ST227 PROJECT MOCK PAPER.

Question 1. The lifetime of a mechanical system is modelled by the following mortality intensity[1]:

$$\mu(t) \equiv \lambda, \quad t \geq 0, \lambda = 10^{-1}.$$

(1) Define in R the survival probability function $(t, x) \to {}_t p_x$ and calculate the probability of surviving the next 5 years for a 15-year-old system.[2]

General wear and tear in mechanical systems mean the memoryless property of constant mortality is a questionable one. We overcome that by introducing a time-dependent term:

$$\tilde{\mu}(t) = \lambda + \gamma \log(\log(e + t)), \quad t \geq 0, \quad \lambda = 10^{-1}, \gamma = 1.5$$

(2) Consider a 15-year-old machine, of which the remaining lifetime T_{15} follows the mortality function $\tilde{\mu}$.

   (a) Define in R the density function for T_{15}. This definition may involve a numerical integral.
   (b) Calculate the expected remaining lifetime for this individual.
   (c) Define in R the cumulative distribution function of T_{15}. This definition may involve a numerical integral. Discuss how you would find the 95-th percentile of T_{15}.[3]

Question 2. This questions is divided into parts. Both parts use the same data set of fully observed lifetimes given below:

64	75	29	45	67	65	77	90	65	55
80	67	72	46	64	28	68	75	49	94

(1) Let us suppose that this data set comes from a gamma distribution with shape-rate parametrisation, i.e:

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x}, \quad x > 0. \tag{1}$$

   (a) Using the results:

$$\mathbb{E}(X) = \frac{\alpha}{\beta}, \quad \mathrm{Var}(X) = \frac{\alpha}{\beta^2}, \tag{2}$$

   derive the method of moment estimators for α and β.

   (b) Using your MMEs above as the initial values for `optim`, derive the MLE for α and β.

(2) We propose a lifetime model with the following mortality intensity function:

$$\mu(t) = \alpha \times \lambda^\alpha \times t^{\alpha - 1}, \quad t \geq 0.$$

   (a) Derive algebraically the probability density function for lifetime and write down the joint-likelihood of the given sample.

[1] Of course, machines are not "mortal". In engineering, this is referred to equivalently as the hazard function. We shall refer to it as mortality intensity, though, to avoid unnecessary jargon.

[2] I am aware that this can be computed explicitly by hand. This is for your R practice, however.

[3] That is, a value α such that $\mathbb{P}(T_{15} \leq \alpha) = 0.95$.

(b) Using the optim method in R, numerically obtain the maximum likelihood estimators of the model parameters.[4]

Question 3. Cancer patients who are in remission are observed and the number of days until the symptoms reappear is recorded.[5] Some records have been right-censored. The data set is provided in a spreadsheet named cancer.xlsx and the columns therein are:

- time: the time until reappearance of symptoms in number of days.
- event: an indicator variable taking value 0 if the record has been right-censored and 1 if fully observed.
- fullyObserved: logical variable indicating whether the record has been fully observed.
- sex: categorical variable with value 0 for male (the reference group) and 1 for female.

(1) Calculate the Kaplan-Meier estimate for survival probabilities.

(2) Using the Greenwood's formula:

$$\mathrm{Var}(\hat{\mathrm{S}}(\mathrm{t})) \approx (\hat{S}(t))^2 \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}, t \in [t_{(k)}, t_{(k+1)}),$$

calculate the variance of $\hat{S}(t)$ at the fully observed times $t_{(i)}, i = 1, 2, \ldots$.

Question 4. In this question, we will fit a Cox Proportional Hazard model on the same data set in Question 3, with time as the response variable and sex as the categorical covariate.

(1) By using the survival package or otherwise, calculate the MLE for the Cox Proportional Hazard Model.
(2) Based on the output you have generated, perform the z-test, Score test, and Likelihood Ratio test on the following hypotheses:

$$H_0 : \beta = 0, \quad \text{vs} \quad H_1 : \beta \neq 0.$$

[4] You may have to play around with the initial parameters a bit. This model is quite stable, meaning you don't need very good initial parameters for it to converge.

[5] What if a patient is completely cured of cancer? We will need a model that allows for $T = \infty$. For the purpose of this exercise, though, we will ignore this subtlety and proceed as usual.