4. Survival likelihood and parametric survival models

September 1 – 25, 2015

This lecture

- The general form of survival likelihood
- Parametric survival models
  - The hazard rate depends on time through (a small number of) parameters
- Parametric families of (continuous) survival distributions
  - Exponential
  - Weibull
  - More complex distributions
- Data exploration, model choice and model checking

Basic concepts revisited

- Let X_i be the time of event of interest and C_i the (random) censoring time. For each i, observed time $T_i = \min(X_i, C_i)$, censoring indicator $d_i = 1_{\{X_i \leq C_i\}}$, and a covariate vector Z_i are observed.
- The cumulative distribution function and the corresponding survival function of X_i are denoted as $F_i(t) = \mathrm{P}(X_i \leq t)$ and $S_i(t) = 1 - F_i(t) = \mathrm{P}(X_i > t)$.
- Hazard rate is defined as the momentary probability of event, given survival up to that moment:

$$\lambda_i(t) = \lim_{h \to 0} \frac{\mathrm{P}(X_i \in [t, t+h) \mid X_i \geq t)}{h}$$

Basic concepts revisited cont.

- Cumulative hazard is defined as

$$\Lambda_i(t) = \int_0^t \lambda_i(t)dt$$

- Cumulative hazard and the survival function have the following relation:

$$S_i(t) = \exp\{-\Lambda_i(t)\}$$

,

- The density function has the expression

$$f_i(t) = \lambda_i(t)S_i(t)$$

Parametric models of the hazard: notation

- In parametric modelling, the hazard rate $\lambda(t;\theta)$ is allowed to depend on time through a model with parameters θ
- We use the following notation (however, sometimes omitting to explicitly write the dependence on θ):
  - Survival function

$$S(t;\theta) = \exp(-\int_0^t \lambda(u;\theta)du) = \exp(-\Lambda(t;\theta)),\ \ t \geq 0$$

  - Cumulative hazard $\Lambda(t;\theta) = \int_0^t \lambda(u;\theta)du,\ \ t \geq 0$
  - Density function $f(t;\theta) = \lambda(t;\theta)S(t;\theta),\ \ t \geq 0$

Survival data and censoring

Assume the following data from a follow-up of a cohort of N individuals from time 0 to time T_{max}:

- observed times $t_i, \ i = 1, ..., N$
- censoring indicators $d_i, \ i = 1, ..., N$
- In summary, this mean the data are $(t_i, d_i), \ i = 1, \ldots, N$

In what follows we will assume uninformative censoring

- This means that the (random) censoring time is assumed to be independent of the parameters of interest θ (i.e. those defining the distribution of the event of interest)

Likelihood contributions under uninformative censoring

- Let $h(t;\phi)$ be the hazard of the censoring process, $\lambda(t;\theta)$ is the hazard of the event of interest, and θ and ϕ are separate
- Censoring can then be omitted from the likelihood expression of parameter(s) θ because the likelihood expression factorises completely
- If the $d_i = 1$, i.e., the event of interest occurs at time t_i, the whole likelihood contribution (for parameters θ and ϕ) is

$$
\begin{aligned}
L_i(\theta,\phi;t_i) &= \lambda(t_i;\theta)\exp\{-\int_0^{t_i}(\lambda(u;\theta)+h(u;\phi))du\} \\
&= \left[\lambda(t_i;\theta)\exp\{-\int_0^{t_i}\lambda(u;\theta)du\}\right] \times \exp\{-\int_0^{t_i} h(u;\phi)du\} \\
&\equiv \lambda(t_i;\theta)S(t_i;\theta) \times L^c_\phi(\phi;t_i)
\end{aligned}
$$

- If the $d_i = 0$, i.e., censoring occurs at time t_i:

$$
\begin{aligned}
L_i(\theta,\phi;t_i) &= h(t_i;\phi)\exp\{-\int_0^{t_i}(\lambda(u;\theta)+h(u;\phi))du\} \\
&= \exp\{-\int_0^{t_i}\lambda(u;\theta)du\} \times \left[h(t_i;\phi)\exp\{-\int_0^{t_i} h(u;\phi)du\right. \\
&\equiv S(t_i;\theta) \times L_\phi(\phi;t_i)
\end{aligned}
$$

Survival likelihood

Under uninformative censoring and for statistically independent individuals, the likelihood then is a product over individual contributions (omitting notation for θ):

$$
\begin{aligned}
L(\theta; \{(t_i, d_i); i = 1, \ldots, N\}) &= \prod_{i=1}^N \left(\lambda(t_i)^{d_i} S(t_i)\right) \\
&= \left(\prod_{i=1}^N \lambda(t_i)^{d_i}\right) \exp(-\sum_{i=1}^N \int_0^{t_i} \lambda(u) du) \\
&= \left(\prod_{i=1}^N \lambda(t_i)^{d_i}\right) \exp(-\int_0^{T_{max}} Y(u)\lambda(u) du),
\end{aligned}
$$

where $Y(u) = \sum_{i=1}^N Y_i(u) \equiv \sum_{i=1}^N \mathbf{1}(t_i \geq u)$ is size of the risk set at time u. Here $Y_i(u) = 1$ if individual i is still in the risk set (under observation) at time $u-$ (i.e. just before u).

The risk process

- Technically, the risk indicator $Y_i(t)$ and the risk set $Y(t)$ are
  - left-continuous stochastic processes
  - and so, they are *predictable*, e.g. $\lim_{u \to t-} Y(u) = Y(t)$

N.B. The individual that fails at time t is still included in the risk set at the failure time.

N.B. For standard survival data, in which everyone enters the risk set at time 0, the risk process $Y(t)$ is non-increasing. This is not true for more general models with late entry, i.e. when individuals may enter the risk set at at a later time than the time origin.

Left truncation and late entry

- Left truncation is another pattern of incomplete observation, usually tractable in the analysis
- It emerges when individuals may enter the study (i.e. the risk set) later than the time origin
- Likelihood contribution of an individual with entry at (random) time v_i and exit at time t_i with censoring indicator d_i:

$$\begin{aligned}\lambda(t_i)^{d_i} \mathrm{P}(X > t_i | X > V_i) &= \lambda(t_i)^{d_i} \frac{S(t_i)}{S(v_i)} \\ &= \lambda(t_i)^{d_i} \frac{\exp\{-\int_0^{t_i} \lambda(u) du\}}{\exp\{-\int_0^{v_i} \lambda(u) du\}} \\ &= \lambda(t_i)^{d_i} \exp\{-\int_{v_i}^{t_i} \lambda(u) du\}\end{aligned}$$

- This means that the only change in the general survival likelihood needed is to define the risk process as $Y(t) = \sum_{i=1}^{N} \mathbf{1}(v_i < u \leq t_i)$

(1) Exponential distribution

This is the simplest of all, with a *constant hazard rate* λ

- $S(t;\lambda) = \exp(-\lambda t),\ t > 0$
- $f(t;\lambda) = \lambda S(t),\ t > 0$
- $\Lambda(t;\lambda) = \lambda t,\ t > 0$

If $X \sim \text{Exp}(\lambda)$,

- mean $\text{E}(X) = 1/\lambda$
- variance $\text{Var}(X) = 1/\lambda^2$.
- coefficient of variation (the ratio of standard deviation to mean) is 1

Exponential distribution cont.: No memory!

- A constant conditional rate of occurrence means that there is no memory, i.e., the *mean residual life time* (rml) is always the same!

  $$\text{rml}(t) = \text{E}(X - t | X > t) = \text{E}(X) = 1/\lambda$$

N.B. Despite the simplistic assumption, the exponential distribution proves to be useful in many applications (cf. the use of piecewise constant hazards in non-parametric methods).

The log-likelihood under the exponential model

- The log-likelihood is

$$l(\lambda) = \log L(\lambda) = \sum_{i=1}^{N} d_i \log(\lambda) - \lambda \sum_{i=1}^{N} t_i = D \log(\lambda) - \lambda Y$$

where D is the total number of observed failures and Y is the total person-time.

- Together (D, Y) are the sufficient statistics for λ.
- The ML estimate is $\hat{\lambda} = D/Y$

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)

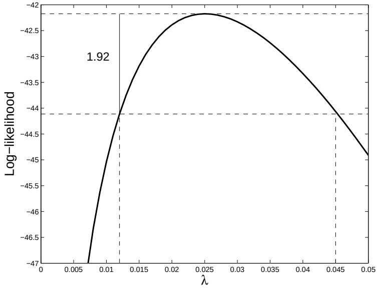
- The suffficient statistics ($D = 9$, $Y = 359$), from which the ML estimate

$$\hat{\lambda} = 9/359 = 0.025 \text{ (per week)}$$

Cox and Oakes: Analysis of Survival Data, Table 1.1

Confidence intervals

- 95% CI based on the log-likelihood function (see Figure): [0.012,0.045]
- 95% CI based on the normal approximation of the log-likelihood for parameter $\beta = \log \lambda$ and the standard error for β (see formula below): [0.013,0.048]



$$\left(\left[\frac{\partial^2 l}{\partial \beta^2} \right]_{\hat{\beta}} \right)^{-1/2} = \left(\frac{1}{D}\right)^{1/2} = 0.333$$

(2) Weibull distribution

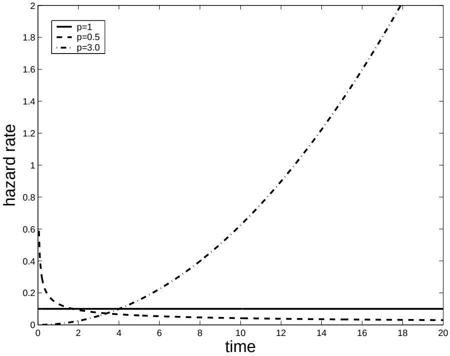
- The Weibull hazard is defined in terms of two parameters: α ("scale") and γ ("shape")
- Properties:

$$\begin{aligned}
\lambda(t;\alpha,\gamma) &= \alpha^{-1}\gamma(t/\alpha)^{\gamma-1},\ t>0 \\
S(t;\alpha,\gamma) &= \exp(-(t/\alpha)^{\gamma}),\ t>0 \\
\Lambda(t;\alpha,\gamma) &= (t/\alpha)^{\gamma},\ t>0
\end{aligned}$$

N.B. There are many alternative parameterisations of the Weibull distribution.

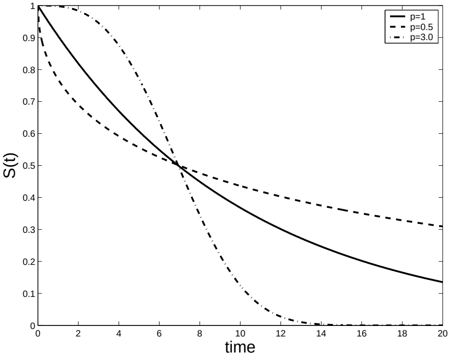
Weibull hazard

Depending on the shape parameter γ, the Weibull hazard can either increase or decrease with time:



Weibull survival function

The corresponding survival functions are:



Weibuill examples (cf. above)

- all three distributions have the same median (6.93) in this example
- when $\gamma = 1$, the distribution is *exponential*, i.e., the hazard rate is constant
- when $\gamma > 1$, the hazard rate is increasing with time
- when $\gamma < 1$, the hazard rate is decreasing with time

Likelihood estimation under the Weibull model

- The log-likelihood from a censored sample is (u denotes summation over the uncensored failure times only):

$$l(\alpha,\gamma) = D\log(\gamma) - D\gamma\log(\alpha) + (\gamma-1)\sum_u \log(t_i) - (1/\alpha)^\gamma \sum {t_i}^\gamma$$

- Even in the absence of censoring, there is no sufficient statistics for (α, γ)
- The score equations:

$$U_\alpha = \frac{\partial l}{\partial \alpha} = -\frac{D\gamma}{\alpha} + \gamma(1/\alpha)^{\gamma+1} \sum t_i^\gamma = 0$$

$$U_\gamma = \frac{\partial l}{\partial \gamma} = \frac{D}{\gamma} - D\log(\alpha) + \sum_u \log(t_i) - (1/\alpha)^\gamma \sum t_i^\gamma \log(t_i/\alpha) = 0$$

Likelihood estimation cont.

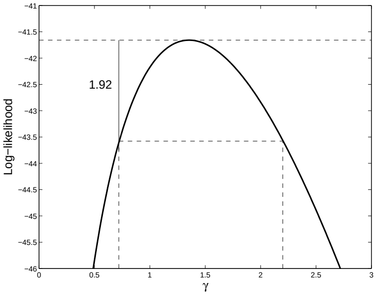
- From $U_\alpha = 0$, for a given value of γ, the ML estimator of α is $\alpha(\gamma) = \left((\sum t_i^\gamma)/D\right)^{1/\gamma}$
- Substituting $\alpha(\gamma)$ to $U_\gamma = 0$:

$$0 = \frac{D}{\gamma} + \sum\nolimits_u \log(t_i) - D \frac{\sum\limits_u t_i^\gamma \log(t_i)}{\sum t_i^\gamma}$$

- This can be solved numerically to find $\hat{\gamma}$ (and then $\hat{\alpha}$)
  - In the leukaemie example, $(1/\hat{\alpha}, \hat{\gamma}) = (0.030, 1.35)$

Profile likelihood

- Substituting $\alpha(\gamma)$ to the log-likelihood, we obtain the profile likelihood for γ
- This can be used to test for exponentiality:
  - The 95% confidence interval for γ is [0.70,2.20]
  - Exponentiality cannot be rejected (see also next slide)



Likelihood ratio test

- The reasoning above was based on the fact that the (profile) confidence interval included value 1
- Alternatively, a likelihood ratio test can be used (Weibull vs. the Exponential as a nested model):

$$-2\log\frac{L_1(\hat{\alpha})}{L_2(\hat{\alpha},\hat{\gamma})}\sim\chi^2_1$$

under the exponential model as the null hypothesis. Here L_1 refers to the exponential and L_2 to the Weibull model.

More complex distributions

- Survival distributions for time X can sometimes be defined through $\log(X) = \alpha + \sigma W$, where W has a defined distribution
- Different choices for the distribution of W lead to different models for X:
  - log-normal distribution, if $W \sim \text{Normal(0,1)}$ (i.e. the standard normal distribution)
  - log-logistic distribution, if W has the logistic density

$$\frac{\exp^w}{(1 + \exp^w)^2}$$

  - gamma distribution, if $\sigma = 1$ and the density of W

$$\frac{\exp(kw - \exp^w)}{\Gamma(k)}$$

- For some of these distributions, the hazard and/or the survival function are cumbersome to compute

Log-logistic distribution

- The density function is

$$f(x) = \lambda p(\lambda x)^{p-1}\left[1 + (\lambda x)^p\right]^{-2}, \quad \text{where } \lambda = \exp(-\alpha), \; p = \sigma^{-1}$$

- Here λ is a rate parameter and p a shape parameter
- Function *Llogis* in package *eha* (uses scale as 1/rate)
- Example (with rate 2.0 and shape 3):