Proportional Hazards Regression: Special Topics

Dave Harrington

May 14 - 18, 2018

Introduction

Diagnostics for the Cox model

Correlated event times

Marginal models for correlated event times

Conditional (Frailty) models for correlated event times

Models for repeated events

Introduction

A DISCLAIMER

The topics in this unit are treated only briefly, and are intended to serve as an introduction in a short course.

Each topic could be a substantial part of an advanced course on regression and model fitting with the PH model.

See texts such as *Therneau and Grambsch* or *Klein and Moeschberger*.

Diagnostics for the Cox model

Types of model diagnostics for Cox model

Simple plots for models with two groups

Graphical diagnostics based on residuals for larger models

Formal goodness-of-ft tests

SIMPLE PLOTS FOR MODELS WITH TWO GROUPS

Suppose a covariate Z is coded 0, 1. In the *PH* model,

$$\Lambda(t; Z = 1) = e^{\beta}\Lambda(t; Z = 0), \text{ so}$$

$$\log\left[\frac{\Lambda(t; Z = 1)}{\Lambda(t; Z = 0)}\right] = \log(\Lambda(t; Z = 1)) - \log(\Lambda(t; Z = 0)) = \beta$$

If the model is correct, then

- a simple plot of the difference of log cumulative hazard functions should be approximately constant

- plots of the two log cumulative hazard functions should have the same separation

Since
$$S(t) = \exp(-\Lambda(t)),$$
$$\log(\Lambda(t)) = \log(-\log(S(t)))$$

The plot of $\log(-\log(S(t)))$ vs t is called the complementary log-log survival plot.[1]

---

[1] In R, the "cloglog" option in survfit plots $\log(-\log(S(t)))$ vs $\log(t)$.

EXAMPLE: PBT01

In a stratified model, the simple diagnostic plot should be done within strata.

```
library(survival)
library(eventtimedata)
data("pbt01")
cycle.1 = subset(pbt01, cycle.of.resp == "cycle.1")
cycle.1.km = survfit(Surv(survival,died) ~ treatment, data=cycle.1)

plot(cycle.1.km, fun = "cloglog", conf.int=FALSE, lty = 2:3,
     axes=FALSE, col = 3:4, xlab="Months", ylab="Log(-Log(Survival))",
     main="Cycle.1 Strata",
     cex = 0.6,
     cex.main = 0.8)
axis(1)
axis(2)
legend("bottomright", lty=2:3, col = 3:4, c("AMBT", "Control"))
```

Cycle.1 Strata
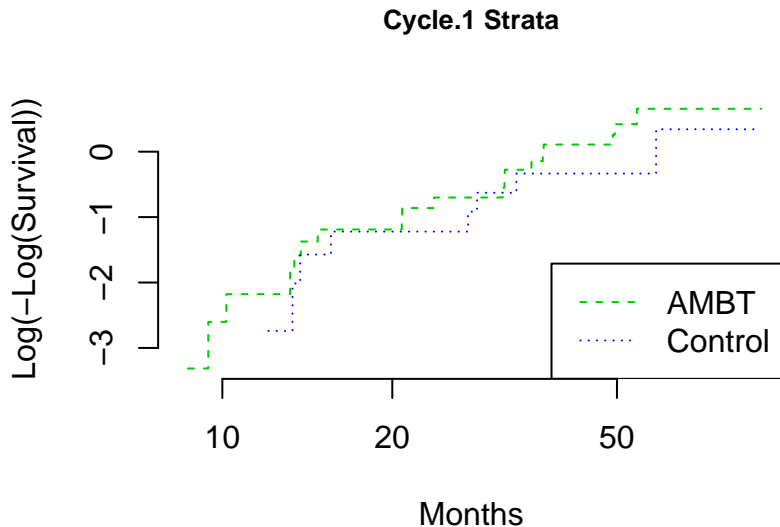
Cycle.2 Strata
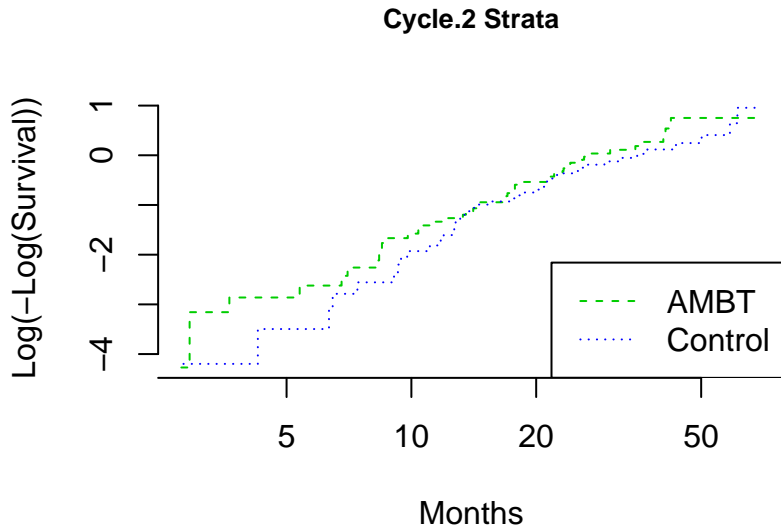
More diagnostics

Several types of residuals are used with the Cox PH model.

- Martingale residuals

- Generalized (Cox-Snell) residuals

- Schoenfeld and weighted Schoenfeld residuals

- Scaled "dfbetas" residuals (approximate influence on parameter estimates from each case)

All of these are easily available in the `survival` package using the residuals() after a PH fit, illustrated in the examples in this unit.

Martingale residuals: The counting process approach

The martingale residuals arose from the counting process approach to survival data.

Suppose

- X_i is the observed follow-up time for a case.

- T_i is the underlying failure time.

- U_i is the censoring time.

- $\delta_i = I\{T_i \leq U_i\}$ is the indicator for failure.

- $N_i(t) = I\{T_i \leq t, \delta_i = 1\}$

- $Y_i(t) = I\{X_i \geq t\}$

Martingale residuals: The counting process approach...

It turns out that

$$M_i(t) = N_i(t) - \int_0^t Y_i(t) \exp[\beta \mathbf{Z}_i(t)] \lambda_0(u) du$$

is a zero mean stochastic process with independent increments (i.e., a martingale).

The martingale residuals m_i are

$$m_i = \delta_i - \widehat{\Lambda}_i(X_i),$$

where for case i,

- δ_i is the observed indicator of failure.

- X_i is the observed follow-up time.

- $\widehat{\Lambda}_i$ is the estimated covariate specific cumulative hazard.

Martingale residuals are a form of *observed - expected* and have properties similar to residuals from linear models.

Properties of martingale residuals

- $E(m_i) = 0$

- $\sum_{i=1}^{n} m_i = 0$

- $\text{var}(m_i) \approx E_i = \Lambda_i(X_i)$

- $\text{cov}(m_i, m_j) \approx 0$. Since the residuals must sum to zero there is a small negative correlation.

A bit more surprising:

Assume one dimensional covariate, and that the true model is

$$\lambda_i(t) = \lambda_0(t) e^{f(Z_i)}.$$

Then it can be shown that

$$E(m_i) \approx cf(Z_i)$$

where the constant c depends on the amount of censoring.
(Therneau, Grambsch, and Fleming, 1990)

PROPERTIES OF MARTINGALE RESIDUALS...

Implications of the last slide:

- A scatterplot of Z versus m should suggest the functional form of the covariate f, since c will only affect the scale (labeling) of the vertical axis of the plot.

These are not as easy to interpret as residuals from linear models since the m_i are skewed and do not have constant variance.

- Martingale residuals are calculated using

  - `residuals(phmodel)`, `type = "martingale"`, where `phmodel` is defined as the `coxph()` model fit.

- It is customary to add a LOESS or LOWESS curve to the plot.

Null model vs. partial residuals

Martingale residuals can be calculated

- in a null model with no covariates,

- This is the analogue of a scatterplot in regression before model estimation.

- or in a model with covariates to examine possible transformations after an initial model.

As in linear regression, there are advantages and disadvantages to either approach.

Null models are fit using `coxph(Surv() ~ 1)`.

# Null model residuals against age in the Rossi data

```
library(survival)
library(eventtimedata)
data(rossi)
rossi.phmodel = coxph(Surv(week, arrest) ~ 1, data = rossi)
mart.resid = residuals(rossi.phmodel, type = "martingale")
plot(rossi$age, mart.resid)
lines(lowess(rossi$age, mart.resid), col = "red")
```

Partial martingale residuals against age in the Rossi data

```r
library(survival)
library(eventtimedata)
data(rossi)
rossi.phmodel = coxph(Surv(week, arrest) ~ fin + age + race + mar,
                      data = rossi)
mart.resid = residuals(rossi.phmodel, type="martingale")
plot(rossi$age, mart.resid)
lines(lowess(rossi$age, mart.resid), col = "red")
```
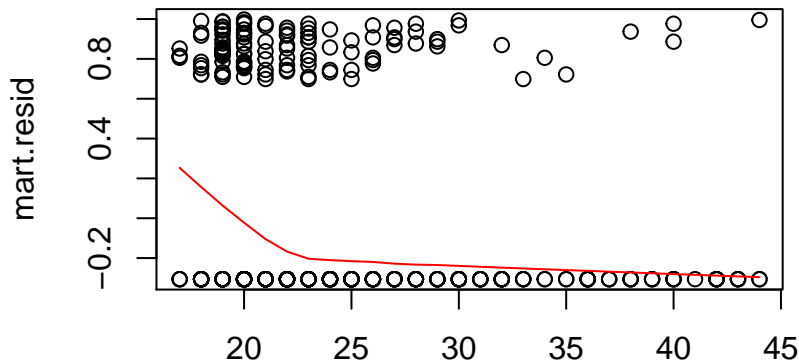
Partial martingale residuals against age in the Rossi data
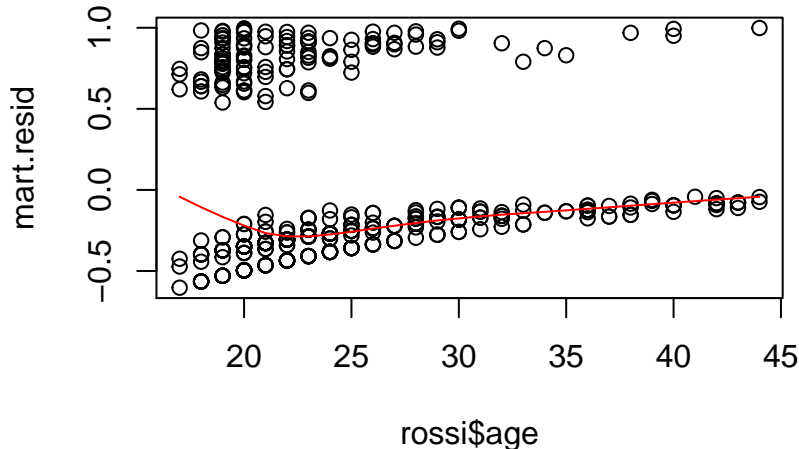
Except for younger ages, there is no evident transformation.

In the null model, what causes the line of residuals just below $y = -0.2$? (Lab question)

Cox-Snell residuals: background

If T_i (the survival time for the i-th individual) has survivorship function $S_i(t)$, then

- the transformed random variable $S_i(T_i)$ will have a uniform distribution on $[0, 1]$,

- and $-\log[S_i(T_i)]$ should be from a unit exponential distribution.

More mathematically stated:

$$\text{If} \quad T_i \sim S_i(t)$$
$$\text{then} \quad S_i(T_i) \sim \text{Uniform}[0, 1]$$
$$\text{and} \; -\log S_i(T_i) \sim \text{Exponential}(1)$$

Proof in the labs.

Generalized (Cox-Snell) residuals

The implication of the last result:

- If the model is correct, the estimated cumulative hazard for each individual at the time of death or censoring should behave like a censored sample from a unit exponential.

- This quantity, $\widehat{\Lambda}_i$, is called the *generalized* or *Cox-Snell* residual.

Calculating the Cox-Snell residuals in R

Not difficult to program R to program and plot the Cox-Snell residuals, but not necessary.

Recall the *martingale* residuals

$$m_i = \delta_i - \widehat{\Lambda}_i(T_i),$$

where for case i, when there are no time-dependent covariates, so

$$cs_i = \delta_i - m_i = \widehat{\Lambda}_i(T_i),$$

where cs_i is the Cox-Snell residual for case i.

Cox-Snell residuals in the Rossi data

```r
library(survival)
library(eventtimedata)
data(rossi)
rossi.phmodel = coxph(Surv(week, arrest) ~ fin + age + race + mar,
                      data = rossi)
mart.resid = residuals(rossi.phmodel, type="martingale")
cs.resid = rossi$arrest - mart.resid
cs.surv = survfit(Surv(cs.resid, rossi$arrest) ~ 1,
                               type = "fleming-harrington")
plot(cs.surv$time, -log(cs.surv$surv))
abline(a = 0, b = 1, col = "red")
```

Cox-Snell residuals in the Rossi data
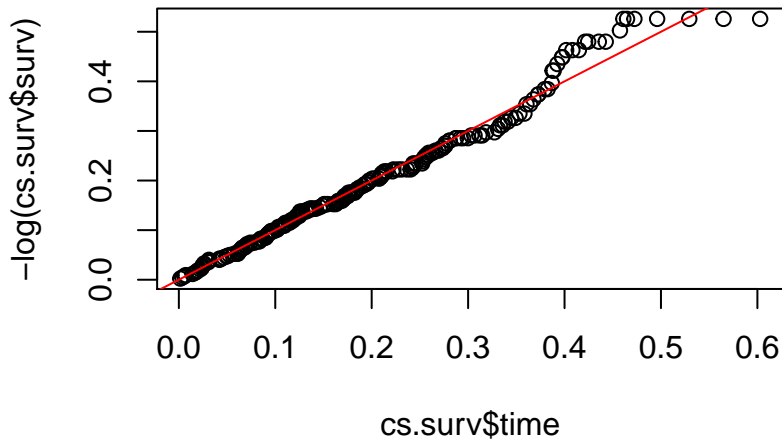
Cox-Snell residuals . . .

The Cox-Snell residuals are theoretically appealing because the notion shows up more generally in non-linear models.

However, they are not widely used because they do not help diagnose reasons for an ill-fitting model.

The martingale and Schoenfeld residuals are more practially useful.

Schoenfeld residuals: testing the PH assumption

The R function `cox.zph` provides a *p*-value for a test of the hypothesis that PH is the right model, as well as plots.

- The plots show *scaled Schoenfeld residuals* for each covariate, plotted against time.

- Under *PH*, the smoothing line should be approximately horizontal.

- These have become more popular than either the Cox-Snell or martingale residuals.

Derivation not shown here, application on the next slide.

Schoenfield residuals in the Rossi data

```
library(survival)
library(eventtimedata)
data(rossi)
rossi.phmodel = coxph(Surv(week, arrest) ~ fin + age + race + mar,
                      data = rossi)
rossi.zph = cox.zph(rossi.phmodel)
rossi.zph
```

```
##                       rho   chisq       p
## finyes             0.0088 0.00899 0.92446
## age               -0.2131 6.97001 0.00829
## raceother          0.1230 1.71919 0.18980
## marnot married    -0.1301 2.05051 0.15215
## GLOBAL                 NA 9.83392 0.04332
```
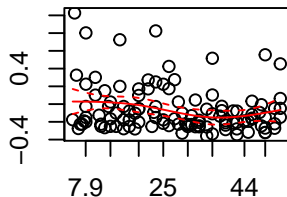
Schoenfield residuals in the Rossi data
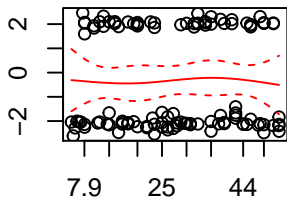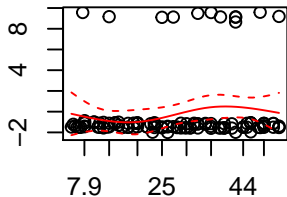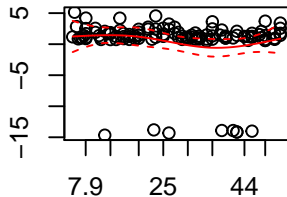
DFBETA RESIDUALS

In an estimated linear model $y = \mathbf{Xb} + e$, it is easy to calculate the approximate change in the estimated coefficients $\mathbf{b} - \mathbf{b}_{-i}$ if case i is dropped, using the *hat matrix*.

- These changes are often referred to as *dfbeta* residuals and help identify cases with high *leverage*.

- A similar (approximate) calculation can be done for the PH model using changes in the score function if a case is dropped (details not shown here).

The residuals() function used on a PH model fit can be used to calculate two versions of dfbeta residuals:

- dfbeta: changes in the estimated coefficients
    - easier to see true influence

- dfbetas: changes in estimated coefficients divided by their standard errors
    - easier to detect 'large changes', since these can be thought of as approximately $N(0, 1)$

Unstandardized dfbeta residuals for age in the rossi data

```r
library(survival)
library(eventtimedata)
data(rossi)

rossi.phmodel = coxph(Surv(week, arrest) ~ fin + age + race + mar,
                      data = rossi)
unstand.dfbeta = residuals(rossi.phmodel, type = "dfbeta")
rossi$caseid = 1:length(rossi$age)
colnames(unstand.dfbeta) = names(rossi.phmodel$coef)

plot(rossi$caseid, unstand.dfbeta[, "age"], xlab = "case id",
     ylab = "age dfbeta", cex = 0.5)
abline(h = 0, lty = 2, col "red")
```

Unstandardized dfbeta residuals for age in the Rossi data
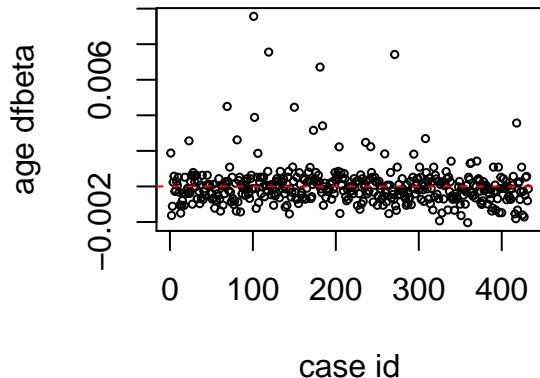
case id

Standardized dfbeta residuals for age in the rossi data

```r
library(survival)
library(eventtimedata)
data(rossi)
rossi.phmodel = coxph(Surv(week, arrest) ~ fin + age + race + mar,
                      data = rossi)
stand.dfbeta = residuals(rossi.phmodel, type = "dfbetas")
rossi$caseid = 1:length(rossi$age)
colnames(stand.dfbeta) = names(rossi.phmodel$coef)

plot(rossi$caseid, stand.dfbeta[,"age"], xlab = "case id",
     ylab = "age dfbeta standard", cex = 0.5)
abline(h = 0, lty = 2, col = "red")
```

STANDARDIZED DFBETA RESIDUALS FOR AGE IN THE ROSSI DATA
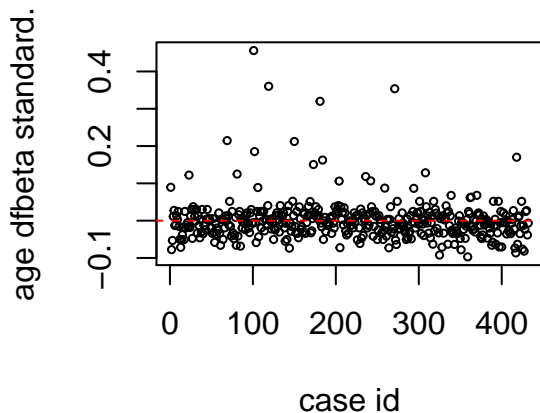
Correlated event times

Causes of correlated failure time data

- Natural clustering of subjects in families, villages, treatment centers

- Shared unmeasured covariates

- Multiple event times from individual subjects

We begin with a dataset...

Diabetic Retinopathy Study (DRS)

Diabetic retinopathy

- Complication associated with diabetes mellitus

- Abnormalities in the microvasculature within eye retina.

- Leading cause of new cases of blindness in patients under 60 years of age in US

Diabetic Retinopathy Study (DRS)...

DRS begun in 1971 to study effectiveness of laser therapy in delaying onset of blindness.

- Eligible patients had DR and visual acuity of 20/100 or better

- One eye of each patient randomly selected for therapy; other eye observed without therapy

- Main endpoint: visual acuity less than 5/200 at two consecutive 4-month follow-ups

- 1,742 patients registered for the study.[2]

- Subset used here consists of 197 patients that make up a 50% subsample of high-risk patients as defined in the study.[3]

---

[2]Study originally reported in Murphy and Patz (1978).

[3]These 197 patients analyzed in Huster, 1989. Data in `drs` dataset in `eventtimedata` package.

THE FIRST 8 RECORDS

```
library(survival)
library(eventtimedata)
data(drs)
drs[1:8,]

##   id obstime fail          age      tx
## 1  5   46.23    0 older than 20   laser
## 2  5   46.23    0 older than 20 control
## 3 14   42.50    0 20 or younger   laser
## 4 14   31.30    1 20 or younger control
## 5 16   42.27    0 20 or younger   laser
## 6 16   42.27    0 20 or younger control
## 7 25   20.60    0 20 or younger   laser
## 8 25   20.60    0 20 or younger control
```

Some features of the data

Some important features of this dataset:

- Strong intra-pair (within-person) correlation

- No natural ordering of event times within the pairs

Two ways to think about capturing the effect:

- *Marginal Model*: estimating the 'average', or population effect of treatment, at least for the population represented by these cases

- *Conditional Model*: estimating subject-specific effect (i.e., the within-subject difference in outcomes for the two eyes), conditional on some possible unobserved subject effect

Marginal models for correlated event times

Background on marginal models

Approach is a natural extension of generalized estimating equations used with correlated longitudinal data.

- Most popular method due to Wei, Lin, and Weissfeld, 1989 (*WLW*).

- *DRS* is a good concrete example to keep in mind.

Main ideas

- Assume K different possibly correlated failure times, $(T_1, \ldots, T_K)$.

- In general, there would be a joint survivor function

$$S(t_1, \ldots, t_K \mid \mathbf{Z}) = \Pr(T_1 > t_1, \ldots, T_K > t_K \mid \mathbf{Z})$$

  depending on a covariate vector $\mathbf{Z}$.

- General approach of modeling and estimating S difficult. The most progress has been made with parametric models.

Marginal approach

- Specify a model for marginal survivor function of T_k,

$$S_k(t_k \mid \mathbf{Z'_k}),$$

where $\mathbf{Z'_k} = (\mathbf{Z_{1k}}, \ldots, \mathbf{Z_{pk}})$ is a p-dimensional covariate vector associated with failure type k.

- Estimate the S_k's (or at least the effect of $\mathbf{Z_k}$ on S_k) as if $T_1, \ldots, T_K$ are independent.

- Establish (through some general theory) that $\widehat{S}_k$ (or estimated regression effect) is unbiased.

- Compute a 'corrected' variance estimate for any estimated parameters that will account for the dependence among failure types.

Marginal model with the Cox model

Marginal proportional hazards model of *WLW*:

- Model hazard of T_k as

$$\lambda_0(t)\exp(\beta'\mathbf{Z_k}) \text{ or } \lambda_{0k}(\mathbf{t})\exp(\beta'\mathbf{Z_k})$$

- Construct partial likelihood function which assumes that failure types define independent subgroups or *strata*.

- Compute a robust estimate of the $\text{cov}(\widehat{\beta})$, using a usual sandwich type estimator.

Details coming later. . .

Notation in the DRS data

- T_1 denotes failure time for treated eye, T_2 for eye given control treatment.

- $\lambda_{01}(t) = \lambda_{02}(t) = \lambda_0(t)$

- $\mathbf{Z}'_\mathbf{k} = $ (treatment, age at onset) $= (Z_{1k}, Z_{2k})$

- It is natural to assume no biological difference between eyes, but if that were not the case, one could stratify on eye type or add a third covariate that denoted eye type.

- Treatment by eye type interactions would allow for different treatment effects by eye type.

STRUCTURE OF THE DRS DATASET

Datasets fit with WLW usually take one of two forms

- Paired data, with no obvious way to distinguish between members of the pair (at least from the data)

- Several identified failure types within an individual that are used to create strata, with a strata denoting a particular failure type.

The `drs` dataset is of the first type.

- Fit an unstratified model, assuming same baseline failure for each eye.

- Calculate a robust variance to account for the within-pair correlation, using the `cluster` term in the model.

```
library(survival)
library(eventtimedata)
data(drs)

coxph(formula = Surv(obstime, fail) ~ tx + age + cluster(id),
      data = drs)
```

```
## Call:
## coxph(formula = Surv(obstime, fail) ~ tx + age + cluster(id),
##     data = drs)
##
##                      coef exp(coef) se(coef) robust se     z      p
## txlaser           -0.7789    0.4589   0.1689    0.1485 -5.25 1.6e-07
## ageolder than 20   0.0539    1.0554   0.1621    0.1786  0.30   0.76
##
## Likelihood ratio test=22.5  on 2 df, p=1.31e-05
## n= 394, number of events= 155
```

THE ROLE OF AGE IN THE DRS

Confounding should be minimized in a randomized trial.

- This can be assessed with a marginal model that includes only treatment.

```
## Call:
## coxph(formula = Surv(obstime, fail) ~ tx + cluster(id), data = drs)
##
##           coef exp(coef) se(coef) robust se    z       p
## txlaser -0.777     0.460    0.169     0.147 -5.27 1.4e-07
##
## Likelihood ratio test=22.4  on 1 df, p=2.25e-06
## n= 394, number of events= 155
```

What about an age by treatment interaction?

Randomized trials do not preclude interaction effects.

- One might suspect that the treatment acts differently on younger versus older participants.

```
## Call:
## coxph(formula = Surv(obstime, fail) ~ tx * age + cluster(id),
##     data = drs)
##
##                             coef exp(coef) se(coef) robust se     z      p
## txlaser                   -0.425     0.654    0.218     0.185 -2.30 0.0216
## ageolder than 20           0.341     1.407    0.199     0.196  1.74 0.0813
## txlaser:ageolder than 20  -0.846     0.429    0.351     0.304 -2.79 0.0053
##
## Likelihood ratio test=28.5  on 3 df, p=2.87e-06
## n= 394, number of events= 155
```

SOME CONSEQUENCES OF THIS MODELING STRATEGY

Need not specify correlation structure among failure types.

Does not exploit correlation structure that may be evident in some data.

Requires that each observational unit (usually subject) have same number of failure types.

Some failure types may produce dependent censoring for other types.

- Why might that happen?

THE ROBUST VARIANCE ESTIMATOR

The robust variance is based on a "sandwich" estimate

$$V = \mathcal{I}^{-1} B \mathcal{I}^{-1},$$

where

- $\mathcal{I}^{-1}$ is the usual variance estimate of a Cox model, the inverse of the information matrix $\mathcal{I}$.
- B is a correction factor.

This form of a variance estimate is widely used with generalized estimating equations and has been variously named

- Huber sandwich estimator
- White's estimate
- Horvitz-Thompson estimator
- Infinitesimal or approximate jacknife
- WLE estimate

Robust variance estimate in the Cox model

For a Cox model, the *WLW* estimator becomes, after some calculations, $D'D$, where D is the matrix of *dfbeta* residuals from `coxph`.

The robust variance estimate is (essentially) a jackknife estimate of the covariance of the coefficient estimates.

The jackknife statistic was one of the early versions of 'bootstrap' estimates of variance calculated through resampling.

The robust variance estimate is approximate, since the dfbeta residuals are approximate changes in coefficients when cases are dropped.

Conditional (Frailty) models for correlated event times

FRAILTY MODELS

Suppose data is made up of J clusters, and that the jth cluster has n_j members.

Assume that conditional on an unobservable variable ψ_j in cluster j, the failure times within a cluster are independent with hazard

$$\lambda_j(t \mid \mathbf{Z_j}) = \psi_\mathbf{j} \lambda_\mathbf{0}(\mathbf{t}) \exp(\beta' \mathbf{Z_j}).$$

ψ is called the frailty term.

Assume that the cluster-specific frailties $\psi_j (j = 1, \ldots, J)$ are independent and identically distributed, and that failure times in different clusters are independent.

ESTIMATION FOR FRAILTY MODELS

Specify the marginal distribution for ψ.

- Gamma frailty – Clayton (1978)

- Positive stable frailty – Hougaard (1986a, 1986b, 1987)

- Lognormal – McGilchrist and Aisbett (1991)

Write down joint likelihood for (T, ψ) as $[T|\psi][\psi]$, then integrate out ψ. More later on other ways to do this.

Estimation strategy outlined above suggests that one must have a parametric model for T, but coxph can also incorporate frailty terms.

Parametric frailty models can be fit using survreg in the survival package, or the more extensive package parfm.

CONSEQUENCES OF THE FRAILTY MODELS

Intellectually appealing.

Closely related to stratified models.

Regression parameter β must be interpreted only within clusters.

- In fact, PH model must be interpreted within cluster.

- Theory shows that only the positive stable frailty distributions lead to marginal proportional hazards model.

Computational problems are substantial, but progress is being made.

For cluster size > 2, basic model $\longrightarrow$ symmetric correlation structure. Work has been done to relax that.

Adding frailty terms to coxph with drs

```
library(survival)
library(eventtimedata)
data(drs)

coxph(formula = Surv(obstime, fail) ~ tx + age +
        frailty(id, distribution = "gamma"), data = drs)
```

```
## Call:
## coxph(formula = Surv(obstime, fail) ~ tx + age + frailty(id,
##     distribution = "gamma"), data = drs)
##
##                               coef se(coef)     se2   Chisq DF       p
## txlaser                    -0.9111   0.1744  0.1712 27.3081  1 1.7e-07
## ageolder than 20            0.0410   0.2207  0.1657  0.0345  1   0.853
## frailty(id, distribution                           113.8026 84   0.017
##
## Iterations: 6 outer, 30 Newton-Raphson
##     Variance of random effect= 0.851   I-likelihood = -850.8
## Degrees of freedom for terms= 1.0  0.6 84.0
## Likelihood ratio test=201  on 85.6 df, p=2.77e-11  n= 394
```

What about the age by treatment interaction examined earlier?

```
## Call:
## coxph(formula = Surv(obstime, fail) ~ tx * age + frailty(id,
##     distribution = "gamma"), data = drs)
##
##                               coef se(coef)    se2   Chisq   DF      p
## txlaser                     -0.506    0.225  0.221   5.034  1.0 0.0248
## ageolder than 20             0.397    0.259  0.205   2.346  1.0 0.1256
## frailty(id, distribution                        122.555 88.6 0.0098
## txlaser:ageolder than 20    -0.985    0.362  0.355   7.414  1.0 0.0065
##
## Iterations: 6 outer, 31 Newton-Raphson
##      Variance of random effect= 0.926   I-likelihood = -847
## Degrees of freedom for terms= 1.0  0.6 88.6  1.0
## Likelihood ratio test=218  on 91.1 df, p=1.86e-12  n= 394
```

Models for repeated events

INTRODUCTION

Correlated failure time data may arise in repeated event data.

- In drs, the correlated failure times were measured on separate eyes and both measured from time 0.

Perhaps best illustrated with the cgd dataset in the survival package.

CGD is a heterogeneous group of rare inherited disorders characterized by recurrent infections beginning in childhood.

CGD DATASET

Data from 1986 trial conducted by International CGD Study Group.[4]

- Randomized, double-blind trial comparing placebo with recombinant γ interferon (rIFN-g), each administered 3 times daily for a year.

- 128 patients: 65 on placebo, 63 on rIFN-g.

- Primary outcome: time to first serious infection.

- Data collected on all serious infections until the end of follow-up, 400 days or less for nearly all patients.

- Trial reported in *NEJM* (1991) **324: 509-516**.

---

[4]Data appears in Appendix D.2 of Fleming and Harrington.

Data appears in two formats in the package survival.

The dataset cgd0 is the data as collected and as it appears in Fleming and Harrington.

```
## Warning in data("cgd0"): data set 'cgd0' not found

##   id treat age futime etime1 etime2 etime3
## 1 1     1  12    414    219    373     NA
## 2 2     0  15    439      8     26    152
## 3 3     1  19    382     NA     NA     NA
## 4 4     1  12    388     NA     NA     NA
## 5 5     0  17    383    246    253     NA
## 6 6     1  44    364     NA     NA     NA
```

CGD DATASET...

The dataset cgd has been transformed to the multi-line per case format used for correlated failure time data in R.

```
##   id   treat age tstart enum tstop status
## 1  1  rIFN-g  12      0    1   219      1
## 2  1  rIFN-g  12    219    2   373      1
## 3  1  rIFN-g  12    373    3   414      0
## 4  2 placebo  15      0    1     8      1
## 5  2 placebo  15      8    2    26      1
```

Some simple summary statistics

Observed events in placebo group:

- 30 out of 65 individuals with at least one event
- 56 total infections

Observed events in treatment group:

- 14 out of 63 individuals with at least one event
- 20 total infections

Modeling options

There are three common approaches to data like these:

- Analyze time to first event, ignoring the subsequent events

- Multiplicative intensity (MI) model, also called the Andersen-Gill (AG) model

- Marginal models for the infection times

- Conditional (frailty) models

The Multiplicative Intensity (MI) model

The MI model is a natural consequence of proportional hazards regression and Poisson regression.

Two approaches available:

- Intuitive ideas based on Poisson process, analysis of binary data

- More formal approach based on martingale theory

Simple approach first. Suppose an event may be observed repeatedly in same subject with continued follow-up.

- CGD data is a good concrete example to keep in mind.

- Let $N(t)$ be the number of observed events in a given subject by time t.

- In the Cox model for failure time data, $N(t) \leq 1$.

The Multiplicative Intensity (MI) model...

If

- T is time to an event,

- C is potential censoring time, and

- $\mathbf{Z}' = (\mathbf{Z_1}, \ldots, \mathbf{Z_p})$ is p-dimensional covariate vector,

    - For simplicity, assume the covariates are not time-dependent.

then ...

$$\lambda(t \mid \mathbf{Z})\mathbf{dt} = P(t \leq T < t + dt, T \leq C \mid \mathbf{Z}, \mathbf{T} \geq \mathbf{t}, \mathbf{C} \geq \mathbf{t})$$
$$= \lambda_0(t) \exp(\beta' \mathbf{Z}).$$

Counting processes

In the counting process model,

$$dN(t) = N(t) - N(t-)$$

is a binary variable analogous to $I(t \leq T < t + dt, T \leq C)$.

Let

- $\mathcal{F}_t$ represent the 'history' of the observable data over $[0, t]$, and

- $Y(t) = 1$ if subject is at risk for an observable failure at time t and 0 otherwise.

It is reasonable to model

$$E\{dN(t) \mid \mathcal{F}_{t-}\} = Y(t)\lambda_0(t) \exp(\beta' \mathbf{Z}) \mathbf{dt}.$$

Counting processes. . .

Relationship on previous slide specifies 'incremental' mean for N.

For second moment, let $dA(t) = Y(t)\lambda_0(t)\exp(\beta \mathbf{Z})dt$. Then

$$\text{Var}\{dN(t) \mid \mathcal{F}_{t-}\} = dA(t)[1 - dA(t)] \approx dA(t).$$

With this specification, $N(t)$ is Poisson-like with mean and variance function $A(t)$.

We can think of the observable data as Poisson-like count data with

- Variable amounts of follow-up per person.

- An event rate that may depend not only on covariates but time-dependent functions of the history of the observables. Note that dependence is multiplicative with respect to a common baseline rate function λ_0.

- Possible gap times in the time-at-risk.

Martingale specification of the MI/AG model

This approach is due to Andersen & Gill (1982); discussed in detail in ABGK (1993) and Fleming & Harrington (1991).

We only give terminology here:

A process M is called a martingale with respect to a specification of history, $\mathcal{F}_t, t \geq 0$ if

1. $M(t)$ is adapted to $\mathcal{F}_t$ for each t,

2. $\mathsf{E}|M(t)| < \infty$ for all $t < \infty$, and

3. $\mathsf{E}\{M(t+s)|\mathcal{F}_t\} = M(t)$ a.s. for all $s \geq 0, t \geq 0$.

Condition (3) implies that $\mathsf{E}\{M(t) - M(u)|\mathcal{F}_u\} = 0$ for all $u \leq t$,

This is sometimes written informally as $\mathsf{E}\{dM(t)|\mathcal{F}_{t-}\} = 0$.

Martingale approach

A *martingale model* for a counting process $N(t)$ is a specification of a time-dependent rate function $A(t)$, called the compensator for N, and a history of observables $\mathcal{F}_t$ such that

$$M(t) = N(t) - A(t)$$

is an $\mathcal{F}_t$-martingale.

Consequences of modeling strategy

- Standard partial likelihood methods lead to unbiased and asymptotically normal estimators.

- Parameter estimates have simple interpretation.

- Much of technology for standard Cox model (time-dependent covariates, diagnostics, etc.) applies here.

- The centered process $M(t)$ has independent increments, a strong assumption about dependence structure. This implies that the inter-event times within an individual are independent.

- Requires common baseline hazard within strata, and time origin must be chosen carefully.

FITTING THE AG MODEL

Fitting the AG model in R uses the multi-line `t-start`, `t-stop` form of a survival dataset.

Start with a simple model

```r
library(survival)
data(cgd)
coxph(Surv(tstart, tstop, status) ~ treat + age
      + cluster(id), data=cgd)
```

FITTING THE AG MODEL

```
## Call:
## coxph(formula = Surv(tstart, tstop, status) ~ treat + age + cluster(
##     data = cgd)
##
##                coef exp(coef) se(coef) robust se    z      p
## treatrIFN-g -1.1201    0.3263   0.2613    0.3099 -3.61 0.0003
## age         -0.0305    0.9699   0.0131    0.0144 -2.11 0.0345
##
## Likelihood ratio test=25.9  on 2 df, p=2.38e-06
## n= 203, number of events= 76
```

Strengths/weaknesses

Strength of the model:

- Easy to interpret

Weaknesses:

- The estimation procedure assumes that the repeated inter-event times are independent.

- Treatment effects are constant across inter-event times.

Using the WLW approach with CGD data

The WLW approach with these data assumes

- Event types are the event number (i.e., first event, second event, etc.) per person.

- Event types define strata.

- Each failure time for event type is measured from time 0 on the original scale.

    - The waiting time for the second event will be larger than for the first event.

    - The waiting time for the third event will be larger than for the first and second events.

    - etc . . .

Details for a similar problem in the lab.

WLW with CGD data...

```
library(eventtimedata)
data(cgd.wlw)
print(cgd.wlw[1:5, c(1,4,13:15)])
```

```
##   id   treat time status enum
## 1  1  rIFN-g  219      1    1
## 2  2 placebo    8      1    1
## 3  3  rIFN-g  382      0    1
## 4  4  rIFN-g  388      0    1
## 5  5 placebo  246      1    1
```

```
n = length(cgd.wlw$time)
print(cgd.wlw[(n-4):n, c(1,4,13:15)])
```

```
##          id   treat time status enum
## 124.6 131  rIFN-g  243      0    7
## 125.6 132 placebo  203      0    7
## 126.6 133  rIFN-g  197      0    7
## 127.6 134 placebo  227      0    7
## 128.6 135 placebo  227      0    7
```

WLW with CGD data...

```r
library(survival)
library(eventtimedata)
data(cgd.wlw)
coxph(Surv(time, status) ~ treat + age + strata(enum)
      + cluster(id), data = cgd.wlw)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ treat + age + strata(enum) +
##      cluster(id), data = cgd.wlw)
##
##                 coef exp(coef) se(coef) robust se     z       p
## treatrIFN-g  -1.4144    0.2431   0.2719    0.3674 -3.85 0.00012
## age          -0.0379    0.9628   0.0134    0.0176 -2.15 0.03144
##
## Likelihood ratio test=37.9  on 2 df, p=5.92e-09
## n= 896, number of events= 76
```