

# STAT 687 - Homework # 4 - Solution

Due 0800, 21 July 2014

## From the text problem # 9.3

Write a computer program to do the following, using the shock absorber data from Example 9.4 to illustrate the program's use:

- (a) Sample with replacement to obtain a sample of size  $n = 38$  shock absorbers from the 38 rows of Appendix Table C.2

We can easily sample, with replacement, from the shockabsorber data set to create a bootstrap sample of 38 observations using `sample()`. Note that in the code below `boot` includes the sampled times as well as the censoring indicator corresponding to that observation

```
boot <- shockabsorber[sample(1:38, 38, replace = TRUE),]
Shockboot.ld <- frame.to.ld(boot,
                           response.column = 1,
                           censor.column = 3)
```

- (b) Compute the Weibull distribution ML bootstrap estimates  $(\hat{\mu}^*, \hat{\sigma}^*)$  for this sample.

Using the bootstrap sample of shock absorbers from part (a) the maximum likelihood bootstrap estimates,  $\hat{\mu}_{MLE}^*, \hat{\sigma}_{MLE}^*$  are found by fitting a Weibull distribution to the times to failure associated with the sampled shock absorbers. In R we can find these values several ways:

**Using the “survival” package:** The survival package requires that we first creat a “Surv” object. Since we’ve already created a “life.data” object, this conversion is easy

```
Shockboot.Surv <- ld.to.surv(Shockboot.ld)
```

We can extract  $\hat{\mu}_{MLE}^*, \hat{\sigma}_{MLE}^*$  from a “survreg” (Survival Regression) object by first running

```
SR <- survival::survreg(Shockboot.Surv~1, dist = "weibull")
```

From the survreg object `SR`, the maximum likelihood parameter values are reported as  $\hat{\mu}_{MLE}^*, \hat{\sigma}_{MLE}^*$  from the smallest extreme value distribution. For this sample these values are

$$\hat{\mu}_{MLE}^*, \hat{\sigma}_{MLE}^* = 10.2272037, 0.3221639$$

**Using the SMRD package:** By using only the function from the SMRD package we can compute  $\hat{\mu}_{MLE}^*, \hat{\sigma}_{MLE}^*$  and present them in a L<sup>A</sup>T<sub>E</sub>X table very quickly using

```
mle.table <- print(mlest(Shockboot.ld, distribution = "Weibull"))$mle.table
print(xtable(mle.table,
             caption = "Weibull parameter estimates for the shock absorber bootstrap sample",
             digits = c(0,4,4,2,2)),
      comment = FALSE,
      caption.placement = "top",
      table.position = 'h',
      type = switch(output, 'html' = 'html', 'pdf' = 'latex'))
```

Comparing the results from both packages shows that they are equivalent.

Table 1: Weibull parameter estimates for the shock absorber bootstrap sample

	MLE	Std.Err.	95% Lower	95% Upper
mu	10.2272	0.1536	9.93	10.53
sigma	0.3222	0.0896	0.19	0.56
Weibull (eta)	27645.0888	4245.6401	20459.42	37354.48
Weibull (beta)	3.1040	0.8630	1.80	5.35

(c) Compute the bootstrap studentized quantile estimate  $Z_{\log(\hat{t}_{0.1}^*)} = [\log(\hat{t}_{0.1}^*) - \log(\hat{t}_{0.1})] / \hat{se}_{\log(\hat{t}_{0.1}^*)}$

The bootstrap studentized quantile estimate (top of pg. 209) for  $p = 0.1$  is expressed as

$$Z_{\log(\hat{t}_{0.1}^*)} = \frac{\log(\hat{t}_{0.1}^*) - \log(\hat{t}_{0.1})}{\hat{se}_{\log(\hat{t}_{0.1}^*)}}.$$

Each bootstrap sample generates a single observation of  $Z_{\log(\hat{t}_{0.1}^*)}$ . This requires first computing  $\log(\hat{t}_{0.1})$  using the set of 38 observations from the original shock absorber test data. The value for  $\log(\hat{t}_{0.1})$  may be found by inverting the Weibull cdf in location-scale form and substituting  $\hat{\mu}_{MLE}$  and  $\hat{\sigma}_{MLE}$ , resulting in

$$\log(\hat{t}_{0.1}) = \hat{\mu}_{MLE} + \hat{\sigma}_{MLE} \times \log(-\log(.9)) = 10.23 + 0.3164(-2.2504) = 9.518 \text{ Kilometers.}$$

Then, in a similar manner,  $\log(\hat{t}_{0.1}^*)$  is computed for each bootstrap sample. For the sample produced in part (a) this value is

$$\log(\hat{t}_{0.1}^*) = \hat{\mu}_{MLE}^* + \hat{\sigma}_{MLE}^* \times \log(-\log(.9)) = 10.227 + 0.322(-2.2504) = 9.502 \text{ Kilometers.}$$

Finally,  $\hat{se}_{\log(\hat{t}_{0.1}^*)}$  is computed by first using Equation (8.12) to find

$$\hat{se}_{\hat{t}_{0.1}^*} = \hat{t}_{0.1}^* \left[ \hat{Var}(\hat{\mu}^*) + 2 \times \log(-\log(0.9)) \times \hat{Cov}(\hat{\mu}^*, \hat{\sigma}^*) + \log(-\log(0.9))^2 \times \hat{Var}(\hat{\sigma}^*) \right]^{0.5}$$

Then,  $\hat{se}_{\log(\hat{t}_{0.1}^*)}$  is found using the Delta Method a second time as shown in Section 7.3.3, where Equation 7.12 shows that

$$\hat{se}_{\log(\hat{t}_{0.1}^*)} = \frac{\hat{se}_{\hat{t}_{0.1}^*}}{\hat{t}_{0.1}^*}.$$

The variance and covariance terms in the above equation can be produced from either the “survreg” object or the “life.data” object.

```
print(xtable(SR$var,
             caption = "Covariance matrix (from survreg) for the shock absorber bootstrap sample",
             digits = c(0,7,7)),
      comment = FALSE,
      caption.placement = "top",
      table.placement = 'h',
      type = switch(output, 'html' = 'html', 'pdf' = 'latex'))
```

Table 2: Covariance matrix (from survreg) for the shock absorber bootstrap sample

	(Intercept)	Log(scale)
(Intercept)	0.0235858	0.0286421
Log(scale)	0.0286421	0.0773053

```
vcv <- print(mlest(Shockboot.ld, distribution = "Weibull"))$vcv

print(xtable(vcv,
  caption = "Covariance matrix (from SMRD) for the shock absorber bootstrap sample",
  digits = c(0,7,7)),
  comment = FALSE,
  caption.placement = "top",
  table.placement = 'h',
  type = switch(output, 'html' = 'html', 'pdf' = 'latex'))
```

Table 3: Covariance matrix (from SMRD) for the shock absorber bootstrap sample

	mu	sigma
mu	0.0235858	0.0092275
sigma	0.0092275	0.0080235

Therefore,

$$\hat{s}e_{\log(\hat{t}_{0.1}^*)} = [0.0235858 + 2 \times (-2.2504)(0.0092275) + (-2.2504)^2 \times 0.0080235]^{0.5} = 0.151$$

and for this sample a single observation is produced

$$Z_{\log(\hat{t}_{0.1}^*)} = \frac{9.502 - 9.518}{0.151} = -0.10596$$

- (d) Repeat steps (a) to (c) 2000 times to get 2000 values of  $Z_{\log(\hat{t}_{0.1}^*)}$ . Make a histogram of these values to see the bootstrap distribution of  $Z_{\log(\hat{t}_{0.1}^*)}$ . Find the 0.25 and .975 quantiles of this distribution.

The histogram below illustrates the approximate distribution of the 2000  $Z_{\log(\hat{t}_{0.1}^*)}$  observations.

Finding the 2.5% and 97.5% quantile values of the  $Z_{\log(\hat{t}_{0.1}^*)}$  can then be accomplished using

```
quantile(Z_boot_t.q,c(.025,.975))
```

which results in quantile values of  $-1.4817924$  and  $1.7937902$ .

- (e) Use the results in part (d) to compute a bootstrap confidence interval for  $t_{.1}$ . Compare with the interval in Table 9.2. Why are they not exactly the same? What could be done to assure better agreement in such a comparison?

Finally, a 95% confidence interval for  $\hat{t}_{0.1}$  may be produced using Equation 8.11 where the upper and lower bounds of this interval were calculated as

$$[\underline{t}_{0.1}, \overline{t}_{0.1}] = [7058.2908408, 1.0779019 \times 10^4]$$

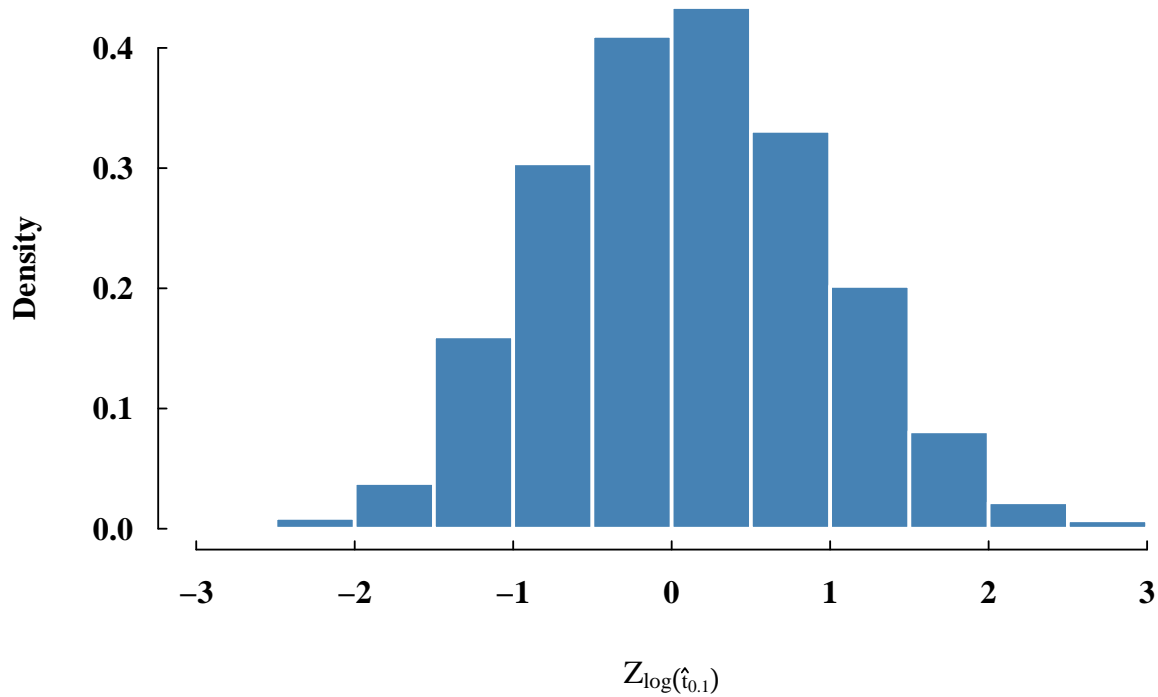


Figure 1: Histogram of nonparametric bootstrap studentized quantile estimates (2000 samples)

These upper and lower confidence limits differ from those shown in Table 9.2 because the table limits were found using 10,000 bootstrap samples rather than 2,000. Further, the table limits were found using a parametric bootstrap procedure wherein the confidence intervals are based on a larger range of observations than is possible using the nonparametric approach.

Interestingly, by re-running the nonparametric analysis with 10,000 bootstrap samples did not result in a significant change in the upper and lower confidence limits.

## From the text problem # 10.6

A reliability engineer wants to run a life test to estimate the .05 quantile of the fatigue life distribution of a metal component used in a switch. The engineer has to choose a sample size that will allow estimation to be precise enough so that the lower endpoint of a 95% confidence interval for the quantile will be about one-half of the ML estimate. It will be possible to test each specimen until about 100-thousand cycles, when it is expected that about 15% of the specimens will have failed. It is expected that about 5% will have failed after about 40-thousand cycles.

- (a) Use the information above on Weibull probability paper to obtain “planning values” for the Weibull parameters.

For this problem it is simpler if we avoid using the plotting paper and instead calculate the planning values  $\beta^\square$  and  $\eta^\square$  directly from the Weibull cdf where

$$0.15 = 1 - \exp \left[ - \left( \frac{100000}{\eta^\square} \right)^{\beta^\square} \right] \quad \text{and} \quad 0.05 = 1 - \exp \left[ - \left( \frac{40000}{\eta^\square} \right)^{\beta^\square} \right]$$

Solving the first equation for  $\eta^\square$  results in

$$\eta^\square = \frac{100000}{(-\log[.85])^{\frac{1}{\beta^\square}}}$$

substituting this expression for  $\eta^\square$  into the second Weibull cdf equation above results in a function where  $\beta^\square$  is the only unknown

$$.95 = \exp \left[ - \left( 40000 \times \frac{(-\log[.85])^{\frac{1}{\beta^\square}}}{100000} \right)^{\beta^\square} \right].$$

Solving the above expressions for  $\beta^\square$  and  $\eta^\square$  give

$$\beta^\square = \frac{\log \left( \frac{\log(.95)}{\log(.85)} \right)}{\log(0.4)} \approx 1.26 \quad \text{and} \quad \eta^\square \approx 423612$$

Alternatively, we can express these values in location scale form as

$$\sigma^\square = 0.793 \quad \mu^\square = 12.956$$

- (b) Determine the sample size needed to achieve the desired precision

In critical safety items there is interest in knowing the “safe life” of a system or component. Often “safe life” is defined as half of the expected life. In this problem we are asked to find the required sample size such that the lower limit of the 95% confidence interval for  $t_{.05}$  is half of the expected value for  $t_{.05}$ . This can be accomplished using Equation 10.8

$$n = \frac{z_{(1-\alpha/2)}^2 V_{\log(\hat{t}_p)}^\square}{[\log(R_T)]^2}$$

where

$$z_{(1-\alpha/2)}^2 = 1.96^2 = 3.8416$$

$$[\log(R_T)]^2 = \left[ \log \left( \frac{\hat{t}_{.05}}{\hat{t}_{.05}} \right) \right]^2 = \left[ \log \left( \frac{\hat{t}_{.05}}{\frac{1}{2}\hat{t}_{.05}} \right) \right]^2 = [\log(2)]^2 = 0.4804.$$

The variance component  $V_{\log(\hat{t}_p)}^\square = V_{\hat{\mu}}^\square + [\Phi^{-1}(p)]^2 V_{\hat{\sigma}}^\square + 2\Phi^{-1}(p)V_{(\hat{\mu}, \hat{\sigma})}^\square$  can be read from Table 10.5, following the curve for  $p_c = .15$  as the fraction of units planned to survive at the conclusion of the test (i.e. right censored observations).

From the table, the variance factor appears to be  $\approx 15$ . This value can also be found with the SMRD package, using the command

```
#single.quantile.var.fact(quantile, fraction surviving at end of test, "sev")
#single.quantile.var.fact(.05, .15, "sev")
```

where the command results in  $V_{\log(\hat{t}_{.05})}^\square \times \left( \frac{1}{\sigma^\square} \right)^2 \approx 14.6$ . Thus,

$$V_{\log(\hat{t}_{.05})}^\square = 14.6 \times .795^2 = 9.224$$

Finally, the required sample size may be computed as

$$n = \frac{9.224 \times 3.8416}{0.4804} \approx 74 \text{ units.}$$

## From the text problem # 10.9

Section 10.5.1 shows how to use the elements of the Fisher information matrix to compute variance factors for test planning in a life test with a single censoring time. In Appendix Table C.20, use the values of  $f_{11}$ ,  $f_{22}$ , and  $f_{12}$  in the row corresponding to  $\zeta_c = 1$ , and show how to compute the quantities  $\frac{1}{\sigma^2} V_{\bar{\mu}}$ ,  $\frac{1}{\sigma^2} V_{\bar{\sigma}}$ ,  $\frac{1}{\sigma^2} V_{(\bar{\mu}, \bar{\sigma})}$ ,  $\rho(\bar{\mu}, \bar{\sigma})$ ,  $\frac{1}{\sigma^2} V_{\bar{\mu}|\sigma}$ , and  $\frac{1}{\sigma^2} V_{\bar{\sigma}|\mu}$  in that row.

The  $f_{ij}$  elements in Appendix C.20 are the elements of the scaled Fisher information matrix, that is

$$\begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} = \sigma^2 \begin{bmatrix} E \left\{ -\frac{\partial^2 \mathcal{L}(\mu, \sigma)}{\partial \mu^2} \right\} & E \left\{ -\frac{\partial^2 \mathcal{L}(\mu, \sigma)}{\partial \mu \partial \sigma} \right\} \\ E \left\{ -\frac{\partial^2 \mathcal{L}(\mu, \sigma)}{\partial \mu \partial \sigma} \right\} & E \left\{ -\frac{\partial^2 \mathcal{L}(\mu, \sigma)}{\partial \sigma^2} \right\} \end{bmatrix} = \begin{bmatrix} .96841 & -0.11490 \\ -0.11490 & 1.56779 \end{bmatrix}$$

The variance terms in Table C.20 are then found from

$$\begin{aligned} \frac{1}{\sigma^2} \begin{bmatrix} V_{\bar{\mu}} & V_{(\bar{\mu}, \bar{\sigma})} \\ V_{(\bar{\mu}, \bar{\sigma})} & V_{\bar{\sigma}} \end{bmatrix} &= \frac{n}{\sigma^2} \begin{bmatrix} Avar(\bar{\mu}) & Acov(\bar{\mu}, \bar{\sigma}) \\ Acov(\bar{\mu}, \bar{\sigma}) & Avar(\bar{\sigma}) \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix}^{-1} \\ &= \frac{1}{f_{11}f_{22} - f_{12}^2} \begin{bmatrix} f_{22} & -f_{12} \\ -f_{12} & f_{11} \end{bmatrix} \\ &= \frac{1}{0.96841 \times 1.56779 - (-0.11490)^2} \begin{bmatrix} 0.96841 & 0.11490 \\ 0.11490 & 1.56779 \end{bmatrix} \\ &= \begin{bmatrix} 1.04168 & 0.07634 \\ 0.07634 & 0.64344 \end{bmatrix} \end{aligned}$$

The asymptotic correlation is then computed as

$$\rho(\bar{\mu}, \bar{\sigma}) = \frac{V_{(\bar{\mu}, \bar{\sigma})}}{\sqrt{V_{\bar{\mu}} V_{\bar{\sigma}}}} = \frac{0.07634}{\sqrt{1.04168 \times 0.64344}} = 0.09325$$

Finally, the scaled asymptotic variances for either a known  $\mu$  or a known  $\sigma$  are

$$\frac{n}{\sigma^2} Avar(\bar{\mu}|\sigma) = \frac{1}{\sigma^2} V_{\bar{\mu}|\sigma} = [f_{11}]^{-1} = [0.96841]^{-1} = 1.03262$$

$$\frac{n}{\sigma^2} Avar(\bar{\sigma}|\mu) = \frac{1}{\sigma^2} V_{\bar{\sigma}|\mu} = [f_{22}]^{-1} = [1.56779]^{-1} = 0.63784$$

### From the text problem # 10.18

Refer to Appendix Table C.20. What can you say about the effect that censoring has on the correlation between ML estimators  $\hat{\mu}$  and  $\hat{\sigma}$ ?

With no censoring we can see from the table that  $\hat{\mu}$  and  $\hat{\sigma}$  are independent. As the number of censored observations increases the correlation between  $\hat{\mu}$  and  $\hat{\sigma}$  becomes stronger. The intuition for this result is that with a very small percentage of units failing, the extrapolated estimate of  $\exp(\mu) = t_{0.5}$  is highly dependent.



## From the text problem # 11.1

Wilk, Gnanadesikan and Hyuett (1962b) give the number of weeks until failure for a sample of 34 transistors subjected to accelerated conditions. The reported failure times, with the number of ties shown in parentheses, were 3, 4, 5, 6(2), 7, 8(2), 9(3), 10(2), 11(3), 13(5), 17(2), 19(2), 25, 29, 33, 42(2), 52. The other 3 transistors had not failed at the end of 52 weeks.

- (a) Use ML to estimate the parameters of the BISA, IGAU, gamma and lognormal distributions to these data using discrete data likelihood. Plot all of these estimates on lognormal probability paper and compare the different estimates. Describe any important differences you see in the estimates.

As discussed in class, the problem does state how frequently the transistors were inspected. My solution (and therefore your solution) assumes that the transistors were inspected every week, thus each observed failure occurs during a one week interval. In RSplida, once the “life.data” object has been created the maximum likelihood parameter estimates for each of the requested distributions can be found with the following code.

For the above distributions the code returns the following parameter values.

Birnbaum-Sanders:  $(\theta, \beta) = (14.39088, 0.91967)$

Inverse-Gaussian:  $(\theta, \beta) = (20.87752, 0.92822)$

Gamma:  $(\theta, \kappa) = (13.00663, 1.51559)$

Lognormal:  $(\mu, \sigma) = (2.63022, 0.86299)$

The plot comparing these distributions on lognormal probability paper is shown below

- 
- (b) Redo the gamma distribution analyses assuming that the failures occurred at exactly the reported time. Are there differences of practical importance in this example?

By considering the failures times as exact results in the following plot comparing the same distributions as were compared in the plot from part (a) of this problem.

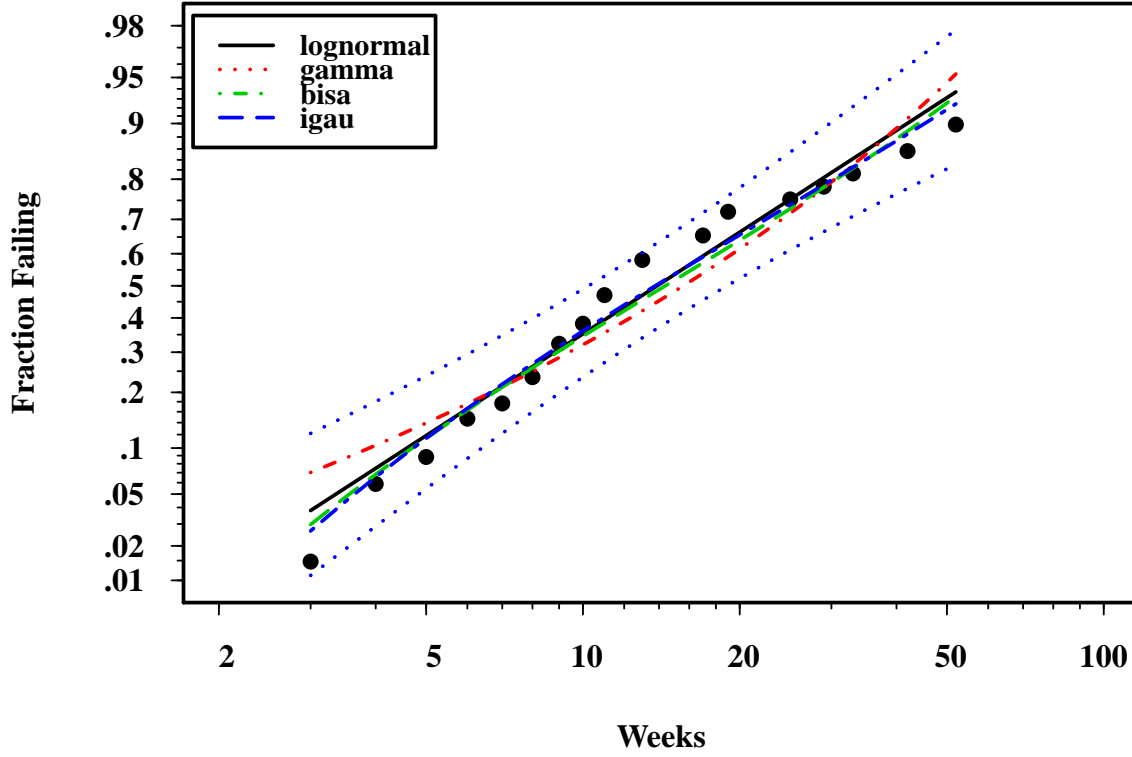


Figure 2: Plot comparing the lognormal, gamma, Birnbaum-Saunders and inverse gaussian distributions

### From the text problem # 11.12

During a single a company sold 2341 modems. These modems have a 36-month warranty. During the first 24 months, 75 of these modems had been returned for warranty service. Suppose that it is reasonable to assume that this is the number failing out of the 2341 modems that were sold. From previous experience with similar products, it is known that a Weibull distribution with a shape parameter  $\beta = .85$  provides an adequate description of the failure time distribution.

- (a) Show that the conditional probability of failing in the third year of life, given survival up to 2 years, can be expressed as a truncated distribution.

Recall that the conditional probability of event  $A$  occurring, given event  $B$  has already occurred is expressed as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

In this problem  $A$  represents the event that modem  $i$  fails at time  $T_i \in (2, 3]$  years and  $B$  represents the event that  $T_i > 2$ . The conditional probability requested in this problem may then be expressed as

$$P(2 < T_i \leq 3 | T_i > 2) = \frac{P(2 < T_i \leq 3)}{P(T_i > 2)} = \frac{F(3|\theta) - F(2|\theta)}{F(2|\theta)}.$$

This equation can be seen as that of a left truncated distribution where  $\tau^L$ . Replacing  $F(t|\theta)$  with the Weibull cdf gives

$$\frac{e^{(-\frac{2}{\eta})^{0.85}} - e^{(-\frac{3}{\eta})^{0.85}}}{e^{(-\frac{2}{\eta})^{0.85}}} = 1 - \frac{e^{(-\frac{3}{\eta})^{0.85}}}{e^{(-\frac{2}{\eta})^{0.85}}} = 1 - e^{(-\frac{0.7037}{\eta})^{0.85}}.$$

It should be noted that the left-truncated Weibull distribution is a special case in that the resulting model is still in the Weibull family. In general truncated distributions do not remain in the same family as the parent distribution.

- 
- (b) Although the times to failure for the returned modems were not available, it is still possible to compute an estimate of the Weibull cdf using the given value of  $\beta = .85$ . Show how to do this.

Since we are not excluding the 75 modems that were returned from the analysis we must treat them as left censored observations. The remaining  $2341 - 75 = 2266$  modems are then considered as right-censored observations.

Therefore, the likelihood function for  $\eta$ , given the data and the known shape parameter  $\beta = 0.85$  is expressed as

$$\begin{aligned}\mathcal{L}(\eta) &= \prod_{i=1}^{75} \left[ 1 - e^{-(\frac{24}{\eta})^{0.85}} \right] \prod_{i=76}^{2341} \left[ e^{-(\frac{24}{\eta})^{0.85}} \right] \\ &= \left( 1 - e^{-(\frac{24}{\eta})^{0.85}} \right)^{75} \left( e^{-2266(\frac{24}{\eta})^{0.85}} \right).\end{aligned}$$

The log-likelihood function is then expressed as

$$l(\eta) = 75 \times \log \left( 1 - e^{-(\frac{24}{\eta})^{0.85}} \right) - 2266 \left( \frac{24}{\eta} \right)^{0.85}$$

Maximizing the log-likelihood function results in

$$\hat{\eta}_{MLE} \approx 1349 \text{ months}$$

giving an estimate for the Weibull cdf at any value of  $t$ .

If we had chosen to exclude these 75 observations, the appropriate way forward would have been to model the remaining  $2341 - 75 = 2266$  modems as right-censored observations from a left-truncated distribution where  $\tau^L = 24$  months as was found in part (a) of this problem.

- 
- (c) Use the estimate from part (b) to compute an estimate for the number of units that will fail in the third year of operation.

Using the estimate for the Weibull cdf that was found in part (b) of this problem, we can estimate the number of units that will in the third year of life as

$$\exp \left( \frac{24}{1349} \right)^{0.85} - \exp \left( \frac{36}{1349} \right)^{0.85} = 0.0129$$

In R, this can be accomplished using

```
pweibull(36,shape = .85,scale = 1349)-pweibull(24,shape = .85,scale = 1349)
```

```
[1] 0.012882
```

- 
- (d) Suppose that you had learned that 2 years after being sold, between 5% and 10% of the purchased modems had never been put into service. How would you do part (b)?

If we were to reaccomplish the analysis with the updated information that 5% - 10% of the modems sold were never put into service the most appropriate course of action would be to reduce the number of right censored observations. The updated likelihood functions for the 5% and 10% cases would then be expressed as

$$\mathcal{L}(\eta) = \prod_{i=1}^{75} \left[ 1 - e^{-\left(\frac{24}{\eta}\right)^{0.85}} \right] \prod_{i=76}^{2341(.95)} \left[ e^{-\left(\frac{24}{\eta}\right)^{0.85}} \right] \quad 5\% \text{ unused}$$

$$\mathcal{L}(\eta) = \prod_{i=1}^{75} \left[ 1 - e^{-\left(\frac{24}{\eta}\right)^{0.85}} \right] \prod_{i=76}^{2341(.90)} \left[ e^{-\left(\frac{24}{\eta}\right)^{0.85}} \right] \quad 10\% \text{ unused}$$

and the respective log-likelihood functions would be

$$l(\eta) = 75 \times \log \left( 1 - e^{-\left(\frac{24}{\eta}\right)^{0.85}} \right) - 2152.7 \left( \frac{24}{\eta} \right)^{0.85} \quad 5\% \text{ unused}$$

$$l(\eta) = 75 \times \log \left( 1 - e^{-\left(\frac{24}{\eta}\right)^{0.85}} \right) - 2039.4 \left( \frac{24}{\eta} \right)^{0.85} \quad 10\% \text{ unused}$$