Statistics 533 | Midterm Exam | Name ______________________

Spring 1998 | **You must show all of your work**

When asked to explain something, provide an explanation that could be understood by someone who does not have formal training in statistical methods. Your explanations should be clear, but concise.

1. Let θ denote the exponential distribution mean and let $\lambda = 1/\theta$. Also, let $\widehat{\text{se}}_{\widehat{\theta}}$, denote the estimated standard error for $\widehat{\theta}$, the maximum likelihood estimate of θ.

   (a) Use the delta method to derive an expression for $\widehat{\text{se}}_{\widehat{\lambda}}$ as a function of $\widehat{\text{se}}_{\widehat{\theta}}$ where $\widehat{\lambda} = 1/\widehat{\theta}$.

   (b) What is the practical interpretation of λ?

   (c) Is λ a probability or is it related to a probability? *Explain.*

2. Provide an intuitive explanation for the reason that bootstrap (simulation) based confidence intervals would be expected to provide a better approximation to the nominal coverage probability than the simpler normal-approximation intervals. Draw a picture to help your explanation.

3. Provide an intuitive explanation for the reason that likelihood ratio based confidence intervals would be expected to provide a better approximation to the nominal coverage probability than the simpler normal-approximation intervals. Draw a picture to help your explanation.

4. Refer to problem 9. The Weibull distribution is a popular model that is frequently used to describe a time to failure distribution. The Weibull cdf is $F(t) = 1 - \exp[-(t/\alpha)^{\beta}]$.

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

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

   (c) Explain how your would plot the nonparametric estimate on the probability paper (i.e., where would you plot points), in hte case where failures are reported at specific times?

5. Refer to problems 9 and 4.

   (a) Give an expression for the Weibull distribution likelihood for the case where the failures are taken to be at exactly the reported times.

   (b) Given an expression for the Weibull distribution likelihood for the case where the failures are known to have been rounded to the nearest 10 hours.

(c) *List* potential advantages or disadvantages of the two different representations for the likelihood.

-
-
-

6. If we plan a life test to run until 5% of 100 units will fail and if we are interested in estimating $t_{.1}$, the time at which 10% of the population of such units will fail, we will be extrapolating when we make the inference. Consider that the same life test might also be used to to estimate $t_{.001}$, the time at which .1% of the population will have failed.

   (a) Explain (and perhaps draw a picture to help with the explanation) why this is also extrapolation.

   (b) What has to be done to design a life test to estimate $t_{.001}$ that will *not* involve extrapolation?

7. The negative Hessian matrix (matrix of second derivatives and mixed partial derivatives) of the loglikelihood function plays an important role in parametric model inference. When evaluated at the ML estimate, this matrix is known as the "observed" information matrix. The expected value of the negative Hessian matrix of the loglikelihood is knonw as the "Fisher" information matrix.

   (a) Explain the important role (i.e., application) of the observed information matrix and the intuition behind its use.

   (b) Explain the important role (i.e., application) of the "Fisher" information matrix and the intuition behind its use.

8. Refer to Table C.20 (copy attached). A life test is planned initially to run for 500,000 cycles and the responsible engineers feel that life can be described by a lognormal distribution with $\sigma^{\square} = .5$ with approximately 8% failing at 500,000 cycles. Management has asked for an estimate and a 95% confidence interval for $t_{.5}$ for this distribution. In planning the test, it is thought that the sample size should be large enough such that the confidence interval upper endpoint should be approximately 50% larger than the ML estimate of $t_{.5}$.

   (a) Derive an expression for the planning value for μ as a function of the information given above?

   (b) What is the approximate sample size that is needed to estimate $t_{.5}$ for this distribution, according to the specified criterion, assuming that both μ and σ will be estimated using the method of maximum likelihood?

   (c) Why is the needed sample size in part (a) so large? How large would the sample size need to be if the test were run long enough so that all of the units failed?

   (d) What is the approximate sample size that is needed to estimate $t_{.5}$ for this distribution, according to the specified criterion, assuming that μ is to be estimated using the method of maximum likelihood but that $\sigma = .5$ is given? Is it important to know whether σ is given or not? Explain.

9. 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 $t_1, t_2, \ldots, t_{20}$. The data, however, were really interval censored, because of the rounding to the nearest 10 hours. The nonparametric estimate would differ slightly depending on whether we recognize the interval censoring or assume that the reported times were exact failures. What form will the nonparametric estimator take for each data representation? Draw a picture and explain.