STA 312s19 Assignment Four[1]

The paper and pencil part of this assignment is not to be handed in. It is practice for Quiz 4 on Feb. 4th. The R parts of Questions 5 and 6 may be handed in as part of the quiz. **Bring hard copy of your printout to the quiz**. Do not write anything on your printout in advance except possibly your name and student number.

Unless otherwise noted, T is a continuous random variable with $P(T>0)=1$, density $f(t)$ and cumulatve distribution function $F(t)=P(T\leq t)$.

1. The survival function is $S(t)=P(T>t)$. Prove $E(T)=\displaystyle\int_0^\infty S(t)\,dt$.
2. The hazard function is denoted by $h(t)$, and defined on the formula sheet. Starting with the definition, prove $h(t)=\frac{f(t)}{S(t)}$.
3. Prove $S(t)=e^{-\int_0^t h(x)\,dx}$. You may use anything on the formula sheet except the fact you are proving.
4. Let T have a Weibull distribution with parameters $\alpha>0$ and $\lambda>0$.
   (a) Derive the survival function $S(t)$ for $t>0$.
   (b) What is the hazard function $h(t)$ for $t>0$?
   (c) For what values of α and λ is $h(t)$ increasing? Decreasing?
   (d) What happens to $h(t)$ as $t\rightarrow\infty$?
5. Let the continuous random variable T have hazard function $h(t)=(t-2)^2$ for $t>0$, so that the risk of failure decreases at first, and then increases without bound.
   (a) What is the survival function $S(t)$ for $t>0$? Show your work.
   (b) What is the density $f(t)$ for $t>0$? Show a little work.
   (c) Using R, make a plot of $f(t)$. Bring hard copy to the quiz, including the R code that generated the plot.

[1]This assignment was prepared by Jerry Brunner, Department of Mathematical and Computational Sciences, University of Toronto. The LaTeX source code is available from the course website: http://www.utstat.toronto.edu/~brunner/oldclass/312s19

6. Let X have an exponential distribution with $\lambda = 1$, and let $Y = \log(X)$. The distribution of Y is called the (standard) Gumbel, or extreme value distribution. It has an important role in the analysis survival data, and also is used to model the rare events like natural disasters.

   (a) Where is the density of Y non-zero?

   (b) Find the probability density function of Y. Show your work.

   (c) Using R, make a plot of the standard Gumbel density. Bring hard copy to the quiz, including the R code that generated the plot.

   (d) What is the median of Y? The answer is a number. Use your calculator.

   (e) The *mode* of a continuous distribution is the point where the density is highest. What is the mode of Y? Show your work.

   (f) What is the survival function $S(y)$? Show your work.

   (g) The expected value of Y is surprisingly difficult. Trust me that if you try to use the definition of expected value, you will not be able to do the integral. Instead, use moment-generating functions. Recall that the moment-generating function of Y is $M(t) = E(e^{Yt})$, and $M'(0) = E(Y)$.

      i. Derive the moment-generating function of Y. Show your work.

      ii. Differentiate with respect to t and set $t = 0$. Using R's `digamma` function, get a numerical answer. Please put the one line of calculation on the same printout as Question 6c. You can check your answer by giving this to Wolfram Alpha: `integral of y*exp(y-e^y) from y = minus infinity to infinity`. A minor bonus is that we find the expected value to be $-\gamma$, where γ is the Euler-Mascheroni constant. Who knew?

7. Let $Z \sim N(0, 1)$, and let $X = \sigma Z + \mu$, where $\sigma > 0$. Find the density of X. Show your work. Identify the distribution by name. It is on the formula sheet.

8. Let the continuous random variable Z have density $f(z)$, and let $X = \sigma Z + \mu$, where $\sigma > 0$. Show that the density of X is $f_x(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. The quantity μ is called a *location parameter*, and σ is called a *scale parameter*.

9. Let Z have the standard extreme value distribution of Question 6, and let $X = \sigma Z + \mu$. Give the density of X. This is a Gumbel (extreme value) distribution with location μ and scale σ. Is μ the expected value?

10. Let T have a Weibull distribution with parameters $\alpha > 0$ and $\lambda > 0$, and let $Y = \log T$.

   (a) Find the density of Y; show your work. Do not forget to specify where the density is non-zero.

   (b) Now re-parameterize, meaning express the parameters in a different, equivalent way. Instead of the parameters α and λ, we will have μ and σ. Let $\sigma = \frac{1}{\alpha}$ and $\mu = -\log \lambda$. Write the density of Y in terms of μ and σ. Simplify, and compare your answer to Question 9.

   The lesson here is that the log of a Weibull is an extreme value (Gumbel) distribution. So if you believe the distribution of a set of failure time data could be Weibull (a popular choice), you can log-transform the data and apply a Gumbel model. The Gumbel distribution may be preferable because the parameters μ and σ are easy to interpret.

11. This question is about the meaning of μ and σ in the Gumbel distribution. You can use your answers to earlier questions to make it easier. Show your work when necessary.

   Let Y have density
$$f(y) = \frac{1}{\sigma} \exp\left\{ \left(\frac{y-\mu}{\sigma} \right) - e^{\left(\frac{y-\mu}{\sigma}\right)} \right\}.$$

   (a) What is the survival function $S(y)$?

   (b) What is the hazard function $h(y)$?

   (c) What is the mode?

   (d) What is the median?

   (e) What is the expected value? Write your answer in terms of γ, the Euler-Mascheroni constant.

   (f) The variance of a standard Gumbel is $\frac{\pi^2}{6}$, though this is not easy to show. How do you know that the variance of a general Gumbel (with density given at the beginning of this question) is proportional to σ^2?

Bring your printouts from Questions 5 and 6 to the quiz. Do not write anything on your printout(s) in advance except possibly your name and student number.