

Statistics 520 Computing Notes

Basic Likelihood Estimation in Practice

These notes illustrate some of the basic computational procedures involved in making use of maximum likelihood estimation and inference. Here, we focus on fundamental one and two sample problems.

1 An Illustratory Problem

In problems involving time-to-event data, such as survival analysis, random variables represent the time of the event (e.g., death) for sampling units, starting at some appropriately defined origin. A popular choice of distribution to model a set of independent and identically distributed random variables in such problems is the Weibull distribution. The probability density and distribution functions of a random variable T that follows a Weibull distribution are, for $\alpha > 0$ and $\beta > 0$,

$$\begin{aligned} f(t|\alpha, \beta) &= \beta^\alpha \alpha t^{\alpha-1} \exp[-(\beta t)^\alpha]; \quad t > 0, \\ F(t|\alpha, \beta) &= \begin{cases} 0 & t < 0 \\ 1 - \exp[-(\beta t)^\alpha] & t > 0. \end{cases} \end{aligned} \quad (1)$$

The expected value and variance are,

$$\begin{aligned} E(T) &= \frac{1}{\beta} \Gamma\left(\frac{\alpha+1}{\alpha}\right), \\ \text{var}(T) &= \frac{1}{\beta^2} \left[\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \Gamma\left(\frac{\alpha+1}{\alpha}\right)^2 \right]. \end{aligned} \quad (2)$$

An important function in survival analysis is the survival function $S(t)$, defined as the probability that $T > t$. For the Weibull model,

$$S(t|\alpha, \beta) = 1 - F(t|\alpha, \beta) = \exp[-(\beta t)^\alpha]. \quad (3)$$

Another important part of survival analysis is the hazard function, defined as the probability of death in an infinitesimal interval at time t given survival up to time t , which is $f(t)/S(t)$. For the Weibull distribution,

$$h(t|\alpha, \beta) = \frac{f(t|\alpha, \beta)}{S(t|\alpha, \beta)} = \alpha\beta^\alpha t^{\alpha-1}; \quad t > 0. \quad (4)$$

The Weibull model has an increasing hazard function if $\alpha > 1$ and a decreasing hazard function if $\alpha < 1$. If $\alpha = 1$ the hazard function is a constant and, in fact, the Weibull distribution reduces to an exponential distribution.

On the course web page in the Data module is a file `weibulldat.txt` that contains data from a hypothetical survival study with two treatments. The first column of this file is named `trt` having values 1 and 2. The second column contains observed time to death in some appropriate units (perhaps years). The values are recorded to three decimal places. Our objective is to fit Weibull distributions to data from the two treatments using maximum likelihood, test a full model having separate Weibull distributions for the two treatment groups against a reduced model having a common Weibull distribution for the two groups, estimate expected lifetimes for the two treatments with associated 95% confidence intervals, and produce estimated survival functions, along with 95% pointwise confidence intervals. In the Computing Functions module of the course web page is a file `weibulllikfctns.txt` that contains the functions we will use to accomplish these objectives.

2 Likelihood Estimation

We certainly could use the generic Newton-Raphson algorithm provided on the course web page to locate maximum likelihood estimates in this problem, but we will first illustrate the use of the black-box optimizers `optim` and `nlm` in base R. To make use of these functions we will need to write a separate function that evaluates the (negative) log likelihood for a one sample Weibull model. By independence, the

log likelihood for a sample is

$$\ell(\alpha, \beta) = \sum_{i=1}^n \ell_i(\alpha, \beta),$$

where,

$$\ell_i(\alpha, \beta) = \alpha \log(\beta) + \log(\alpha) + (\alpha - 1) \log(t_i) - (\beta t_i)^\alpha. \quad (5)$$

An R function to evaluate this likelihood suitable for use with the R functions `nlm` and `optim` is as follows.

```
weibullnegloglik<-function(pars,dat){
#negative log likelihood for one sample Weibull model
#for use with optim and nlm
#pars is alpha, beta
#dat is vector of responses
#
alp<-pars[1]; bet<-pars[2]
lis<-alp*log(bet) + log(alp) + (alp-1)*log(dat) - (bet*dat)^alp
llik<-sum(lis)
nllik<-(-1)*llik
return(nllik)
}
```

2.1 Estimation Using Black Box Optimizers

To use this function in conjunction with `nlm` with starting values $\alpha = 1$ and $\beta = 0.6$, for example, we would issue the command,

```
restrt1<-nlm(weibullnegloglik,c(1, 0.6),hessian=T,dat=weibdat1)
restrt1a<-optim(c(1,0.6),weibullnegloglik,hessian=T,dat=weibdat1)
```

There are several things we want to check for in the output from `nlm` or `optim`. One is the convergence code, which should be 1 for `nlm` and 0 for `optim`. Another is the gradient (only produced by `nlm`), which should be very small, near 0. There are several additional steps we can take to ensure that the algorithm has worked as intended.

1. Run the algorithm several more times with different starting values. The algorithm is not guaranteed to converge for any starting values, but one wants to ensure it is working is to make sure that, for any starting values that result in convergence, the convergence is to the same place.
2. Run `optim` as well as `nlm` to verify that the convergence of these two different algorithms is to the same place.

One of the outputs of either `nlm` or `optim` if you specify the `hessian=T` option is the matrix of second derivatives of the objective function. If the objective function was the log likelihood, that would correspond to the negative of the observed information matrix, as we discussed in class. But the objective function we gave to these optimizers was the negative log likelihood, which means the hessian here is equal to the observed information. To obtain 95% Wald theory intervals for the parameters, then, simple requires,

```
Invinf<-solve(restrt1$hessian)
restrt1$estimate-1.96*sqrt(diag(Invinf))
restrt1$estimate+1.96*sqrt(diag(Invinf))
```

or, if we had used `optim`

```
Invinf<-solve(restrt1$hessian)
restrt1$par-1.96*sqrt(diag(Invinf))
restrt1$par+1.96*sqrt(diag(Invinf))
```

Suppose the response variables for treatment 1 are in the object `weibdat1`. We can obtain a histogram of the data with the estimated density overlain as follows.

```
hist(weibdat1,prob=T,xlab='Time',ylab='Density',main='')
ts<-seq(0,3.5,0.001)
fs<-dweibull(ts,rest1$par[1],1/rest1$par[2])
lines(ts,fs)
```

Notes:

1. Always use the option `prob=T` with `hist` so that the histogram is on a probability scale.
2. Defining the range and spacing of points at which to evaluate the density may require some trial and error. Find the minimum and maximum values in the data and choose a reasonably small interval.
3. The functions `dweibull`, `rweibull`, `pweibull` use a parameterization for which the second parameter (my β) is the reciprocal, which is why the command `dweibull` is written the way it is above.

Plots of the data along with fitted density functions are given in Figure 1.

2.2 Estimation Using Newton-Raphson

An alternative for maximum likelihood estimation is provided by the generic Newton-Raphson algorithm contained in the file `newtraph.txt` in the Computing Functions module on the course web page. There is also a document that details the syntax, output, and use of this function in the file `usingnewtonraphson.pdf` in the Computing Notes module. Essentially, what is necessary to make use of the generic Newton-Raphson algorithm provided is to write a function that evaluates the log

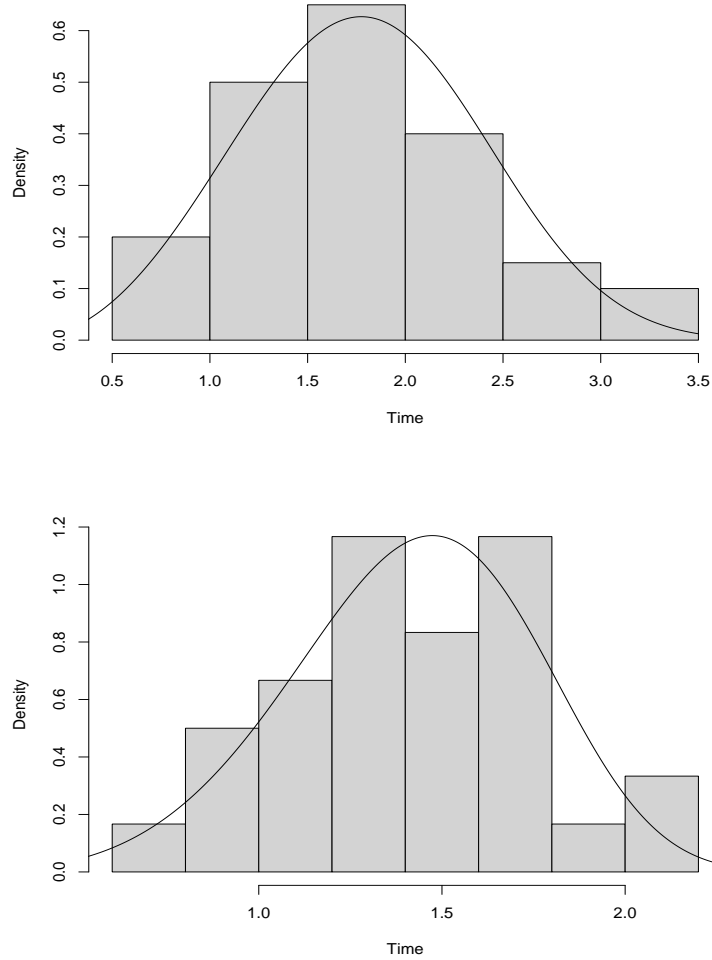


Figure 1: Histograms of data and fitted Weibull density functions for two treatments.

likelihood, first derivatives, and second derivatives. The necessary form of the output of this derivative function is detailed in the document on using **newtraph**. A function called **weibullders** is included in the file **weibulllikfctns.txt** file in the Computational Functions module of the course web page. This is an example of the type of derivative function that we need to use **newtraph**. To understand this

function, first note that the first and second derivatives of (5) are,

$$\begin{aligned}
\frac{\partial \ell_i}{\partial \alpha} &= \log(\beta) + \frac{1}{\alpha} + \log(t_i) - (\beta t_i)^\alpha \log(\beta t_i), \\
\frac{\partial \ell_i}{\partial \beta} &= \frac{\alpha}{\beta} - \alpha(\beta t_i)^{\alpha-1} t_i, \\
\frac{\partial^2 \ell_i}{\partial \alpha^2} &= \frac{-1}{\alpha^2} - (\beta t_i)^\alpha [\log(\beta t_i)]^2, \\
\frac{\partial^2 \ell_i}{\partial \beta^2} &= \frac{-\alpha}{\beta^2} - \alpha(\alpha-1)(\beta t_i)^{\alpha-2} t_i^2, \\
\frac{\partial^2 \ell_i}{\partial \alpha \partial \beta} &= \frac{1}{\beta} - \alpha(\beta t_i)^{\alpha-1} \log(\beta t_i) t_i - \frac{1}{\beta} (\beta t_i)^\alpha.
\end{aligned} \tag{6}$$

The function **weibullders** evaluates these derivatives at supplied parameter values and returns them in the form required by **newtraph**. First, we might try running **weibullders** at parameter values returned by **nlm** or **optim**. We can compare the gradient and hessian matrices produced by these functions. Then we can use **weibullders** in a call to **newtraph**,

```
nrests1<-newtraph(weibullders,weibulldat1,c(1,0.6))
nrests2<-newtraph(weibullders,weibulldat2,c(1,0.6))
```

The function **newtraph** outputs the inverse observed information matrix as the third element in a list. This matrix can be used to construct Wald theory intervals that can be compared to those already computed from the output of the black box optimizers.

2.3 Constructing a Likelihood Ratio Test

To construct a likelihood ratio test for difference between the treatment groups, suppose the response data are in **weibdat1** for treatment 1 and **weibdat2** for treatment 2. Estimation results are produced as already shown for treatment 1,

```
restrt1<-nlm(weibullnegloglik,c(1, 0.6),hessian=T,dat=weibdat1)
restrt2<-nlm(weibullnegloglik,c(1, 0.6),hessian=T,dat=weibdat2)
restrtall<-nlm(weibullnegloglik,c(1, 0.6),hessian=T,dat=c(weibdat1,weibdat2))
```

Maximized log likelihoods for a full model in which the treatment groups have different distributions, and a reduced model in which the treatment groups have the same distribution are obtained as,

```
llikF<-(-1)*restrt1$minimum - restrt2$minimum
llikR<-(-1)*restrtall$minimum
```

and then a likelihood ratio test statistic and associated p -value are obtained as,

```
T<-(-2)*(llikR-llikF)
p<-1-pchisq(T,2)
```

3 Additional Intervals

As given previously, the expected value of a Weibull distribution with parameters α and β is,

$$E(T) = \mu = \frac{1}{\beta} \Gamma \left(\frac{\alpha + 1}{\alpha} \right),$$

and $\hat{\mu}$ is this same expression evaluated at $\hat{\alpha}$ and $\hat{\beta}$. To compute a Wald interval for this expected value we make use of the delta method as follows.

$$\begin{aligned} \frac{\partial \mu}{\partial \alpha} &= \frac{1}{\beta} \Gamma' \left(\frac{\alpha + 1}{\alpha} \right) \left(\frac{-1}{\alpha^2} \right), \\ \frac{\partial \mu}{\partial \beta} &= -\frac{1}{\beta^2} \Gamma \left(\frac{\alpha + 1}{\alpha} \right). \end{aligned}$$

We want

$$D = \left(\frac{\partial \mu}{\partial \alpha}, \frac{\partial \mu}{\partial \beta} \right) \Big|_{\hat{\alpha}, \hat{\beta}}$$

Here, $\Gamma'(\cdot)$ is the derivative of the gamma function,

$$\Gamma'(\gamma) = \int_0^\infty u^{\gamma-1} \log(u) \exp(-u) du,$$

which we can compute either using the numerical integration function `integrate` in R, or by exploiting the fact that the `digamma` function returns the derivative of

$\log\{\Gamma(\cdot)\}$, namely, $\Gamma'(\gamma)/\Gamma(\gamma)$. Suppose we have saved the estimated value of α in an object named `ahat`, the estimated value of β in `bhat` and the observed inverse information in `Invinf`. Then we can compute the derivative with respect to α evaluated at the estimate and the estimated variance of $\hat{\mu}$ as,

```
aarg<-(ahat+1)/ahat
dmudalp<-(1/bhat)*digamma(aarg)*gamma(aarg)*(-1/ahat^2)
dbet<-(-1/bhat^2)*gamma(aarg)
D<-c(dalp,dbet)
vmu<-D%%Invinf%%D
```

A function named `muint` is available on the course web page that conducts these computations.

In a manner similar to that used in estimation of μ , we can see that, at each value of t , the survival function is simply a function of α and β ,

$$S(t|\alpha, \beta) = \exp[-(\beta t)^\alpha].$$

Then,

$$\begin{aligned}\frac{\partial S}{\partial \alpha} &= -\exp[-(\beta t)^\alpha][-(\beta t)^\alpha] \log(\beta t) \\ \frac{\partial S}{\partial \beta} &= \exp[-(\beta t)^\alpha][-\alpha(\beta t)^{\alpha-1}]t.\end{aligned}$$

We can then use the delta method at individual values of t in the same manner as for the expected value to obtain pointwise confidence bands for the survival function.

An R function to accomplish this is,

```
survconfband<-function(ahat,bhat,Invinf,ts){
#pointwise confidence band for Weibull survival function
#ahat and bhat are estimates, ts are evaluation times
#Infinf is inverse observed information
```

```

#(not necessarily observed)
#
S<-exp(-(bhat*ts)^ahat)
dalp<-S*(-(bhat*ts)^ahat)*log(bhat*ts)
dbet<-S*(-ahat*(bhat*ts)^(ahat-1))*ts
T<-length(ts)
lows<-NULL; ups<-NULL
cnt<-0
repeat{
  cnt<-cnt+1
  tS<-S[cnt]
  tda<-dalp[cnt]
  tdb<-dbet[cnt]
  tD<-c(tda,tdb)
  tv<-tD%%Invinf%%tD
  tlow<-tS-1.96*sqrt(tv)
  tup<-tS+1.96*sqrt(tv)
  lows<-c(lows,tlow)
  ups<-c(ups,tup)
  if(cnt==T) break
}
res<-data.frame(S=S,low=lows,up=ups)
return(res)
}

```

A plot of the estimated survival curves for the two treatment groups along with 95% pointwise confidence bands as computed by this function are presented in Figure 2.

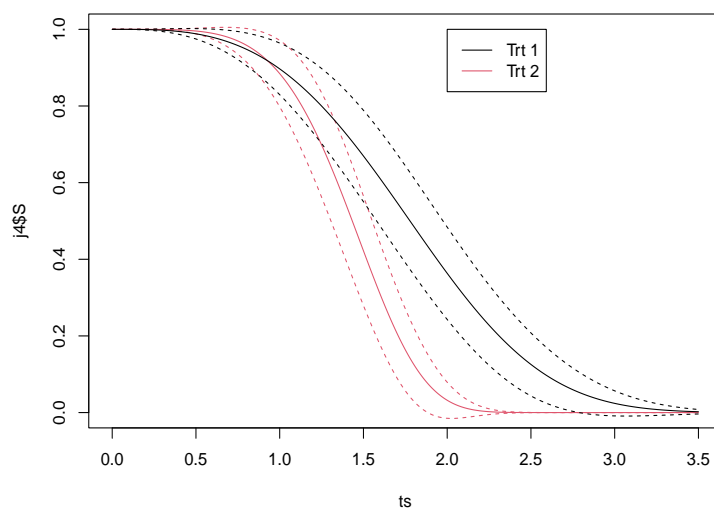


Figure 2: Estimated survival curves and 95% confidence bands for the two treatments.