Statistics 533 | Midterm Exam | Name Key
Spring 2002 | 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. You do not need to do any complicated calculations. Set things up and show me that you know how to do the calculations. Leave requested numerical answers as fractions.

1. Suppose that time to failure T for Component-A has the cdf

$$\Pr(T \le t) = F(t) = \exp\left[-\left(\frac{\delta}{t}\right)^{\beta}\right], \quad t > 0. \tag{1}$$

This is known as the Frèchet distribution of maxima and is a distribution that is in the log-location-scale family.

(a) Derive an expressions for the quantile of T and for the quantile of $\log(T)$.

$p = \exp[-(\delta/t_p)^\beta] \Rightarrow -\log(p) = (\delta/t_p)^\beta \Rightarrow [-\log(t_p)]^{\frac{1}{\beta}} = \frac{\delta}{t_p} \Rightarrow$

$t_p = \delta/[-\log(p)]^{\frac{1}{\beta}}$

$y_p = \log(t_p) = \log(\delta) - \frac{1}{\beta}\log[-\log(p)]$

(b) Show that the distribution of $\log(T)$ has a location-scale distribution.

Let $Y = \log(T)$, $y = \log(t)$, $t = \exp(y)$ $\qquad \sigma = 1/\beta$

$\Pr(T \le t) = \Pr(\log(T) \le y) = \exp\left[-\left(\frac{\delta}{\exp(y)}\right)^\beta\right] \qquad \mu = \log(\delta)$

Note: could also argue from (a)

$= \exp[-\exp(\beta(\log(\delta) - \log(t)))] = \exp\left[-\exp\left(-\frac{\log(t) - \mu}{\sigma}\right)\right]$

(c) Show how you would transform time and fraction failing to construct a probability plot for the Frèchet distribution.

Using (a) $\log(t_p) = \log(\delta) - \frac{1}{\beta}\log[-\log(p)]$

is a linear function relating Transformed Time and Transformed probability.

2. Suppose that you have an expression $\mathrm{Var}[\log(\widehat{\theta})]$, where $\widehat{\theta}$ is the ML estimator of the scalar parameter θ. Use the delta method to get an expression for the approximate (large-sample) variance of $\widehat{\theta}$.

let $\hat{\delta} = \log(\hat{\theta})$ and then $\hat{\theta} = g(\hat{\delta}) = \exp(\hat{\delta})$

$\mathrm{Var}(g(\theta)) = \left(\frac{dg}{d\theta}\right)^2 \mathrm{Var}(\log(\hat{\theta}))$

$= [\exp(\delta)]^2 \mathrm{Var}(\log(\hat{\theta})) = \theta^2 \mathrm{Var}(\log(\hat{\theta}))$

3. Refer to the distribution in question 1.

(a) Derive an expression for the probability density function for T.

$$f(t) = \frac{dF(t)}{dt} = \frac{d \exp\left[-\left(\frac{\delta}{t}\right)^{\beta}\right]}{dt} = \beta\left(\frac{\delta}{t}\right)^{\beta-1}\frac{\delta}{t^2}\exp\left[-\left(\frac{\delta}{t}\right)^{\beta}\right]$$

$$= \left(\frac{\beta}{t}\right)\left(\frac{\delta}{t}\right)^{\beta}\exp\left[-\left(\frac{\delta}{t}\right)^{\beta}\right]$$

(b) Derive an expression for the hazard function for T.

$$h(t) = \frac{f(t)}{1-F(t)} = \frac{\left(\frac{\beta}{t}\right)\left(\frac{\delta}{t}\right)^{\beta}\exp\left[-\left(\frac{\delta}{t}\right)^{\beta}\right]}{1-\exp\left[-\left(\frac{\delta}{t}\right)^{\beta}\right]}$$

(c) Suppose that n of these components are put on test and run until a fixed time t_c. Write down an expression for the log likelihood for the resulting data, assuming that the exact time of failure is recorded.

$$\mathcal{L}(\beta, \delta) = \sum_{i=1}^{r} \log\left[f(t_i;\beta,\delta)\right] + (n-r)\log\left[1-\exp\left[-\left(\frac{\delta}{t_c}\right)^{\beta}\right]\right]$$

4. In some situations (e.g., a life test with Type II censoring, assuming a log-location-scale distribution) there are "exact" confidence interval procedures, called so because such procedures have exactly the nominal coverage probability. Suppose that an exact 95% confidence interval for $t_{.10}$ is computed from a sample in such a situation.

(a) Explain why it is that one *cannot* clam "The probability that the interval contains $t_{.10}$ is .95."

Because after an interval has been constructed, it is either covering $t_{.10}$ or not (probability 1 or 0).

(b) What is a relevant, correct statement (using the word "probability") that one could make in such a situation?

It is the procedure that has probability .95 of covering the true value of $t_{.10}$.

5. Suppose that a component has an exponential failure time distribution with mean θ and that n of these components are put on test and run until r ($r \le n$) failures occur. This is known as Type II censoring. Then total time on test is defined as $TTT = t_{(1)} + \cdots + t_{(r)} + (n-r)t_{(r)}$. It is possible to show that $W = 2(TTT)/\theta$ has a chisquare distribution with $2r$ degrees of freedom.

   (a) Explain (no derivation needed; use the information given above) why W can be called a pivotal quantity.

   W has a completely specified distribution with no unknown parameters.

   (b) Using the pivotal property of W, derive an expression for a $100(1-\alpha)\%$ confidence interval for θ.

   $W = \frac{2TTT}{\theta}$

   $$\Pr\left(\chi^2_{(\alpha/2, 2r)} < \frac{2TTT}{\theta} < \chi^2_{(1-\alpha/2, 2r)}\right) = 1-\alpha$$

   $$\Pr\left(1/\chi^2_{(1-\alpha/2, 2r)} < \theta/(2TTT) < \chi^2_{(\alpha/2, 2r)}\right) = 1-\alpha$$

   $$\Pr\left(2TTT/\chi^2_{(1-\alpha/2, 2r)} < \theta < 2TTT/\chi^2_{(\alpha/2, 2r)}\right) = 1-\alpha$$

6. Give an intuitive explanation or reason why the distribution of W in question 5 is not pivotal when a test is stopped at a particular point in time (instead after a particular number of failures).

   Because the number of failures r is random, the distribution of W will depend on the number of failures, which will not be parameter free.

7. Applications in reliability often require extrapolation. Fitting a parametric failure-time distribution (e.g., Weibull or lognormal) allows extrapolation in time. For example, by extrapolating toward the upper tail of the distribution, one can estimate the fraction failing at 5 years, based on only two years of data. Briefly describe a practical example of when extrapolation in time toward the lower tail of the distribution is required.

   If we test n units, but wish to estimate the p quantile where $p < 1/n$, we need extrapolate into the lower tail of the distribution.

8. If the random variable T has a lognormal distribution, then $1/T$ also has a lognormal distribution. Show why this is true.

   $$\Pr(T<t) = F(t) = \Phi_{nor}\left[\frac{\log(t)-\mu_T}{\sigma}\right]$$

   let $U = 1/T$ (not this is decreasing)

   $$\Pr(U<u) = 1 - \Phi_{nor}\left[\frac{\log(\frac{1}{u})-\mu_T}{\sigma}\right] = 1 - \Phi_{nor}\left[-\frac{\log(u)-\mu_U}{\sigma}\right]$$

   $\mu_u = -\mu_T$

   $$= \Phi_{nor}\left[\frac{\log(u)-\mu_U}{\sigma}\right]$$ due to symmetry

9. The Weibull distribution is widely used to describe the time to failure in certain areas of application. Provide a physical/statistical explanation that might explain the success of the Weibull distribution in explaining the failure time distribution for certain kinds of failure mechanisms.

The Weibull distribution is on of the limiting distributions of minima. It is often used to describe systems that have many components that fail when the "weakest link" fails.

10. Refer to Figure 10.6 from Meeker and Escobar (1998). A copy of this figure is attached. What simple rule does this figure suggest for choosing the length of a life test?

Testing longer will reduce the variance but there are diminsling returns when the fraction failing exceeds the quantile of interest.

11. Briefly explain how simulation can be used to to help plan a life test.

Simulation is useful to allow visualation of the variability expected in repeated sampling. It also does not require the use of large-sample approximations.

12. For a sample of size n with no censoring (i.e., exact failure times $t_1, \ldots, t_n$ recorded for all units in the sample), the likelihood can be written as

$$L(\theta) = \prod_{i=1}^{n} f(t_i; \theta)$$

In this case, the likelihood is *not* a probability. The likelihood is, however, under certain circumstances, approximately proportional to the probability of the data. Explain.

The probability of the data can be written as

$C \prod_{i=1}^{n} \left[F(t_i + \Delta) - F(t_i - \Delta) \right]$ but

$\lim_{\Delta \to 0} \left[F(t_i + \Delta) - F(t_i - \Delta) \right] / \Delta \to f(t_i)$

13. In a life test of n units, how may failures must be observed in order to be able to estimate the two parameters of a Weibull distribution?

1, as long as a censoring time exceeds the failure time.

14. If the shape parameter of a two-parameter Weibull distribution can be assumed to be known, and a life test of n units is run for t_c hours without any failures, what can one say about the weibull fraction failing at time t_e ($t_e < t_c$)?

A conservative lower confidence bound on η is $\underset{\sim}{\eta} = \left(\dfrac{\sum t_i^{\beta}}{-\log(\alpha)} \right)^{1/\beta}$

Thus a conservative upper confidence bound on $F(t)$ is $\tilde{F}(t) = 1 - \exp\left[-\left(\dfrac{t_e}{\underset{\sim}{\eta}} \right)^{\beta} \right]$