

STA 312f23 Assignment Five¹

The paper and pencil part of this assignment is not to be handed in. It is practice for Quiz 5 on October 20th. The R part may be handed in as part of the quiz. **Bring hard copy of your printout(s) to the quiz.** Do not write anything on your printout(s) in advance except possibly your name and student number.

1. Consider a random sample of right censored data from an exponential distribution with parameter $\lambda > 0$. In addition to the data T_1, \dots, T_n , we have indicators $\delta_1, \dots, \delta_n$, where $\delta_j = 1$ if T_j is a failure time, and $\delta_j = 0$ if T_j is a censoring time.
 - (a) Derive the maximum likelihood estimate of λ . Show your work. Include the second derivative test. Circle your final answer. It is on the lecture slides. What happens if *all* the observations are censored?
 - (b) Give a formula for the estimated asymptotic variance of $\hat{\lambda}$. Show your work. Circle your answer.
 - (c) Give a 95% confidence interval for λ . Your answer is two formulas, one for the lower confidence limit and one for the upper confidence limit.
 - (d) Give a formula for the MLE of the survival function $S(t)$. Your answer is a formula that could be computed from the sample data for any given t .
 - (e) Using the one-variable delta method (see formula sheet), derive a 95% confidence interval for the survival function $S(t)$, for a particular time t . As above, your answer is two formulas, one for the lower confidence limit and one for the upper confidence limit.
 - (f) Now do the same thing a different way. Writing your confidence interval for λ as $0.95 \approx P(A < \lambda < B)$, multiply by $-t$ and take the exponential function to obtain a different confidence interval for $S(t)$, one that is also valid. Is it possible for these confidence limits to be outside the range zero to one?
2. The file <http://www.utstat.toronto.edu/brunner/data/legal/expo.data2.txt> contains a random sample from an exponential distribution with right censoring. Using R as a high-powered calculator (no numerical search yet),
 - (a) Give the maximum likelihood estimate of λ . The answer is a number. You are calculating your answer to Question 1a.
 - (b) What is the estimated asymptotic variance of $\hat{\lambda}$? The answer is a number. You are calculating your answer to Question 1b.
 - (c) Give a 95% confidence interval for λ . Your answer is two numbers, the lower confidence limit and the upper confidence limit. You are calculating your answer to Question 1c.
 - (d) Based on your answers to Questions 1d and 1e, calculate your estimated $S(t)$ and the confidence interval for the following values of t : `seq(from=0,to=3,by=0.1)`. Bind $t, \hat{S}(t)$ the lower confidence limit and the upper confidence limit into a matrix with 4 columns.
 - (e) Now do the same thing, using your confidence limits from Question 1f. The result is another 4×31 matrix. Which one do you like more and why?

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/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>

3. Using the data from Question 2,
- Obtain the maximum likelihood estimate of λ by a numerical search. Make sure to request the Hessian. Compare your answer to Question 2a.
 - Obtain the estimated variance of $\hat{\lambda}$ from the output of your search. If you think about it, you will realize that $-\ell''(\hat{\lambda})$ is a number on your printout. Display the estimated variance on your printout. Compare your answer to Question 2b.

After that, what you'd do would be the same as in Question 2. You don't have to do it twice.

4. A log-normal random variable with parameters μ and σ^2 has the property that if you take its natural log, you get a normally distributed random variable with expected value μ and variance σ^2 . A log-normal random variable is positive valued and right skewed. It is often a good model for survival times.

You have already seen the log-normal distribution in Question 1 of Assignment 3. To repeat, its density is

$$f(t|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \left\{ \frac{(\log(t) - \mu)^2}{2\sigma^2} \right\} \frac{1}{t} I(t > 0)$$

When there are censored data, maximum likelihood estimation might seem to present a problem, because you cannot do the integral that defines the survival function $S(t)$. So, you need to do the whole thing numerically.

The file <http://www.utstat.toronto.edu/brunner/data/legal/lognorm1.data.txt> contains a random sample from a log-normal distribution with right censoring.

- Obtain the maximum likelihood estimates of μ and σ^2 by a numerical search. Your answers are numbers. Make sure to request the Hessian. There is an easy way to do this, and several variations of the hard way. I recommend doing it the easy way. The main hint is `help(dlnorm)`. You may be wondering about starting values. One possibility is to ignore the censoring, and start with the MLEs for uncensored data. The answer will not be correct for censored data, but it could be close enough to be useful. See your answer to Question 1 of Assignment 3. When I did this, I didn't bother with the difference between n and $n - 1$.
- Obtain the estimated asymptotic variance-covariance matrix, and display it on the printout that you bring to the quiz.
- Give a 95% confidence interval for μ . Of course you need to display the code that generated your confidence interval.
- Show that the median of the log-normal distribution is e^μ . Do it the easy way, using the fact that the median of a normal distribution is μ . I'll start you out. Let T be log-normal, so $X = \log T \sim N(\mu, \sigma^2)$. Of course, $T = e^X$. Then, $P(X > \mu) = \frac{1}{2} \implies \dots$
- Going back to the numerical data, give a confidence interval for the median. Do *not* use the delta method this time.

Bring your printout for Questions 2, 3 and 4 to the quiz. All the requested numbers and the code that produced them should appear on your printout. Do not write anything on your printout in advance except possibly your name and student number.