STK4080/9080 SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 16: Multivariate frailty 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

Multivariate survival data (Chapter 7 in ABG)

Two major types of data:

- *Clustered survival data*
- *Recurrent (repeated) events data*

Three major analysis types:

- *Frailty models* (main interest in the course)
- *Marginal models* (will be briefly mentioned)
- *Dynamic models* (will not be covered in the course)

Example of clustered survival data

Duration of amalgam fillings in teeth (Aalen, Bjertness and Sønju, Statistics in Medicine, 1995)

- Clusters correspond to each of m patients
- ith patient has n_i fillings, that may fail, $i = 1, \ldots, m$
- T_{ij} = time to failure of jth filling of ith patient, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$
- Some patients may have a larger risk of failure than others, possibly due to varying dental hygiene or other factors. Thus: Observations within each cluster, $(T_{i1}, \ldots, T_{in_i})$ are *dependent*
- Observations are possibly right-censored

Clustered data from recurrent events: Small bowel motility (Aalen and Husebye, Statistics in Medicine 1991)

no.	completely observed periods									censored
1	112	145	39	52	21	34	33	51		54
2	206	147								30
3	284	59	186							4
4	94	98	84							87
5	67									131
6	124	34	87	75	43	38	58	142	75	23
7	116	71	83	68	125					111
8	111	59	47	95						110
⋮	⋮									⋮
17	162	141	107	69						39
18	106	56	158	41	41	168				13
19	147	134	78	66	100					4

Each person forms a cluster. Observations within each cluster: duration of MMC periods (last is censored)

Modeling of clustered survival data:
Shared frailty model

- Consider m independent individuals (clusters)
- Observations for ith individual (cluster): $(\widetilde{T}_{ij}, D_{ij})$, $j = 1, \ldots, n_i$, where D_{ij}, $j = 1, \ldots, n_i$ are censoring statuses (1 for failure, 0 for censoring)
- Each individual (cluster) has a *separate frailty variable* Z_i, such that conditonal on Z_i, the observations in the ith cluster are independent with hazard rate $Z_i\alpha(t)$
- Here $\alpha(t)$ is a basic hazard rate common to all individuals (assume, for now, no covariates)
- The Z_i are i.i.d. realizations of a random variable $Z > 0$, usually with expected value 1

Likelihood function for shared frailty model

History of ith cluster: $H_i = (\widetilde{T}_{ij}, D_{ij}, j = 1, \ldots, n_i)$.

For given value of the frailty Z_i, the contribution $P(H_i)$ to likelihood from ith cluster is (use formula for right censored data):

$$P(H_i \mid Z_i) = \prod_{j=1}^{n_i} \left[(Z_i \alpha(\widetilde{T}_{ij}))^{D_{ij}} \exp(-Z_i A(\widetilde{T}_{ij})) \right]. \tag{1}$$

where $A(t) = \int_o^t \alpha(s) ds$.

Since Z_i is unobserved, we need the unconditional value obtained by taking the expectation with respect to Z_i:

$$P(H_i) = \left\{ \prod_{j=1}^{n_i} \alpha(\widetilde{T}_{ij})^{D_{ij}} \right\} \mathrm{E}_{Z_i} \left\{ Z_i^{D_{i\bullet}} \exp(-V_i Z_i) \right\}. \tag{2}$$

where $D_{i\bullet} = \sum_{j=1}^{n_i} D_{ij}$ is the number of uncensored observations for cluster i, and $V_i = \sum_{j=1}^{n_i} A(\widetilde{T}_{ij})$.

Likelihood function for shared frailty model

Recall Laplace transform of Z:

$$\begin{aligned} \mathcal{L}(c) &= E\{\exp(-cZ)\} \\ \mathcal{L}^{(r)}(c) &= (-1)^r E\{Z^r \exp(-cZ)\} \end{aligned}$$

Then

$$P(H_i) = \left\{ \prod_{j=1}^{n_i} \alpha(\widetilde{T}_{ij})^{D_{ij}} \right\} (-1)^{D_{i\bullet}} \mathcal{L}^{(D_{i\bullet})}(V_i)$$

is the likelihood contribution of cluster i. The total likelihood of m independent clusters is the product of these, and by taking log:

$$\log L = \sum_{i=1}^{m} \left[\sum_{j=1}^{n_i} D_{ij} \log(\alpha(\widetilde{T}_{ij})) + \log\{(-1)^{D_{i\bullet}} \mathcal{L}^{(D_{i\bullet})}(V_i)\} \right] \qquad (3)$$

Can maximize this over a set of parameters to derive estimates.

Exercise 1

Simplify the expression for the log likelihood in the case where

- $\alpha(t) = \lambda$ is constant
- the frailties Z_i are Gamma$(1/\delta, 1/\delta)$ (i.e. gamma-distributed with expected value 1 and variance δ).

What is the sufficient data for this analysis?

Hint:

From subsection 7.2.4 on page 258 in book we have

$$\mathcal{L}(c) = \{1 + \delta c\}^{-1/\delta}.$$

$$\mathcal{L}^{(s)}(c) = \delta^s (-1)^s \{1 + \delta c\}^{-1/\delta - s} \prod_{q=1}^{s} (\frac{1}{\delta} + q - 1).$$

Maximum likelihood estimation for small bowel motility data (ABG Example. 7.1, p. 280)

- Weibull-model: $\alpha(t) = bt^{k-1}$ for $k > 0$
- Frailty Z is gamma-distributed with $E(Z) = 1,\ Var(Z) = \delta$
- $\mathcal{L}(c) = (1 + \delta c)^{-1/\delta}$

Maximum likelihood estimates (standard errors using standard asymptotic theory):

$$\begin{aligned} \log(\hat{b}) &= -10.0\ (1.0) \\ \hat{k} &= 2.28\ (0.22) \\ \hat{\delta} &= 0.146\ (0.12) \end{aligned}$$

Thus the Z_i can be expected to vary in $1 \pm 2\sqrt{\hat{\delta}}$ (why?) i.e. approximately 1 ± 0.75.

Or - is there a significant frailty effect?

Likelihood ratio test for frailty effect of small bowel data (see ABG p. 282

Will test

$$H_0 : \delta = 0 \text{ vs } H_1 : \delta > 0$$

$2\times$ (difference in log-likelihood of full model minus log likelihood of model with $\delta = 0$) $= 2.58$.

Approximate p-value (χ^2-distribution with 1 degree of freedom) is 0.108.

Because test is one-sided ($\delta = 0$ vs $\delta > 0$) is p-value 5.4%.

Thus: Close to a significant frailty effect.

Empirical Bayes estimate of individual frailties

Recall for Gamma(η, ν): Density $\propto z^{\eta-1}e^{-\nu z}$, expected value η/ν, variance η/ν^2. Assume frailty Z_i is Gamma$(1/\delta, 1/\delta)$.

Want to find conditional distribution of Z_i given data H_i from ith cluster:

Bayes' formula: $f(z_i|H_i) \propto$

$$P(H_i|Z_i = z_i)f_{Z_i}(z_i) \propto z_i^{D_{i\bullet}}e^{-V_i z_i}z_i^{(1/\delta)-1}e^{-z_i/\delta} = z_i^{D_{i\bullet}+(1/\delta)-1}e^{-(V_i+(1/\delta))z_i}$$

since from earlier, $P(H_i \mid Z_i) = \left\{\prod_{j=1}^{n_i}(Z_i\alpha(\widetilde{T}_{ij}))^{D_{ij}}\right\}\exp(-Z_i V_i)$,

where $V_i = \sum_{j=1}^{n_i} A(\widetilde{T}_{ij})$.

Thus $Z_i|H_i$ is Gamma$(D_{i\bullet} + (1/\delta), V_i + (1/\delta))$, so

$$E(Z_i|H_i) = \frac{D_{i\bullet} + (1/\delta)}{V_i + (1/\delta)}$$

These can be used to estimate the (unobserved) Z_i by using $\hat{\delta}$ and inserting estimated parameters in V_i.

Small bowel motility: Empirical Bayes estimates for frailties

Individual	Empirical Bayes estimate $\hat{Z} = \mathrm{E}(Z \mid \text{history})$ (and $\mathrm{SD}(Z \mid \text{history})$)	Estimated inter-event time estimate of $\mathrm{E}(T \mid \text{history})^*$ (and $\mathrm{SD}(T \mid \text{history})$)
1	1.46 (0.38)	89 (43)
2	0.72 (0.24)	123 (61)
3	0.55 (0.18)	137 (68)
4	1.08 (0.34)	103 (51)
5	0.93 (0.33)	111 (55)
6	1.40 (0.35)	90 (43)
7	1.07 (0.31)	102 (50)
8	1.13 (0.34)	100 (49)
9	0.91 (0.28)	110 (54)
10	0.85 (0.29)	114 (57)
11	1.27 (0.37)	95 (46)
12	1.11 (0.34)	101 (49)
⋮	⋮	⋮

* *See formula (7.10) in book.*

Maximum likelihood estimation for duration of amalgam fillings (ABG Example. 7.2 p. 283)

- Weibull-model: $\alpha(t) = bt^{k-1}$ for $k > 0$
- Frailty Z is gamma-distributed with $E(Z) = 1,\ Var(Z) = \delta$

Maximum likelihood estimates (Standard errors):

$$\begin{aligned} \log(\hat{b}) &= -4.21\ (0.25) \\ \hat{k} &= 0.43\ (0.10) \\ \hat{\delta} &= 0.85\ (0.31) \end{aligned}$$

Thus the Z_i have estimated standard deviation $\sqrt{0.85} = 0.92$ which is quite high.

P-value for frailty effect (testing of $\delta = 0$) is 0.35% (one-sided).

Next page:

Plots of Kaplan-Meier curves for selected separate clusters, with estimated Weibull survival function $\exp(-\hat{Z}_i \hat{b} t^{\hat{k}}/\hat{k})$

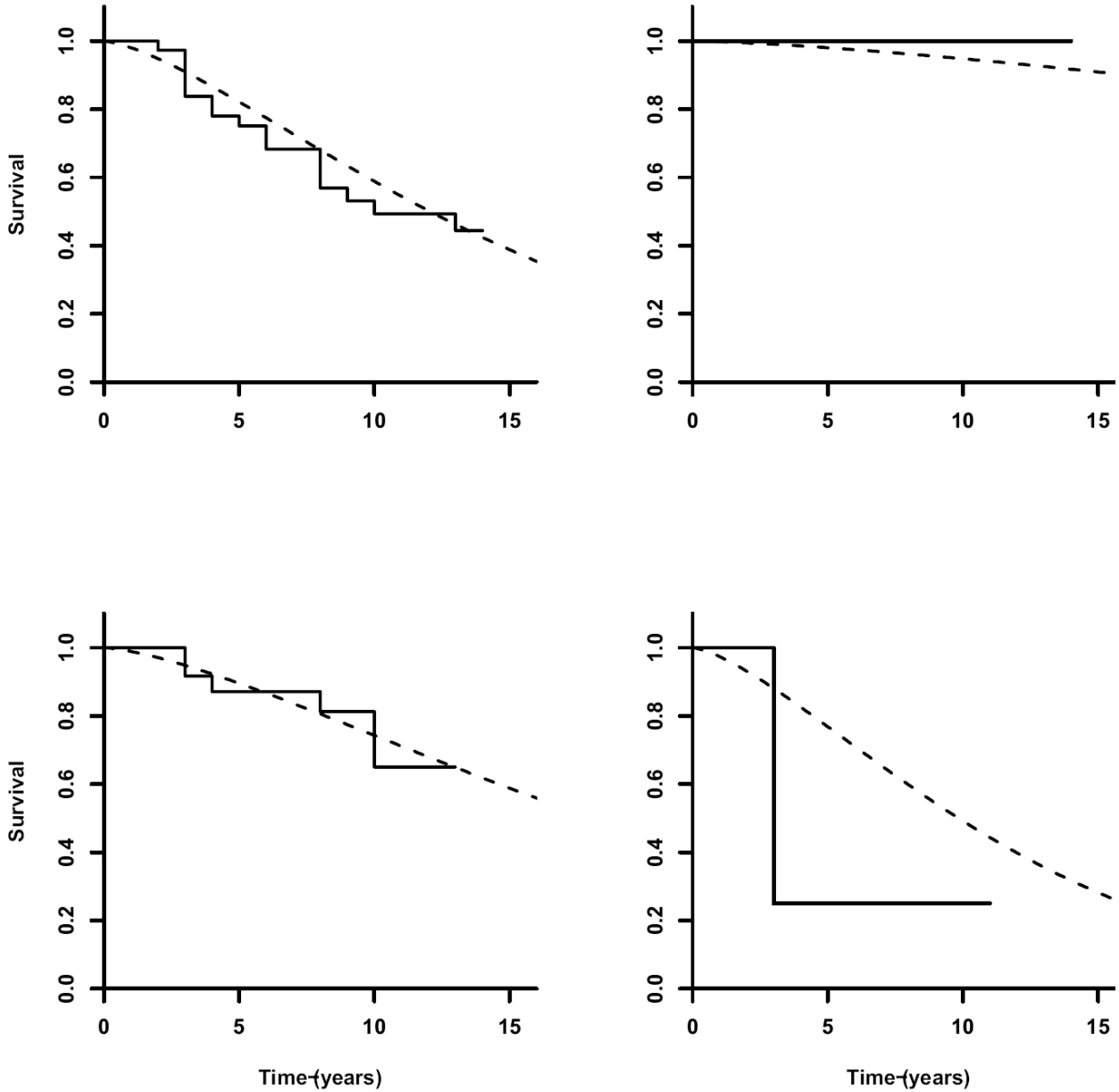


Fig. 7.1. *Survival of amalgam fillings for four selected patients. The step functions are individual Kaplan-Meier survival curves. These are compared to empirical Bayes survival functions computed from the frailty model*

Shared frailty **with** covariates

Assume

- ith cluster has frailty Z_i (as before)
- Lifetime T_{ij} for jth observation in ith cluster has
  - covariate vector X_{ij}
  - hazard conditional on Z_i,

$$Z_i \alpha(t) e^{\beta^T X_{ij}} = Z_i \alpha(t) \exp(\beta_1 X_{1ij} + \ldots + \beta_p X_{pij})$$

- $\alpha(t)$ is baseline hazard
- $\beta_1, \ldots, \beta_p$ are (usually) the interesting parameters
- Assume, e.g., $\alpha(t) = bt^{k-1}$ to get a purely *parametric* model
- Or, assume a nonparametric $\alpha(t)$ to get a *Cox-model with shared frailties* (the option of R)

Example of clustered data with covariates

The Diabetic Retinopathy Study, DRS (See ASAUR Example 9.2 p. 114)

- 197 patients with diabetic retinopathy
- Treatment: Laser photocoagulation randomly assigned to one eye of each patient
- Observe time to severe visual loss ("blindness") for each eye (time may be censored)
- There are 197 clusters with 2 observations per cluster

Data

T_{i1} = time to blindness of the *untreated* eye of the ith person

T_{i2} = time to blindness of the *treated* eye of the ith person

(T_{i1}, T_{i2}) are observations in the ith cluster (patient) - these are clearly dependent (and may be censored)

Note: $n_i = 2$ for all clusters.

DRS data (one line of data for each eye)

```
> diabetes=read.table("https://folk.ntnu.no/bo/STK4080/
diabetes.txt",header=T)
> head(diabetes)
```

id	time	status	trteye	treat	adult	agedx
5	46.24967	0	2	1	2	28
5	46.27553	0	2	0	2	28
14	42.50684	0	1	1	1	12
14	31.34145	1	1	0	1	12
16	42.30098	0	1	1	1	9
16	42.27406	0	1	0	1	9

- **id**: patient code
- **time**: survival time: time (in months) to blindness or censoring
- **status**: status for eye (0=censored; 1=blindness)
- **trteye**: random eye selected for treatment (1=right; 2=left)
- **treat**: treatment of eye (0=untreated eye; 1=treated eye)
- **adult**: type of diabetes (1= juvenile, 2= adult)
- **agedx**: age at diagnosis of diabetes

Note: If one eye gets blind, the other is still at risk (see id=14).

Exercise 2

Consider the DRS data. Assume that treatment (`treat`) is the only covariate "X", with value 0 for untreated eye and 1 for treated eye.

Let $H_i = \{(T_{i1}, D_{i1}), (T_{i2}, D_{i2})\}$ be the history of the ith patient.

Find simple expressions for $P(H_i|Z_i)$, $P(H_i)$ and the log likelihood for all the data.

Hint:

- *The hazard for an untreated eye (T_{i1}) is $Z_i\alpha(t)$*
- *the hazard for a treated eye (T_{i2}) is $Z_i\alpha(t)e^{\beta}$*

Modify the earlier expressions for the case without covariates.

R-output for RDS data with **frailty** (see ASAUR p. 120)

```
library(survival)
diabetes.frail = coxph(Surv(time, status) ~ treat +
as.factor(adult) + treat:as.factor(adult) + frailty(id),
data=diabetes)
```

```
> summary(diabetes.frail)
Call:
coxph(formula = Surv(time, status) ~ treat + as.factor(adult) +
    treat:as.factor(adult) + frailty(id), data = diabetes)

  n= 394, number of events= 155

                          coef    se(coef) se2    Chisq  DF     p
treat                    -0.5054 0.2255   0.2207   5.03  1.00 0.0250
as.factor(adult)2         0.3972 0.2591   0.2053   2.35  1.00 0.1300
frailty(id)                                      122.55 88.57 0.0098
treat:as.factor(adult)2 -0.9859 0.3618   0.3553   7.43  1.00 0.0064

                          exp(coef) exp(-coef) lower .95 upper .95
treat                        0.6032     1.6577    0.3878    0.9384
as.factor(adult)2            1.4876     0.6722    0.8952    2.4721
treat:as.factor(adult)2      0.3731     2.6803    0.1836    0.7581

Iterations: 6 outer, 31 Newton-Raphson
     Variance of random effect= 0.925909   I-likelihood = -847
Degrees of freedom for terms=  1.0  0.6 88.6  1.0
Concordance= 0.86  (se = 0.86 )
Likelihood ratio test= 218.4  on 91.12 df,   p=2e-12
```

A variant of frailty models: Cox regression models with **mixed effects** (ASAUR section 9.1.4, p. 120)

- Lifetime T_{ij} for jth observation in ith cluster has
  - covariate vector $\mathbf{X}_{ij}$
  - hazard, conditional on the value of a mixed effect U_i,

$$\alpha(t)e^{\boldsymbol{\beta}^T\mathbf{x}_{ij}+U_i} = \alpha(t)\exp(\beta_1 X_{1ij} + \ldots + \beta_p X_{pij} + U_i)$$

- $\alpha(t)$ is baseline hazard
- $\beta_1, \ldots, \beta_p$ are (usually) the interesting parameters
- The U_i are **mixed effects** defined for each cluster i, assumed to be i.i.d. and, for example, $N(0, \sigma^2)$
- Note that $Z_i = e^{U_i}$ can then be interpreted as a **frailty** variable

The above setup is used in the R-package `coxme`. For the RDS data, the appropriate commands are:

```
library(coxme)
diabetes.me = coxme(Surv(time, status) ~ treat +
as.factor(adult) + treat:as.factor(adult) + (1 | id),
  data=diabetes)
```

R-output for DRS-data with mixed effects

```
library(coxme)
diabetes.me = coxme(Surv(time, status)   treat + as.factor(adult)
+ treat:as.factor(adult) + (1 | id), data=diabetes)
```

```
> summary(diabetes.me)
Cox mixed-effects model fit by maximum likelihood
  Data: diabetes
  events, n = 155, 394
  Iterations= 7 39
                    NULL Integrated    Fitted
Log-likelihood -867.9511  -847.3837 -761.3231

                   Chisq    df          p   AIC     BIC
Integrated loglik  41.13  4.00 2.5207e-08 33.13   20.96
 Penalized loglik 213.26 77.99 1.6542e-14 57.28 -180.07

Model:  Surv(time, status) ~ treat + as.factor(adult) + treat:as.factor(adult) + (1 | id)
Fixed coefficients
                             coef exp(coef)  se(coef)     z      p
treat                  -0.4997742 0.6066676 0.2254101 -2.22 0.0270
as.factor(adult)2       0.3994717 1.4910367 0.2456771  1.63 0.1000
treat:as.factor(adult)2 -0.9680873 0.3798088 0.3616371 -2.68 0.0074

Random effects
 Group Variable  Std Dev   Variance
 id    Intercept 0.9171924 0.8412418
```

Marginal modeling by rate functions (ABG Chapter 8)

- Recall **shared frailty model**: Hazard rate conditional on frailty are

$$\alpha_{ij}(t|Z_i) = Z_i\alpha(t)e^{\beta^T X_{ij}} = Z_i\alpha(t)\exp(\beta_1 X_{1ij} + \ldots + \beta_p X_{pij})$$

- **Marginal model:** The idea is to model directly the *population hazard*, i.e., the hazard of a randomly drawn T_{ij}. This function is called the *rate function* and is assumed to be of the ordinary Cox-form

$$r_{ij}(t) = \alpha(t)e^{\beta^T X_{ij}} = \alpha(t)\exp(\beta_1 X_{1ij} + \ldots + \beta_p X_{pij})$$

- This avoids specific modelling of dependence
- Technically: We assume independence among observations, so called "working independence" (as in GEE)
- The lifetimes are, in general, not independent. HOWEVER, treating them as independent still leads to estimators that are consistent for the $\beta-$parameters.
- The method is computationally efficient since one uses the ordinary score equation from Cox' partial likelihood equations.
- **Variance** estimates for the β from Cox regression are, however, not valid. Instead are used so called "sandwich" estimates, which are valid under general dependence (robust estimates).

Exercise 3

Go back to Exercise 2. Modify the $P(H_i)$ and derive the likelihood function to be used in a marginal modelling appraoch for the DRS data.

Specialize to the case where $\alpha(t) = \lambda$. Derive the resulting estimates for λ and β.

Hint: The Z_i can now be set equal to 1.

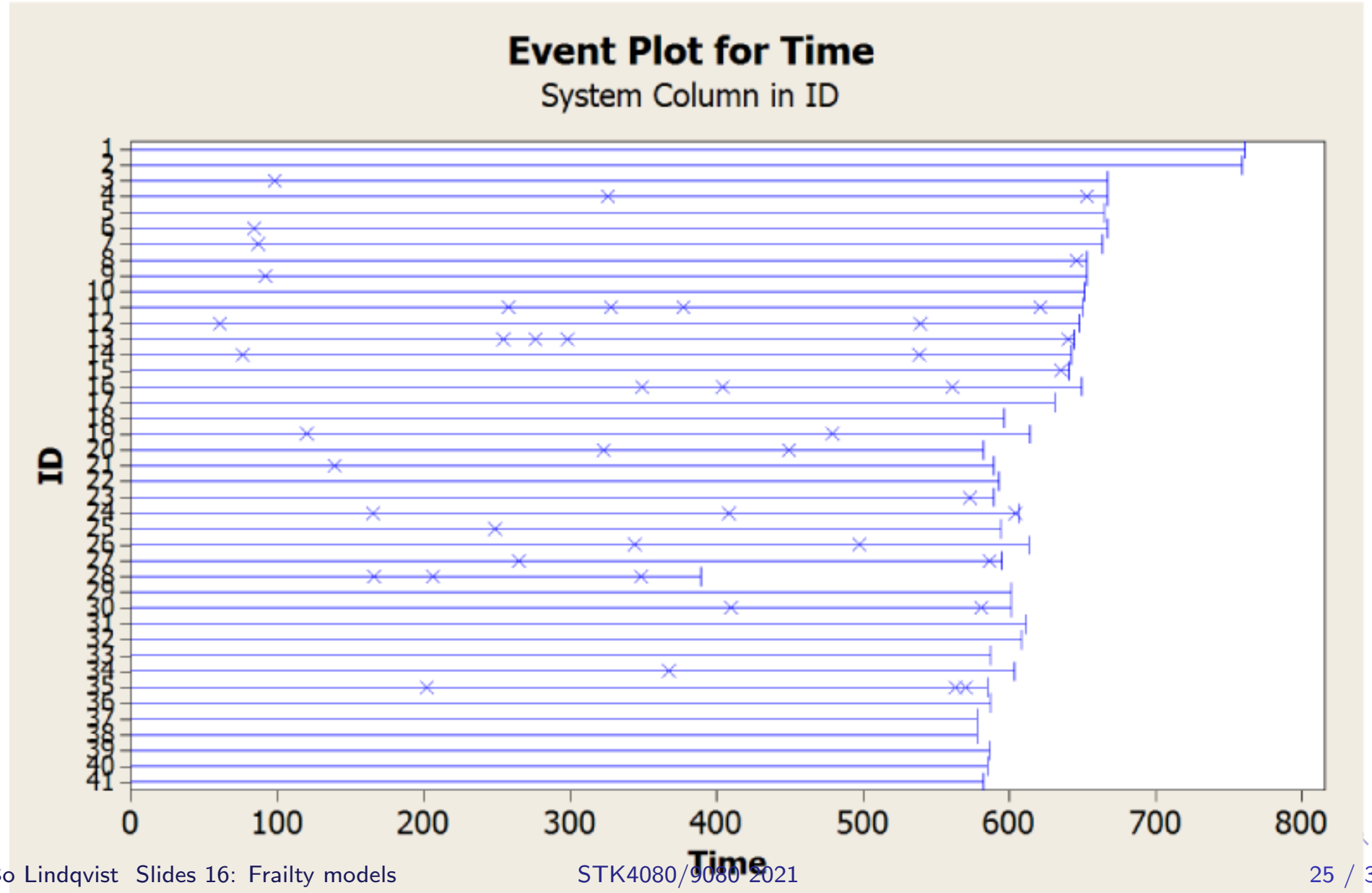
R-output for RDS data with marginal model

```
diabetes.marg = coxph(Surv(time, status) ~ treat +
as.factor(adult) + treat:as.factor(adult) + cluster(id),
data=diabetes)
```

```
> summary(diabetes.marg)
Call:
coxph(formula = Surv(time, status) ~ treat + as.factor(adult) +
    treat:as.factor(adult), data = diabetes, cluster = id)

  n= 394, number of events= 155

                           coef exp(coef) se(coef) robust se      z Pr(>|z|)
treat                   -0.4248    0.6539   0.2177    0.1850 -2.296   0.0217 *
as.factor(adult)2        0.3412    1.4067   0.1992    0.1957  1.743   0.0813 .
treat:as.factor(adult)2 -0.8464    0.4290   0.3509    0.3035 -2.788   0.0053 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                        exp(coef) exp(-coef) lower .95 upper .95
treat                      0.6539     1.5293    0.4550    0.9396
as.factor(adult)2          1.4067     0.7109    0.9585    2.0644
treat:as.factor(adult)2    0.4290     2.3311    0.2366    0.7777

Concordance= 0.613  (se = 0.019 )
Likelihood ratio test= 28.49  on 3 df,   p=3e-06
Wald test            = 34.93  on 3 df,   p=1e-07
Score (logrank) test = 28.44  on 3 df,   p=3e-06,   Robust = 30.31  p=1e-06

  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the Wald and robust score tests do not).
```

Recurrent events data: Nelson's valve data

Valve replacements of 41 diesel engines (Nelson, 1995)



Basic recurrent event models

Important distinction between recurrent event models is the *choice of basic time scale*

- Time measured from an initial event (e.g. birth, onset of disease)
  - Models: Poisson processes and their generalizations
  - Example: Valve data
- Time measured between events
  - Models: Renewal processes and their generalizations
  - Example: Aalen-Husebye data

EXAMPLE: The bladder cancer study:

- 86 patients with superficial bladder tumors
- Tumors were removed, and the patients randomized to placebo or treatment by thiotepa
- Patients were followed up, and the recurrence of tumors were registered (in the data used here, patients are censored after four recurrences)

Bladder cancer data

id	trt	fu	no	size	intno	start	stop	rec
1	1	1	1	3	1	0	1	0
2	1	4	2	1	1	0	4	0
...								
5	1	10	4	1	1	0	6	1
5	1	10	4	1	2	6	10	0
...								
59	2	21	1	1	1	0	17	1
59	2	21	1	1	2	17	19	1
59	2	21	1	1	3	19	21	0
...								

Note: One line for each time interval between events

id patient number
trt treatment (1=placebo; 2=thiotepa)
fu total time of follow-up
no initial number of tumors (8 denotes 8+)
size initial size of largest tumor (in cm)
intno number of the time interval
start start of the time interval
stop: end of the time interval
rec recurrence at end of the time interval (0=no; 1=yes)

Recurrent events: Frailty model with covariates

- ith individual (patient) has frailty Z_i
- ith individual has
  - covariate vector X_i
  - intensity conditional on Z_i,

$$Z_i \alpha(t) e^{\beta^T X_i} = Z_i \alpha(t) \exp(\beta_1 X_{1i} + \ldots + \beta_p X_{pi})$$

- $\alpha(t)$ is baseline intensity, either parametric or nonparametric
- R estimates a "Cox-model", i.e. assumes a nonparametric $\alpha(t)$.

R-output for bladder data with frailty

```
> bladder <- read.table("bladder.txt",header=T)
> bladderfit.frail <- coxph(Surv(start,stop,rec)~trt+frailty(id),data=bladder)
> summary(bladderfit.frail)
Call:
coxph(formula = Surv(start, stop, rec) ~ trt + frailty(id), data = bladder)

  n= 178
              coef   se(coef) se2   Chisq DF   p
trt           -0.396 0.335    0.202  1.4   1.0 0.24000
frailty(id)                           96.3  50.8 0.00012

    exp(coef) exp(-coef) lower .95 upper .95
trt     0.673       1.49     0.349      1.30

Iterations: 5 outer, 20 Newton-Raphson
     Variance of random effect= 1.28   I-likelihood = -439.8
Degrees of freedom for terms=  0.4 50.8
Rsquare= 0.562   (max possible= 0.994 )
Likelihood ratio test= 147  on 51.2 df,   p=3.27e-11
Wald test            = 1.4  on 51.2 df,   p=1
```

Recurrent events: Marginal model with covariates

- ith individual has (external) covariate vector $X_i(t)$ and rate function

$$r_i(t) = \alpha(t)e^{\beta^T X_i(t)} = \alpha(t)\exp(\beta_1 X_{1i}(t) + \ldots + \beta_p X_{pi}(t))$$

- $\alpha(t)$ is baseline intensity, either parametric or nonparametric
- R estimates a "marginal Cox-model", i.e. assuming a nonparametric $\alpha(t)$.

Remarks:

- Regression coefficients are estimated by maximizing the ordinary partial likelihood, ignoring dependencies due to frailties etc.
- The case is similar to the marginal models we considered for clustered data earlier. Thus we get consistent estimates for the β, but variance estimates from ordinary Cox estimations are not vallid.
- "Sandwich" type estimates are used for standard errors in R (called 'robust se' in output – next page).

R-output for bladder data with marginal model

```
> bladderfit.marg <- coxph(Surv(start,stop,rec)~trt+cluster(id),data=bladder
> summary(bladderfit.marg)
Call:
coxph(formula = Surv(start, stop, rec) ~ trt + cluster(id), data = bladder)

  n= 178
       coef exp(coef) se(coef) robust se     z    p
trt -0.373     0.688    0.198     0.283 -1.32 0.19

    exp(coef) exp(-coef) lower .95 upper .95
trt     0.688       1.45     0.395       1.2

Rsquare= 0.02   (max possible= 0.994 )
Likelihood ratio test= 3.68  on 1 df,   p=0.0552
Wald test            = 1.73  on 1 df,   p=0.188
Score (logrank) test = 3.61  on 1 df,   p=0.0575,   Robust = 1.81  p=0.179

  (Note: the likelihood ratio and score tests assume independence of
     observations within a cluster, the Wald and robust score tests do not).
```

Solution to Exercise 1

Simplify the expression for the log likelihood in the case where

- $\alpha(t) = \lambda$ is constant
- the frailties Z_i are Gamma$(1/\delta, 1/\delta)$ (i.e. gamma-distributed with expected value 1 and variance δ).

Solution: Observe first that under this model, $A(t) = \lambda t$ so $V_i = \lambda \sum_{i=1}^{n_i} \tilde{T}_{ij}$. Now putting $\alpha(t) = \lambda$ it follows from p. 7/35 that

$$P(H_i) = \lambda^{D_{i\bullet}} (-1)^{D_{i\bullet}} \mathcal{L}^{(D_{i\bullet})}(\lambda \sum_{i=1}^{n_i} \tilde{T}_{ij}).$$

Here one might insert the expression for the derivatives of the Laplace transform in the gamma case. (We omit this).

It is already seen from the above that it suffices to know, for each cluster i,

- $D_{i\bullet} = \#$ events in ith cluster
- $\sum_{i=1}^{n_i} \tilde{T}_{ij} =$ total time observed ith cluster

Solution Exercise 2

History of ith cluster: $H_i = (\widetilde{T}_{ij}, D_{ij}, x_{ij}, j = 1, \ldots, n_i)$.

For given value of the frailty Z_i, the contribution $P(H_i)$ to likelihood from ith cluster is

$$P(H_i \mid Z_i) = \prod_{j=1}^{n_i} \left[(Z_i \alpha(\widetilde{T}_{ij}) e^{\beta x_{ij}})^{D_{ij}} \exp(-Z_i A(\widetilde{T}_{ij}) e^{\beta x_{ij}}) \right]. \qquad (4)$$

Then take the expectation with respect to Z_i:

$$P(H_i) = \left\{ \prod_{j=1}^{n_i} \left(\alpha(\widetilde{T}_{ij}) e^{\beta x_{ij}} \right)^{D_{ij}} \right\} \mathrm{E}_{Z_i} \left\{ Z_i^{D_{i\bullet}} \exp(-V_i Z_i) \right\}. \qquad (5)$$

where $D_{i\bullet} = \sum_{j=1}^{n_i} D_{ij}$ is the number of uncensored observations for cluster i, and $V_i = \sum_{j=1}^{n_i} A(\widetilde{T}_{ij}) e^{\beta x_{ij}}$.

Solution Exercise 2 (cont.)

Thus

$$P(H_i) = \left\{ \prod_{j=1}^{n_i} \left(\alpha(\widetilde{T}_{ij}) e^{\beta x_{ij}} \right)^{D_{ij}} \right\} (-1)^{D_{i\bullet}} \mathcal{L}^{(D_{i\bullet})}(V_i).$$

is the likelihood contribution of cluster i. The total likelihood of m independent clusters is the product of these, and by taking log:

$$\log L = \sum_{i=1}^{m} \left[\sum_{j=1}^{n_i} D_{ij} \log(\alpha(\widetilde{T}_{ij})) + \sum_{j=1}^{n_i} D_{ij} \beta x_{ij} + \log \left\{ (-1)^{D_{i\bullet}} \mathcal{L}^{(D_{i\bullet})}(V_i) \right\} \right]$$

For the RDS data we have $x_{i1} = 0, x_{i2} = 1$, so

$$V_i = A(\tilde{T}_{i1}) + A(\tilde{T}_{i2}) e^{\beta}$$

Also, the term $\sum_{j=1}^{n_i} D_{ij} \beta x_{ij}$ in $\log L$ can be replaced by just $D_{i2}\beta$.

Solution Exercise 3

Now we put $Z_i \equiv 1$, so

$$P(H_i) = \left\{ \prod_{j=1}^{n_i} \left(\alpha(\widetilde{T}_{ij}) e^{\beta x_{ij}} \right)^{D_{ij}} \right\} \exp\{-V_i\}.$$

is the likelihood contribution of cluster i. The total likelihood of m independent clusters is the product of these, and by taking log:

$$\log L = \sum_{i=1}^{m} \left[\sum_{j=1}^{n_i} D_{ij} \log(\alpha(\widetilde{T}_{ij})) + \sum_{j=1}^{n_i} D_{ij} \beta x_{ij} - V_i \right]$$

For the RDS data we have $x_{i1} = 0, x_{i2} = 1$, so

$$V_i = A(\tilde{T}_{i1}) + A(\tilde{T}_{i2}) e^{\beta}$$

Also, the term $\sum_{j=1}^{n_i} D_{ij} \beta x_{ij}$ in $\log L$ can be replaced by just $D_{i2}\beta$.

Solution Exercise 3 (cont.)

Thus for the RDS data we get

$$\log L = \sum_{i=1}^{m} \left[\sum_{j=1}^{2} D_{ij} \log(\alpha(\widetilde{T}_{ij})) + D_{i2}\beta - A(\tilde{T}_{i1}) - A(\tilde{T}_{i2})e^{\beta} \right]$$

If we specialize to $\alpha(t) = \lambda$, we get

$$\begin{aligned} \log L &= \sum_{i=1}^{m} \left[D_{i\bullet} \log \lambda + D_{i2}\beta - \lambda \tilde{T}_{i1} - \lambda \tilde{T}_{i2} e^{\beta} \right] \\ &= D_{\bullet\bullet} \log \lambda + D_{\bullet 2}\beta - \lambda R_1 - \lambda R_2 e^{\beta} \qquad \text{where } R_j = \textstyle\sum_{i=1}^{n} \tilde{T}_{ij}. \end{aligned}$$

By taking derivatives, equating to 0 and solving the equations, one gets the maximum likelihood estimates

$$\hat{\lambda} = \frac{D_{\bullet 1}}{R_1}, \qquad \exp(\hat{\beta}) = \frac{R_1 D_{\bullet 2}}{R_2 D_{\bullet 1}}$$