STAT 486/886 (Winter 2022)

Assignment 1

The assignment is due on **Jan. 27 (Thursday), 23:00** (time of Kingston Ontario Canada). Please submit to **OnQ**, the dropbox for this assignment.

Guidelines for Preparing Solutions:
For some problems in this course, complete code and output may be very long. Please only include important output and necessary results in the main text of your solutions.
Give descriptions and discussions for your important exploration and findings.
Put long, extra code and output in an **Appendix**, at the end of your assignment.
The Appendix will NOT be marked. Please include it as evidence of your independent work.
Prepare your assignment solutions so that it is easy for readers to follow, without having to search everywhere for your answers from lengthy code and output.
Students submitting identical solutions (multiple sentences, derivation steps or code copied among students or from other resources) will be investigated for violations of academic integrity.

1. The time to relapse, in years, for patients on standard and new treatments for lung cancer is compared using the following regression model:

$$Y = \log(T) = 3 + 0.8Z + 2W,$$

where W has a standard extreme value distribution and $Z = 0$ if standard treatment and 1 if new treatment. Compare the survival probabilities of the two treatments at 1 year and 5 years. (Think of the survival function of Y first, and how it's related to the distribution of W.)

2. Suppose the survival time T has a log normal distribution with parameters μ and σ. Denote the pdf and cdf of the standard normal distribution by $\phi()$ and $\Phi()$.
(a) Find an expression for the hazard function of T. The expression may contain the functions $\phi()$ and $\Phi()$.
(b) It is known that the median and the 95 percentile of T are 100 and 300 respectively. Find the distribution of T in the log normal distribution family (i.e. specify the values of μ and σ for the distribution).
(c) Plot the pdf, survival function and the hazard function for the specific distribution of T you found in (b).
(Please use R (or SAS) for the calculations and plots.)

3. Let T be a failure time having log logistic distribution with parameters μ and σ. In other words, $Y = \log T$ has a Logistic (μ, σ) distribution.

(a) Show that the distribution of T is a log location-scale distribution.
(b) *Re-parameterization.* Show that with new parameters (λ, α), the hazard function and s.f. of T can be written as

$$h(t) = \frac{\alpha\lambda t^{\alpha-1}}{1+\lambda t^{\alpha}}, \text{ and } S(t) = \frac{1}{1+\lambda t^{\alpha}}.$$

Specify the parameters (λ, α) as functions of (μ, σ).

4. Let T be a continuous random variable with survival function $S(t)$. Show that if $E(T)$ exists, then it equals $\int_0^\infty S(t)dt$. (Assume that $\lim_{t\to\infty} tS(t) = 0$.)

5. **This problem is for Stat 886 students only.** The following is a description of the piecewise exponential distribution, which is sometimes used to model survival time. Divide the time axis into k intervals, $[0, \tau_1)$, $[\tau_1, \tau_2)$, ..., $[\tau_{k-1}, \tau_k)$, where $t_0 = 0$ and $\tau_k = \infty$. The hazard rate on each interval is a constant value, specifically,

$$h(t) = \begin{cases} \theta_1, & \text{if } 0 \le t < \tau_1, \\ \theta_2, & \text{if } \tau_1 \le t < \tau_2 \\ \vdots & \\ \theta_{k-1}, & \text{if } \tau_{k-2} \le t < \tau_{k-1} \\ \theta_k, & \text{if } t \ge \tau_{k-1}. \end{cases}$$

(a) Find the survival function for this model.
(b) Suppose T has a piecewise exponential distribution with $k = 3$, $\theta_1 = 1$, $\theta_2 = 1/2$, $\theta_3 = 1/3$, $\tau_1 = 0.5$, $\tau_2 = 1.5$, use R (or SAS) to plot its hazard function and survival function.
(c) A random variable is said to have a continuous distribution if and only if its survival function (or cumulative distribution function) is continuous. Suppose T is from a piecewise exponential distribution, is T a continuous random variable?