1 Definitions and Censoring

1.1 Survival Analysis

We begin by considering simple analyses but we will lead up to and take a look at regression on explanatory factors., as in linear regression part A. The important difference between survival analysis and other statistical analyses which you have so far encountered is the presence of censoring. This actually renders the survival function of more importance in writing down the models.

We begin with a reminder of some **definitions**.

T denotes the positive random variable representing time to event of interest.

Cumulative Distribution function is $F(t) = \Pr(T \leq t)$ with probability density function $f(t) = F^{'}(t)$.

Survival function is

$$S(t) = P(T > t) = 1 - F(t)$$

Note: we use $S(t) = \overline{F}(t)$ throughout.

Hazard function

$$h(t) = \lim_{\delta t \to 0} \left(\frac{\Pr(t \leq T < t + \delta t | T \geq t)}{\delta t} \right)$$

{If T is discrete and positive integer-valued then $h(t) = \Pr(T = t | T \geq t) = \Pr(T = t)/S(t-1)$.}

Cumulative hazard function

$$H(t) = \int_0^t h(s) \mathrm{ds}$$

We have the following **relations** between these functions:

(i)

$$\begin{aligned} h(t) &= \lim_{\delta t \to 0} \left(\frac{S(t) - S(t + \delta t)}{\delta t S(t)} \right) \\ &= -\frac{S'(t)}{S(t)} \\ &= -\frac{d}{dt}(\log S) \end{aligned}$$

(ii)

$$S(t) = \exp(-H(t)), \text{ since } S(0) = 1$$

(iii)

$$f(t) = h(t)S(t)$$

1.2 Censoring and truncation

Right censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. For example, we consider patients in a clinical trial to study the effect of treatments on stroke occurrence. The study ends after 5 years. Those patients who have had no strokes by the end of the year are censored. If the patient leaves the study at time t_e, then the event occurs in (t_e, ∞) .

Left censoring is when the event of interest has already occurred before enrolment. This is very rarely encountered.

Truncation is deliberate and due to study design.

Right truncation occurs when the entire study population has already experienced the event of interest (for example: a historical survey of patients on a cancer registry).

Left truncation occurs when the subjects have been at risk before entering the study (for example: life insurance policy holders where the study starts on a fixed date, event of interest is age at death).

Generally we deal with **right censoring &** sometimes **left truncation**.

Two types of independent **right censoring**:

Type I : completely random dropout (eg emigration) and/or fixed time of end of study no event having occurred.

Type II: study ends when a fixed number of events amongst the subjects has occurred.

1.3 Likelihood and Censoring

If the censoring mechanism is *independent* of the event process, then we have an easy way of dealing with it.

Suppose that T is the time to event and that C is the time to the censoring event.

Assume that all subjects may have an event or be censored, say for subject i one of a pair of observations $\left(\widetilde{t_i}, \widetilde{c_i}\right)$ may be observed. Then since we observe the minimum time we would have the following expression for the likelihood (using independence)

$$L = \prod_{\widetilde{t_i} < \widetilde{c_i}} f(\widetilde{t_i}) S_C(\widetilde{t_i}) \prod_{\widetilde{c_i} < \widetilde{t_i}} S(\widetilde{c_i}) f_C(\widetilde{c_i})$$

Now define the following random variable:

$$\delta = \begin{cases} 1 & \text{if } T < C \\ 0 & \text{if } T > C \end{cases}$$

For each subject we observe $t_i = \min\left(\widetilde{t_i}, \widetilde{c_i}\right)$ and δ_i, observations from a continuous random variable and a binary random variable. In terms of these L becomes

$$L = \prod_i h(t_i)^{\delta_i} S(t_i) \prod_i h_C(t_i)^{1-\delta_i} S_C(t_i)$$

where we have used density = hazard × survival function.

NB If the censoring mechanism is independent (sometimes called non-informative) then we can ignore the second product on the right as it gives us no information about the event time. In the remainder of the course we will assume that the censoring mechanism is independent.

1.4 Data

Demographic v. trial data

The time to event can literally be the age, eg in a life insurance policy. In a clinical trial it will more typically be time from admission to the trial.

Slides show five patients A, B, C, D, E from a Sydney hospital pilot study, concerning treatment of bladder cancer.

Each patient has their own zero time, the time at which the patient entered the study (accrual time). For each patient we record time to event of interest or censoring time, whichever is the smaller, and the status, $\delta = 1$ if the event occurs and $\delta = 0$ if the patient is censored.

2 Non-parametric estimators

Reminder: (informs the argument below)

If there are observations $x_1, \dots, x_n$ from a random sample then we define the empirical distribution function

$$\widehat{F}(x) = \frac{1}{n} \# \{x_i : x_i \leq x\}$$

This is appropriate if no censoring occurs. However if censoring occurs this has to be taken into account.

We measure the pair (X, δ) where $X = \min(T, C)$ and δ is as before

$$\delta = \begin{cases} 1 & \text{if } T < C \\ 0 & \text{if } T > C \end{cases}$$

Suppose that the observations are (x_i, δ_i) for $i = 1, 2 \dots, n$.

$$\begin{aligned} L &= \prod_i f(x_i)^{\delta_i} S(x_i)^{1-\delta_i} \\ &= \prod_i f(x_i)^{\delta_i} \left(1 - F(x_i)\right)^{1-\delta_i} \end{aligned}$$

What follows is a heuristic argument allowing us to find an estimator for S , the survival function, which in the likelihood sense is the best that we can do.

Suppose that there are failure times $(0 <) < t_1 < \dots < t_i < \dots$. Let $s_{i1}, s_{i2}, \cdots, s_{ic_i}$ be the censoring times within the interval $[t_i, t_{i+1})$ and suppose that there are d_i failures at time t_i (allowing for tied failure times). Then the likelihood function becomes

$$\begin{aligned} L &= \prod_{fail} f(t_i)^{d_i} \prod_i \left(\prod_{k=1}^{c_i} (1 - F(s_{ik})) \right) \\ &= \prod_{fail} (F(t_i) - F(t_i -))^{d_i} \prod_i \left(\prod_{k=1}^{c_i} (1 - F(s_{ik})) \right) \end{aligned}$$

where we write $f(t_i) = F(t_i) - F(t_i -)$, the difference in the cdf at time t_i and the cdf immediately before it.

Since $F(t_i)$ is an increasing function, and *assuming that it takes fixed values at the failure time points*, we make $F(t_i -)$ and $F(s_{ik})$ as small as possible in order to maximise the likelihood. That means we take $F(t_i -) = F(t_{i-1})$ and $F(s_{ik}) = F(t_i)$.

This maximises L by considering the cdf $F(t)$ to be a step function and therefore to come from a discrete distrbution, with failure times as the actual failure times which occur. Then

$$L = \prod_{fail} (F(t_i) - F(t_{i-1}))^{d_i} \prod_i (1 - F(t_i))^{c_i}$$

So we have showed that amongst all cdf's with fixed values $F(t_i)$ at the failure times t_i, then the discrete cdf has the maximum likelihood, amongst those with d_i failures at t_i and c_i censorings in the interval $[t_i, t_{i+1})$.

Let us consider the **discrete case** and let

$$\Pr\left(\text{fail at } t_i | \text{survived to } t_i-\right) = h_i$$

Then

$$\begin{aligned} S\left(t_i\right) &= 1 - F\left(t_i\right) = \prod_1^i (1 - h_j), \\ f(t_i) &= h_i \prod_1^{i-1} (1 - h_j) \end{aligned}$$

Finally we have

$$L = \prod_{t_i} h_i^{d_i} (1 - h_i)^{n_i - d_i}$$

where n_i is the number at risk at time t_i. This is usually referred to as the number in the risk set.

Note

$$n_{i+1} + c_i + d_i = n_i$$

2.1 Kaplan-Meier estimator

This estimator for $S(t)$ uses the mle estimators for h_i. Taking logs

$$l = \sum_i d_i \log h_i + \sum_i (n_i - d_i) \log(1 - h_i)$$

Differentiate with respect to h_i

$$\begin{aligned} \frac{\partial l}{\partial h_i} &= \frac{d_i}{h_i} - \frac{n_i - d_i}{1 - h_i} = 0 \\ \implies & \widehat{h}_i = \frac{d_i}{n_i} \end{aligned}$$

So the Kaplan-Meier estimator is

$$\widehat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right)$$

where

$$\begin{aligned} n_i &= \#\{\text{in risk set at } t_i\}, \\ d_i &= \#\{\text{events at } t_i\}. \end{aligned}$$

Note that $c_i = \#\{\text{censored in } [t_i, t_{i+1})\}$. If there are no censored observations before the first failure time then $n_0 = n_1 = \#\{\text{in study}\}$. Generally we assume $t_0 = 0$.

2.2 Nelson-Aalen estimator and new estimator of S

The Nelson-Aalen estimator for the cumulative hazard function is

$$\widehat{H}(t) = \sum_{t_i \le t} \frac{d_i}{n_i} \quad \left(= \sum_{t_i \le t} \widehat{h}_i \right)$$

This is natural for a discrete estimator, as we have simply summed the estimates of the hazards at each time, instead of integrating, to get the cummulative hazard. This correspondingly gives an estimator of S of the form

$$\begin{aligned} \widetilde{S}(t) &= \exp\left(-\widehat{H}(t)\right) \\ &= \exp\left(-\sum_{t_i \le t} \frac{d_i}{n_i}\right) \end{aligned}$$

It is not difficult to show by comparing the functions $1 - x, \exp(-x)$ on the interval $0 \le x \le 1$, that $\widetilde{S}(t) \ge \widehat{S}(t)$.

Invented data set

Suppose that we have 10 observations in the data set with failure times as follows:

$$2, 5, 5, 6+, 7, 12, 14+, 14+, 14+, 14+$$

Here + indicates a censored observation. Then we can calculate both estimators for $S(t)$ at all time points. It is considered unsafe to extrapolate much beyond the last time point, 14, even with a large data set.

2.3 Confidence Intervals

We need to find confidence intervals (pointwise) for the estimators of $S(t)$ at each time point. We differentiate the log-likelihood and use likelihood theory,

$$l = \sum_i d_i \log h_i + \sum_i (n_i - d_i)\log(1 - h_i),$$

differentiated twice to find the Hessian matrix $\left\{\frac{\partial^2 l}{\partial h_i \partial h_j}\right\}$.

Note that since l is a sum of functions of each individual hazard the Hessian must be diagonal.

The estimators $\left\{\widehat{h_1}, \widehat{h_2}, \ldots, \widehat{h_n}\right\}$ are asymptotically unbiased and are asymptotically jointly normally distributed with approximate variance I^{-1}, where the information matrix is given by

$$I = \mathbf{E}\left(-\left\{\frac{\partial^2 l}{\partial h_i \partial h_j}\right\}\right).$$

Since the Hessian is diagonal, the covariances are all asymptotically zero, and coupled with asymptotic normality, this ensures that all pairs $\widehat{h_i}, \widehat{h_j}$ are asymptotically independent.

$$-\frac{\partial^2 l}{\partial h_i^2} = \frac{d_i}{h_i^2} + \frac{n_i - d_i}{(1 - h_i)^2}$$

We use the observed information J and so replace h_i in the above by its estimator $\widehat{h_i} = \frac{d_i}{n_i}$. Hence we have

$$\mathbf{var}\ \widehat{h_i} \approx \frac{d_i\,(n_i - d_i)}{n_i^3}\ .$$

2.3.1 Establishing Greenwood's formula

Reminder: δ method

If the random variation of Y around μ is small (for example if μ is the mean of Y and $\mathbf{var}Y$ has order $\frac{1}{n}$), we use:

$$g(Y) \approx g(\mu) + (Y - \mu)g'(\mu) + \frac{1}{2}(Y - \mu)^2 g''(\mu) + \ldots$$

Taking expectations

$$\begin{aligned}
\mathbf{E}(g(Y)) &= g(\mu) + O\left(\frac{1}{n}\right) \\
\mathbf{var(g(Y))} &= \mathbf{g}'(\mu)^2 \mathbf{var}Y + o\left(\frac{1}{n}\right)
\end{aligned}$$

Derivation of Greenwood's formula for var$(\widehat{S}(t))$

$$\log \widehat{S}(t) = \sum_{t_i \le t} \log\left(1 - \widehat{h_i}\right)$$

But

$$\mathbf{var}\,\widehat{h}_i \approx \frac{d_i\,(n_i-d_i)}{n_i^3} \quad \text{and } \widehat{h}_i \xrightarrow{P} h_i$$

so that, given $g(h_i) = \log(1-h_i)$,

$$g'(h_i) = \frac{-1}{(1-h_i)}$$

we have

$$\begin{aligned} \mathbf{var}\ \log\left(1-\widehat{h}_i\right) &\approx \frac{1}{(1-h_i)^2}\mathbf{var}\,\widehat{h}_i \\ &\approx \frac{1}{(1-\frac{d_i}{n_i})^2}\frac{d_i\,(n_i-d_i)}{n_i^3} \\ &= \frac{d_i}{n_i\,(n_i-d_i)} \end{aligned}$$

Since $\widehat{h}_i, \widehat{h}_j$ are asymptotically independent we can put all this together to get

$$\mathbf{var}\log\left(\widehat{S}(t)\right) = \sum_{t_i \le t} \frac{d_i}{n_i\,(n_i-d_i)}$$

Let $Y = \log\widehat{S}$ and note that we need $\mathbf{var}\left(e^Y\right) \approx \left(e^Y\right)^2 \mathbf{var}Y$, again using the delta-method.

Finally we have *Greenwood's formula*

$$\mathbf{var}\left(\widehat{S}(t)\right) \approx \widehat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{n_i\,(n_i-d_i)}\ .$$

Applying this to the same sort of argument to the Nelson-Aalen estimator and its extension to the survival function we also see

$$\mathbf{var}\ \widehat{H}(t) \approx \sum_{t_i \leq t} \frac{d_i\,(n_i - d_i)}{n_i^3}$$

and

$$\begin{aligned}
\mathbf{var}\widetilde{S}(t) &= \mathbf{var}\left(\exp(-\widehat{H}(t)\right) \\
&\approx \left(e^{-H}\right)^2 \sum_{t_i \leq t} \frac{d_i\,(n_i - d_i)}{n_i^3} \\
&\approx \left(\widetilde{S}(t)\right)^2 \sum_{t_i \leq t} \frac{d_i\,(n_i - d_i)}{n_i^3}
\end{aligned}$$

Clearly these estimates are only reasonable if each n_i is sufficiently large, since they rely heavily on asymptotic calculations.

2.4 Actuarial estimator

The **actuarial estimator** is a further estimator for $S(t)$.It is given as

$$S^*(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i - \frac{1}{2}c_i}\right)$$

The intervals between consecutive failure times are usually of constant length, and it is generally used by actuaries and demographers following a cohort from birth to death. Age will normally be the time variable and hence the unit of time is 1 year.

3 Models: accelerated life model, proportional hazards model

We generally will have heterogeneous data where parameter estimates will be dependent on covariates measured for participants in a study. For example age or sex may have an effect on time to event. A simple example would be where participants fall into two groups such as treatment v. control, smoker v. non-smoker.

There are two popular general classes of model as in the heading above - AL and PH.

3.1 Accelerated Life models

Suppose there are (several) groups, labelled by index i. The accelerated life model has a survival curve for each group defined by

$$S_i(t) = S_0(\rho_i t)$$

where $S_0(t)$ is some baseline survival curve and ρ_i is a constant specific to group i.

If we plot S_i against $\log t$, $i = 1, 2, \ldots, k$, then we expect to see a horizontal shift as

$$S_i(t) = S_0(\mathrm{e}^{\log \rho_i + \log t}) \ .$$

3.1.1 Medians and Quantiles

Note too that each group has a different median lifetime, since, if $S_0(m) = 0.5$,

$$S_i(\frac{m}{\rho_i}) = S_0(\rho_i \frac{m}{\rho_i}) = 0.5,$$

giving a median for group i of $\frac{m}{\rho_i}$. Similarly if the $100\alpha\%$ quantile of the baseline survival function is t_α, then the $100\alpha\%$ quantile of group i is $\frac{t_\alpha}{\rho_i}$.

3.2 Proportional Hazards models

In this model we assume that the hazards in the various groups are proportional so that

$$h_i(t) = \rho_i h_0(t)$$

where $h_0(t)$ is the baseline hazard. Hence we see that

$$S_i(t) = S_0(t)^{\rho_i}$$

Taking logs twice we get

$$\log\left(-\log S_i(t)\right) = \log \rho_i + \log\left(-\log S_0(t)\right)$$

So if we plot the RHS of the above equation against either t or $\log t$ we expect to see a vertical shift between groups.

3.2.1 Plots

Taking both models together it is clear that we should plot

$$\log\left(-\log \widehat{S}_i(t)\right) \text{ against } \log t$$

as then we can check for *AL and PH in one plot.* Generally $\widehat{S}_i$ will be calculated as the Kaplan-Meier estimator for group i, and the survival function estimator for each group will be plotted on the same graph.

(i) If the accelerated life model is plausible we expect to see a horizontal shift between groups.

(ii) If the proportional hazards model is plausible we expect to see a vertical shift between groups.

3.3 AL parametric models

There are several well-known parametric models which have the accelerated life property. These models also allow us to take account of continuous covariates such as blood pressure.

Name	$S(t)$	$h(t)$
Weibull	$\exp(-\,(\rho t)^{\alpha})$	$\alpha\rho^{\alpha}t^{\alpha-1}$
log-logistic	$\frac{1}{1+(\rho t)^{\alpha}}$	$\frac{\alpha\rho^{\alpha}t^{\alpha-1}}{1+(\rho t)^{\alpha}}$
log-normal	1-$\Phi\left(\frac{\log t+\log\rho}{\sigma}\right)$	$\cdots$
exponential	$\exp(-\rho t)$	ρ

Density function

Name	$f(t) = hS$
Weibull	$\alpha\rho^{\alpha}t^{\alpha-1}\mathrm{e}^{-(\rho t)^{\alpha}}$
log-logistic	$\frac{\alpha\rho^{\alpha}t^{\alpha-1}}{(1+(\rho t)^{\alpha})^{2}}$
log-normal	$\frac{1}{t\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\left(\log t+\log\rho\right)^2\right)$
exponential	$\rho\mathrm{e}^{\rho t}$

Remarks:

(i) Exponential is a submodel of Weibull with $\alpha = 1$

(ii) log-normal is derived from a normal distribution with mean $-\log\rho$ and variance σ^2. In this distribution $\alpha = \frac{1}{\sigma}$ has the same role as in the Weibull and log-logistic.

(iii) The **shape** parameter is α. The **scale** parameter is ρ.

Shape in the hazard function $h(t)$ is important.

Weibull $\cdots \begin{cases} h \text{ monotonic increasing } \alpha>1 \\ h \text{ monotonic decreasing } \alpha<1 \end{cases}$

log-normal $\cdots \quad h \longrightarrow 0$ as $t \longrightarrow 0, \infty$, one mode only

log-logistic $\cdots$ see problem sheet 5.

Comments:

a) to get a "bathtub" shape we might use a mixture of Weibull's. This gives high initial probability of an event, a period of low hazard rate and then increasing hazard rate for larger values of t.

b) to get an inverted "bathtub" shape we may have a mixture of log-logistics, or possibly a single log-normal or single log-logistic.

To check for appropriate parametric model (given AL checked)

There are some distributional ways of testing for say Weibull v. log-logistic etc., but they involve generalised F-distributions and are not in general use.

We can do a simple test for Weibull v. exponential as this simply means testing a null hypothesis $\alpha = 1$, and the exponential is a sub-model of the Weibull model. Hence we can use the likelihood ratio statistic which involves

$$2\log\widehat{L}_{weib} - 2\log\widehat{L}_{\exp} \sim \chi^2(1), \text{ asymptotically.}$$

3.3.1 Plots for parametric models

However most studies use plots which give a rough guide from shape. We should use a **straightline fit** as this is the fit which the human eye spots easily.

1) Exponential - $S = \mathrm{e}^{-\rho t}$, plot $\log S$ v. $\log t$

2) **Weibull** - $S = \mathrm{e}^{-(\rho t)^{\alpha}}$, plot $\log(-\log S)$ v. $\log t$

3) **log-logistic** - $S = \frac{1}{1+(\rho t)^{\alpha}}$, plot $\cdots$ see problem sheet 5

4) **log-normal** - $S = 1 - \Phi\left(\frac{\log t + \log \rho}{\sigma}\right)$, plot $\Phi^{-1}(1-S)$ v. $\log t$ or equivalently $\Phi^{-1}(S)$ v. $\log t$

In each of the above we would estimate S with the Kaplan-Meier estimator $\widehat{S}(t)$, and use this to construct the plots.

3.3.2 Regression in parametric AL models

In general studies each observation will have measured explanatory factors such as age, smoking status, blood pressure and so on. We need to incorporate these into a model using some sort of generalised regression. It is usual to do so by making ρ a function of the explanatory variables.

For each observation (say individual in a clinical trial) we set the scale parameter $\rho = \rho(\beta.x)$, where $\beta.x$ is a linear predictor composed of a vector x of known explanatory variables (covariates) and an unknown vector β of parameters which will be estimated. The most common link function is

$$\log \rho = \beta.x \ , \ \text{ equivalently } \ \rho = \mathrm{e}^{\beta.x} \ .$$

The idea is to mirror ordinary linear regression and find a baseline distribution which does not depend on ρ, similar to looking at the error term in least squares regression.

To give a derivation we will restrict to the Weibull distribution, but similar arguments work for all AL parametric models. We have

$$\begin{aligned} S(t) &= \mathrm{e}^{-(\rho t)^\alpha} = \Pr(T > t) \\ &= \Pr(\log T > \log t) \\ &= \Pr(\alpha\,(\log T + \log \rho) > \alpha\,(\log t + \log \rho)) \end{aligned}$$

Now let $Y = \alpha\,(\log T + \log \rho)$ and $y = \alpha\,(\log t + \log \rho)$.

$$\begin{aligned} \Pr(Y &> y) = S_Y(y) \\ &= S(t) \\ &= \mathrm{e}^{-(\rho t)^\alpha} \\ &= \exp(-\mathrm{e}^y) \end{aligned}$$

Hence we have

$$\log T = -\log \rho + \frac{1}{\alpha} Y, \ \text{ where } \ S_Y(y) = \exp(-\mathrm{e}^y)$$

The distribution of Y is independent of the parameters ρ and α. And in the case of the Weibull distribution its distribution is called the **extreme value distribution** and is as above.

In general we will write $\log T = -\log \rho + \frac{1}{\alpha} Y$ for all AL parametric models, and Y has a distribution in each case which is independent of the model parameters.

Name	$S(t)$	Y
Weibull	$\exp(-\,(\rho t)^{\alpha})$	$\log T = -\log \rho + \frac{1}{\alpha}Y$
log-logistic	$\frac{1}{1+(\rho t)^{\alpha}}$	$\log T = -\log \rho + \frac{1}{\alpha}Y$
log-normal	1-$\Phi\left(\frac{\log t+\log \rho}{\sigma}\right)$	$\log T = -\log \rho + \sigma Y$

as before $\alpha = \frac{1}{\sigma}$, for the log-normal.

Name	$S_Y(y)$	distribution
Weibull	$\exp(-\mathrm{e}^{y})$	extreme value
log-logistic	$\frac{1}{1+\mathrm{e}^{y}}$	logistic distribution
log-normal	$1-\Phi(y)$	N(0,1)

3.3.3 With real data (assuming right censoring only)

Censoring is assumed to be independent mechanism and is sometimes referred to as non-informative.

The **shape parameter** α is assumed to be the same for each observation in the study.

There are often very many covariates measured for each subject in a study.

A row of data will have perhaps:-

response - event time t_i , status δ_i (=1 if failure, =0 if censored)

covariates - age, sex, systolic blood pressure, treatment, and so a mixture of categorical variables and continuous variables amongst the covariates.

Suppose that Weibull is a good fit. Then

$$\begin{aligned} S(t) &= \mathrm{e}^{-(\rho t)^{\alpha}} \quad \text{and} \quad \rho = \mathrm{e}^{\beta.x} \\ \beta.x &= b_0 + b_1 x_{age} + b_2 x_{sex} + b_3 x_{sbp} + b_4 x_{trt} \end{aligned}$$

where b_0 is the intercept and all regression coefficients b_i are to be estimated, as well as estimating α. Note this model assumes that α is the same for each subject. We have not shown, but could have, interaction terms such as $x_{age} * x_{trt}$. This interaction would allow a different effect of age according to treatment group.

Suppose subject j has covariate vector x_j and so scale parameter

$$\rho_j = \mathrm{e}^{\beta.x_j}\ .$$

This gives a likelihood

$$\begin{aligned} L(\alpha,\beta) &= \prod_j \left(\alpha \rho_j^{\alpha} t_j^{\alpha-1}\right)^{\delta_j} \mathrm{e}^{-\left(\rho_j t_j\right)^{\alpha}} \\ &= \prod_j \left(\alpha \mathrm{e}^{\alpha\beta.x_j} t_j^{\alpha-1}\right)^{\delta_j} \mathrm{e}^{-\left(\mathrm{e}^{\beta.x_j} t_j\right)^{\alpha}}. \end{aligned}$$

We can now look for mle's for α and all components of the vector β, giving estimators $\widehat{\alpha}$, $\widehat{\beta}$ together with their standard errors ($=\sqrt{\text{var}\widehat{\alpha}}$, $\sqrt{\text{var}\widehat{\beta}_j}$) calculated from the observed information matrix (see problem sheet 5).

As already noted we can test for $\alpha = 1$ using

$$2 \log \widehat{L}_{weib} - 2 \log \widehat{L}_{\exp} \sim \chi^2(1), \text{ asymptotically.}$$

Packages allow for Weibull, log-logistic and log-normal models, sometimes others.

In recent years, a semi-parametric model has been developed in which the baseline survival function S_0 is modelled non-parametrically, and each subject has time t scaled to $\rho_j t$. This model is beyond the scope of this course.