STK4080 SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 14: Parametric survival models

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

https://www.ntnu.no/ansatte/bo.lindqvist
bo.lindqvist@ntnu.no

University of Oslo, Autumn 2019

What is a parametric model?

A model for a lifetime T is called *parametric* if it is given on the form $f(t;\theta)$, $F(t;\theta)$, etc., for functions which are “fixed” except for a parameter value θ which is allowed to vary in some prespecified interval or area. *Note:* θ may be a vector of several parameters.

Examples:

- $f(t;\theta) = \frac{1}{\theta}e^{-t/\theta}$, $F(t;\theta) = 1 - e^{-t/\theta}$; defined for all $\theta > 0$
  – *Exponential distribution with hazard (scale)* θ
- $f(t;a,b) = \frac{1}{\theta}e^{-(t/b)^a}$, $F(t;a,b) = 1 - e^{-(t/b)^a}$
  – *Weibull-distribution with shape=a and scale=b.*
  Here, $\boldsymbol{\theta} = (a,b)$ *is a vector.)*

Aim: *To estimate or test hypotheses about the true* θ *or* $\boldsymbol{\theta}$ *in a sample (possibly censored) of observations of* T.

Recall: Main censoring types

Lifetime data typically include *censored* data, meaning that:

- some lifetimes are known to have occurred only within certain intervals.
- The remaining lifetimes are known exactly.

Categories of censoring:

- right censoring
- left censoring
- interval censoring

Representation of censored data

Assume we have data for n units with *potential lifetimes* $T_1, T_2, \cdots, T_n \sim f(t;\theta)$.

Noncensored unit: Record the failure time T_i (*ideal case*)

Censored unit: Exact lifetime T_i is not recorded; all we know is that $T_i \in [a, b]$ for an interval of times.

Here

- a is the observed time, and $b = \infty$ for *right censorings*
- $a = 0$, while b is the observed time for *left censorings*
- $0 < a < b < \infty$ for an *interval censoring* between the observed interval limits a and b

Data for censored data may typically be represented as follows:

Unit no	start variable	end variable	Frequency (optional)
1	a_1	b_1	f_1
2	a_2	b_2	f_2
3	a_3	b_3	f_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$

An **uncensored** observation is entered by letting both a_i and b_i equal the observed lifetime.

Interval censored data can be analysed in R both nonparametrically and parametrically by the package `icenReg` and probably several other packages (*will not be considered in lectures*).

Likelihood construction: Illustrative example

Obs. type	Lower bound a_i	Upper bound b_i	Likelihood contribution
Exact lifetime	1.7	1.7	$f(1.7;\theta)$
Right cens.	2.0	∞	$1 - F(2.0;\theta)$
Left cens.	0	0.5	$F(0.5;\theta)$
Interval cens.	1.0	1.5	$F(1.5;\theta) - F(1.0;\theta)$

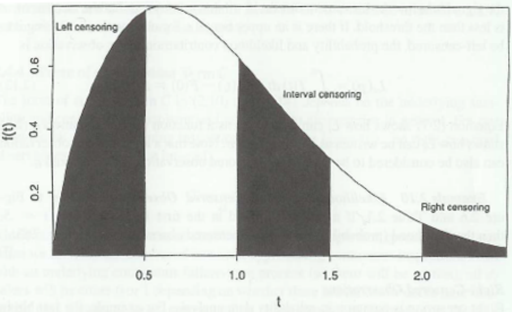


Figure 2.6. Likelihood contributions for different kinds of censoring.

Under the simplifying assumption that the lifetimes are independent and the censoring times are non-random, we obtaine the likelihood function

$$\begin{aligned}
L(\theta) &= \text{Probability of gettting the observed data under parameter } \theta \\
&= P_\theta(T_1 \in [a_1, b_1] \cap \cdots \cap T_n \in [a_n, b_n]) \\
&= P_\theta(T_1 \in [a_1, b_1]) \cdots P_\theta(T_n \in [a_n, b_n]) \\
&= (F(b_1;\theta) - F(a_1;\theta)) \cdots (F(b_n;\theta) - F(a_n;\theta)) \\
&= \prod_{i=1}^{n} (F(b_i;\theta) - F(a_i;\theta))
\end{aligned}$$

Recall $L(\theta) = \prod_{i=1}^{n}(F(b_i;\theta) - F(a_i;\theta))$.

- *Right censoring:* Here $b_i = \infty$, so the contribution to likelihood function is

$$F(\infty;\theta) - F(a_i;\theta) = 1 - F(a_i;\theta) = S(a_i,\theta)$$

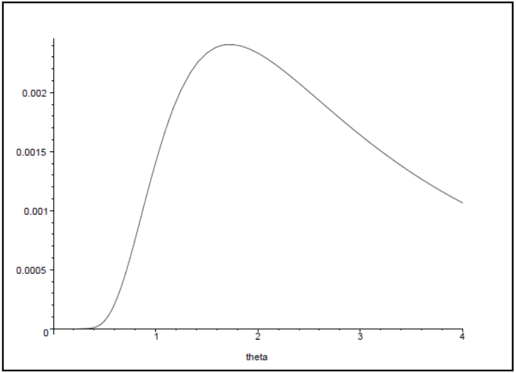
- *Left censoring:* Here $a_i = 0$, so contribution to likelihood is

$$F(b_i;\theta) - F(0;\theta) = F(b_i,\theta)$$

- *Interval censoring:* Contribution is $F(b_i;\theta) - F(a_i;\theta)$
- *Exact observed lifetime:* Then $a_i = b_i$. Write instead $b_i = a_i + h$, so contribution is $F(a_i + h;\theta) - F(a_i;\theta) \approx f(a_i)h$. Let contribution be just $f(a_i)$.

LIKELIHOOD FOR MODEL $f(t;\theta) = (1/\theta)e^{-t/\theta}$

$$L(\theta) = (\frac{1}{\theta}e^{-1.7/\theta}) \cdot (e^{-2.0/\theta}) \cdot (1 - e^{-0.5/\theta}) \cdot (e^{-1.0/\theta} - e^{-1.5/\theta})$$



Maximum likelihood estimate: $\hat{\theta} = 1.725$

With standard notation:

$$L(\theta) = \prod_{i:D_i=1} f(\tilde{T}_i;\theta) \cdot \prod_{i:D_i=0} S(\tilde{T}_i;\theta)$$

Log-location-scale models

A lifetime T has a *log-location-scale* family of distributions if $\log T$ has a *location-scale* family i.e.

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

where U has a “standardized” distribution centered around 0, with values in $(-\infty, +\infty)$. See next page for definitions.

- if $U \sim N(0,1)$, then $T \sim \text{lognormal}(\mu, \sigma)$
- if $U \sim \textit{logistic}(0,1)$, then $T \sim \text{log-logistic}(\mu, \sigma)$
- if $U \sim \textit{Gumbel}(0,1)$, then $T \sim \textit{Weibull}(a, b)$ with

$$\log b = \mu, \; 1/a = \sigma$$

Below are given, respectively, cdf and pdf of some "standardized" distributions, for $-\infty < u < \infty$.

Generic: $\Psi(u) = P(U \leq u),\ \psi(u) = \Psi'(u)$

Normal: $\Phi(u) = \int_{-\infty}^{u} \phi(x)dx, \quad \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$

Logistic: $H(u) = \frac{e^u}{1+e^u},\ h(u) = \frac{e^u}{(1+e^u)^2}$

Gumbel: $G(u) = 1 - e^{-e^u},\ g(u) = e^{u-e^u}$

Distribution of T

Let T be distributed as a log-location-scale family with

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

Then

$$\begin{aligned} F_T(t) &= P(T \le t) = P(\log T \le \log t) \\ &= P(\mu + \sigma U \le \log t) = P\big(U \le \frac{\log t - \mu}{\sigma}\big) \\ &= \Psi\big(\frac{\log t - \mu}{\sigma}\big) \end{aligned}$$

Thus

$$R_T(t) = 1 - \Psi\big(\frac{\log t - \mu}{\sigma}\big)$$

and

$$f_T(t) = \psi\big(\frac{\log t - \mu}{\sigma}\big) \cdot \frac{1}{\sigma t}$$

Likelihood for data from a general log-location-scale family:

$$L(\mu,\sigma) = \prod_{i:\delta_i=1} \psi\left(\frac{\log y_i - \mu}{\sigma}\right) \cdot \frac{1}{\sigma y_i} \cdot \prod_{i:\delta_i=0} \left(1 - \Psi\left(\frac{\log y_i - \mu}{\sigma}\right)\right)$$

and log-likelihood is

$$\ell(\mu,\sigma) = \sum_{i:\delta_i=1} \left(\log \psi\left(\frac{\log y_i - \mu}{\sigma}\right) - \log\sigma - \log y_i\right) + \sum_{i:\delta_i=0} \log\left(1 - \Psi\left(\frac{\log y_i - \mu}{\sigma}\right.\right.$$

Now the **observed information matrix** is

$$I(\hat{\mu},\hat{\sigma}) = \begin{bmatrix} -\frac{\partial^2 \ell(\mu,\sigma)}{\partial \mu^2} & -\frac{\partial^2 \ell(\mu,\sigma)}{\partial \mu \partial \sigma} \\ -\frac{\partial^2 \ell(\mu,\theta)}{\partial \mu \partial \sigma} & -\frac{\partial^2 \ell(\mu,\sigma)}{\partial \sigma^2} \end{bmatrix}$$

and

$$\left[I(\hat{\mu},\hat{\sigma})\right]^{-1} = \begin{bmatrix} \widehat{Var(\hat{\mu})} & \widehat{Cov(\hat{\sigma},\hat{\mu})} \\ \widehat{Cov(\hat{\mu},\hat{\sigma})} & \widehat{Var(\hat{\sigma})} \end{bmatrix}$$

Fractiles ξ_p for log-location scale families

Recall definition:

$$P(T \leq \xi_p) = p$$

$$p = P(T \leq \xi_p) = P(\log T \leq \log \xi_p) = \Psi\left(\frac{\log \xi_p - \mu}{\sigma}\right)$$

From this,

$$\Psi^{-1}(p) = \frac{\log \xi_p - \mu}{\sigma}$$

$$\log \xi_p = \mu + \sigma \Psi^{-1}(p)$$

$$\xi_p = e^{\mu + \sigma \Psi^{-1}(p)}$$

where $\Psi^{-1}(p)$ has to be calculated for each model, see next page.

T is **lognormal**: $\Phi^{-1}(p)$ is in our tables of standard normal distribution.
Particular percentiles:
Median : $\xi_{0.5} = e^{\mu+\sigma\Phi^{-1}(0.5)} = e^{\mu}$ as $\Phi^{-1}(0.5) = 0$
$\xi_{0.25} = e^{\mu+\sigma\Phi^{-1}(0.25)} = e^{\mu-0.675\sigma}$
$\xi_{0.75} = e^{\mu+\sigma\Phi^{-1}(0.25)} = e^{\mu+0.675\sigma}$

T is **Weibull**: Here we need $G^{-1}(p)$. Solving $G(u) = 1 - e^{-e^u} = p$ we get $u = G^{-1}(p) = \log(-\log(1-p))$ and hence

$$\begin{aligned}\xi_p &= e^{\mu+\sigma\log(-\log(1-p))} = e^{\log b+\frac{1}{a}\log(-\log(1-p))} \\ &= e^{\log b+\log[(-\log(1-p))^{1/a}]} \\ &= b\cdot(-\log(1-p))^{1/a}\end{aligned}$$

Example: `survreg(Surv(time, censor) ~ x1 + x2, dist="weibull", data=blabla.dat)`

NOTE: There are multiple ways to parameterize a Weibull distribution. The survreg function embeds it in a general *location-scale family*, which is a different parameterization than the one used by the `rweibull` function, which often leads to confusion:

- survreg's scale = 1/(rweibull shape), i.e. $\sigma = 1/a$.
- survreg's intercept = log(rweibull scale), i.e. $\mu = \log b$.

Try the example:

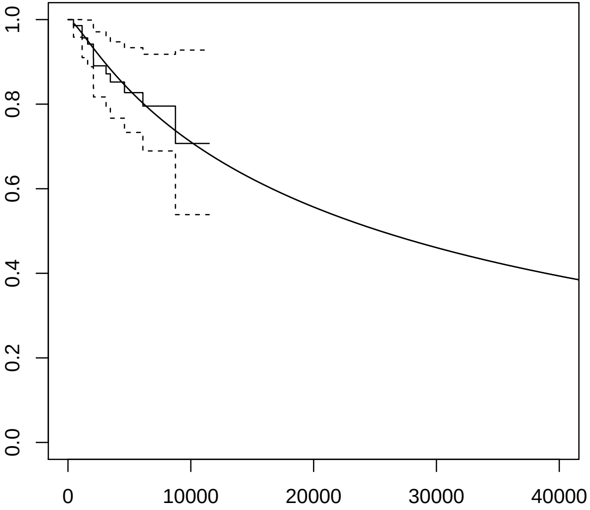
```
y <- rweibull(1000, shape=2, scale=5)
survreg(Surv(y)~1, dist="weibull")
```

```
fandat=read.table("https://folk.ntnu.no/bo/SIF5075/data
/NelsonFanDataTxt.txt",header=T)
names(fandat)
fit.logn=survreg(Surv(Y,D)~1, dist="lognormal",
data=fandat)
# Note that plotting of estimated survival function is not
# part of 'survreg' Here is a trick:
pct = 1:98/100
ptime = predict(fit.logn, type='quantile', p=pct, se=TRUE)
plot(x=ptime$fit[1,], y=1-pct, type="l")
```

Try also:

```
fit.KM=survfit(Surv(Y,D) ~ 1, data=fandat)
plot(fit.KM, xlim=c(0,40000))
lines(x=ptime$fit[1,], y=1-pct, type="l")
```

KM-plot and estimated lognormal plot for fan data



Parametric survival regression with `survreg`: Right censored data

Consider first data with a single fixed covariate: *Data:*

$$(\tilde{T}_1, D_1, x_1), (\tilde{T}_2, D_2, x_2), \ldots, (\tilde{T}_n, D_n, x_n)$$

Here $(\tilde{T}_i, D_i)$ *are, as before, the observed time and censoring status for unit* i*. Now in addition we have information on a covariate value* x_i *for each unit.*

Model:

- $\log T_i = \beta_0 + \beta_1 x_i + \sigma U_i; i = 1, \cdots, n$, where $x_1, \cdots, x_n$ are the covariates, and $U_1, \cdots, U_n$ are i.i.d., e.g., $N(0, 1)$, Gumbel$(0, 1)$, Logistic$(0, 1)$, etc.
- There may also be right censoring at time C_i, so we observe $(\tilde{T}_i, D_i)$ where $\tilde{T}_i = \min(T_i, C_i)$, and $D_i = I(\tilde{T}_i = T_i)$.

The extension from earlier implies estimation of β_0, β_1,σ *instead of earlier* μ, σ.

Likelihood function for regression data

We need the density and survival function of a pair (T, x):

$$f(t; \beta_0, \beta_1, \sigma) = \psi\left(\frac{\log t - \overbrace{(\beta_0 + \beta_1 x)}^{\mu \text{ before}}}{\sigma}\right)\frac{1}{\sigma t}$$

$$S(t; \beta_0, \beta_1, \sigma) = 1 - \Psi\left(\frac{\log t - \overbrace{(\beta_0 + \beta_1 x)}^{\mu \text{ before}}}{\sigma}\right)$$

So likelihood is $L(\beta_0, \beta_1, \sigma) =$

$$\prod_{i:D_i=1} \psi\left(\frac{\log \tilde{T}_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\frac{1}{\sigma \tilde{T}_i} \cdot \prod_{i:D_i=0} \left(1 - \Psi\left(\frac{\log \tilde{T}_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\right)$$

This is maximized w.r.t parameters, β_0, β_1, σ (`survreg` does it!)

$$I(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}) = \begin{bmatrix} -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_0^2} & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_0 \partial \beta_1} & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_0 \partial \sigma} \\ \cdot & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_1^2} & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_1 \partial \sigma} \\ \cdot & \cdot & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \sigma^2} \end{bmatrix}$$

inserted the estimated parameters. Further,

$$I(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})^{-1} = \begin{bmatrix} \widehat{Var}\hat{\beta}_0 & \cdot & \cdot \\ \cdot & \widehat{Var}\hat{\beta}_1 & \cdot \\ \cdot & \cdot & \widehat{Var}\hat{\sigma} \end{bmatrix}$$

where as usual the entries outside the diagonal are estimated covariances.

Model:

$$\begin{aligned}\log T &= \overbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}^{\mu} + \sigma U \\ &= \beta_0 + \boldsymbol{\beta}' \mathbf{x} + \sigma U\end{aligned}$$

$$\text{where} \qquad \boldsymbol{\beta} = \begin{pmatrix}\beta_1 \\ \beta_2 \\ \vdots \\ \beta_k\end{pmatrix}; \qquad \mathbf{x} = \begin{pmatrix}x_1 \\ x_2 \\ \vdots \\ x_k\end{pmatrix}$$

With data from n units:

$(y_i, \delta_i, x_{1i}, \cdots, x_{ki})$ or $(y_i, \delta_i, \mathbf{x}_i)$ for $i = 1, 2, \ldots, n$. Lifetimes satisfy:

$$\begin{aligned}\log T_i &= \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \sigma U_i \\ &= \beta_0 + \boldsymbol{\beta}' \mathbf{x}_i + \sigma U_i\end{aligned}$$

where $U_1, U_2, \ldots, U_n$ are i.i.d $\sim \Psi$. We can extend the observed information matrix to $(\beta_0, \cdots, \beta_k, \sigma)$

Standardized residuals

Recall model:

$$\log T_i = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_i + \sigma U_i$$

which implies

$$U_i = \frac{\log T_i - \beta_0 - \boldsymbol{\beta}' \mathbf{x}_i}{\sigma}$$

Recall also that $U_1, U_2, \cdot, U_n$ are i.i.d $\sim \Psi$, and define *standardized residuals* by

$$\hat{U}_i = \frac{\log \tilde{T}_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}' \mathbf{x}_i}{\hat{\sigma}}$$

Now the $\hat{U}_i$ should behave like a *right censored* set from the standard distribution, $N(0,1)$, Gumbel(0,1), Logistic(0,1) etc.

NOTE: If T has survival function $S(t)$ and cumulative hazard $A(t)$, then $A(T) = -\log S(T) \sim$ exponential(1)

Application here: Since $\log T_i = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_i + \sigma U_i$, we have

$$S_{T_i}(t) = 1 - \Psi\left(\frac{\log t - \beta_0 - \boldsymbol{\beta}' \mathbf{x}_i}{\sigma}\right)$$

and hence

$$V_i \equiv -\log S_{T_i}(T_i) = -\log\left[1 - \Psi\left(\frac{\log T_i - \beta_0 - \boldsymbol{\beta}' \mathbf{x}_i}{\sigma}\right)\right] \sim \text{expon}(1)$$

Cox-Snell residuals are now defined as

$$\begin{aligned} \hat{V}_i &= -\log\left[1 - \Psi\left(\frac{\log \tilde{T}_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}' \mathbf{x}_i}{\hat{\sigma}}\right)\right] \\ &= -\log\left[1 - \Psi\left(\text{standardized residuals}\right)\right] \end{aligned}$$

which should behave as a set of right-censored observations from expon(1) if the model is correctly specified.

Parametric counting process models

Consider counting processes

$$N_i(t);\, i = 1, 2, \ldots, n$$

that count the occurrences of an event of interest for n individuals.

Let the *intensity* process involve a parameter $\boldsymbol{\theta}$:

$$\lambda_i(t; \boldsymbol{\theta});\, i = 1, 2, \ldots, n$$

Recall that

$$\lambda_i(t; \boldsymbol{\theta})dt = P(dN_i(t) = 1 | \mathcal{F}_{t-})$$

General likelihood for parametric counting processes

Note that in general, the processes $N_i(t)$ are not independent due to various censoring mechanisms (e.g., type II censoring ...) Earlier we derived a likelihood for censored data assuming fixed censoring times. Now we will consider the general case.

Introduce the aggregated processes

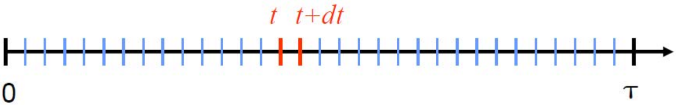
$$N_\bullet(t) = \sum_{i=1}^{n} N_i(t) \quad \text{and} \quad \lambda_\bullet(t;\boldsymbol{\theta}) = \sum_{i=1}^{n} \lambda_i(t;\boldsymbol{\theta})$$

and note that

$$P(dN_\bullet(t) = 1|\mathcal{F}_{t-}) = \lambda_\bullet(t;\boldsymbol{\theta})dt$$

(It should be noted that the λ_i-functions are in general *stochastic*, being functions of the history $\mathcal{F}_{t-}$).

General likelihood...



Divide the study time interval $[0, \tau]$ into small intervals $0 = t_0 < t_1 < \cdots < t_K = \tau$, each of length dt. Using the multiplicative probability rule we can then write $P(\text{data}) =$

$$
\begin{aligned}
&= \prod_{k=0}^{K-1} P(\text{data in } [t_k, t_k + dt) | \mathcal{F}_{t_k -}) \\
&= \prod_{k=0}^{K-1} \{ P(\text{events of interest in } [t_k, t_k + dt) | \mathcal{F}_{t_k -}) \\
&\times \ P(\text{other data in } [t_k, t_k + dt) | \text{events of interest in } [t_k, t_k + dt), \mathcal{F}_{t_k -}) \} \\
&\propto \prod_{k=0}^{K-1} P(\text{events of interest in } [t_k, t_k + dt) | \mathcal{F}_{t_k -})
\end{aligned}
$$

We will consider the partial likelihood

$$\text{Partlik} = \prod_{k=0}^{K-1} P(\text{events of interest in } [t_k, t_k + dt)|\mathcal{F}_{t_k-})$$

Conditional on the past, $\mathcal{F}_{t-}$, the occurrence of the events of interest in $[t, t+dt)$ can be considered as a single multinomial trial with $n+1$ possible outcomes: $\{dN_i(t) = 1\}, i = 1, 2, \ldots, n$; and $\{dN_\bullet = 0\}$. The conditional probability of the outcome is therefore

$$\begin{aligned} & P(\text{events of interest in } [t, t+dt)|\mathcal{F}_{t-}) \\ = & \left\{\prod_{i=1}^{n} P(dN_i(t) = 1|\mathcal{F}_{t-})^{dN_i(t)}\right\} P(dN_\bullet(t) = 0|\mathcal{F}_{t-})^{1-dN_\bullet(t)} \\ = & \left\{\prod_{i=1}^{n} (\lambda_i(t;\boldsymbol{\theta})dt)^{dN_i(t)}\right\} \{1 - \lambda_\bullet(t;\boldsymbol{\theta})dt\}^{1-dN_\bullet(t)} \end{aligned}$$

The partial likelihood now becomes a product-integral of these factors.

General likelihood

$$\text{Partlik} = \prod_{0<t\leq\tau} \left\{ \prod_{i=1}^{n} (\lambda_i(t;\boldsymbol{\theta})dt)^{dN_i(t)} \right\} \{1 - \lambda_\bullet(t;\boldsymbol{\theta})dt\}^{1-dN_\bullet(t)}$$

- The first part is just a product over the jump times of the counting processes.
- The exponent $1 - dN_\bullet(t)$ equals 1 for all but a finite number of time points t and can be replaced by 1.
- The dt will cancel on forming likelihood ratios and can be deleted.

Thus the partial likelihood may be given as

$$\begin{aligned} L(\boldsymbol{\theta}) &= \left\{ \prod_{0<t\leq\tau} \prod_{i=1}^{n} \lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)} \right\} \prod_{0<t\leq\tau} (1 - \lambda_\bullet(t;\boldsymbol{\theta})dt) \\ &= \left\{ \prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)} \right\} \exp\left\{ \int_0^\tau \lambda_\bullet(t;\boldsymbol{\theta})dt \right\} \end{aligned}$$

Likelihood for **right censored lifetimes**

Recall $L(\boldsymbol{\theta}) = \left\{\prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)}\right\} \cdot \exp\left\{-\int_0^\tau \lambda_\bullet(t;\boldsymbol{\theta})dt\right\}$

Suppose for the ith individual we have $\lambda_i(t;\boldsymbol{\theta}) = Y_i(t)\alpha(t;\boldsymbol{\theta})$. Then (since with right censored lifetimes there is at most one event for each individual)

$$\prod_{0<t\leq\tau} \lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)} = \alpha(\tilde{T}_i;\boldsymbol{\theta})^{D_i}$$

$$\begin{aligned}\exp\{-\int_0^\tau \lambda_\bullet(t;\boldsymbol{\theta})dt\} &= \exp\{-\sum_{i=1}^{n}\int_0^\tau Y_i(t)\alpha(t;\boldsymbol{\theta})dt\} \\ &= \exp\{-\sum_{i=1}^{n}\int_0^{\tilde{T}_i} \alpha(t;\boldsymbol{\theta})dt\}\end{aligned}$$

Thus $L(\boldsymbol{\theta})$ equals the likelihood that we have found before:

$$\prod_{i=1}^{n}\left\{\alpha(\tilde{T}_i;\boldsymbol{\theta})^{D_i}\exp\{-\int_0^{\tilde{T}_i}\alpha(t;\boldsymbol{\theta})dt\}\right\} = \prod_{i:D_i=1} f(\tilde{T}_i;\theta) \cdot \prod_{i:D_i=0} S(\tilde{T}_i;\theta)$$

Recall $L(\boldsymbol{\theta}) = \left\{\prod_{i=1}^{n}\prod_{0<t\leq\tau}\lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)}\right\}\cdot\exp\left\{-\int_0^\tau \lambda_\bullet(t;\boldsymbol{\theta})dt\right\}$

Suppose that n processes with the same intensity $\alpha(t;\boldsymbol{\theta})$ are observed, where the ith process, $N_i(t)$, is observed on the time interval $[0,\tau_i]$, with events at times $T_{i1},\ldots,T_{iN_i}$. For the ith process we have $\lambda_i(t;\boldsymbol{\theta}) = I(t\leq\tau_i)\alpha(t;\boldsymbol{\theta})$, so with $\tau = \max\{\tau_i\}$,

$$\prod_{0<t\leq\tau}\lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)} = \prod_{k=1}^{N_i(\tau_i)}\alpha(T_{ik};\boldsymbol{\theta})$$

$$\exp\{-\int_0^\tau \lambda_\bullet(t;\boldsymbol{\theta})dt\} = \exp\{-\sum_{i=1}^{n}\int_0^{\tau_i}\alpha(t;\boldsymbol{\theta})dt\}$$

Thus $L(\boldsymbol{\theta})$ equals

$$\prod_{i=1}^{n}\left\{\left(\prod_{k=1}^{N_i(\tau_i)}\alpha(T_{ik};\boldsymbol{\theta})\right)\exp\left\{-\int_0^{\tau_i}\alpha(t;\boldsymbol{\theta})dt\right\}\right\}$$

Statistical inference

Recall $L(\boldsymbol{\theta}) = \left\{\prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)}\right\} \cdot \exp\left\{-\int_0^\tau \lambda_\bullet(t;\boldsymbol{\theta})dt\right\}$

Log-likelihood:

$$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^\tau \log \lambda_i(t;\boldsymbol{\theta})dN_i(t) - \int_0^\tau \lambda_\bullet(t;\boldsymbol{\theta})dt$$

Score functions:

$$U_j(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j}\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^\tau \frac{\partial}{\partial\theta_j} \log \lambda_i(t;\boldsymbol{\theta})dN_i(t) - \int_0^\tau \frac{\partial}{\partial\theta_j}\lambda_\bullet(t;\boldsymbol{\theta})dt$$

It may be shown that the score functions $U_j(\boldsymbol{\theta})$ are stochastic integrals w.r.t. martingales when evaluated at the true value of the parameter.

This is key to prove that MLE $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \theta_q)$ enjoys "the usual" large sample properties.

The MLE may be found by maximizing the log-likelihood or by solving the likelihood equations (numerically, if needed)

As indicated, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \theta_q)$ is asymptotically normally distributed around its true value with a covariance matrix that may be estimated by

$$\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}$$

where $\mathbf{I}(\hat{\boldsymbol{\theta}})$ is the observed information matrix with elements

$$i_{hj}(\boldsymbol{\theta}) = -\frac{\partial}{\partial \theta_h} U_j(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \theta_h \partial \theta_j} \ell(\boldsymbol{\theta})$$

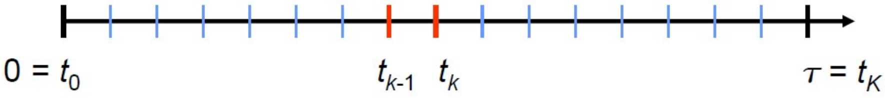
Alternatively we may use the expected information matrix (see the ABG-book for details).

The likelihood ratio, score and Wald tests apply as usual.

Consider now a model with covariates and proportional hazards:

$$\lambda_i(t;\boldsymbol{\theta},\boldsymbol{\beta}) = Y_i(t)\alpha_0(t;\boldsymbol{\theta})\exp(\boldsymbol{\beta}^T\mathbf{x}_i)$$

and piecewise constant baseline hazard $\alpha_0(t;\boldsymbol{\theta})$:



$$\alpha_0(t;\boldsymbol{\theta}) = \theta_k \quad \text{for} \quad t_{k-1} < t \leq t_k$$

Introduce:

$$\begin{aligned} O_{ik} &= N_i(t_k) - N_i(t_{k-1}) \\ R_{ik} &= \int_{t_{k-1}}^{t_k} Y_i(u)du \end{aligned}$$

$$L(\boldsymbol{\theta}) = \left\{ \prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)} \right\} \cdot \exp\left\{ -\int_0^{\tau} \lambda_\bullet(t;\boldsymbol{\theta})dt \right\}$$

$$= \left\{ \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{t_{k-1}<t\leq t_k} (\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} Y_i(t))^{\Delta N_i(t)} \right\} \exp\left\{ -\sum_{i=1}^{n} \sum_{k=1}^{K} \int_{t_{k-1}}^{t_k} \theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} Y_i(t)dt \right\}$$

$$= \prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ \left(\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} \right)^{O_{ik}} \cdot \exp\left(-\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik} \right) \right\}$$

$$\propto \prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ \left(\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik} \right)^{O_{ik}} \cdot \exp\left(-\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik} \right) \right\}$$

The likelihood is proportional to the likelihood of "independent Poisson variables" O_{ik} with "parameters"

$$\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik}$$

Estimation procedure

Recall expression: $\prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ \left(\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik}\right)^{O_{ik}} \cdot \exp\left(-\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik}\right) \right\}$

- Fit model by GLM-software treating the O_{ik} as "independent Poisson variables" with "parameters"

$$\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik} = \exp\left\{\psi_k + \boldsymbol{\beta}^T \mathbf{x}_i + \log R_{ik}\right\}$$

  with $\psi_k = \log \theta_k$.
- Use logarithmic link and $\log R_{ik}$ as offsets.
- Time interval number is treated as a categorical covariate
- An individual contributes one record to the data file for each time interval when at risk

Assume the covariate vectors $\mathbf{x}_i$ can only attain L distinct values: $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(L)}$

$$L(\boldsymbol{\beta},\boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{\ell=1}^{L}\left\{\left(\theta_k e^{\boldsymbol{\beta}^T\mathbf{x}_i}\right)^{O_k^{(\ell)}}\cdot \exp\left(-\theta_k e^{\boldsymbol{\beta}^T\mathbf{x}_i}R_k^{(\ell)}\right)\right\}$$

$$\propto \prod_{k=1}^{K}\prod_{\ell=1}^{L}\left\{\left(\theta_k e^{\boldsymbol{\beta}^T\mathbf{x}_i}R_k^{(\ell)}\right)^{O_k^{(\ell)}}\cdot \exp\left(-\theta_k e^{\boldsymbol{\beta}^T\mathbf{x}_i}R_k^{(\ell)}\right)\right\}$$

where

$$O_k^{(\ell)} = \sum_{i:\mathbf{x}_i=\mathbf{x}^{(\ell)}} O_{ik} = \sum_{i:\mathbf{x}_i=\mathbf{x}^{(\ell)}} (N_i(t_k) - N_i(t_{k-1}))$$

$$R_k^{(\ell)} = \sum_{i:\mathbf{x}_i=\mathbf{x}^{(\ell)}} R_{ik} = \sum_{i:\mathbf{x}_i=\mathbf{x}^{(\ell)}} \int_{t_{k-1}}^{t_k} Y_i(u)du$$

Recall Poisson regression in GLM

Assume $Y_i \sim Poisson(m_i \exp(\psi + \boldsymbol{\beta}^T \mathbf{x}_i))$ for $i = 1, \ldots, n$.

This means that

$$\log E(Y_i) = \log m_i + \psi + \boldsymbol{\beta}^T \mathbf{x}_i$$

The parameters ψ and $\boldsymbol{\beta}$ can be estimated in R by:

- Generalized linear model: `glm`
- with Poisson-family: `family=poisson`
- Need "`offset`" for $\log m_i$
- May also fit other link functions than default log-link

Example: Respiratory cancer for the Montana smelter workers

Here is the first part of the data (that has 114 entries)

```
> montana
   agegr calper hireper arsexp personyrs alldeaths respcanc allcanc circdis
1      1      1       1      1   3075.27        15        2       5       5
2      1      1       1      2    485.83         2        0       0       0
3      1      1       1      3    478.18         5        0       1       2
4      1      1       1      4    337.29         5        0       0       1
5      1      1       2      1   3981.61        27        2       2      10
6      1      1       2      2    656.06         5        1       2       1
7      1      1       2      3    190.34         2        0       0       1
8      1      1       2      4     12.46         1        0       0       1
9      1      2       1      1    936.75         4        0       1       2
10     1      2       1      2    194.58         1        1       1       0
11     1      2       1      3    164.87         2        1       1       0
12     1      2       1      4    121.00         0        0       0       0
13     1      2       2      1  10740.68        85        2       4      32
14     1      2       2      2   1696.77         9        1       2       4
15     1      2       2      3    870.52        14        0       1       6
16     1      2       2      4    224.00         0        0       0       0
17     1      3       2      1  12451.29       101        7      13      38
18     1      3       2      2   2511.97        15        0       2       5
19     1      3       2      3    868.35        12        0       2       5
```

Using R

```
# Read data:
path="http://www.uio.no/studier/emner/matnat/math/STK4080/h14/
montana.txt"
montana=read.table(path,header=T)
# Using Poisson regression in GLM
fit=glm(respcanc ~  offset(log(personyrs))+factor(agegr)+
factor(hireper)+factor(arsexp)-1,family=poisson,data=montana)
summary(fit)
```

Output

```
Call:
glm(formula = respcanc ~ offset(log(personyrs)) + factor(agegr) +
    factor(hireper) + factor(arsexp) - 1, family = poisson, data = montana)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7947  -0.8633  -0.2523   0.6103   2.3930

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
factor(agegr)1    -7.7419     0.2506 -30.894  < 2e-16 ***
factor(agegr)2    -6.3032     0.1598 -39.440  < 2e-16 ***
factor(agegr)3    -5.4580     0.1353 -40.346  < 2e-16 ***
factor(agegr)4    -5.2600     0.1544 -34.071  < 2e-16 ***
factor(hireper)2  -0.2916     0.1329  -2.194  0.02826 *
factor(arsexp)2    0.8195     0.1579   5.190 2.11e-07 ***
factor(arsexp)3    0.5727     0.2062   2.778  0.00547 **
factor(arsexp)4    0.8972     0.1815   4.943 7.70e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 246430.18  on 114  degrees of freedom
Residual deviance:    117.52  on 106  degrees of freedom
AIC: 352.31
```

Special case: No covariates

In the previous slide, let $L = 1$ and $\mathbf{x}^{(1)} = \mathbf{0}$. The model is then a model for i.i.d. individual lifetimes with piecewise constant hazard. Now

$$L(\boldsymbol{\theta}) = \prod_{k=1}^{K} \left\{ \theta_k^{O_k} \cdot \exp\left(-\theta_k R_k\right) \right\}$$

where

$$\begin{aligned} O_k &= \sum_{i=1}^{n} (N_i(t_k) - N_i(t_{k-1})) \\ &= N_\bullet(t_k) - N_\bullet(t_{k-1}) = \text{ “(total) occurrence” in } (t_{k-1}, t_k] \\ R_k &= \sum_{i=1}^{n} \int_{t_{k-1}}^{t_k} Y_i(u)du = \int_{t_{k-1}}^{t_k} Y_\bullet(u)du = \text{ “(total) exposure” in } (t_{k-1}, t_k] \end{aligned}$$

The log-likelihood is

$$\ell(\boldsymbol{\theta}) = \sum_{k=1}^{K} (O_k \log \theta_k - \theta_k R_k)$$

Statistical inference (piecewise constant hazard, no covariates)

Recall $\ell(\boldsymbol{\theta}) = \sum_{k=1}^{K} (O_k \log \theta_k - \theta_k R_k)$.

The score function hence are

$$U_k(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_k} \ell(\boldsymbol{\theta}) = \frac{O_k}{\theta_k} - R_k$$

so the MLE become

$$\hat{\theta}_k = \frac{O_k}{R_k} = \frac{\text{“occurrence”}}{\text{“exposure”}}$$

for $k = 1, 2, \ldots, K$, i.e. the so called *occurrence/exposure* rates.