Part IIA: Introduction to parametric survival models and the survival likelihood

- The exponential model
  - Properties of the contant-hazard model
  - Maximum likelihood estimation of the hazard
  - Intro to exponential regression models and Poisson regression
- Survival likelihood under the general hazard model
  - Survival (Poisson) likelihood
  - Cumulative hazards
- Examples of survival models
  - Weibull
  - Log-logistic

The exponential model

- The exponential model is the simplest survival model
- It is characterised by one parameter, a constant hazard λ
- An analogy to radioactive decay
  - Each nucleus has, irrespective of how long it has remained intact, a constant rate λ to decay
  - Starting with N nuclei at time $t = 0$, the number of nuclei remaining at time t is $N \exp(-\lambda t)$
- In other words, the survival function of the exponential distribution is $S(t) = \mathrm{P}(T > t) = \exp(-\lambda t)$, $t \geq 0$
- Although the exponential model may appear too simplistic in most applications, it is sometimes useful on its own right and quite often serves as a building block for more realistic models (e.g. models with piesewise constant hazards; see part IV)

Characteristics of the exponential distribution

- The survival function is $S(t) = \exp(-\lambda t)$, $t \geq 0$
- The density function is $f(t) = \lambda \exp(-\lambda t)$, $t \geq 0$

- The expected time until failure is $1/\lambda$
  - This is called *scale* ($= 1/$hazard) and sometimes used as the parameter of the exponential model
  - The dimension of hazard is “1/time” and that of scale ”time”
- The hazard does not depend on the history of the individual, not even on how long the individual has survived so far
  - This is why the distribution is said to be memoryless
  - The expected remaining lifetime at any(!) time t is always the same, i.e. $1/\lambda$
- N.B. Terms *hazard*, *hazard rate* or even *rate* are often used interchangeably

Survival likelihood under the exponential model

- Next we construct the likelihood function for a constant hazard λ, based on a follow-up of a study cohort with the following observations
  - occurrence times (failures and censorings) $t_i,\ i = 1, ..., N$
  - failure (event) indicators $d_i,\ i = 1, ..., N$
- Assuming independent observations across different individuals, the likelihood function based on the above data is a product over individual contributions:

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{N} L_i(\lambda) = \prod_{i=1}^{N} \lambda^{d_i} S(t_i) = \lambda^{\sum_{i=1}^{N} d_i} \times \prod_{i=1}^{N} \exp(-\lambda t_i) \\ &= \lambda^D \exp(-\lambda \sum_{i=1}^{N} t_i) = \lambda^D \exp(-\lambda Y) \end{aligned}$$

- $L(\lambda)$ is often called the Poisson likelihood (see page 14)

Maximum likelihood estimation

- Above, $Y = \sum_{i=1}^{N} t_i$ is the total person-time in the study cohort
  - It is the time spent 'at risk' for failure in the study cohort under observation.
- D is the total number of failures observed in the study cohort.
- A sufficient summary (statistics) of the data to estimate a constant hazard λ thus is
  - person-time $Y = \sum_{i=1}^{N} t_i$
  - the total number of failures $D = \sum_{i=1}^{N} d_i$
- The maximum likelihood estimate is easily found from the log likelihood $\log L(\lambda) = D \log(\lambda) - \lambda Y$:

$$\frac{d \log(L(\lambda))}{d\lambda} = \frac{D}{\lambda} - Y = 0 \;\Rightarrow \hat{\lambda} = D/Y$$

Standard error of the hazard

- Next, the standard error is derived for the log hazard $\beta = \log(\lambda)$

$$\beta = \log(\lambda) \;\Leftrightarrow \exp(\beta) = \lambda$$

- The log-likelihood of parameter β is

$$\log(L(\beta)) = D\beta - Y\exp(\beta)$$

- The derivatives are

$$\frac{d}{d\beta^2}(\log(L)) = D - Y\exp(\beta)$$
$$\frac{d^2}{d\beta^2}(\log(L)) = -Y\exp(\beta)$$

- The standard error is found to be
$S_\beta = \sqrt{1/(Y\exp(\hat{\beta})} = 1/\sqrt{Y\exp(\log(D/Y))} \;=\; 1/\sqrt{D}$

Example: leukaemia remission

- Time of remission in the treatment group of leukaemia patients ($*$ = censored):

$6^*, 6, 6, 6, 7, 9^*, 10^*, 10, 11^*, 13, 16, 17^*, 19^*, 20^*, 22, 23, 25^*, 32^*,$
$32^*, 34^*, 35^*$ (weeks)

- The sufficient data summary is $(D = 9, Y = 359)$ from which the point estimate of the rate is $\hat{\lambda} = D/Y = 9/359 = 0.025$ (per week)

- 90% CI based on the normal approximation of the log likelihood for parameter $\beta (= \log \lambda)$ and its standard error $S_\beta = 1/\sqrt{D}$:

$$\log(\hat{\lambda}) \pm 1.645/\sqrt{D}$$

- This leads to the following 90% confidence interval for λ:

$$\exp\left(\log(\lambda) \pm 1.645/\sqrt{D}\right) = 0.025 \overset{\times}{\div} \exp(1.645/\sqrt{9}) = [0.015, 0.043]$$

About parameter transformations

- There is often need to transform parameters estimates (including confidence intervals) between different one-to-one parameterisations of the model
- Above we calculated the confidence interval of the constant rate for logarithm, before transforming it back to the "absolute" scale
- N.B. Most statistical programmes tend to output logarithms of parameters
- Another example, should we be interested in the scale parameter (scale $= 1/$rate):
  - The hazard rate is 0.025 (per week), with a 90% confidence interval [0.015,0.043]
  - The point estimate of the scale parameter is $1/0.025 = 40$ (weeks) and the 90% confidence interval is $[1/0.043, 1/0.015]$ $= [23.3, 66.7]$

Exponential regression

- The hazard can be allowed to depend on explanatory variables (covariates) $Z_1, \ldots, Z_k$
- The hazard is then specified as follows:

$$\lambda(t; Z_1, \ldots, Z_k) \quad = \lambda(Z_1, \ldots, Z_k) = \lambda_0 \exp(Z_1\beta_1 + \ldots + Z_k\beta_k)$$

- The baseline hazard λ_0 is constant over time, pertaining to those with baseline values of the explanatory variables
- For any set of explanatory variables, the hazard remains a constant but a different from the baseline
- We return to regression models in general in Part 2A of the lectures

Poisson likelihood

- It was already mentioned that the survival likelihood is often called Poisson likelihood
- This is based on the notion that considering the observed number of failures, D, as the observation when a fixed total of Y time units of person-time (unit-time) has accrued in the study cohort, the distribution of D is proportional to the Poisson distribution with expectation λY:

$$P(D = d; \lambda) = \lambda^D \exp(-\lambda Y) \propto \frac{(\lambda Y)^D}{D!} \exp(-\lambda Y)$$

- It follows that exponential regression models can be fitted as Poisson regression because the expected counts of events in individuals with explanatory variables $Z_1, \ldots, Z_k$ is $\lambda(Z_1, \ldots, Z_k) Y(Z_1, \ldots, Y_k)$, where $Y(Z_1, \ldots, Y_k)$ is the person-time (unit-time) accrued in those individuals (see part IV)

Hazard revisited

- Recall that, in general, the hazard function is defined as the rate of change of the conditional failure probability:

$$\lambda(t) = \lim_{h \to 0} \frac{\mathrm{P}(T \in [t, t+h[| T \geq t)}{h}$$

- Assuming that h is short, the conditional failure probability over the time interval $[t, t+h[$, given survival until t, is

$$\pi_t = \mathrm{P}(T \in [t, t+h[| T \geq t) \simeq \lambda(t)h$$

- In the exponential model, the hazard (rate) is constant: $\lambda(t) = \lambda$ for all times t

Survival function and density function

- A survival model (i.e. distribution) can be defined terms of three alternative and equivalent ways: the hazard function $\lambda(t)$, the survival function $S(t)$ or the density function $f(t)$
- It was shown in Part I lectures that the following identities hold between the three functions:

$$
\begin{aligned}
S(t) &= \exp(-\int_0^t \lambda(u)du) \\
f(t) &= -\frac{dS(t)}{dt} = \lambda(t)S(t)
\end{aligned}
$$

- In particular, knowing any one function, the two other functions can be calculated

Survival models

- There are a large number of possible models for survival times
- In addition the exponential (constant and piece-wise constant hazard) models, some of the more commonly used ones include
  - Weibull distribution
  - Log-logistic distribution
  - Log-normal distribution
  - Gamma distribution
  - And more distributions: see e.g. Kalbfleisch and Prentice
- These distributions and others aim to add more flexibility in how the hazard may vary over time

Weibull model

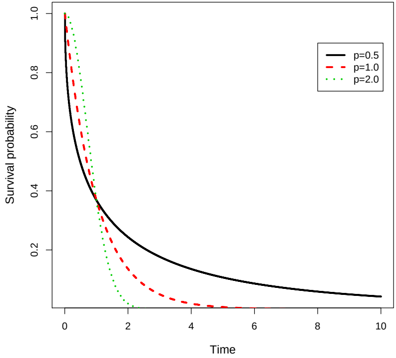
- The Weibull model is a more flexible family of survival distributions than the exponential model
- With one more parameter, it can account for hazard rates that either increase or decrease over time
  - Such models are required if, for example, the failure rate increases as the individual/unit grows older
- The model has two parameters. We use the following parameterisation for the hazard $\lambda(t)$ in terms of parameters α (rate) and p (shape):

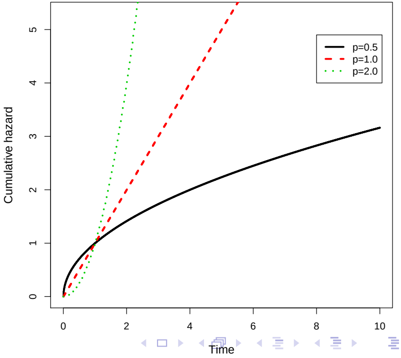
$$\lambda(t) = p\alpha^p t^{p-1}, \; t > 0$$

- If $p > 1 (< 1)$, the hazard increases (decreases) over time. If $p = 1$, we have the standard exponential model
- Warning: there are many alternative ways to parameterise the Weibull distribution. Always check which parameterisation is being used.

The survival function under the Weibull model

- The survival function is now $S(t) = \exp(-(\alpha t)^p)$ and the cumulative hazard $\Lambda(t) = (\alpha t)^p$
- The mean and the median of the Weibull distribution are $\Gamma(1 + 1/p)/\lambda$ and $(\ln 2)^{1/p}/\lambda$
- In the figures below, $\alpha = 1$ and the shape parameter is given three alternative values: $p = 0.5, 1.0, 2.0$





Checking the Weibull assumption

- Taking the log-log transformation of the survival function under the Weibull model we obtain

$$\log(-\log(S(t))) = p\log(\alpha) + p\log(t)$$

- In other words, the log-log transformed survival function is linear over logarithmic time under the Weibull model
  - Remember also that $-\log S(t)$ is the cumulative hazard
- This can be used to data exploration and/or model checking, based on the Kaplan-Meier estimate of the survival function

Log-logistic model

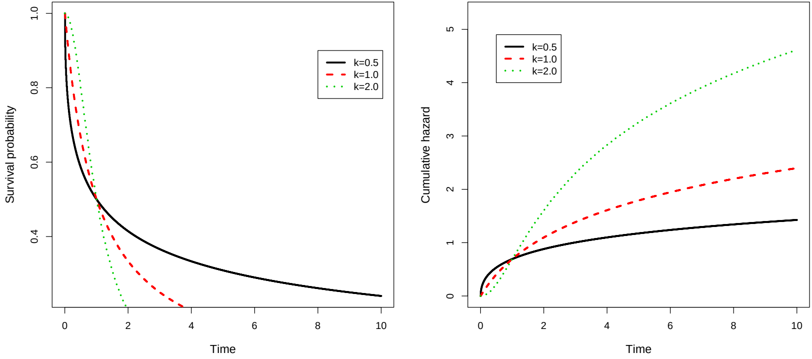
- The model has two parameters, rate α and shape k
- The hazard function with shape k and rate α is defined as

$$f(t) = \frac{k\alpha(\alpha t)^{(k-1)}}{\left[1 + (\alpha t)^k\right]^2}$$

- The survival function is simple (next page)
- The mean is $(k/\alpha)\pi/\sin(\pi/k)$ (if $k > 1$)
- The median is $1/\alpha$

The survival function in the log-logistic model

- The survival function is now $S(t) = 1/[1 + (\alpha t)^k]$ and the cumulative hazard $\Lambda(t) = \log[1 + (\alpha t)^k]$



Survival likelihood: the general case

- In what follows we take that $\lambda(t)$ is the parameter of the survival distribution
- Later, the explicit dependence of $\lambda(t)$ on explanatory variables will be made explicit (see, however, also p. 13 for the exponential model)
- Survival data are presented as (t_i, d_i), $i = 1, ..., N$
- The likelihood function for $\lambda(t)$ is again a product over individual contributions:

$$\begin{aligned} L(\lambda(t)) &= \prod_{i=1}^N L_i(\lambda(t)) = \prod_{i=1}^N \lambda(t_i)^{d_i} S(t_i) \\ &= \prod_{i=1}^N \left[\lambda(t_i)^{d_i)} \exp(-\int_0^{t_i} \lambda(u)du)\right] \\ &= \prod_{i=1}^N \left[\lambda(t_i)^{d_i)}\right] \times \exp(-\int_0^{\infty} \lambda(u)Y(u)du) \end{aligned}$$

where $Y(u) = \sum_{i=1}^N \mathbf{1}(t_i \geq u)$ is the size of the risk set at time u

More on censoring

- Censoring: individuals leave the study cohort still facing the risk of the event of interest
- Above we assumed that right censoring was *uninformative*:
  - the censored individuals are statistically alike with those remaining in the study
  - in particular, the reason of leaving the study is independent of the event of interest
- If censoring is *informative*, the analysis may lead to biased estimates
  - For example, if those that remain in the study are healthier, estimates of survival may be biased upwards
  - It is sometimes very difficult to know whether censoring could be informative if the censoring mechanism is not under the control of the investigator
- N.B. When the study is predefined to end at a certain time, censoring can safely be taken to be uninformative