

1. Recall the heat exchanger data shown in the following figure.

	Year 1	Year 2	Year 3	
Plant 1	100 1	99 2	97 2	95
Plant 2	100 2	98 3	95	
Plant 3	100 1	99		

	Year 1	Year 2	Year 3	
All Plants	300 4	197 5	97 2	95
Conditional Failure Probability	4/300	5/197	2/97	

Uncracked tubes: 99 95 95

- (a) Write down the likelihood for these data in terms of multinomial probabilities (π 's).
- (b) Construct a table showing how to compute the nonparametric estimate of $F(t)$. To save time, you can leave your answer in terms of fractions.
- (c) How many parameters are there in your model for these data? Explain why the estimate in part 1b called a "nonparametric estimate?"

2. Hazard function estimates are often reported in units of FITs [computed as 10^9 times $h(t)$]. Why?

3. In parametric model inference, the negative Hessian matrix (matrix of second derivatives and mixed partial derivatives) plays an important role. The expected value of the negative Hessian is also important.
 - (a) Explain the important role (i.e., application) of the negative Hessian matrix and the intuition behind its use.

 - (b) Explain the important role (i.e., application) of the expected value of the Hessian matrix and the intuition behind its use.

4. An important assumption in the analysis of censored data is that censoring is “fair.” Explain this concept and give a new example (i.e., one not used in class or on previous assignments) of a censoring mechanism that would not be fair.

5. Given an expression for $\widehat{se}[F(t)]$, use the delta method to derive an expression for $\widehat{se}\{\text{logit}[F(t)]\}$.

6. The Weibull distribution is a popular model to describe a time to failure distribution.
 - (a) What are the advantages of assuming that the Weibull shape parameter is known?

 - (b) What is the potential danger of assuming that the Weibull shape parameter is known?

 - (c) What can one do to protect against this danger?

7. The Weibull cdf is $F(t) = 1 - \exp[-(t/\alpha)^\beta]$. Derive an expression for the hazard function of this distribution.

8. A random sample of $n = 100$ items was put on test at a particular point in time and the test was run until $t_c = 1000$ hours at which point $r = 20$ units had been reported as failing. The reported failure times were rounded up to the nearest hour. Engineers believe that, based on physical knowledge of the failure mechanism, that the one-parameter exponential distribution can be used to describe time to failure. The exponential cdf is $F(t) = 1 - \exp(-t/\theta)$.

(a) For these data, write down an expression for the loglikelihood based on the “correct likelihood.”

(b) For these data, write down an expression for the loglikelihood based on the “density approximation for the correct likelihood” for observations reported as exact failures.

(c) Write down the first derivative of the loglikelihood from part 8b.

(d) Use part 8c to derive an expression for the ML estimate for this model.

9. Consider the situation in problem 8.

(a) List the steps that you would use to make a plot with exponential plotting scales for an exponential probability plot (say starting with ordinary graph paper and a calculator).

(b) Show why the exponential cdf will plot as a straight line on your exponential probability paper.

(c) Explain how you would plot the nonparametric estimate on the probability paper (i.e., where would you plot points)?

10. Refer to problem 8. The data were really interval censored, because of the rounding. The nonparametric estimate would differ slightly depending on whether we recognize the interval censoring or assume that the reported times were exact failures. Draw a picture and explain.