XVII Summer School of the Master's degree in Statistics and Operations Research

Multi-state models: Rates, risks, and pseudo-values

III: **Direct regression models for marginal parameters**

Per Kragh Andersen and Henrik Ravn

https://multi-state-book.github.io/barcelona2024/

III: Direct regression models for marginal parameters

- The Cox regression model for survival data, re-visited
- Competing risks: the Fine-Gray model
- Recurrent events: the Lawless-Nadeau/LWYY and Ghosh-Lin models
- IPCW: ELOS and direct binomial regression

Recall: marginal parameters in multi-state models

- $Q_h(t) = P(V(t) = h) = E(I(V(t) = h)), h \in \mathcal{S}$ state occupation probabilities
- $\varepsilon_h(t_0) = E(\int_0^{t_0} I(V(t) = h)dt) = \int_0^{t_0} Q_h(t)dt$ *expected length of stay* in state $h \in \mathcal{S}$ in $[0, t_0]$

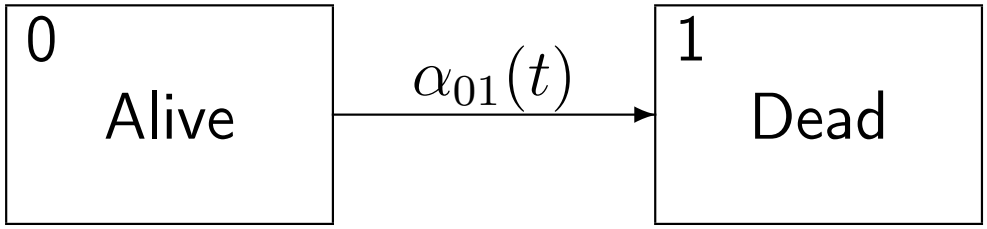
  For recurrent events also:

- $\mu(t) = E(N(t)) = \int_0^t S(u)\alpha^*(u)du$, expected number of events in $[0, t]$ (where $S(t) = P(T > t)$ is the marginal survival function).

Non-parametric inference for these parameters was based on plug-in of non-parametric intensity estimators. The same may, in principle, be done for regression models. However, this activity:

- does not provide regression parameters that directly quantify the association between covariate and marginal parameter
- relies on correct specification of *all* intensity models

The Cox model for survival data, re-visited



Survival function,
$Q_0(t) = S(t) = P(\text{state 0 time } t) = P(T > t) = \exp\big(-\int_0^t \alpha_{01}(u)du\big)$.

The Cox model is a linear model for the log(hazard) (and the log(cumulative hazard)) and, thereby, a linear model for $\log(-\log(Q_0(t)))$. It follows that β-parameters can be interpreted as affecting the state occupation probability $Q_0(t)$ on the log-minus-log survival scale (the scale of the *link function*).

The Cox model for survival data, re-visited

The Cox partial likelihood score equation, Eq. (3.17), is $U(\beta) = 0$ where

$$U(\beta) = \sum_i \int_0^\infty (Z_i - E(\beta, u))\, dN_{01i}(u)$$

and $E(\beta, u) = \frac{\sum_j Z_j Y_{0j}(u) \exp(\beta^\mathsf{T} Z_j)}{\sum_j Y_{0j}(u) \exp(\beta^\mathsf{T} Z_j)}$.

With solution $\widehat{\beta}$, the cumulative baseline intensity is estimated by the Breslow estimator, Eq. (3.18):

$$\widehat{A}_{01}(t) = \int_0^t \frac{\sum_i dN_i(u)}{\sum_i Y_{0i}(u) \exp(\widehat{\beta}^\mathsf{T} Z_i)}.$$

The fact that the hazard model also gives a model for the marginal parameter $Q_0(t) = S(t)$ is due to the *one-to-one correspondence* between intensity and state occupation probability in the simple two-state model.

Inference inspired by partial likelihood

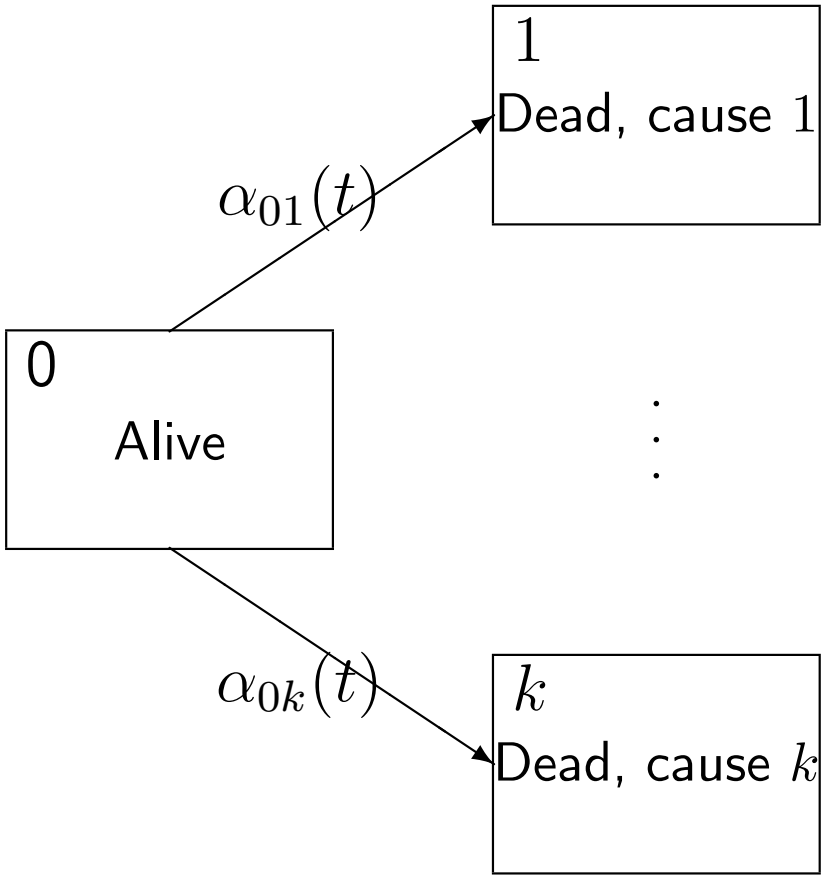
In what follows, we will show that for

- competing risks,
- recurrent events without competing risks,
- recurrent events with competing risks,

regression models for *all time points* may be analyzed using methods inspired by the Cox partial likelihood.

This is because (modified) versions of the Cox score equation still provide unbiased estimating equations in these situations.

Competing risks



The Fine-Gray model

The *Fine-Gray* model is a direct model for the cumulative incidence:

$$Q_h(t \mid Z) = F_h(t \mid Z) = P(T \leq t, D = h \mid Z), h = 1, \ldots, k.$$

Recall from a Cox model for all-cause mortality that:

$$\log(-\log(1 - F(t \mid Z))) = \log(A_0(t)) + \beta^{\mathsf{T}} Z.$$

Fine & Gray (1999, *JASA*) studied the similar model for a cumulative incidence:

$$\log(-\log(1 - F_h(t \mid Z)) = \log(\widetilde{A}_{0h}(t)) + \widetilde{\beta}_j^{\mathsf{T}} Z.$$

The Fine-Gray model

This is a model for the hazard for the improper random variable

$$T_h^* = T \cdot I(D = h) + \infty \cdot I(D \neq h) = \inf_t \{V(t) = h\},$$

i.e., for

$$\widetilde{\alpha}_h(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log(1 - F_h(t)).$$

That is, the transformation which for all-cause mortality takes us from cumulative risk to hazard is used for a cumulative incidence in a competing risks model.

The resulting $\widetilde{\alpha}_h(t)$ is denoted the *sub-distribution hazard* and the Fine-Gray model is thus a proportional sub-distribution hazards model.

The Fine-Gray model

A problem is that, while the hazard function $\alpha(t) = \alpha_{01}(t)$ has the useful 'rate' interpretation:

$$\alpha(t) \approx P(\text{die before } t + dt \mid \text{ alive } t)/dt, \quad dt > 0 \text{ small},$$

and so has the cause-specific hazard $\alpha_h(t) = \alpha_{0h}(t)$:

$$\alpha_h(t) \approx P(\text{die from cause } h \text{ before } t+dt \mid \text{ alive } t)/dt, \quad dt > 0 \text{ small},$$

the sub-distribution hazard has *not*. Thus, for $dt > 0$ (small),

$$\widetilde{\alpha}_h(t) \approx P(\text{die from cause } h \text{ before } t + dt \mid$$
$$either \text{ alive at } t \ \ or \text{ dead from a competing cause by } t)/dt.$$

The Fine-Gray model

The model for the sub-distribution hazard is:

$$\widetilde{\alpha}_h(t \mid Z) = \widetilde{\alpha}_{0h}(t) \exp(\widetilde{\beta}_1 Z_1 + ... + \widetilde{\beta}_p Z_p),$$

but, while a 'sub-distribution hazard' sounds like a hazard, it is not! Therefore, the resulting parameters $\exp(\widetilde{\beta})$ in the Fine-Gray model have a rather indirect interpretation as 'sub-distribution hazard ratios'.

Anyway, the model is being used quite a bit and it is, indeed, useful by giving parameters that directly link the cumulative incidence to covariates (on the log-minus-log cumulative incidence scale).

Math

With no censoring, Fine and Gray defined the cause h 'risk set'

$$\widetilde{R}_h(t) = \{i : (T_i \geq t) \text{ or } (T_i \leq t, D_i \neq h)\}$$

and $\widetilde{\beta}_h$ is estimated by the partial likelihood score equation

$$U_h(\widetilde{\beta}_h) = \sum_i \int_0^\infty \Big(Z_i - \frac{\sum_{j \in \widetilde{R}_h(t)} Z_j \exp(\widetilde{\beta}_h^\mathsf{T} Z_j)}{\sum_{j \in \widetilde{R}_h(t)} \exp(\widetilde{\beta}_h^\mathsf{T} Z_j)}\Big) dN_{hi}(t) = 0$$

corresponding to replacing times of failures from causes other than h by $+\infty$.

With known (e.g., 'administrative') censoring (at C_i), the cause h risk set is replaced by

$$\widetilde{R}_h(t) = \{i : (\min(T_i, C_i) \geq t) \text{ or } (T_i \leq t, D_i \neq h, C_i \geq t)\}.$$

Math (ctd.)

- To identify this 'risk set', we need to know the times C of censoring for subjects who failed
- With general censoring, an Inverse Probability of Censoring Weighted (IPCW) score equation is used and to use this, a model for censoring is needed
- In the simplest case, one uses the 'Kaplan-Meier for censoring', that is, estimating $P(C > t)$. (In this analysis 'failures are censorings')
- If censoring depends on covariates then a model for $P(C > t \mid Z)$ is needed for the weights, e.g., a Cox model
- A Breslow-type estimator for the cumulative baseline sub-distribution hazard and asymptotic theory were also discussed by Fine and Gray (1999)

The Fine-Gray model

The Fine-Gray model provides parameters describing the relationship between the covariates and the cause h risk. For example, for a binary covariate Z_1 with an estimated regression coefficient $\widetilde{\beta}_1 > 0$ it follows that for all values, Z_2^0, for the other covariates in the model we have that

$$\widehat{F}_h(t \mid Z_1 = 1, Z_2^0) > \widehat{F}_h(t \mid Z_1 = 0, Z_2^0).$$

The positive regression coefficient has the *qualitative* meaning that individuals with $Z_1 = 1$ have a uniformly increased cause h cumulative incidence compared to those with $Z_1 = 0$.

However, the resulting estimates $\exp(\widetilde{\beta}_h)$ are sub-distribution hazard ratios, so the *quantitative* meaning of the regression coefficient is not simple.

Example: the PBC-3 trial

PBC-3 trial in liver cirrhosis: Estimated coefficients (and SD) from Fine-Gray models for death without transplantation and transplantation.

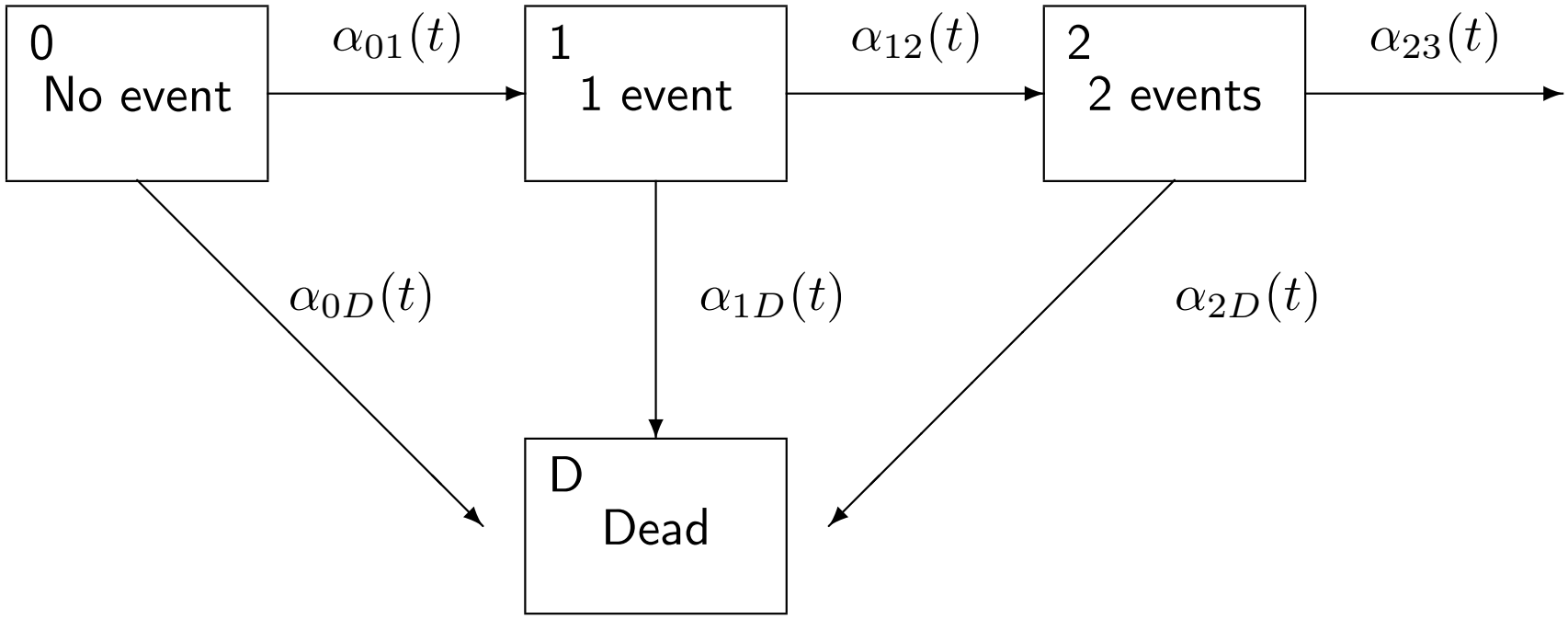
		Death without transplantation		Transplantation	
Covariate		$\widehat{\beta}$	SD	$\widehat{\beta}$	SD
Treatment	CyA vs. placebo	-0.353	0.260	-0.409	0.368
Albumin	per 1 g/L	-0.061	0.031	-0.070	0.033
$\log_2$(Bilirubin)	per doubling	0.616	0.089	0.619	0.101
Sex	male vs. female	-0.415	0.317	-0.092	0.580
Age	per year	0.087	0.016	-0.075	0.017

R code, Fine-Gray model

```
library(survival)
# Transplantation
transdat<-finegray(Surv(days,dc)~ .,data=pbc3,etype=1)
summary(coxph(Surv(fgstart,fgstop,fgstatus)~tment+albumin+...,
        data=transdat,weight=fgwt))

# Death without transplantation
deathdat<-finegray(Surv(days,dc)~ .,data=pbc3,etype=2)
summary(coxph(Surv(fgstart,fgstop,fgstatus)~tment+albumin+...,
        data=deathdat,weight=fgwt))
```

Recurrent events



Recurrent events

The main marginal parameter is $\mu(t) = E(N(t))$, and a multiplicative regression model for this is (Eq. (4.12)):

$$\mu(t \mid Z) = \mu_0(t) \exp(\beta^{\mathsf{T}} Z).$$

Without competing risks, this is the Lawless-Nadeau (1995)/ Lin-Wei-Yang-Ying (2000) model, and β-parameters may be estimated by solving exactly the Cox partial likelihood 'score' equations, which are still unbiased, though no longer likelihood-based.

For this reason, *robust* ('sandwich') variance estimators must be used, rather than those based on the second order derivatives of a log likelihood.

However, situations without competing risks are rare (if an event recurs, it is not all cause mortality and, hence, mortality will 'always' be a competing risk).

Recurrent events with competing risks

In this situation, the model $\mu(t \mid Z) = \mu_0(t)\exp(\beta^{\mathsf{T}} Z)$ is the Ghosh-Lin (2002) regression model.

Estimation follows to a large extent what we saw for the Fine-Gray model in the sense that Ghosh and Lin first studied the case where censoring times C_i are known for everybody - even for those who died and, next, an IPCW method was derived. For the first case the estimating equation is:

$$U(\beta) = \sum_i \int_0^\infty \big(Z_i - \bar{Z}^U(t)\big) I(C_i \geq t) dN_i(t) = 0$$

where $\bar{Z}^U(t)$ is the average

$$\bar{Z}^U(t) = \frac{\sum_j I(C_j \geq t) Z_j \exp(\beta^{\mathsf{T}} Z_j)}{\sum_j I(C_j \geq t) \exp(\beta^{\mathsf{T}} Z_j)}.$$

The Ghosh-Lin model

In the more common case where censoring times are not known for all, a model for $P(C > t \mid Z)$ (possibly without Z) is needed. Let $\widehat{G}(t)$ be the estimator for the censoring distribution, e.g., Kaplan-Meier if C is independent of Z or a Cox-model-based estimator otherwise.

Ghosh and Lin introduced the same weights

$$w_i(t) = I(C_i \geq \min(T_i, t))\widehat{G}(t)/\widehat{G}(\min(\widetilde{T}_i, t))$$

as those used by Fine and Gray and showed that:

$$E(w_i(t)) \approx G(t)$$

which is the expectation of $I(C_i \geq t)$ and, therefore, this indicator in the previous estimating equation can be replaced by $w_i(t)$.

The Ghosh-Lin model

The resulting estimating equation, Eq. (5.30), is

$$U(\beta) = \sum_i \int_0^\infty \big(Z_i - \bar{Z}^G(t)\big) w_i(t) dN_i(t) = 0$$

with

$$\bar{Z}^G(t) = \frac{\sum_j w_j(t) Z_j \exp(\beta^\mathsf{T} Z_j)}{\sum_j w_j(t) \exp(\beta^\mathsf{T} Z_j)}.$$

Finally, (sandwich-type) variances and the baseline mean function can be estimated, the latter by a Breslow-type estimator:

$$\widehat{\mu}_0(t) = \sum_i \int_0^t \frac{w_i(u) dN_i(u)}{\sum_j w_j(u) \exp(\widehat{\beta}^\mathsf{T} Z_j)}.$$

Example: Repeated episodes in affective disorder

Recurrent episodes in affective disorders: Estimated ratios between mean numbers of psychiatric episodes between patients with bipolar vs. unipolar diagnosis (c.i.: confidence interval).

Model	Mortality treated as	$\exp(\widehat{\beta})$	95% c.i.
LWYY model	Censoring	1.52	(1.07, 2.17)
Ghosh-Lin model	Competing risk	1.96	(1.34, 2.87)

Estimated hazard ratio between bipolar and unipolar patients in a Cox model for the marginal mortality rate is 0.410 with 95% confidence limits from 0.204 to 0.825.

R code, recurrent events

```
library(survival)
# (Incorrect) LWYY model
fitLWYY <- coxph(Surv(prev, stop, status==1) ~ factor(bip) +
        cluster(id), data = affective)
        summary(fitLWYY)
# Ghosh-Lin model
library(mets)
fitGL <- recreg(Event(prev, stop, status) ~ factor(bip) +
        cluster(id), data = affective, cause = 1, cens.code = 0,
        death.code = 2)
summary(fitGL)
# Cox model for mortality
fitdeath<-coxph(Surv(prev,stop,status==2) ~ factor(bip),
          data = affective, ties="breslow")
```

Inverse probability of censoring weighting (IPCW)

Consider a target parameter for which regression analysis is intended which is the expectation of a single random variable, say Y, e.g.,

- t_0-RMST: $\varepsilon_0(t_0) = E(\min(T, t_0)) = \int_0^{t_0} S(u)du$,
- expected time lost due to cause h before t_0: $\varepsilon_h(t_0) = E(t_0 - \min(T_h, t_0)) = \int_0^{t_0} F_h(u)du$ (with $T_h = \inf_t\{V(t) = h\}$, time of entry ($\leq \infty$) into state h),
- state h occupation probability at time t_0: $Q_h(t_0) = E(I(V(t_0) = h))$.

In the (unrealistic) case of *no censoring*, a generalized linear model $g(E(Y \mid Z)) = \beta^{\mathsf{T}} Z$ with link function g could be analyzed by solving the unbiased GEE, $U(\beta) = 0$, where

$$U(\beta) = \sum_i A(\beta, Z_i)(Y_i - g^{-1}(\beta^{\mathsf{T}} Z_i)).$$

Inverse probability of censoring weighting (IPCW)

The idea in IPCW GEE is to replace this (impossible) GEE by a weighted equation, Eq. (5.18), based on the completely observed subjects

$$U(\beta) = \sum_i A(\beta, Z_i) w_i (Y_i - g^{-1}(\beta^\mathsf{T} Z_i))$$

where the weight, w_i, is based on an estimate of the probability of having a complete observation of the random variable Y_i at time t_0:

$$G_i(\min(t_0, T_i)) = P(C_i > \min(t_0, T_i) \mid Z_i),$$

i.e., either the time point t_0 or the time, T_i of death ('reaching an absorbing state') should be before the time C_i of censoring for subject i. The weight is then

$$w_i = I(C_i > \min(t_0, T_i))/\widehat{G}_i(\min(t_0, T_i)).$$

Inverse probability of censoring weighting (IPCW)

This approach has been used by

- Scheike and Zhang (2007) and Scheike et al. (2008) for state occupation probabilities (in competing risks model)
- Tian et al. (2014) for the t_0-restricted mean survival time
- Conner and Trinquart (2021) for the cause-specific time lost before t_0 in a competing risks model

Some of these ideas have been implemented in `R` (`mets`, `survRM2` packages and others) but we will neither exemplify this, nor base any exercises on it. Rather, we will later discuss how *pseudo-values* may be used for such purposes.

Exercises

1. Fit Fine-Gray models for the cumulative incidences of stroke and death without stroke adjusting for ESVEA, sex, age, and systolic blood pressure.

   Are the (adjusted) associations between ESVEA and the cumulative incidences statistically significant?

2. Consider the data on recurrent episodes in affective disorder.

   Estimate non-parametrically the mean number of episodes, $\mu(t)$, in $[0, t]$ for unipolar and bipolar patients, taking the mortality into account.

   Estimate, incorrectly, the same mean curves by treating death as censoring and compare with the correct curves.

3. Continuing the previous exercise, fit Ghosh-Lin models for the expected number of episodes, $\mu(t)$, in $[0, t]$ taking the mortality into account and adjusting for initial diagnosis (bipolar vs. unipolar) and calendar year of diagnosis. Fit, incorrectly, LWYY models for the same expectations by treating death as censoring and compare with the correct analysis.