Some likelihood theory, with application to estimation for exponential and Weibull distributed censored data

TMA4275 Spring 2013
Bo Lindqvist

Suppose we have right-censored survival data of the form $(y_1, \delta_1), \ldots, (y_n, \delta_n)$, where the y_i are observed times and δ_i are the censoring indicators, with $\delta_i = 1$ for an observed failure and $\delta_i = 0$ for a censored observation.

Consider the modeling of the lifetimes by a parametric distribution with probability density $f(t;\theta)$ and survival function $R(t;\theta)$. Here θ may well be a vector, i.e. there may be more than one unknown parameter in the model. In order to make statistical inference about θ we need the *likelihood function* for θ based on our data.

The likelihood function is given by

$$L(\theta) = \prod_{i:\delta_i=1} f(y_i;\theta) \prod_{i:\delta_i=0} R(y_i;\theta).$$

One parameter: exponential distribution

Suppose θ is a single parameter. As an example, assume exponential lifetimes, with

$$f(t;\theta) = \frac{1}{\theta} e^{-\frac{t}{\theta}}, \;\; R(t;\theta) = e^{-\frac{t}{\theta}}.$$

Then we get

$$L(\theta) = \frac{1}{\theta^r} e^{-\frac{s}{\theta}}, \tag{1}$$

where

$$\begin{aligned} r &= \sum_{i=1}^{n} \delta_i = \text{ number of observed failures,} \\ s &= \sum_{i=1}^{n} y_i = \text{ total time on test.} \end{aligned}$$

The *log-likelihood* is given by

$$l(\theta) =_{def} \ln L(\theta) = -r\ln(\theta) - \frac{s}{\theta}. \tag{2}$$

The maximum likelihood estimator (MLE) $\hat{\theta}$ for θ is found by solving the likelihood equation

$$l'(\theta) = 0.$$

Here the equation is

$$l'(\theta) = -\frac{r}{\theta} + \frac{s}{\theta^2} = 0$$

so the MLE for θ is

$$\hat{\theta} = \frac{s}{r}. \tag{3}$$

The general theory of maximum likelihood estimation tells us that $\hat{\theta}$ is approximately (more precisely: *asymptotically*) unbiased, with a *standard error* (i.e. standard deviation) which can be estimated as described in the following.

First we define the observed information about θ in $\hat{\theta}$,

$$I(\hat{\theta}) =_{\text{def}} -l''(\hat{\theta}),$$

the double derivative of the log likelihood, computed at the MLE $\hat{\theta}$.

The theory then says that we can estimate the variance of $\hat{\theta}$ by

$$\widehat{Var(\hat{\theta})} = \frac{1}{I(\hat{\theta})}.$$

In our example we have $l''(\theta) = \frac{r}{\theta^2} - \frac{2s}{\theta^3}$, so the observed information is

$$\begin{aligned}
I(\hat{\theta}) &= -l''(\hat{\theta}) \\
&= -\frac{r}{\hat{\theta}^2} + \frac{2s}{\hat{\theta}^3} \\
&= -\frac{r}{(\frac{s}{r})^2} + \frac{2s}{(\frac{s}{r})^3} \\
&= -\frac{r^3}{s^2} + \frac{2r^3}{s^2} \\
&= \frac{r^3}{s^2} \\
&= \frac{r}{\hat{\theta}^2}
\end{aligned}$$

It follows that $\widehat{Var(\hat{\theta})} = \frac{1}{I(\hat{\theta})} = \frac{\hat{\theta}^2}{r}$, so the estimated standard deviation of the estimator $\hat{\theta}$ (*standard error*) is

$$\widehat{\text{SD}(\hat{\theta})} = \sqrt{\widehat{Var(\hat{\theta})}} = \frac{\hat{\theta}}{\sqrt{r}}.$$

Likelihood confidence interval and likelihood tests for θ

The log likelihood $l(\theta)$ can also be used to derive a confidence interval for θ. Likelihood theory tells us that

$$W(\theta) = 2\big(l(\hat{\theta}) - l(\theta)\big) \approx \chi_1^2 \tag{4}$$

if θ is the true value behind our data. Here χ_1^2 means the chi-square distribution with 1 degree of freedom.

Thus from a table of the χ^2-distribution we conclude that

$$P(2\big(l(\hat{\theta}) - l(\theta)\big) \leq 3.84) \approx 0.95,$$

which can also be phrased as

$$P(l(\theta) \geq l(\hat{\theta}) - 1.92) \approx 0.95.$$

Thus a confidence interval for θ can be found in the figure of page 74 in Slides 3 as the set of θ for which $l(\theta) \geq -12.63 - 1.92 = -14.55$. This gives the interval from 2.1 to 12.8 which is hence an approximate 95% confidence interval.

The result (4) can also be used to test the null hypothesis that θ has a prespecified value θ_0, e.g. $\theta_0 = 10$. Then under the null hypothesis, $W(\theta_0)$ is approximately χ_1^2 and we should then reject the null hypothetsis if $W(\theta_0) > 3.84$ (if we use 5% significance level). With the same data as above we get $W(10) = 2(-12.63 - (-13.81)) = 2.36$, so we do not reject the hypothesis.

Note also that in general we can test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ with significance level 5% if we have a 95% confidence interval for θ. We then reject the null hypothesis if θ_0 is not in the confidence interval. In the case considered in the previous paragraph, it is easily seen that the suggested test is of this kind.

Two parameters: Weibull distribution

Suppose now that there are two parameters in the model, and let us use the Weibull distribution for illustration.

Now

$$f(t;\theta,\alpha) = \frac{\alpha}{\theta^\alpha} t^{\alpha-1} e^{-(\frac{t}{\theta})^\alpha}, \quad R(t;\theta,\alpha) = e^{-(\frac{t}{\theta})^\alpha},$$

and the likelihood function and the log-likelihood functions become

$$\begin{aligned} L(\theta,\alpha) &= \frac{\alpha^r}{\theta^{r\alpha}}\Big(\prod_{\delta_i=1} y_i\Big)^{\alpha-1} e^{-\frac{1}{\theta^\alpha}\sum_{i=1}^n y_i^\alpha} \\ l(\theta,\alpha) &= r\ln\alpha - \alpha r\ln\theta + (\alpha-1)\sum_{i:\delta_i=1}\ln y_i - \frac{1}{\theta^\alpha}\sum_{i=1}^n y_i^\alpha. \end{aligned} \tag{5}$$

The maximum likelihood estimators $\hat\theta, \hat\alpha$ are found by maximizing $l(\theta,\alpha)$. In practice this is done by solving the likelihood equations (two equations in two unknowns),

$$\frac{\partial l(\theta,\alpha)}{\partial\theta} = 0, \qquad \frac{\partial l(\theta,\alpha)}{\partial\alpha} = 0.$$

This gives the equations

$$\begin{aligned} 1. \quad \frac{\partial l(\theta,\alpha)}{\partial\theta} &= -\frac{\alpha r}{\theta} + \frac{\alpha}{\theta^{\alpha+1}}\sum_{i=1}^n y_i^\alpha = 0 \\ 2. \quad \frac{\partial l(\theta,\alpha)}{\partial\alpha} &= \frac{r}{\alpha} - r\ln\theta + \sum_{i=1}^n \delta_i \ln y_i - \sum_{i=1}^n \big(\frac{y_i}{\theta}\big)^\alpha \ln\big(\frac{y_i}{\theta}\big) = 0 \end{aligned}$$

Here we may for example solve eq. 1 for θ as a function of α and substitute this into eq. 2. From eq. 1 we get

$$\begin{aligned} \frac{\alpha r}{\theta} &= \frac{\alpha}{\theta^{\alpha+1}}\sum_{i=1}^n y_i^\alpha \\ \theta^\alpha &= \frac{\sum_{i=1}^n y_i^\alpha}{r} \\ \theta &= \Big(\frac{\sum_{i=1}^n y_i^\alpha}{r}\Big)^{1/\alpha}. \end{aligned}$$

Note that this solution is the maximum likelihood estimator of θ if the value of α is *known*. We denote it by

$$\hat\theta(\alpha) = \Big(\frac{\sum_{i=1}^n y_i^\alpha}{r}\Big)^{1/\alpha}. \tag{6}$$

In particular we see that if $\alpha = 1$, then we get the estimator (3) of θ for the exponential model. In fact, $L(\theta, 1)$ is seen to equal the likelihood function (1) for the exponential model.

As noted above, in order to find $\hat\theta, \hat\alpha$, we may substitute the solution $\hat\theta(\alpha)$ into equation 2 and thereby get an equation with just one unknown, namely α. The solution of this is $\hat\alpha$, the MLE of α, while $\hat\theta = \hat\theta(\hat\alpha)$ is the MLE of θ.

The profile (log) likelihood function

We have shown how we in principle can find the MLE of θ and α. Now we shall see that we may also use $\hat{\theta}(\alpha)$ in (6) to obtain a likelihood function for α alone, called the *profile log likelihood function*. We then substitute $\hat{\theta}(\alpha)$ into the log likelihood function $l(\theta,\alpha)$ to get

$$\tilde{l}(\alpha) = l(\hat{\theta}(\alpha),\alpha) \equiv \max_{\theta} l(\theta,\alpha).$$

In the computation of $\tilde{l}(\alpha)$ we thus answer the question: "How big can we get $l(\theta,\alpha)$ for a given α"? Note that $\tilde{l}(1) = l(\hat{\theta}(1),1)$ equals the maximum value of the log likelihood $l(\theta)$ in (2) for the exponential distribution.

For the Weibull case we compute $\tilde{l}(\alpha)$ as follows,

$$\begin{aligned}
\tilde{l}(\alpha) &= r\ln\alpha - \alpha r \ln\left[\left(\frac{\sum y_i^{\alpha}}{r}\right)^{1/\alpha}\right] + (\alpha-1)\sum \delta_i \ln y_i - \frac{1}{\left(\frac{\sum y_i^{\alpha}}{r}\right)}\sum_{i=1}^{n} y_i^{\alpha} \\
&= r\ln\alpha - \alpha r\frac{1}{\alpha}\ln\left(\sum y_i^{\alpha}\right) + \alpha r\frac{1}{\alpha}\ln r + (\alpha-1)\sum \delta_i y_i - r \\
&= r\ln\alpha - r\ln\left(\sum y_i^{\alpha}\right) + r\ln r + (\alpha-1)\sum \delta_i \ln y_i - r.
\end{aligned}$$

We may plot this as a function of α to get $\hat{\alpha}$ at the maximum of the curve. An example is given on page 79 in Slides 3. Here $\hat{\alpha} = 0.978$. Then we get $\hat{\theta}$ given as $\hat{\theta}(\hat{\alpha}) = 6.880$. Further, we estimate the mean time to failure by

$$\widehat{\text{MTTF}} = \hat{\theta}\,\Gamma\left(\frac{1}{\hat{\alpha}}+1\right) = 6.88\,\Gamma\left(\frac{1}{0.978}+1\right) = 6.88\Gamma(2.0225) = 6.9469$$

where we used the formula for the expected value of the Weibull distribution.

Computation of standard errors of $\hat{\theta}$, $\hat{\alpha}$.

Recall that for the case of a single parameter θ we defined the observed information $I(\hat{\theta}) = -l''(\hat{\theta})$ and obtained the estimator $\widehat{Var}(\hat{\theta}) = 1/I(\hat{\theta})$.

In the case of two parameters θ and α we define the *observed information matrix* to be

$$I(\hat{\theta},\hat{\alpha}) =_{def} \begin{bmatrix} -\frac{\partial^2 l(\theta,\alpha)}{\partial\theta^2} & -\frac{\partial^2 l(\theta,\alpha)}{\partial\theta\partial\alpha} \\ -\frac{\partial^2 l(\theta,\alpha)}{\partial\alpha\partial\theta} & -\frac{\partial^2 l(\theta,\alpha)}{\partial\alpha^2} \end{bmatrix}_{\theta=\hat{\theta},\alpha=\hat{\alpha}}$$

This matrix is also known under the name of *Hessian matrix*.

The theory of maximum likelihood says that

$$\left[I(\hat{\theta},\hat{\alpha})\right]^{-1} = \begin{bmatrix} \widehat{Var(\hat{\theta})} & \widehat{Cov(\hat{\theta},\hat{\alpha})} \\ \widehat{Cov(\hat{\alpha},\hat{\theta})} & \widehat{Var(\hat{\alpha})} \end{bmatrix}$$

which means that by inverting the observed information matrix we get a matrix with the estimated variances of the parameters on the diagonal. We furthermore get estimated covariances outside the diagonal. These are used for computation of estimated variances of functions of both θ and α.

From the estimated variances we compute standard errors by taking square roots, and we may then compute either standard confidence intervals or *standard confidence intervals for positive parameters*. The latter are given as

$$\hat{\theta}e^{\pm 1.96\frac{\widehat{\text{SD}(\hat{\theta})}}{\hat{\theta}}}, \qquad \hat{\alpha}e^{\pm 1.96\frac{\widehat{\text{SD}(\hat{\alpha})}}{\hat{\alpha}}}$$

The inverse of the observed information matrix ca n also be used to compute standard errors and confidence intervals for, e.g., MTTF, median and, more generally, percentiles t_p. For this we need the covariance of $\hat{\theta},\hat{\alpha}$. We will not give details here, but only note that Minitab does the computation for us!

Likelihood confidence interval for α

The profile log likelihood $\tilde{l}(\alpha)$ can be used in the same manner as the log likelihood for a single parameter to obtain a confidence interval for α. More precisely, likelihood theory tells us that

$$W(\alpha) = 2\big(\tilde{l}(\hat{\alpha}) - \tilde{l}(\alpha)\big) \approx \chi_1^2 \tag{7}$$

if α is the true value. Thus an approximate 95% confidence interval for α can be found in the figure of page 79 in Slides 3 as the set of α for which $\tilde{l}(\alpha) \geq -14.584 - 1.92 = -16.50$ (which unfortunately is not covered by the graph).

Testing for exponentiality

The result (7) can also be used to test the null hypothesis that the data come from an exponential distribution. This is done by testing

$$H_0 : \alpha = 1 \quad \text{vs.} \quad H_1 : \alpha \neq 1,$$

and using the following test statistic:

$$W(1) = 2\big(\tilde{l}(\hat{\alpha}) - \tilde{l}(1)\big)$$

which is approximately χ_1^2 if the null hypothesis is true. Hence we reject the null hypothesis at 5% significance level if $W(1) \geq 3.84$.

Note that

$$W(1) = 2\big[\underbrace{l(\hat{\theta}, \hat{\alpha})}_{\text{max value inWeibull model}} - \underbrace{l(\hat{\theta}(1), 1)}_{\text{max value in exponential model}} \big]$$

where we can find the two maxima in the Minitab output for, respectively, Weibull-distribution and exponential distribution. In the present case we get $l(\hat{\theta}, \hat{\alpha}) = -14.576$, and $l(\hat{\theta}(1), 1) = -14.577$. Thus, $W(1) = 2(-14.576 - (-14.577)) = 0.002$ which is much too small to reject the null hypothesis! The conclusion is hence that there is not enough evidence to conclude that the data are not exponentially distributed.

Note that we may use the same idea as above to test for other values of α, i.e. we may test $H_0 : \alpha = \alpha_0$ for any given number α_0 by rejecting the null hypothesis if $W(\alpha_0) \geq 3.84$. (But then we will not find the value $W(\alpha_0)$ so easily from the Minitab outputs.)

Probability plot for Weibull-distribution

Suppose $T \sim$ Weibull(θ, α). Then we get

$$\begin{aligned} R(t) &= e^{-\left(\frac{t}{\theta}\right)^\alpha} \\ -\ln R(t) &= \left(\frac{t}{\theta}\right)^\alpha \\ \ln(-\ln R(t)) &= \alpha \ln t - \alpha \ln \theta. \end{aligned} \tag{8}$$

It follows from this that, for any t, the point $\big[\ln t, \ln(-\ln R(t))\big]$ is on the line

$$y = \alpha x - \alpha \ln \theta$$

which we here consider as a line in the "ordinary" (x, y) coordinate system. Thus α is the slope of the line, while $-\alpha \ln \theta$ is the intercept.

Suppose now that we have a right-censored data set $(y_1, \delta_1), \ldots, (y_n, \delta_n)$, where $t_{(1)} < \ldots < t_{(k)}$ are the observed *failure* times (i.e. uncensored times). Then we can compute the Kaplan-Meier estimator $\hat{R}(t)$ for $R(t)$.

The *Weibull probability plot* is a plot of the points $\big[\ln t_{(i)}, \ln(-\ln \hat{R}(t_{(i)}))\big]$ in the (x, y) coordinate system. If the Weibull model is the correct one, then these points will tend to be close to a straight line since $\hat{R}(t)$ then is supposed to be close to the underlying Weibull survival function (8). If, on the other hand, the data are not Weibull-distributed, then $\hat{R}(t)$ may be far from a Weibull survival function and consequently the points may not follow a straight line.

A modified KM-estimator is sometimes used, defined by

$$\hat{\hat{R}}(t_{(i)}) = \frac{\hat{R}(t_{(i)}) + \hat{R}(t_{(i-1)})}{2}.$$

This gives a somewhat more “smooth” KM-curve.

Minitab plots the points $\big[\ln t_{(i)}, \ln(-\ln \hat{\hat{R}}(t))\big]$ together with the straight line $y = \hat{\alpha} x - \hat{\alpha} \ln \hat{\theta}$, where the $\hat{\alpha}$ and $\hat{\theta}$ are the estimated parameters based on a Weibull-distribution. The Weibull model is hence considered to be a good model for the data if the points are close to this line.

Exponential Distribution

Minitab uses essentially the same probability plot for the exponential distribution as for the Weibull distribution. But since $\alpha = 1$ for the exponential distribution, the plotted line is $y = x - \ln \hat{\theta}$, where $\hat{\theta}$ is the estimate of θ obtained from the exponential model.

Remark on Minitab’s probability plots

Actually, a slightly different modified KM-estimator than the $\hat{\hat{R}}(t)$ is used as default by Minitab. Further, Minitab plots a 95% confidence band around the estimated line to help in the judgement of when the points are “close enough” to the line.