

STA 312f23 Assignment Four¹

The paper and pencil parts of this assignment are not to be handed in. They are practice for Quiz 4 on Oct. 6th. The R input and output for Question 1 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 cumulative distribution function $F(t) = P(T \leq t)$.

1. This is a repeat of Question 3(e) from Assignment 3. For background, see the your answers to the earlier parts of the question.

The file <http://www.utstat.toronto.edu/brunner/data/legal/Weibull.data1.txt> contains a random sample from a Weibull distribution. Read it with `scan()`.

- (a) Use R's `optim` function to find the MLE. The answer is a pair of numbers from your printout.
- (b) Calculate a 95% confidence interval for α . The answer is a pair of numbers from your printout. My lower confidence limit is 1.780622.
- (c) Give a point estimate of the expected value for the Weibull data. The answer is a number that you calculate with R. It should appear on your printout. The answer should be fairly close to the sample mean. Why?
- (d) Calculate a 95% confidence interval for the expected value. Hint: to apply the delta method, you will need the derivative of the gamma function. See `help(digamma)`. Because of the Central Limit Theorem, the `t.test` function yields a confidence interval for μ that is close to the right answer.
- (e) Give a point estimate of the median for the Weibull data. The answer is a number that you calculate with R. It should appear on your printout. The answer should be fairly close to the sample median.
- (f) Give a 95% confidence interval for the median. The answer is a pair of numbers on your printout, the lower confidence limit and the upper confidence limit. My lower confidence limit is 3.182938.
- (g) This is the last part of the question, and you should do it last. Using the usual 0.05 significance level, test whether the expected value equals the median. What do you conclude? There is more than one right way to do this. I did it several ways. You only need to do it one way. Using the delta method, I got $Z = 9.92$. Another approach is to observe that the mean equals the median if and only if

¹This assignment was prepared by Jerry Brunner, Department of Mathematical and Computational Sciences, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/). Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/brunner/oldclass/312f23>

α has a particular value. I found that value numerically, and tested $H_0 : \alpha = \alpha_0$, obtaining $Z = -22.71$ and $G^2 = 381.89$. Note that the Z -tests are 2-sided, so the difference in sign is not a problem.

2. The survival function is $S(t) = P(T > t)$. Prove $E(T) = \int_0^\infty S(t) dt$.
3. The hazard function is denoted by $h(t)$, and defined on the formula sheet. Starting with the definition, prove $h(t) = \frac{f(t)}{S(t)}$.
4. Prove $S(t) = e^{-\int_0^t h(x) dx}$. You may use anything on the formula sheet except the fact you are proving.
5. 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 $t \rightarrow \infty$?
6. 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.

7. Let X have an exponential distribution with $\lambda = 1$, and let $Y = -\log(X)$. The distribution of Y is another version of the (standard) Gumbel, or extreme value distribution². The extreme value distribution has an important role in the analysis survival data, and also is used to model the rare events like natural disasters.
- Where is the density of Y non-zero?
 - Find the probability density function of Y . Show your work.
 - What is the median of Y ? The answer is a number. Use your calculator.
 - The *mode* of a continuous distribution is the point where the density is highest. What is the mode of Y ? Show your work.
 - What is the survival function $S(y)$?
 - 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 $M'(0) = E(Y)$.
 - Derive the moment-generating function of Y . Show your work.
 - Differentiate with respect to t and set $t = 0$. Using R's `digamma` function, get a numerical answer. 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 γ , where γ is the Euler-Mascheroni constant. Who knew? By the way, chatGPT got the same wrong answer (zero) three times in a row.
8. 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.
9. 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*.
10. Let Z have the standard extreme value distribution of Question 7, 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?

²In Assignment One, you saw another version of the Gumbel distribution, in which $Y = \log(X)$.

11. Let T have a Weibull distribution with parameters $\alpha > 0$ and $\lambda > 0$, and let $Y = -\log T$.
- Find the density of Y ; show your work. Do not forget to specify where the density is non-zero.
 - 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 10.

The lesson here is that the (minus) 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.

12. 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

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

- What is the survival function $S(y)$?
- The hazard function $h(y)$ does not simplify. Forget it.
- What is the mode?
- What is the median?
- What is the expected value? Write your answer in terms of γ , the Euler-Mascheroni constant.
- 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 printout from Question 1 to the quiz. Do not write anything on your printout in advance except possibly your name and student number.