MAS3301 / MAS8311 Biostatistics
Part II: Survival

M. Farrow
School of Mathematics and Statistics
Newcastle University

Semester 2, 2009-10

8 Parametric models

8.1 Introduction

In the last few sections (the KM estimate and hypothesis tests) we haven't made any assumptions about the form of the underlying hazard function eg. constant, decaying, increasing. In the next few sections we introduce the idea of assuming some specific functional form for the hazard function, and fitting it to the data. We start off by looking at different exploratory plots of the estimated survivor function that we use to assess which parametric assumptions are appropriate.

8.2 Identification of distributions

8.2.1 Exponential

The exponential distribution has the following density, hazard, and survivor functions:

$$f(t) = \lambda \exp(-\lambda t)$$
$$h(t) = \lambda$$
$$S(t) = \exp(-\lambda t)$$

where λ is a positive constant.

It follows that

$$S(t) = \exp(-\lambda t)$$
$$\text{so } -\log[S(t)] = \lambda t.$$

We plot

$$-\log[\hat{S}(t)] \quad \text{against } t.$$

If the lifetime distribution underlying the data is exponential, then the plot will approximately be a straight line through the origin, gradient λ.

8.2.2 Weibull

The Weibull distribution has the following density, hazard, and survivor functions:

$$f(t) = \lambda\gamma t^{\gamma-1} \exp(-\lambda t^{\gamma})$$
$$h(t) = \lambda\gamma t^{\gamma-1}$$
$$S(t) = \exp(-\lambda t^{\gamma})$$

where λ, γ are positive constants.

It follows that

$$S(t) = \exp(-\lambda t^{\gamma})$$
$$\text{so } \log(-\log[S(t)]) = \log\lambda + \gamma \log t.$$

We plot

$$\log\left(-\log[\hat{S}(t)]\right) \quad \text{against } \log t.$$

If the lifetime distribution is Weibull, then the plot will approximately be a straight line gradient γ intercept $\log\lambda$. If the distribution is exponential then the gradient will be approximately 1.

You can plot the data in this way using R as follows:

```
library(survival)
my.surv <- Surv(times,cens)
surv.fn <- survfit(my.surv)
plot(log(surv.fn$time), log(-log(surv.fn$surv)), type="S")
```

8.2.3 Log-logistic

The log-logistic distribution has the following density, hazard, and survivor functions:

$$f(t) = \frac{\gamma\rho(\rho t)^{\gamma-1}}{(1+(\rho t)^{\gamma})^2}$$
$$h(t) = \frac{\gamma\rho(\rho t)^{\gamma-1}}{1+(\rho t)^{\gamma}}$$
$$S(t) = \frac{1}{1+(\rho t)^{\gamma}}$$

where γ, ρ are positive constants.

What sort of plot could we use to recognize a survivor function with this sort of distribution?

$$\frac{1}{S(t)} - 1 = (\rho t)^{\gamma} \quad \text{so}$$
$$\log\left(\frac{1}{S(t)} - 1\right) = \gamma \log \rho + \gamma \log t.$$

We plot

$$\log\left(\frac{1}{\hat{S}(t)} - 1\right) \quad \text{against} \ \log t.$$

If the lifetime distribution is log-logistic then this should approximately give a straight line with intercept $\gamma \log \rho$ and gradient γ.

Example: Given the parametric form

$$S(t) = \frac{1}{(1 + \lambda^2 t^2)^{1/2}}$$

for a survivor function, suggest a suitable plot of the estimated survivor function to assess whether this form is suitable and describe how an estimate for λ can be obtained from the graph.

Plot $1/\hat{S}^2$ vs t^2. The graph will have intercept 1 and gradient λ^2 if the data are distributed with the given survivor function.

8.3 Some other lifetime distributions

8.3.1 The log normal distribution

Let $X \sim N(\mu, \sigma^2)$ be a normal random variable. If $T = \exp X$ and $t = \exp x$ then

$$\begin{aligned} f_T(t) &= f_X(x) \times \left|\frac{dx}{dt}\right| \\ &= \frac{1}{t\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(\log t - \mu)^2 \end{aligned}$$

Unfortunately the survivor and hazard functions can only be expressed in terms of integrals:

$$S(t) = \Pr(T > t) = \Pr(X > \log t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$$

where Φ is the standard normal distribution function

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) \ du.$$

Clearly this isn't very convenient. The log-logistic distribution is easier to use and is very similar.

8.3.2 The gamma distribution

You might also see the gamma distribution used for lifetime random variables:

$$f_T(t) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha t^{\alpha-1} \exp(-\lambda t)$$

where Γ is the gamma function (the constant is used to make sure the density integrates to 1). Again, this isn't very convenient to use since the survivor function can only be expressed using integrals. The Weibull distribution is similar and simpler to use.

8.4 The log linear representation and R

Lemma. *Suppose T has a Weibull distribution with*

$$S(t) = \exp(-\lambda t^\gamma).$$

If we define a random variable E by $E = \lambda T^\gamma$ then

$$\log T = -\frac{1}{\gamma} \log \lambda + \frac{1}{\gamma} \log E$$

and $E \sim \exp(1)$.

Proof: The definition of E gives

$$\log E = \log \lambda + \gamma \log T \qquad (\dagger)$$

and therefore

$$\log T = -\frac{1}{\gamma} \log \lambda + \frac{1}{\gamma} \log E.$$

Furthermore

$$\begin{aligned} \Pr(E > \eta) &= \Pr(\lambda T^\gamma > \eta) = \Pr(T > (\eta/\lambda)^{1/\gamma}) \\ &= S([\eta/\lambda]^{1/\gamma}) = \exp(-\eta) \end{aligned}$$

and so $E \sim \exp(1)$. □

Transformations of this kind are quite useful, and we refer to equation (†) as the *log-linear representation* of T.

For the log-logistic distribution the log-linear representation has the form

$$\log T = -\log \rho + \frac{1}{\gamma} \log E$$

where E is a 'standard' log-logistic random variable (one with $\gamma = \rho = 1$).

Proof:

$$T = \exp\left\{-\log\rho + \frac{1}{\gamma}\log E\right\} = \frac{1}{\rho}E^{1/\gamma}$$

so

$$E = (\rho T)^\gamma$$

and

$$\begin{aligned}\Pr(E > \eta) &= \Pr[(\rho T)^\gamma > \eta] = \Pr\left(T > \frac{1}{\rho}\eta^{1/\gamma}\right)\\ &= S\left(\frac{1}{\rho}\eta^{1/\gamma}\right) = \frac{1}{1+\eta}\end{aligned}$$

In general the transformation has the form

$$\log T = \mu + \sigma \log E$$

for some constants μ, σ usually referred to as the *intercept* and *scale parameter* respectively.

Why is this useful? R fits parametric models using this sort of representation. This is carried out using the `survreg` function: it fits models of the form

$$\log T = \mu + (\text{a term that depends on covariates}) + \sigma \log E.$$

Compare this to a normal regression formula:

$$Y = \mu + \beta X + \sigma\epsilon$$

where $\epsilon \sim N(0, 1)$ is an error term. You can see how `survreg` is carrying out a similar task to regular regression. We'll see in the practicals how to use the function, and we'll refer back to the little bit of theory presented here.

9 Maximum likelihood for parametric models

In the last section we looked at how exploratory plots of the estimated survivor function can suggest a specific functional form for the underlying lifetime distribution. In this section we consider how to fit the parameters to the observed data. The way we do this is with maximum likelihood. If you look back at your notes for MAS2305 you'll see how to estimate the mean and standard deviation for a set of data assumed to be normally distributed. We're going to do exactly the same kind of thing here for survival data: we'll just be using different distributions and we'll have to take censoring into account.

9.1 Likelihood functions

9.1.1 Notation

Probability density function:	$f(t)$
Distribution function:	$F(t)$
Survivor function:	$S(t) = 1 - F(t)$
Hazard function:	$h(t) = f(t)/S(t)$
Data:	n independent subjects, of whom
	n_R have right-censored times,
	n_Q have left-censored times,
	n_I have interval-censored times,
	n_D have observed failure times.
	$n = n_D + n_R + n_Q + n_I$
Define sets:	$R = \{i : \text{right censored }\}$
	$Q = \{i : \text{left censored }\}$
	$I = \{i : \text{interval censored }\}$
	$D = \{i : \text{failure (death) time observed }\}$
	$A = D \cup R \cup Q \cup I$. ('A' is for 'all'.)

Denote sums and products over these sets using, e.g., $\sum_{i \in D} t_i$, $\prod_{i \in D} t_i$, $\sum_{i \in R} t_i$ etc.

9.1.2 All survival times observed

$$
\begin{aligned}
D &= A \\
n_D &= n
\end{aligned}
$$

Likelihood:

$$L = \prod_{i \in A} f(t_i)$$

Log likelihood

$$l = \sum_{i \in A} \log[f(t_i)]$$

9.1.3 Some observations right-censored

$$\begin{aligned} A &= D \cup R \\ n &= n_D + n_R \end{aligned}$$

Likelihood:

$$L = \prod_{i \in D} f(t_i) \prod_{i \in R} S(t_i)$$

Log likelihood

$$l = \sum_{i \in D} \log[f(t_i)] + \sum_{i \in R} \log[S(t_i)]$$

The contribution of a right-censored observation to the likelihood is $S(t_i)$, the probability that the patient is still alive at time t_i.

9.1.4 Left-censoring

Similarly, a left-censored observation at time t_i would contribute $F(t_i) = 1 - S(t_i)$ to the likelihood, the probability that the subject dies before time t_i.

9.1.5 Interval-censoring

Likewise, an interval-censored observation, in the interval (t_{Li}, t_{Ui}) contributes

$$[F(t_{Ui}) - F(t_{Li})] = [S(t_{Li}) - S(t_{Ui})] = \int_{t_{Li}}^{t_{Ui}} f(t)dt$$

to the likelihood, the probability that the subject dies in the interval (t_{Li}, t_{Ui}).

9.1.6 Example

Given a survivor function of the form

$$S(t) = \exp[-(at + bt^2)]$$

write down and simplify the log-likelihood function for a set of right-censored data.

Start off by calculating $f(t)$:

$$f(t) = -\frac{dS(t)}{dt} = (a + 2bt)\exp[-(at + bt^2)]$$

Then

$$L(a, b) = \prod_{i \in D} \left((a + 2bt_i)\exp[-(at_i + bt_i^2)]\right) \times \prod_{i \in R} \exp[-(at_i + bt_i^2)]$$

and

$$\begin{aligned} l(a, b) &= \sum_{i \in D} \log(a + 2bt_i) - \sum_{i \in D}(at_i + bt_i^2) - \sum_{i \in R}(at_i + bt_i^2) \\ &= \sum_{i \in D} \log(a + 2bt_i) - \sum_{i \in D \cup R} (at_i + bt_i^2) \end{aligned}$$

9.2 Exponential distribution

Suppose the lifetime distribution is exponential, so that

$$f(t) = \lambda e^{-\lambda t}, \quad \text{and } S(t) = e^{-\lambda t}.$$

Let's suppose we have n_D failure times and n_R right censored times so that

$$A = D \cup R, \quad n = n_R + n_D.$$

The likelihood is

$$\begin{aligned} L &= \prod_{i \in D} f(t_i) \prod_{i \in R} S(t_i) \\ &= \prod_{i \in D} \lambda e^{-\lambda t_i} \prod_{i \in R} e^{-\lambda t_i} \\ &= \lambda^{n_D} \exp\left(-\lambda \sum_{i \in D \cup R} t_i\right) \end{aligned}$$

and

$$l = \log(L) = n_D \log \lambda - \lambda \sum_{i \in D \cup R} t_i.$$

Differentiating gives
$$\frac{\partial l}{\partial \lambda} = \frac{n_D}{\lambda} - \sum_{i \in D \cup R} t_i, \qquad \frac{\partial^2 l}{\partial \lambda^2} = -\frac{n_D}{\lambda^2}.$$
The second derivative is negative, so solving $\partial l / \partial \lambda = 0$ gives a maximum. The MLE is
$$\hat{\lambda} = \frac{n_D}{\sum_{i \in D \cup R} t_i}.$$

The mean of the exponential distribution with parameter λ is $1/\lambda$, so the maximum likelihood estimate of the mean is
$$\hat{\mu} = \frac{\sum_{i \in D \cup R} t_i}{n_D}.$$
Similarly, the median t_m is the solution to
$$1/2 = F(t_m) = \exp(-\lambda t_m).$$
Rearranging gives $t_m = (\log 2)/\lambda$, so
$$\hat{t}_m = \hat{\mu} \log 2.$$
Standard theory about MLEs shows that the approximate large-sample variance of $\hat{\lambda}$ is
$$\mathrm{var}(\hat{\lambda}) \approx \left[-E\left(\frac{\partial^2 l}{\partial \lambda^2} \right) \right]^{-1}.$$
Substituting in the value for the second derivative we obtained above gives
$$\mathrm{var}(\hat{\lambda}) \approx \frac{\lambda^2}{n_D} \approx \frac{\hat{\lambda}^2}{n_D}.$$
An approximate 95% CI for λ is therefore $\hat{\lambda} \pm 1.96 \times \hat{\lambda}/\sqrt{n}_D$, assuming $\hat{\lambda}$ is normally distributed.

10 Maximum likelihood for some other distributions

10.1 The Weibull Distribution

Suppose that a lifetime random variable adopts a Weibull distribution with
$$f(t) = \lambda \gamma t^{\gamma - 1} \exp(-\lambda t^{\gamma}), \quad \text{and } S(t) = \exp(-\lambda t^{\gamma}).$$
Let's suppose we have n_D failure times and n_R right censored times so that
$$A = D \cup R, \quad n = n_R + n_D.$$
The likelihood is

$$
\begin{aligned}
L &= \prod_{i\in D} f(t_i) \prod_{i\in R} S(t_i) \\
&= \prod_{i\in D} \{\lambda\gamma t_i^{\gamma-1}\exp(-\lambda t_i^{\gamma})\} \prod_{i\in R} \exp(-\lambda t_i^{\gamma}) \\
&= \lambda^{n_D}\gamma^{n_D}\left(\prod_{i\in D} t_i^{\gamma-1}\right)\exp\left(-\lambda \sum_{i\in D\cup R} t_i^{\gamma}\right)
\end{aligned}
$$

and

$$
l = \log(L) = n_D \log\lambda + n_D\log\gamma + (\gamma-1)\sum_{i\in D}\log t_i - \lambda\sum_{i\in D\cup R} t_i^{\gamma}.
$$

In order to carry out the differentiation, we need the following:

Lemma.

$$
\frac{\partial}{\partial x} t^x = [\log(t)] \times t^x
$$

Proof: From properties of logs:

$$
t^x = e^{x\log(t)}.
$$

Differentiating both sides with respect to x gives

$$
\frac{\partial}{\partial x} t^x = [\log(t)] \times e^{x\log(t)} = [\log(t)] \times t^x.
$$

□

Differentiating with respect to λ gives

$$
\frac{\partial l}{\partial \lambda} = \frac{n_D}{\lambda} - \sum_{i\in D\cup R} t_i^{\gamma}.
$$

Solving $\partial l/\partial\lambda = 0$ gives the following relationship between the MLEs:

$$
\hat{\lambda} = \frac{n_D}{\sum_{i\in D\cup R} t_i^{\hat{\gamma}}}.
$$

Differentiating with respect to γ (via the lemma) gives

$$\frac{\partial l}{\partial \gamma} = \frac{n_D}{\gamma} + \sum_{i \in D} \log t_i - \lambda \sum_{i \in D \cup R} t_i^{\gamma} \log t_i.$$

Setting $\partial l / \partial \gamma = 0$ gives a second relationship between the MLEs:

$$0 = \frac{n_D}{\hat{\gamma}} + \sum_{i \in D} \log t_i - \hat{\lambda} \sum_{i \in D \cup R} t_i^{\hat{\gamma}} \log t_i.$$

We can then substitute in the value for $\hat{\lambda}$:

$$0 = \frac{n_D}{\hat{\gamma}} + \sum_{i \in D} \log t_i - n_D \frac{\sum_{i \in D \cup R} t_i^{\hat{\gamma}} \log t_i}{\sum_{i \in D \cup R} t_i^{\hat{\gamma}}}.$$

We cannot solve this equation algebraically so we need a numerical procedure.

The Newton-Raphson algorithm is one way to solve such equations numerically. In practice the Newton-Raphson algorithm is used to solve both equations

$$\frac{\partial l}{\partial \lambda} = 0, \qquad \frac{\partial l}{\partial \gamma} = 0$$

simultaneously. Let

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\gamma} \end{bmatrix}_j \qquad (*)$$

denote the jth approximation to the MLE. We pick some suitable values to start the process off at $j = 0$. For example, you can use a graphical method as we saw in previous lectures to obtain rough estimates for the parameters. To obtain the $(j+1)$th approximation we evaluate

$$\begin{bmatrix} \frac{\partial l}{\partial \lambda} \\ \frac{\partial l}{\partial \gamma} \end{bmatrix}_j \qquad \text{and} \qquad \begin{bmatrix} \frac{\partial^2 l}{\partial \lambda^2} & \frac{\partial^2 l}{\partial \lambda \partial \gamma} \\ \frac{\partial^2 l}{\partial \lambda \partial \gamma} & \frac{\partial^2 l}{\partial \gamma^2} \end{bmatrix}_j$$

using the values $(*)$ for λ, γ. At iteration $(j+1)$, the approximation is then

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\gamma} \end{bmatrix}_{j+1} = \begin{bmatrix} \hat{\lambda} \\ \hat{\gamma} \end{bmatrix}_j - \begin{bmatrix} \frac{\partial^2 l}{\partial \lambda^2} & \frac{\partial^2 l}{\partial \lambda \partial \gamma} \\ \frac{\partial^2 l}{\partial \lambda \partial \gamma} & \frac{\partial^2 l}{\partial \gamma^2} \end{bmatrix}_j^{-1} \begin{bmatrix} \frac{\partial l}{\partial \lambda} \\ \frac{\partial l}{\partial \gamma} \end{bmatrix}_j$$

where

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{n_D}{\lambda^2}$$
$$\frac{\partial^2 l}{\partial \lambda \partial \gamma} = -\sum_{i \in D \cup R} t_i^\gamma \log t_i$$
$$\frac{\partial^2 l}{\partial \gamma^2} = -\frac{n_D}{\gamma^2} - \lambda \sum_{i \in D \cup R} t_i^\gamma (\log t_i)^2.$$

Standard MLE theory gives that, asymptotically,

$$\operatorname{var}\begin{bmatrix}\hat{\lambda} \\ \hat{\gamma}\end{bmatrix} \approx \left[-E\begin{bmatrix}\frac{\partial^2 l}{\partial \lambda^2} & \frac{\partial^2 l}{\partial \lambda \partial \gamma} \\ \frac{\partial^2 l}{\partial \lambda \partial \gamma} & \frac{\partial^2 l}{\partial \gamma^2}\end{bmatrix}\right]^{-1}.$$

We therefore have that when the MLE is found, an approximation to the variance is

$$\operatorname{var}\begin{bmatrix}\hat{\lambda} \\ \hat{\gamma}\end{bmatrix} \approx -\begin{bmatrix}\frac{\partial^2 l}{\partial \lambda^2} & \frac{\partial^2 l}{\partial \lambda \partial \gamma} \\ \frac{\partial^2 l}{\partial \lambda \partial \gamma} & \frac{\partial^2 l}{\partial \gamma^2}\end{bmatrix}^{-1}$$

where the matrix is evaluated at the MLE. This is a useful by-product of the Newton-Raphson algorithm.

Note that, if the value of γ is known, then

$$\hat{\lambda} = \frac{n_D}{\sum_{i \in D \cup R} t_i^\gamma}$$

and

$$\operatorname{var}(\hat{\lambda}) \approx -\left(\frac{d^2 l}{d\lambda^2}\right)^{-1} = \frac{\lambda^2}{n_D} \approx \frac{\hat{\lambda}^2}{n_D} = \frac{n_D}{(\sum_{i \in D \cup R} t_i^\gamma)^2}.$$

10.2 Log-logistic

$$f(t) = \frac{\gamma\rho(\rho t)^{\gamma-1}}{(1 + (\rho t)^\gamma)^2}$$
$$S(t) = \frac{1}{1 + (\rho t)^\gamma}$$

With the same censoring assumptions as above you get a likelihood function

$$L = \prod_{i \in D} f(t_i) \prod_{i \in R} S(t_i).$$

It's straightforward to substitute in the formulae for f and S, differentiate, and come up with a Newton-Raphson scheme, although the calculation isn't short. As for the Weibull distribution, initial estimates of the parameters can be obtined via a graphical method.

10.3 Gamma and Log-normal

Since the survivor functions cannot be written down in a closed form they have to be evaluated numerically, and maximizing the likelihood cannot be performed using a straightforward Newton Raphson algorithm.

Tutorial Examples 3

1. The survival time for individuals is modelled as a random variable T. Define the hazard function $h(t)$ associated with T in terms of a limit. From the definition prove that

$$h(t) = -\frac{d}{dt}\log[S(t)]$$

   where $S(t) = \text{Prob}(T \geq t)$.

2. A study was performed to investigate the survival times of patients following heart bypass surgery. It was found that 50% of patients lived for 24 months or longer, while 30% of patients lived for 60 months or longer. Assuming the survival time for the population has a Weibull distribution, obtain approximate estimates of the parameters λ and γ. Then estimate:

   (a) the probability of an individual living 72 months or longer, and

   (b) the probability of an individual dying between 24 and 60 months, given she was alive at 24 months.

3. The time until relapsing back to injecting heroin is recorded for drug addicts taking part in a study of replacement therapy.

   The response variables are the time until relapse to injecting and the censoring status (some times are right-censored). There are two explanatory variables: p, a binary variable which takes value 1 if the addict has had a previous prison sentence, and 0 otherwise; and x, the dose of the replacement therapy given.

   The parameter estimates, when both explanatory variables are fitted, are as follows.

Variable	Estimated coefficient	Standard error
p	0.327	0.167
x	-0.035	0.006

   (a) What is the estimated hazard ratio for a subject who has not been to prison compared to a subject who has, given that they receive the same dose of the replacement therapy?

   (b) What is the estimated hazard ratio for a subject receiving a dose $x = 30$ and another who receives a dose receiving a dose $x = 50$, given that neither has a previous prison sentence.

   (c) In each case give an approximate 95% confidence interval for the hazard ratio. (Assume that the sample is large enough for the approximation).