Survival Anaylsis Weibull

Victor Feagins

10/27/2021

### Question 1

The following data are the survival times in years of 21 6MP patients whom we studied earlier in the semester. The data are given in LW and in the lecture notes on page 112. Their survival times are given in years. Obtain the MLE of the parameters and median survival times assuming a) one parameter exponential and b) a Weibull distribution and c) compare the two fit plots.

#### Data

Table :First 5 Censored and 5 Uncensored Observations of 6MP Patients

| time | censor |
| --- | --- |
| 0.150 | 0 |
| 0.150 | 0 |
| 0.150 | 0 |
| 0.175 | 0 |
| 0.250 | 0 |
| 0.150 | 1 |
| 0.225 | 1 |
| 0.250 | 1 |
| 0.275 | 1 |
| 0.425 | 1 |

#### Exponential Model

surv.1 <- Surv(df.1$time, abs(df.1$censor - 1))

exp.fit.1 <- survreg(surv.1 ~ 1, dist = "exponential")  
summary(exp.fit.1)

##   
## Call:  
## survreg(formula = surv.1 ~ 1, dist = "exponential")  
## Value Std. Error z p  
## (Intercept) -0.00278 0.33333 -0.01 0.99  
##   
## Scale fixed at 1   
##   
## Exponential distribution  
## Loglik(model)= -9 Loglik(intercept only)= -9  
## Number of Newton-Raphson Iterations: 4   
## n= 21

The median survival time for the exponential fit is 0.6912218 and the MLE estimator for the Intercept is -0.0027816

#### Weibull Model

wei.fit.1 <- survreg(surv.1 ~ 1, dist = "weibull")  
summary(wei.fit.1)

##   
## Call:  
## survreg(formula = surv.1 ~ 1, dist = "weibull")  
## Value Std. Error z p  
## (Intercept) -0.169 0.273 -0.62 0.54  
## Log(scale) -0.303 0.278 -1.09 0.28  
##   
## Scale= 0.739   
##   
## Weibull distribution  
## Loglik(model)= -8.5 Loglik(intercept only)= -8.5  
## Number of Newton-Raphson Iterations: 5   
## n= 21

Fitting a Weibull model we have an Intercept of -0.1694502 and Scale of 0.7386973. This equates to an weibull scale or of 0.8441288 and weibull shape or of 1.3537345.

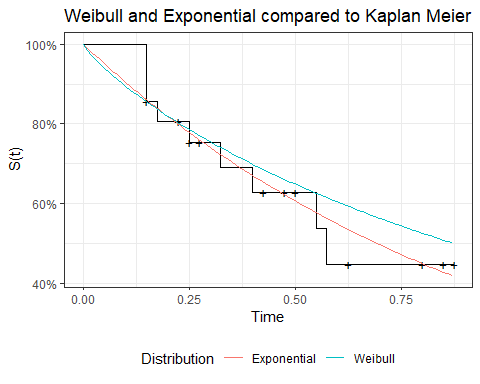
The median survival times is 0.6439126

#### Comparing both models

Let’s plot both models compared to a Kaplan Meier Curve.

km.fit.1 <- survfit(surv.1 ~ 1) # Fits a Kaplan Meier Curve

autoplot(km.fit.1, conf.int = F) +  
 geom\_line(data = predict.exp, mapping = aes(x = x, y = y, col = "Exponential")) +  
 geom\_line(data = predict.wei, mapping = aes(x = x, y = y, col = "Weibull")) +  
 xlim(0,max(df.1$time)) +  
 labs(color = "Distribution",  
 x = "Time",  
 y = "S(t)",   
 title = "Weibull and Exponential compared to Kaplan Meier") +  
 theme(legend.position="bottom")



Looking at the Exponential fit vs Weibull fit we see that the Exponential fits the Kaplan Meier closer then the Weibull.

Examining the -2 log likelihood:

* Exponential: 17.9499304
* Weibull: 16.9175268

Keep in mind these are on the unlogged observations. We see that the Exponential has a higher value indicating a better fit. By looking at the fits and log likelihood we can conclude that the Exponential is a better fit of the data.

### Question 2

Consider the tumor free days in ten animals. Assume the data follow a log-logistic distribution. Their times are given in days. Obtain the MLE of the parameters and median survival times assuming a) a Weibull distribution and b) a Log-logistic distribution and c) compare the two fit plots.

#### Data

Table :First 5 Observations of Tumor Free days in Animals

| time |
| --- |
| 2.0 |
| 3.5 |
| 5.0 |
| 7.0 |
| 9.0 |

#### Weibull Model

surv.2 <- Surv(df.2$time)

wei.fit.2 <- survreg(surv.2 ~ 1, dist = "weibull")  
summary(wei.fit.2)

##   
## Call:  
## survreg(formula = surv.2 ~ 1, dist = "weibull")  
## Value Std. Error z p  
## (Intercept) 2.723 0.270 10.08 <2e-16  
## Log(scale) -0.215 0.244 -0.88 0.38  
##   
## Scale= 0.807   
##   
## Weibull distribution  
## Loglik(model)= -36.1 Loglik(intercept only)= -36.1  
## Number of Newton-Raphson Iterations: 6   
## n= 10

Fitting a Weibull model we have an Intercept of 2.7232078 and Scale of 0.8068817.

The median survival times is 11.3302396

#### Log-Logistic Model

loglog.fit.2 <- survreg(surv.2 ~ 1, dist = "loglogistic")  
summary(loglog.fit.2)

##   
## Call:  
## survreg(formula = surv.2 ~ 1, dist = "loglogistic")  
## Value Std. Error z p  
## (Intercept) 2.288 0.300 7.64 2.2e-14  
## Log(scale) -0.625 0.259 -2.42 0.016  
##   
## Scale= 0.535   
##   
## Log logistic distribution  
## Loglik(model)= -36.3 Loglik(intercept only)= -36.3  
## Number of Newton-Raphson Iterations: 5   
## n= 10

Fitting a Log-Logistic model we have an Intercept of 2.2884703 and Scale of 0.5351704.

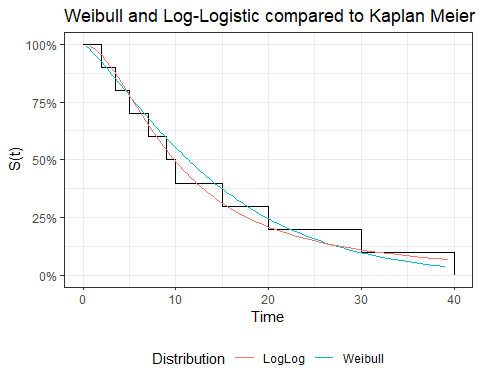
The median survival times is 9.8598432

#### Comparing Both Models

Let’s plot both models compared to a Kaplan Meier Curve.

km.fit.2 <- survfit(surv.2 ~ 1) # Fits a Kaplan Meier Curve

autoplot(km.fit.2, conf.int = F) +  
 geom\_line(data = predict.wei.2, mapping = aes(x = x, y = y, col = "Weibull")) +  
 geom\_line(data = predict.loglog.2, mapping = aes(x = x, y = y, col = "LogLog")) +  
 xlim(0,max(df.2$time)) +  
 labs(color = "Distribution",  
 x = "Time",  
 y = "S(t)",   
 title = "Weibull and Log-Logistic compared to Kaplan Meier") +  
 theme(legend.position="bottom")



In this example the Weibull and Log-Logistic are very similar to each other and both follow the Kaplan Meier Curve closely.

Examining the -2 log likelihood:

* Exponential: 72.2965578
* Log-Logistic: 72.5255399

Keep in mind these are on the unlogged observations. We see that both distributions have similar loglikihood values. I conclude that either distribution would fit the data correctly and that we should consult with the assumptions that distribution makes about the survival process.