \text{if} \quad T \le C \\ C & \text{if} \quad T > C \end{cases}$$

   and the censoring indicator $\Delta = I(T \le C)$.

   (a) Find the probability density function of Y and express it as a function of (θ, c_0).
   (b) Find the survival function of Y and express it as a function of (θ, c_0).
   (c) Find the probability density function of Δ and express it as a function of (θ, c_0).

4. Suppose the failure time variable T has the Exponential($\lambda = 5$) distribution, where $\lambda > 0$ is the hazard parameter, and suppose the censoring variable C has the Uniform$(0, 10)$ distribution. Assume T and C are independent. Define

$$Y = \begin{cases} T & \text{if } \; T \leq C \\ C & \text{if } \; T > C \end{cases}$$

and the censoring indicator $\Delta = I(T \leq C)$.

(a) Find the probability density function of Y.
(b) Find the survival function of Y.
(c) Find the probability density function of Δ.

5. In a simulation study the following design is used: $T_i \sim$ Exponential(θ), $\theta > 0$ (where θ is the constant hazard), and $C_i \sim$ Unif$(0, 5)$. Suppose the pairs $(T_1, C_1), ..., (T_n, C_n)$ are mutually independent, and T_i and C_i are independent, $i = 1, 2, ..., n$. We then generate a set of standard survival data which include i.i.d. $(Y_1, \Delta_1), (Y_2, \Delta_2), \ldots, (Y_n, \Delta_n)$.

(a) Please select a specific value of θ so that the proportion of uncensored data takes about 40% of the whole data set.
(b) Suppose now a value of θ is specified. Describe how you generate the observed data $(Y_1, \Delta_1), (Y_2, \Delta_2), \ldots, (Y_n, \Delta_n)$.