3. Likelihood and the Censored Survival Data

A Data Example

The data below show survival times (in months) of patients with certain disease:

$$3,5,6^*,8,10^*,11^*,15,20^*,22,23,27^*,29,32,35,40,26,28,33^*,21,24^*$$

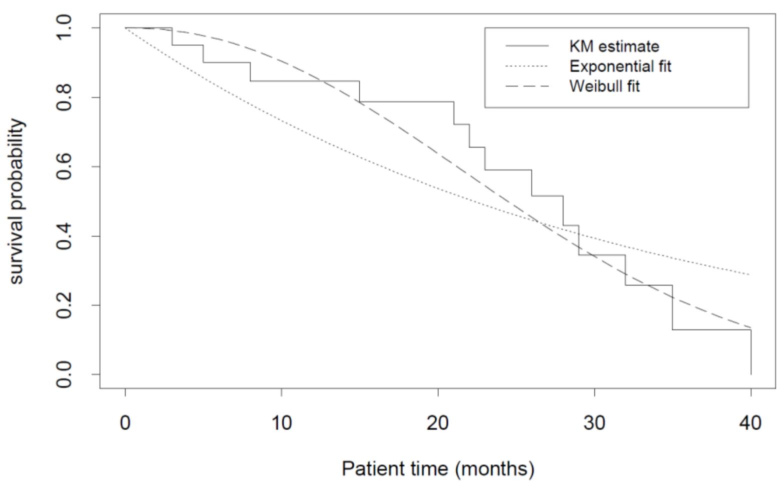
where $*$ indicates right censored data. If we fit exponential model to this data set, we have $D = 13$ and $PT = \sum x_i = 418$, so

$$\hat{\lambda} = \frac{D}{PT} = \frac{13}{418} = 0.0311/month$$

$$se(\hat{\lambda}) = \frac{\hat{\lambda}}{\sqrt{D}} = \frac{0.0311}{\sqrt{13}} = 0.0086$$

A 95% CI of λ is: $\hat{\lambda} \pm z_{0.025} * se(\hat{\lambda}) = 0.0311 \pm 1.96 * 0.0086 = [0.0142, 0.0480]$

The fitted exponential survival function is superimposed to the Kaplan-Meier estimate as shown in the figure below using the following *R* functions:




```
> library(survival)
> example <- read.table(file="chap3_ex1.txt", header=T)
> fit <- survfit(Surv(survtime, status)~1, conf.type=c("plain"),
data=example)
> plot(0, 0, xlim=c(0,40), ylim=c(0,1), xlab="Patient time
(months)", ylab="survival probability", pch=" ");
> lines(fit, lty=1)
> x <- seq(0, 40, by=0.5)
> sx <- exp(-0.0311*x)
> lines(x, sx, lty=2)
```

Obviously, the exponential distribution is a poor fit. In this case, we can choose one of the following options:

1. Use the Kaplan-Meier estimator which makes no assumption regarding the shape of the distribution. In most biomedical applications, the default is to go with the Kaplan-Meier estimator.
2. Choose a more flexible model, such as the Weibull model, which is shown below.

Recall that Weibull model has the following survival function

$$S(t) = e^{-\lambda t^{\alpha}}$$

and the following hazard function

$$\lambda(t) = \alpha\lambda t^{\alpha-1}$$

So the likelihood function of $\theta = (\lambda, \alpha)$ is given by

$$L(\lambda, \alpha; x, \delta) = \prod_{i=1}^{n} \left[\alpha\lambda x_i^{\alpha-1}\right]^{\delta_i} e^{-\lambda x_i^{\alpha}}$$

However, there is no closed form for the MLEs of $\theta = (\lambda, \alpha)$. So we used *Proc Lifereg* in SAS to fit Weibull model.

SAS code

This SAS program fits a Weibull model with two parameters: intercept β_0 and a scale parameter σ. Two parameters we use λ and α are related to β_0 and σ by (the detail will be discussed in Chapter 5).

$$\lambda = e^{-\beta_0/\sigma}; \alpha = \frac{1}{\sigma}$$

Since the MLE of β_0 and σ are $\hat{\beta}_0 = 3.36717$and $\hat{\sigma} = 0.46525$, the MLEs $\hat{\lambda}$ and $\hat{\alpha}$ are

$$\hat{\lambda} = e^{\widehat{\beta}_0/\widehat{\sigma}} = e^{3.36717/0.46525} = 0.00072$$

$$\hat{\alpha} = \frac{1}{\hat{\sigma}} = \frac{1}{0.46525} = 2.149$$

So $\hat{\alpha}$ is the Weibull Shape parameter in the SAS output. However, SAS uses the parameterization $S(t) = e^{-(t/\tau)^{\alpha}}$ for Weibull distribution so that τ is the Weibull scale parameter. Comparing this to our parameterization, we see that

$$\left(\frac{1}{\tau}\right)^{\alpha} = \lambda \;\; \Rightarrow \;\; \tau = \left(\frac{1}{\lambda}\right)^{1/\alpha}$$

The estimate of this Weibull scale parameter is

$$\hat{\tau} = \left(\frac{1}{0.00072}\right)^{1/2.149} = 28.99$$

The fitted Weibull survival function was superimposed to the Kaplan-Meier estimator in the previous figure using the following *R* functions:

```
> alpha <- 1/0.46525
> lambda <- exp(-3.36717/0.46525)
> sx <- exp(-lambda * x^alpha)     # the object "x" was created before
> lines(x, sx, lty=4)
> legend(25,1, c("KM estimate", "Exponential fit", "Weibull fit"), lty=c(1,2,4), cex=0.8)
```

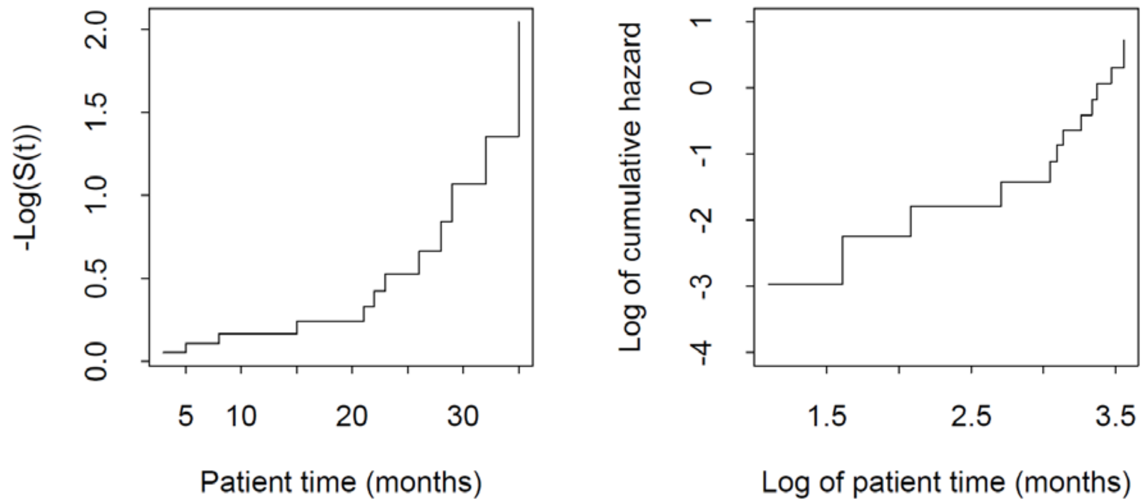
Compared to the exponential fit, the Weibull model fits the data much better (since its estimated survival function tracks the Kaplan-Meier estimator much better than the estimated exponential survival function). In fact, since the exponential model is a special case of the Weibull model (when $\sigma = 1$), we can test $H_0: \alpha = 1$ using the Weibull fit. (Note that $H_0: \alpha = 1$ is equivalent to $H_0: \sigma = 1$.) Since,

$$\left(\frac{\hat{\sigma} - 1}{se(\hat{\sigma})}\right)^2 = \left(\frac{0.46525 - 1}{0.108717}\right)^2 = 24.194$$

and $\mathrm{P}[\chi^2 > 24.194] = 8.7 \times 10^{-7}$, the H_0(exponential model)is rejected.

Note also that $\hat{\alpha} = 2.149 > 1$, so the stimated Weibull model has an increasing hazard function.

We already know that if the exponential distribution fits the data well, plot of $\log[\hat{S}(t)]$ vs t approximates a straight line; On the other hand, if the Weibull distribution fits the data well, plot of $\log\{-\log[S(t)]\}$ vs $\log(t)$ approximates a straight line. The inadequacy of the exponential fit is also demonstrated in the figure below. Here $\hat{S}(t)$ is the KM estimate.




```
> pdf(file="fig4.2.pdf", horizontal = F, height=6, width=8.5, font=3, pointsize=14)
> par(mfrow=c(1,2), pty="s")

> library(survival)
> example <- read.table(file="chap3_ex1.txt", header=T)
> fit <- survfit(Surv(survtime, status)~1, conf.type=c("plain"), data=example)
> plot(fit$time, -log(fit$surv), type="s", xlab=c("Patient time (months)"), ylab=c("-Log(S(t))"));
> plot(log(fit$time), log(-log(fit$surv)), type="s", ylim=c(-4,1), xlab=c("Log of patient time
(months)"), ylab=c("Log of cumulative hazard"))

> dev.off()
```