STK4080/9080 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.edu/employees/bo.lindqvist
bo.lindqvist@ntnu.no
boli@math.uio.no

University of Oslo, Spring 2021

Parametric modeling (ABG Ch.5, ASAUR Ch. 10)

A model for a lifetime T is called *parametric* if it is given on the form $f(t;\boldsymbol{\theta})$, $F(t;\boldsymbol{\theta})$, etc., for functions which are "fixed" except for a parameter value $\boldsymbol{\theta}$ which is allowed to vary in some prespecified interval or area.

Examples:

- $f(t;b) = \frac{1}{b}e^{-t/b}$, $F(t;b) = 1 - e^{-t/b}$; defined for all $\theta > 0$
  – *Exponential distribution with hazard (scale) b.*
  Here, $\boldsymbol{\theta} = b$ *is one-dimensional.*
- $f(t;a,b) = \frac{a}{b}\left(\frac{t}{b}\right)^{a-1} e^{-(t/b)^a}$, $F(t;a,b) = 1 - e^{-(t/b)^a}$
  – *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 value of* $\boldsymbol{\theta}$ *in a sample of observations of* T *(possibly censored).*

Typical method: *Maximum likelihood estimation.*

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 (*type I, type II,...*)
- left censoring
- interval censoring

Special case: **Fixed** censoring times (see also 5.1.2 in ABG for the right censoring case)

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

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

Censored lifetime: 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

Representation of censored data with fixed censoring times

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 may then be 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 the course*).
- The above setup is standard in the package MINITAB.

Likelihood construction for fixed censoring times

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

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

Contributions to likelihood

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

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

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

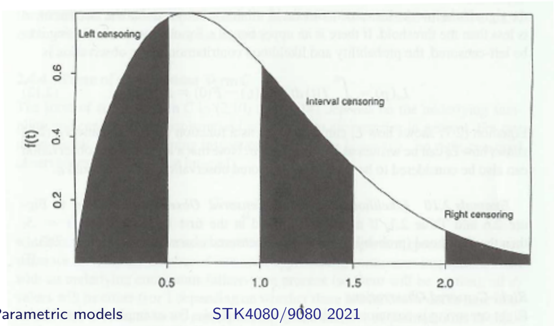
- *Left censoring:* Here $a_i = 0$, so contribution to likelihood is

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

- *Interval censoring:* Contribution is $F(b_i;\boldsymbol{\theta}) - F(a_i;\boldsymbol{\theta})$
- *Exact observed lifetime:* Then $a_i = b_i$. Write instead $b_i = a_i + \Delta$, so contribution is $F(a_i + \Delta;\boldsymbol{\theta}) - F(a_i;\boldsymbol{\theta}) \approx f(a_i;\boldsymbol{\theta})\Delta$. Let contribution be just $f(a_i;\boldsymbol{\theta})$ *(since Δ does not contain information about $\boldsymbol{\theta}$)*.

Likelihood construction: Illustrative example with $n = 4$ observed units

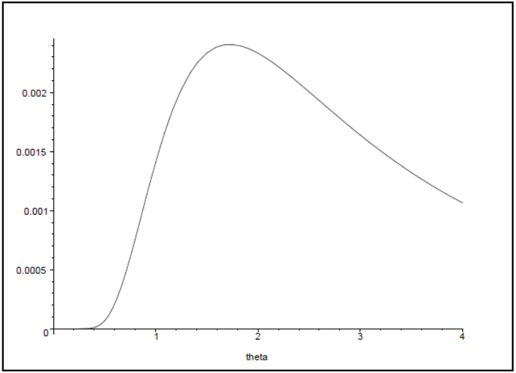
Obs. type	Lower bound a_i	Upper bound b_i	Likelihood contribution
Exact lifetime	1.7	1.7	$f(1.7; \boldsymbol{\theta})$
Right cens.	2.0	∞	$S(2.0; \boldsymbol{\theta})$
Left cens.	0	0.5	$F(0.5; \boldsymbol{\theta})$
Interval cens.	1.0	1.5	$F(1.5; \boldsymbol{\theta}) - F(1.0; \boldsymbol{\theta})$



Likelihood for illustrative example data

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: $\hat{\theta} = 1.725$

Special case:
Right censored data $(\tilde{T}_i, D_i)$ with fixed censoring times

It follows from the presented setup that for right censored data we have

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i:D_i=1} f(\tilde{T}_i;\boldsymbol{\theta}) \cdot \prod_{i:D_i=0} S(\tilde{T}_i;\boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} f(\tilde{T}_i;\boldsymbol{\theta})^{D_i}\, S(\tilde{T}_i;\boldsymbol{\theta})^{1-D_i} \end{aligned}$$

Recall

$f(t;\boldsymbol{\theta}) = \alpha(t;\boldsymbol{\theta}) \exp\{-\int_0^t \alpha(u;\boldsymbol{\theta})du\}; \quad S(t;\boldsymbol{\theta}) = \exp\{-\int_0^t \alpha(u;\boldsymbol{\theta})du\}.$

Thus

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \alpha(\tilde{T}_i;\boldsymbol{\theta})^{D_i} \; \exp\left\{\int_0^{\tilde{T}_i} \alpha(t;\boldsymbol{\theta})dt\right\}$$

Log-location-scale models (= Accelerated Failure Time models, AFT)

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 $(-\infty, +\infty)$.

- 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

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

Log-location scale models: Distributions for U

Recall, $\log T = \mu + \sigma U$.

$P(U \leq u)$ and corresponding density given by:

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

Logistic: $H(u) = \frac{e^u}{1+e^u},\ h(u) = \frac{e^u}{(1+e^u)^2} \quad \Rightarrow T \sim$ log-logistic(μ, σ)

Gumbel: $G(u) = 1 - e^{-e^u},\ g(u) = e^{u-e^u} \quad \Rightarrow T \sim$ Weibull(a, b), with $\log b = \mu,\ 1/a = \sigma$.

General distribution of T with $\log T = \mu + \sigma U$

Let $\Psi(u) = P(U \leq u),\ \psi(u) = \Psi'(u)$. Then

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

Thus

$$\begin{aligned} S_T(t) &= 1 - \Psi\big(\frac{\log t - \mu}{\sigma}\big) \\ f_T(t) &= \psi\big(\frac{\log t - \mu}{\sigma}\big) \cdot \frac{1}{\sigma t} \\ \alpha_T(t) &= \frac{\psi\big(\frac{\log t - \mu}{\sigma}\big) \cdot \frac{1}{\sigma t}}{1 - \Psi\big(\frac{\log t - \mu}{\sigma}\big)} \end{aligned}$$

Likelihood function for right-censored data

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

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

and log-likelihood is

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

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.

Accelerated Failure Time modeling in survival regression

Model:

$$\log T = \overbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_p}^{\mu} + \sigma U$$
$$= \beta_0 + \boldsymbol{\beta}^T \mathbf{x} + \sigma U$$

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

With data from n units:

$(\tilde{T}_i, D_i, \mathbf{x}_i)$ for $i = 1, 2, \ldots, n$. Underlying lifetimes are represented as:

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

where $U_1, U_2, \ldots, U_n$ are i.i.d $\sim \Psi$. We can extend the parametric likelihoods to this situation.

Weibull regression

Special case of AFT models.

Recall: If $T \sim \text{Weibull}(a, b)$, then

$$\begin{aligned} S(t) &= e^{-(\frac{t}{b})^a} \\ \alpha(t) &= \frac{at^{a-1}}{b^a} = ab^{-a}t^{a-1} \\ \log T &= \mu + \sigma W \equiv \log b + \frac{1}{a}W, \\ &\quad \text{where } W \sim \text{Gumbel}(0, 1) \end{aligned}$$

Weibull regression model for a lifetime T and covariate vector $\mathbf{x}$:

$$\log T = \underbrace{\beta_0 + \boldsymbol{\beta}^T \mathbf{x}}_{\log b} + \frac{1}{a}W$$

$$\text{Thus } b = e^{\beta_0 + \boldsymbol{\beta}^T \mathbf{x}}$$

Weibull regression has the proportional hazards property

(From previous page) Weibull regression model can be written:

$$T \sim \text{Weibull}(a, e^{\beta_0 + \boldsymbol{\beta}^T \mathbf{x}})$$

Hence the hazard rate function is

$$\begin{aligned}
\alpha(t|\mathbf{x}) &= a(e^{\beta_0 + \boldsymbol{\beta}^T \mathbf{x}})^{-a} t^{a-1} \\
&= \underbrace{a e^{-a\beta_0} t^{a-1}}_{\alpha_0(t)} \cdot e^{-a\boldsymbol{\beta}^T \mathbf{x}} \\
&= \alpha_0(t)\; e^{\tilde{\boldsymbol{\beta}}' \mathbf{x}}, \qquad \text{where } \tilde{\boldsymbol{\beta}} = -a\boldsymbol{\beta}
\end{aligned}$$

- This is of the form of Cox' proportional hazard, but here the model is **completely parametric**.
- The coefficients in $\boldsymbol{\beta}$ from Weibull regression will always have the **opposite sign** of those of Cox regression (which are $\approx \tilde{\boldsymbol{\beta}}$).
- The Weibull model is the **only** AFT model (log-location-scale model) that has the proportional hazards proprty.

Accelerated Failure Time modeling in R: survreg

Typical use:

```
survreg(Surv(time, censor == 1) ~ x1 + x2, dist="weibull")
```

Alternative distributions, e.g., "exponential", "lognormal" and "loglogistic"

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

- survreg's `scale` $= \sigma = 1/a = 1/($`rweibull` shape$)$
- survreg's `Intercept` $= \mu = \log b = \log($`rweibull` scale$)$,

Try the example:

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

coxph versus survreg in R (ASAUR p. 149-150)

```
> modelAll2.coxph <- coxph(Surv(ttr, relapse) ~ grp + age +
+   employment)
> summary(modelAll2.coxph)

  n= 125, number of events= 89
                     coef exp(coef) se(coef)      z Pr(>|z|)
grppatchOnly      0.60788   1.83654  0.21837  2.784  0.00537 **
age              -0.03529   0.96533  0.01075 -3.282  0.00103 **
employmentother   0.70348   2.02077  0.26929  2.612  0.00899 **
employmentpt      0.65369   1.92262  0.32732  1.997  0.04581 *
---
```

```
> model.pharm.weib <- survreg(Surv(ttr, relapse) ~ grp + age +
+   employment, dist="weibull")
> summary(model.pharm.weib)
                  Value Std. Error     z        p
(Intercept)      2.4024     0.9653  2.49 1.28e-02
grppatchOnly    -1.1902     0.4133 -2.88 3.98e-03
age              0.0697     0.0203  3.43 6.02e-04
employmentother -1.3890     0.5029 -2.76 5.74e-03
employmentpt    -1.3143     0.6132 -2.14 3.21e-02
Log(scale)       0.6313     0.0900  7.02 2.26e-12

Scale= 1.88
Weibull distribution
Loglik(model)= -454.1   Loglik(intercept only)= -466.1
        Chisq= 23.96 on 4 degrees of freedom, p= 8.2e-05
```

Parametric counting process models (ABG Chapter 5)

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

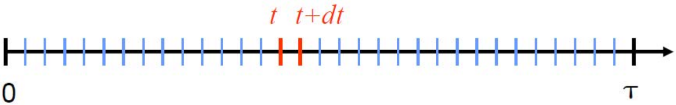
$$N_\bullet(t) = \sum_{i=1}^{n} N_i(t) \quad \text{and} \quad \lambda_\bullet(t;\boldsymbol{\theta}) = \sum_{i=1}^{n} \lambda_i(t;\boldsymbol{\theta})$$

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 $P(\text{data}) =$

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

General likelihood

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)

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

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

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

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

Likelihood for a **non-homogeneous Poisson process**

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 $T_{i1},\ldots,T_{iN_i(\tau_i)}$. 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 the MLE $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)$ enjoys "the usual" large sample properties.

The MLE may be found by maximizing the log-likelihood or by solving the likelihood equations $U_j(\boldsymbol{\theta}) = 0; \;\; j = 1, 2, \ldots, q.$

Statistical inference

As indicated, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)$ 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 ABG Section 5.3 for details – *not in curriculum*).

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

Recall Poisson regression in GLM

Assume Y_i are count data such that

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

The m_i are here typically numbers of initial counts having the same covariate vector $\mathbf{x}_i$. *(More generally they are weights of some kind).*

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 $\log m_i$

R-command might be:

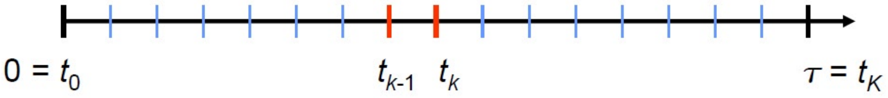
```
glm(Y ~ offset(log(m))+ x, family=poisson)
```

Poisson regression trick

Consider now a model with fixed 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 $\alpha_0(t; \boldsymbol{\theta})$:



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

Introduce:

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

Likelihood

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \left\{ \prod_{i=1}^{n} \prod_{0<t\le\tau} \lambda_i(t;\boldsymbol{\theta})^{\Delta N_i(t)} \right\} \cdot \exp\left\{ -\int_0^{\tau} \lambda_\bullet(t;\boldsymbol{\theta}) dt \right\} \\
&= \left\{ \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{t_{k-1}<t\le t_k} (\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} Y_i(t))^{\Delta N_i(t)} \right\} \exp\left\{ -\sum_{i=1}^{n} \sum_{k=1}^{K} \int_{t_{k-1}}^{t_k} \theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} Y_i(t) dt \right\} \\
&= \prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ \left(\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} \right)^{O_{ik}} \cdot \exp\left(-\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik} \right) \right\} \\
&\propto \prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ \left(\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik} \right)^{O_{ik}} \cdot \exp\left(-\theta_k e^{\boldsymbol{\beta}^T \mathbf{x}_i} R_{ik} \right) \right\}
\end{aligned}
$$

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

where $O_{ik} = N_i(t_k) - N_i(t_{k-1}), \; R_{ik} = \int_{t_{k-1}}^{t_k} Y_i(u)du.$

- 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 (default) and $\log R_{ik}$ as `offset`.
- Time interval number, k, must now be treated as a covariate, represented as a `factor` (see tutorial)
- The ith individual contributes one record to the data file for each time interval, k, when at risk: $(O_{ik}, R_{ik}, \mathbf{x}_i), \; k = 1, \ldots, K$
- This method is an alternative to Cox regression, which *is the limit when the partition of time axis is becoming finer and finer.*

Special case: Poisson regression with categorical covariates

Assume $\mathbf{x}_i$ can attain only L distinct values: $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(L)}$.
Then the likelihood simplifies:

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

where

$$
\begin{aligned}
O_k^{(\ell)} &= \sum_{i:\mathbf{x}_i = \mathbf{x}^{(\ell)}} O_{ik} = \sum_{i:\mathbf{x}_i = \mathbf{x}^{(\ell)}} \left(N_i(t_k) - N_i(t_{k-1})\right) \\
R_k^{(\ell)} &= \sum_{i:\mathbf{x}_i = \mathbf{x}^{(\ell)}} R_{ik} = \sum_{i:\mathbf{x}_i = \mathbf{x}^{(\ell)}} \int_{t_{k-1}}^{t_k} Y_i(u)du
\end{aligned}
$$

We can then use Poisson regression routines with the simplified data:

$$
O_k^{(\ell)} \sim \text{Poisson}\left(e^{\log(\theta_k) + \boldsymbol{\beta}^T \mathbf{x}^{(\ell)} + \log(R_k^{(\ell)})}\right), \quad k = 1, \ldots, K;\ \ell = 1, \ldots, L
$$

Example (see *Tutorial on Weibull and Poisson regression*)

Consider a right-censored sample $(\tilde{T}_i, D_i, x_i)$, $i = 1, \ldots, 23$, of survival times $\tilde{T}_i$ in the interval from 0 to 60, and where x is 0 or 1 (e.g., comparing two groups). Let the time axis be divided into the 6 intervals,

$$(0, 10], (10, 20], \ldots, (50, 60]$$

Let the model be given by hazard $\alpha_0(t; \boldsymbol{\theta}) e^{\beta x}$, where $\alpha_0(t; \boldsymbol{\theta})$ is constant on each of the above intervals, with respective values given by the vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_6)$.

The data needed for using Poisson regression are:
$\mathbf{O}^{(1)} = (4, 1, 2, 1, 1, 1) =$ *counts of events in the 6 intervals when* $x_i = 0$,
$\mathbf{O}^{(2)} = (1, 2, 1, 1, 2) =$ *counts in the 6 intervals when* $x_i = 1$.
$\mathbf{R}^{(1)} = (106, 68, 50, 23, 13) =$ *total exposure times in intervals when* $x_i = 0$
$\mathbf{R}^{(2)} = (109, 84, 61, 41, 27, 9) =$ *total exposure times when* $x_i = 1$.

The model parameters $\theta_1, \ldots, \theta_6, \beta$ are estimated from these (see tutorial).