# Survival Analysis / Modèles de Durée

## Chapitre 3 : Beyond the Cox Model

Gilles de Truchis

Master 2 ESA

# Plan du chapitre

1. Competing Risks
2. Parametric Models
3. Lasso Methods for Selection
4. Survival Analysis with R

## Plan

1. Competing Risks
2. Parametric Models
3. Lasso Methods for Selection
4. Survival Analysis with R

# Clustered survival times

- Until now, we have considered data with a single cause of failure
- Also, we have assumed that survival times were independent

⇒ How to deal with events that are dependent across individuals ?

  - Covid-19 propagation is an example of what we call clustered data

  ⇒ contamination are more likely to occur for people in a same unit

  e.g. children in the same school, employees in the same office, etc.

  - In such a case, survival times within a cluster are more similar to each other than to those from other clusters

  ⇒ the independence assumption no longer holds

⇒ How to deal with an event that can occur repeatedly ?

  - The seizure (crise d'épilespsie) is another example of clustered data

  ⇒ the event may repeat indefinitely per person

# Washington Ashkenazi study : dependent data

- This study examined the mutations of a particular gene (the BRCA)

⇒ Is there an effect of mutations on risk of breast cancer ?

- The study was confined to volunteers from the Ashkenazi population
- Each volunteer was controlled for BRCA mutations
- A subset of 1960 families is available (at most two relatives per family)
- For each volunteer, information of two female relatives are collected
  - age of onset of breast cancer (current age for women without cancer)
- The BRCA mutation status of the volunteer is also collected

## Washington Ashkenazi study

- Here is a subsample of 3 families
  - for each volunteers there are 2 rows

e.g. F#1 consists of 2 first degree female relatives (ages 73 and 40)

... neither of them has ever had breast cancer

... nor the volunteer attached to F#1 have a BRCA mutation

Note 1 The survival variable is age of onset

Note 2 The censoring variable is "brcancer" and "mutant" is the covariate

Note 3 As family members share genetic characteristics, they are not independent

Table – Clustered survival data

| | famID | brcancer | age | mutant |
|---|---|---|---|---|
| 1 | 1 | 0 | 73 | 0 |
| 2 | 1 | 0 | 40 | 0 |
| 7 | 9 | 0 | 89 | 0 |
| 8 | 9 | 1 | 60 | 0 |
| 87 | 94 | 1 | 44 | 1 |
| 88 | 94 | 0 | 45 | 1 |

# Marginal Survival Models (MSM)

- This approach ignores clustered data when estimating the model

⇒ Clusters are accounted for when computing standard errors of $\widehat{\beta}$

- MSM relies on standard Cox model estimation
  - Assume there is one covariate with parameter estimate $\widehat{\beta}$ and $\sigma^2_{\widehat{\beta}} = \mathbb{V}(\widehat{\beta})$
  - $\sigma^2_{\widehat{\beta}}$ can be misleading as it assumes that all subjects are independent

  ⇒ It has to be corrected for the clustering impact

- The correction requires to first define the following score residuals

$$s_{ij} = \delta_{ij}\big(x_{ij} - \bar{x}(t_{ij})\big) - \sum_{t_u \leq t_{ij}} \big(x_i - \bar{x}(t_{ij})\big)e^{x_i\beta}\Big(\widehat{H}_0(t_u) - \widehat{H}_0(t_{u-1})\Big)$$

where we can notice that the first part is the Schoenfeld residuals

- The variance correction is then given by

$$C = \sum_{i=1}^{G}\sum_{j=1}^{n_i}\sum_{m=1}^{n_i} s_{ij}s_{im}, \quad G \text{ and } n_i \text{ are defined in the next slide}$$

where the cluster-adjusted standard error for $\widehat{\beta}$ is $\sigma^*_{\widehat{\beta}} = (\mathbb{V}(\widehat{\beta}) \times C)^{1/2}$

## Cluster-adjusted standard errors

- When there are $q$ covariates in the Cox model, $\beta$ is a vector
- We hence have to apply the correction to the whole estimated covariance matrix of $\beta$
- The score residuals are now $1 \times q$ matrices and C is a $q \times q$ matrix as

  $$C = \sum_{i=1}^{G} \sum_{j=1}^{n_i} \sum_{m=1}^{n_i} s'_{ij} s_{im}$$

  where $G$ is the number of clusters (assumed to be known here) and $n_i$ is the number of failure in the $i$th cluster
- Then, the cluster-adjusted covariance matrix is given by

  $$V^* = \mathbb{V}(\widehat{\beta}) C \mathbb{V}(\widehat{\beta})$$

  the traditional sandwich estimator

⇒ Adjusted standard errors are then derived as follows

$$se(\widehat{\beta}) = \text{diag}(V^*)^{1/2}$$

# Frailty survival models : recall

- Another approach is to generalize to clustered data the likelihood

Recall Under the independence assumption, we may write (see Chapter 1)

$$\mathcal{L}(\beta; x_i) = \prod_{i=1}^{n} f(t_i, \beta)^{\delta_i} S(t_i, \beta)^{1-\delta_i} = \prod_{i=1}^{n} h(t_i, \beta)^{\delta_i} S(t_i, \beta)$$

Recall Under proportional hazards assumption (Cox) it becomes

$$\mathcal{L}(\beta; x_i) = \prod_{i=1}^{n} \left( h_0(t_i) e^{x_i \beta} \right)^{\delta_i} e^{-H_0(t_i) \exp(x_i \beta)}$$

where

$$H_0(t_i) = - \int_0^{t_i} h_0(v) dv$$

is the baseline cumulative hazard

# Frailty survival models : principle

- The idea is to assign each individual in a cluster a common factor

⇒ this common factor is known as frailty or random effect and denoted

$$\omega_i$$

for the $i$th cluster

- Then, for the $j$th subject in the $i$th cluster, the hazard function is

$$h_{ij}(t_{ij}) = h_0(t_{ij})\omega_i e^{x_{ij}\beta}$$

- We allow for $\omega_i$ to vary from one cluster to another

⇒ a common model that governs this variability is a gamma distribution

$$g(\omega, \theta) = \frac{\omega^{1/\theta - 1} e^{-\omega/\theta}}{\Gamma(1/\theta)\theta^{1/\theta}}$$

- An alternative is to use a standard normal distribution

$$h_{ij}(t_{ij}) = h_0(t_{ij})\omega_i e^{x_{ij}\beta} = h_0(t_{ij})e^{x_{ij}\beta + u_i\sigma}, \text{ as } \omega_i = e^{u_i\sigma}$$

such that the random and fixed effects are put on the same level

## Frailty survival models : unfeasible estimation

- Assuming that the frailties $\omega_i$ are observed, the joint likelihood is

  $$\mathcal{L}_{ij}(\beta,\theta;\omega_i,t_{ij},\delta_{ij},x_{ij}) = g(\omega_i,\theta)\left(h_0(t_{ij})\omega_i e^{x_{ij}\beta}\right)^{\delta_{ij}} e^{-H_0(t_{ij})\omega_i \exp(x_{ij}\beta)}$$

  and the full likelihood is

  $$\mathcal{L}_{ij}(\beta,\theta) = \sum_{i=1}^{G}\sum_{j=1}^{n_i}\mathcal{L}_{ij}(\beta,\theta;\omega_i,t_{ij},\delta_{ij},x_{ij})$$

- MLE of $\beta$ and $\theta$ is feasible under assumption that

  $$\omega_i,\;\; t_{ij},\;\; \delta_{ij},\;\; x_{ij}$$

  are observed

- Although we can have an idea of the number of clusters, the frailties

  $$\omega_i$$

  are in general not observed directly

# Frailty survival models : EM algorithm

- In the more realistic case where $\omega_i$ are unknown

... one can use the Expectation-Maximization (EM) algorithm

⇒ It alternates between finding expected values for $\omega_i$ based on current estimates of

$$\beta \text{ and } \theta$$

and using these expected values to find updated estimates for

$$\beta \text{ and } \theta$$

until convergence

- If we use a parametric distribution for

$$f(t, \beta)$$

setting up the EM algorithm is fairly direct

- Generalizing this to the semi-parametric Cox model is more complex

## Example : standard Cox model

- Consider the whole Ashkenazi data set

First Fit the standard Cox model to explain the age of onset of breast cancer

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| mutant (BRCA) | 1.1907 | 3.2895 | 0.1984 | 6.002 | 1.95e-09 |

- The likelihoods of the null versus mutant BRCA models are

# Example : standard Cox model

- Consider the whole Ashkenazi data set

First Fit the standard Cox model to explain the age of onset of breast cancer

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| mutant (BRCA) | 1.1907 | 3.2895 | 0.1984 | 6.002 | 1.95e-09 |

- The likelihoods of the null versus mutant BRCA models are

$$-3579.707 \text{ and } -3566.745$$

respectively and leads to the following LR test statistics

$$LR = 2(-3566.745 + 3579.707) = 25.924$$

that we compare to a $\chi_1^2$ and results in $p < 0.0001$

⇒ this confirms the need of including the BRCA status of the volunteer

## Example : MSM

- We now implement the MSM to account for the clustering
- The clusters are defined through the family ID in the database
- We expect here the coefficient to be the same but the adjusted standard error to be different if the cluster are impacting

| | coef | exp(coef) | se(coef) | robust se | $z$ | $p$ |
|---|---|---|---|---|---|---|
| mutant (BRCA) | 1.1907 | 3.2895 | 0.1984 | 0.2023 | 6.002 | 1.95e-09 |

- The robust standard error is only slightly higher than the unadjusted one

⇒ the effect of clustering within first-degree relatives is small

⇒ the estimation of the MSM reveals that having a first-degree relative with a BRCA mutation increases the hazard of developing breast cancer by a factor of 3.30

# Example : frailty

- Finally we implement the frailty model with a gamma distribution
- We expect here the standard error to be different if the clusters matter

⇒ the coefficient is also likely to vary as the likelihood is modified

| | coef | se(coef) | se2 | Chisq | df | $p$ |
|---|---|---|---|---|---|---|
| mutant | 1.272 | 0.2317 | 0.2004 | 30.13 | 1.0 | 4.0e-08 |
| frailty(famID) | | | | 221.50 | 211.6 | 3.1e-01 |

- Softwares often returns 2 different standard errors
  - the first is directly derived from the Hessian and is generally preferable
  - the second is an alternative estimate based on a variation of the sandwich estimator
- The results are close to those obtained with the MSM and Cox models

⇒ having a first-degree relative with a BRCA mutation increases the hazard of developing breast cancer by a factor of $\exp(1.272) = 3.56$

# Example : frailty

- The likelihoods of the fixed (no cluster) vs random effects models are

$$-3566.745 \text{ and } -3564.622$$

respectively and leads to the following LR test statistics

# Example : frailty

- The likelihoods of the fixed (no cluster) vs random effects models are

  $$-3566.745 \text{ and } -3564.622$$

  respectively and leads to the following LR test statistics

  $$LR = 2(-3564.622 + 3566.745) = 4.246$$

  that we compare to a $\chi^2_1$ and results in $p = 0.03934$

- When comparing the null model with the random effects model we have

## Example : frailty

- The likelihoods of the fixed (no cluster) vs random effects models are

$$-3566.745 \text{ and } -3564.622$$

respectively and leads to the following LR test statistics

$$LR = 2(-3564.622 + 3566.745) = 4.246$$

that we compare to a $\chi^2_1$ and results in $p = 0.03934$

- When comparing the null model with the random effects model we have

$$-3579.707 \text{ and } -3564.622$$

respectively which leads to the following LR test statistics

$$LR = 2(-3579.707 + 3566.745) = 30.17$$

and that we compare to a $\chi^2_1$ and results in $p < 0.00001$

## Cause-specific hazards

- Until now we have considered a single, well-defined outcome
- In some study we may face multiple causes of failure

e.g. an employee can quit the job for different reasons : fired, retirement, ...

- A naive solution is to focus on a particular type of failure

... and treat the others as a type of censoring

- This is questionable as censoring relies on an independence assumption

⇒ What we face here are competing risks, and we have to examine them

Note 1 Interpretation of survival analyses in the presence of competing risks will always be subject to at least some ambiguity due to uncertainty about the degree of dependence among the competing outcomes

Note 2 For a particular subject, we observe only one cause of failure

## Kaplan-Meier estimation with competing risks

- Consider first the naive solution : for each type of failure

... while considering others as a type of censoring

- As presumably, the independence assumption is violated, we can question the consequences on Kaplan-Meier estimation

Note Conversely to Cox, KM estimator considers that censoring occurs first

- We illustrate this issue with the prostate cancer data (see Chapter 1)

⇒ focus on patients ages 80+, stage T2, poorly differentiated

Note old patients, with grade 3 advanced cancer

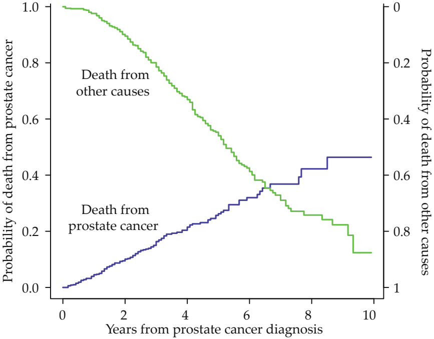
Table – Cancer prostate data for patients ages 80+

| | grade | stage | ageGroup | survTime | $\delta$ (status) | $\Delta$ (other) | $1-\Delta$ (prost) |
|---|---|---|---|---|---|---|---|
| 13 | poor | T2 | 80+ | 21 | 0 | 0 | 0 |
| 38 | poor | T2 | 80+ | 105 | 0 | 0 | 0 |
| 41 | poor | T2 | 80+ | 2 | 1 | 0 | 1 |
| 47 | poor | T2 | 80+ | 67 | 2 | 1 | 0 |
| 78 | poor | T2 | 80+ | 2 | 0 | 0 | 0 |
| 93 | poor | T2 | 80+ | 60 | 2 | 1 | 0 |

# Example : Kaplan-Meier and competing risks

- In Table 2, when $\delta = 2$ we create a new censoring variable $\Delta$

$\Rightarrow$ we apply twice 2 the KME : $\delta = 2$ as censored and $\delta = 1$ as censored



Note 1 At 10 years, e.g., the $\mathbb{P}$(of dying of prostate cancer) is 0.46 versus 0.88

Note 2 If one assume those 2 probabilities to be independent there is no issue

Note 3 If there are not, as they sums to $1.34 > 1$, this reveals a severe bias

Note 4 Unfortunately, this hypothesis cannot be tested from the data

## The cumulative incidence functions

- How to formally address this issue in a non-parametric framework ?
- To develop a formal model to accommodate competing risks,

... assume that there are $K < \infty$ distinct causes of failure

- Also assume that the subject can experience at most one of the $K$ causes

- Then, for each cause of interest, we defined as sub-distribution function

$$F_j(t) = \mathbb{P}(T \leq t, C = j) = \int_0^t h_j(u)S(u)du$$

also known as cumulative risk (or incidence) function for the $j$th cause

- It is increasing as any cumulative distribution function

... but goes, in the limit, to the probability of failure from the $j$th cause rather than to 1

$$F_j(\infty) = \mathbb{P}(C = j)$$

# The cause-specific hazard

- The cause-specific hazard is hence defined conditionally to $C = j$

$$h_j = \lim_{\delta \to 0} \left( \frac{\mathbb{P}(t < T < t + \delta, C = j | T > t)}{\delta} \right)$$

- One can obtain the whole hazards function as follows

$$h(t) = \sum_{j=1}^{K} h_j(t)$$

⇒ The risk of failure at a particular time is simply the sum of the risks of all specific causes at that time

- Now assume that we have $D$ distinct ordered failure times $t_1, t_2, \ldots, t_D$
- We may estimate the hazard at the $i$th time $t_i$ using

$$\widehat{h}(t_i) = d_i / n_i$$

and the cause-specific hazard for the $k$th type cause as

$$\widehat{h}_k(t_i) = d_{ik} / n_i$$

i.e. the # of events of type $k$ at $t_i$ divided by the # of subjects at risk

# Estimating cause-specific hazards

- The sum over all cause-specific hazards is estimated as

$$\widehat{h}(t_i) = n_i^{-1} \sum_{j=1}^{K} d_{ik}$$

- The probability of failure from any cause at $t_i$ is

$$\widehat{S}(t_{i-1}) \times \widehat{h}(t_i)$$

and hence, for a particular cause $k$ we have

$$\widehat{S}(t_{i-1}) \times \widehat{h}_k(t_i)$$

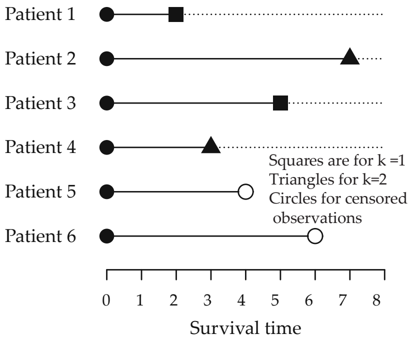
from which we obtain an estimate of the cumulative incidence function

$$\widehat{F}_k(t) = \sum_{t_i \leq t} \widehat{S}(t_{i-1}) \times \widehat{h}_k(t_i)$$

# Example : estimation of the cumulative incidence function

- Consider the following artificial data and compute $\widehat{F}_k(t)$ given that

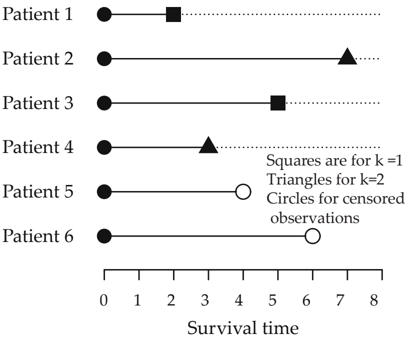
$$\widehat{S}(0,2,3,5,7) = (1, 0.833, 0.667, 0.444, 0.000)'$$



# Example : estimation of the cumulative incidence function

- Consider the following artificial data and compute $\widehat{F}_k(t)$ given that

$$\widehat{S}(0, 2, 3, 5, 7) = (1, 0.833, 0.667, 0.444, 0.000)'$$



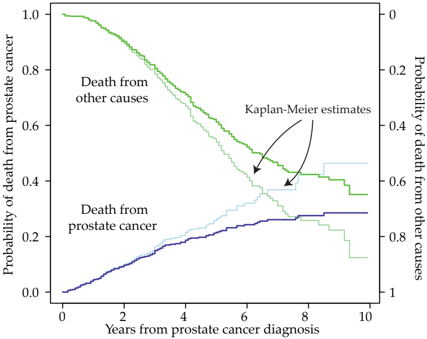
| $t_i$ | $n_i$ | $d_{i1}$ | $d_{i2}$ | $d_i$ | $\widehat{S}(t_{i\cancel{-1}})$ | $\widehat{h}_1(t_i)$ | $\widehat{h}_2(t_i)$ | $\widehat{F}_1(t_i)$ | $\widehat{F}_2(t_i)$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 6 | 0 | 0 | 0 | 1 | / | / | 0.000 | 0.000 |
| 2 | 6 | 1 | 0 | 1 | 0.833 | 1/6 | 0 | 0.167 | 0.000 |
| 3 | 5 | 0 | 1 | 1 | 0.667 | 0 | 1/5 | 0.167 | 0.167 |
| 5 | 3 | 1 | 0 | 1 | 0.444 | 1/3 | 0 | 0.389 | 0.167 |
| 7 | 1 | 0 | 1 | 1 | 0.000 | 0 | 1 | 0.389 | 0.611 |

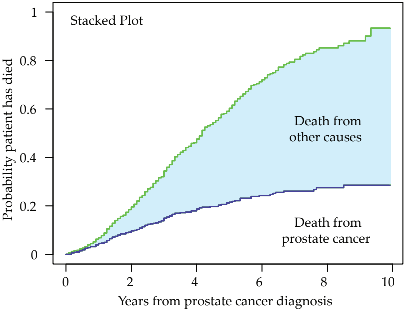
# Example : cumulative incidence function for the prostate cancer data

- An estimate of $\widehat{F}_k(t)$ for the prostate cancer data gives

| $t_i$ | $\widehat{S}(t_{i-1})$ | $\widehat{F}_1(t_i)$ | $\widehat{F}_2(t_i)$ |
|---|---|---|---|
| 1 | 0 | 1.000 | 0.00000 |
| 2 | 1 | 0.994 | 0.00000 |
| 3 | 2 | 0.988 | 0.00602 |
| 4 | 3 | 0.984 | 0.00848 |
| 5 | 4 | 0.983 | 0.00973 |
| 6 | 5 | 0.978 | 0.01477 |

- When comparing with KME, we see that $\widehat{F}_1(t_i)$ and $\widehat{F}_2(t_i)$ never cross





## Regression methods for cause-specific hazards

- Capturing the influence of covariates is challenging in the semi-parametric model of Cox

⇒ How to define the $h_k(t_i)$ on which the covariates should operate ?

- In the spirit of the naive method for the KME, one can consider other causes as censoring and vice versa

- When fitting the Cox model for prostate cancer death we obtain

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| gradepoor | 1.2199 | 3.3867 | 0.1004 | 12.154 | 2e-16 |
| ageGroup70-74 | -0.2860 | 0.7513 | 0.2595 | -1.102 | 0.2704 |
| ageGroup75-79 | 0.4027 | 1.4958 | 0.2257 | 1.784 | 0.0744 |
| ageGroup80+ | 0.9728 | 2.6454 | 0.2148 | 4.529 | 5.92e-06 |

Note 1 Patients having poorly differentiated disease have much worse prognosis than do patients with moderately differentiated disease

Note 2 The hazard of dying from prostate cancer increases with increasing age of diagnosis (the reference is the youngest age group, 65-69)

# Regression methods for cause-specific hazards

- When fitting the Cox model for death from other causes we obtain

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| gradepoor | 0.28104 | 1.32451 | 0.05875 | 4.784 | 1.72e-06 |
| ageGroup70-74 | 0.09462 | 1.09924 | 0.12492 | 0.757 | 0.44879 |
| ageGroup75-79 | 0.31330 | 1.36793 | 0.11709 | 2.676 | 0.00746 |
| ageGroup80+ | 0.79012 | 2.20367 | 0.11204 | 7.052 | 1.76e-12 |

Note 1 Patients with poorly differentiated cancer have a higher risk of death from non-prostate-cancer related disease than do those with moderately differentiated disease

Note 2 The log hazard ratio is much smaller than with prostate cancer death as the outcome (0.28104)

⇒ This suggests that cancer grade wouldn't have any effect on death from non-prostate-cancer causes

Note 3 These results are highly suspect as they rely on the independence assumption

# The Fine-Gray method for cause-specific hazards

- A solution that can overcome this issue is to set

$$h_k(t) = \lim_{\delta \to 0} \left( \frac{\mathbb{P}(t < T_k < t + \delta | E)}{\delta} \right)$$

i.e. to define the effects of covariates on the cause specific hazards where

$$E = \Big( (T_k > t \text{ or } (T_{k'} \leq t \text{ and } k' \neq k) \Big)$$

denotes the conditional event

- The effects of the covariates enter the sub-distribution hazard as follows

⇒ the conditioning set specifies not only $T_k > t$ but also allows other events

... in which case we must have $T_{k'} \leq t$

⇒ the risk set includes not only those currently alive and at risk for the $k$th event type but also those who failed earlier of causes of type $k'$

# The Fine-Gray method and the model of Cox

- The Fine-Gray framework meets the proportional hazard models by setting

$$h_k(t) = -\frac{\delta \log(1 - F_k(t))}{\delta t}$$

- A proportional Cox-type equation is then apply to sub-distribution hazard

$$h_k(t, x, \beta) = h_{0,k}(t)e^{x\beta}$$

$\Rightarrow$ the sub-distribution hazard for a subject with covariate $x$ is proportional to a baseline sub-distribution function $h_{0,k}(t)$

- To apply this approach to the prostate cancer dataset we need to reshape the covariates as

| | (Intercept) | gradepoor | ageGroup70-74 | ageGroup75-79 | ageGroup80+ |
|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 0 |
| 2 | 1 | 1 | 0 | 1 | 0 |
| 3 | 1 | 1 | 0 | 1 | 0 |
| 4 | 1 | 1 | 0 | 0 | 1 |
| 5 | 1 | 0 | 0 | 1 | 0 |
| 6 | 1 | 0 | 0 | 1 | 0 |

# Example : the Fine-Gray method

- We first fit the Fine-Gray model with the prostate cancer as death cause

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| gradepoor | 1.132 | 3.102 | 0.101 | 11.20 | 0.00000 |
| ageGroup70-74 | -0.272 | 0.762 | 0.253 | -1.08 | 0.28000 |
| ageGroup75-79 | 0.367 | 1.443 | 0.219 | 1.67 | 0.09400 |
| ageGroup80+ | 0.799 | 2.224 | 0.208 | 3.85 | 0.00012 |

- Second, we estimate the model for death from other causes

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| gradepoor | 0.126 | 1.13 | 0.0584 | 2.154 | 3.1e-02 |
| ageGroup70-74 | 0.103 | 1.11 | 0.1252 | 0.824 | 4.1e-01 |
| ageGroup75-79 | 0.273 | 1.31 | 0.1176 | 2.323 | 2.0e-02 |
| ageGroup80+ | 0.667 | 1.95 | 0.1128 | 5.917 | 3.3e-09 |

Note 1 Again we see that poorly differentiated patients have higher risk for death from other causes

Note 2 The risk ratio being 0.126 the effect size is smaller than we obtained with the naive method (0.281)

Note 3 The estimated effect on death from prostate cancer of having poorly differentiated disease is similar for both methods

# Comparing the effects of covariates on different causes of death

- One could be interested in comparing the effect of the grade and the age on both causes of death

e.g. the risk of death increases with age but can differ from one cause to another

- To answer this question we have to transform the data

⇒ we create for each patient several rows, one for each cause of death

| | id | from | to | trans | Tstart | Tstop | time | censored | grade | ageGroup |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 2 | 1 | 0 | 27 | 27 | 0 | mode | 70-74 |
| 2 | 1 | 1 | 3 | 2 | 0 | 27 | 27 | 0 | mode | 70-74 |
| 3 | 2 | 1 | 2 | 1 | 0 | 38 | 38 | 0 | poor | 75-79 |
| 4 | 2 | 1 | 3 | 2 | 0 | 38 | 38 | 1 | poor | 75-79 |
| 5 | 3 | 1 | 2 | 1 | 0 | 13 | 13 | 0 | poor | 75-79 |
| 6 | 3 | 1 | 3 | 2 | 0 | 13 | 13 | 0 | poor | 75-79 |

- We can ignore the first columns until the one labeled “trans”

⇒ “trans” indicates the death cause : 1 (prostate cancer), 2 (otherwise)

- The next important column is “time” as it display the survival times

# Example : comparing the effects of covariates on different causes of death

- Here is a summary of the numbers of events of each type for the dataset

| from/to | event-free | prostate | other | no event | total entering |
|---|---|---|---|---|---|
| event-free | 0 | 410 | 1345 | 4165 | 5920 |

- Now we can stratify on cause of death using “trans” and get estimates of

... the effect of “grade” on cause of death under the assumption that they affect

1 both causes equally

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| gradepoor | 0.515 | 1.673 | 0.050 | 10.372 | 2.0e-16 |
| ageGroup70-74 | 0.027 | 1.027 | 0.112 | 0.238 | 0.81210 |
| ageGroup75-79 | 0.332 | 1.394 | 0.104 | 3.198 | 0.00139 |
| ageGroup80+ | 0.833 | 2.301 | 0.099 | 8.396 | 2.0e-16 |

Note This first model is not really useful as we expect that cancer grade affects prostate cancer death differently than it does death from other causes

# Example : comparing the effects of covariates on different causes of death

2 or the “grade” status affects both causes differently

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| gradepoor | 1.239 | 3.451 | 0.100 | 12.391 | 2.0e-16 |
| factor(trans)2 | NA | NA | 0.000 | NA | NA |
| ageGroup70-74 | 0.026 | 1.027 | 0.112 | 0.235 | 0.81431 |
| ageGroup75-79 | 0.333 | 1.395 | 0.104 | 3.201 | 0.00137 |
| ageGroup80+ | 0.833 | 2.301 | 0.099 | 8.394 | 2.0e-16 |
| gradepoor :<br>factor(trans)2 | -0.963 | 0.382 | 0.116 | -8.327 | 2.0e-16 |

- The estimate for “grade” (1.239) is the effect of grade on prostate cancer death, and is similar to what we got earlier (see S25)

- However, the last row is an estimate for the difference between the effect on prostate cancer death and death from other causes

⇒ -0.963, represents the additional effect of poor grade on risk of death from other causes relative to its effect on prostate cancer death

Note 1 Specifically, the hazard of death from other causes is $\exp(-0.963) = 0.381 (< 1)$ times the hazard of death from prostate cancer

# Example : comparing the effects of covariates on different causes of death

- Regarding the age, here are the results we obtain

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| gradepoor | 1.220 | 3.387 | 0.100 | 12.154 | 2.0e-16 |
| ageGroup70-74 | -0.286 | 0.751 | 0.260 | -1.102 | 0.2704 |
| ageGroup75-79 | 0.403 | 1.496 | 0.226 | 1.784 | 0.0744 |
| ageGroup80+ | 0.973 | 2.645 | 0.215 | 4.529 | 5.92e-06 |
| trans2 | NA | NA | 0.000 | NA | NA |
| gradepoor :trans2 | -0.939 | 0.391 | 0.116 | -8.072 | 6.66e-16 |
| ageGroup70-74 :trans2 | 0.380 | 1.463 | 0.288 | 1.322 | 0.1863 |
| ageGroup75-79 :trans2 | -0.089 | 0.914 | 0.254 | -0.351 | 0.7252 |
| ageGroup80+ :trans2 | -0.183 | 0.833 | 0.242 | -0.754 | 0.4508 |

Note None of these differences are statistically significant

⇒ we conclude that there is no difference in the effect of age on the two death causes, after adjusting for grade

# Plan

1 Competing Risks

2 Parametric Models

3 Lasso Methods for Selection

4 Survival Analysis with R

# The parametric approach

- In what follows, we develop further the MLE section of Chapter 1
- Non-parametric (e.g. KME) and semi-parametric (e.g. Cox model) approaches are powerful

but they accommodate complex censoring and truncation less directly

⇒ In the parametric framework, the standard likelihood theory applies

but its validity depends on the appropriateness of the selected model

- Here we essentially review
  - the exponential distribution
  - the Weibull distribution
  - the log-normal distribution
  - the log-logistic distribution

# The exponential distribution

- In Ch. 1, we saw that the simple distribution to work with is the exponential one
- It has constant hazard function $h(t) = \lambda$ ($\Rightarrow$ memory-less property)

⇒ The risk of facing the event of interest is the same at any point in time

i.e. neither declines nor increases in time

Recall The p.d.f and survival functions are

$$f(t;\lambda) = \lambda e^{\lambda t} \text{ and } S(t;\lambda) = e^{-\lambda t}$$

- In general, it is not flexible enough but it can help in some specific applications

⇒ power and size calculations

⇒ The Weibull distribution, of which the exponential distribution is a special case, offers more flexibility

# The Weibull distribution

Recall The hazard and survival functions are

$$h(t) = \alpha\lambda t^{\alpha-1} \text{ and } S(t) = e^{-(\lambda t)^\alpha}$$

- In view of introducing covariates in the parametric model, let define

$$\mu = -log\lambda \text{ and } \sigma = 1/\alpha$$

a location and scale parameter for the distribution

⇒ One can hence rewrite the hazard and survival functions as

$$h(t) = \frac{1}{\sigma} e^{-\mu/\sigma} t^{1/\sigma-1} \text{ and } S(t) = e^{-e^{-\mu/\sigma} t^{1/\sigma}}$$

Note Obviously, when $\sigma = 1$, this reduces to the exponential distribution

# Diagnostic tool for the Weibull distribution

- Consider now the $g(u) = \log(-\log(u))$ transformation function for $S(t)$

$$g(S(t_i)) = \alpha \log(\lambda) + \alpha \log(t_i) = -\frac{\mu}{\sigma} + \frac{1}{\sigma} \log(t_i)$$

- This will allow for assessing how well a set of survival data follow a Weibull distribution

1. First compute the KME $\widehat{S}(t_i)$ and define

$$y_i = g(\widehat{S}(t_i))$$

2. Then, plot $y_i$ versus $\log(t_i)$ and fit the linear equation

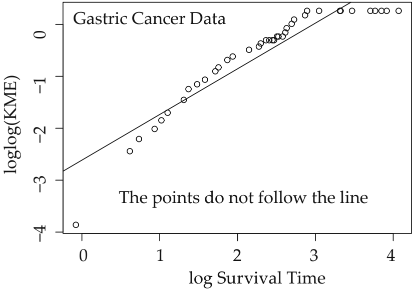
$$y = b + m \log t$$

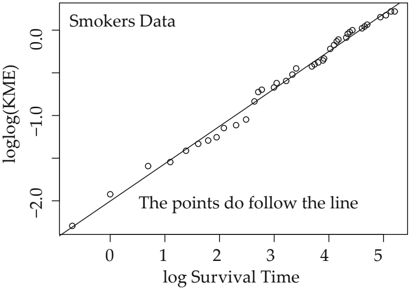
where $m = 1/\sigma$ and $b = -\mu/\sigma$

⇒ If the plotted points fall along this fitted line, a Weibull distribution should approximate well the distribution of the data

## Example : diagnostic tool for the Weibull distribution

- Consider the some databases introduced in Chapter 1





- For the second data set, the Weibull distribution seems plausible

- The fitted straight line parameters are : $b = -2.0032$ and $m = 0.4385$

⇒ Weibull scale and location parameter estimates are :

$$\widehat{\mu} = -b/m = 2.0032/0.4385 = 4.568 \text{ and } \widehat{\sigma} = 1/m = 1/0.4385 = 2.280$$

## MLE of Weibull parameters for a single group of survival data

- The linear approach is limited but provides good entries for the MLE
- The log-likelihood function is (see Ch. 1)

$$\ell(\lambda,\alpha) = \sum_{i=1}^{n} \Big(\delta_i \log h(t_i) + \log S(t_i)\Big)$$

- Substituting the expressions for $h(t_i)$ and $S(t_i)$ we get

$$\ell(\lambda,\alpha) = d \log\alpha + d\alpha\log\lambda + (\alpha-1)\sum_{i=1}^{n}\delta_i\log t_i - \lambda^\alpha\sum_{i=1}^{n} t_i^\alpha$$

  with $d = \sum_{i=1}^{n}\delta_i$

- The expression can of course be expressed in terms of $\mu$ and $\sigma$
- Once implemented and applied to the smokers data, we obtain

$$\widehat{\mu}_{MLE} = 4.656329 \text{ and } \widehat{\sigma}_{MLE} = 2.041061$$

⇒ The results are not so far from the linear approach

Note In general, the standard errors are computed for $\widehat{\mu}_{MLE}$ and $\log\widehat{\sigma}_{MLE}$

$$\widehat{\sigma}_{\mu} = 0.2170 \text{ and } \widehat{\sigma}_{\log\sigma} = 0.0919$$

## Profile Weibull likelihood

- Suppose that a survival random variable $T$, follows a Weibull distribution

$$T \sim \text{Weib}(\alpha)$$

- For a given value of $\alpha$ one can define a new random variable

$$T^* = T^\alpha \sim \exp(\lambda^\alpha)$$

- In such a case (see Ch. 1), the analytic solution of the MLE is known

$$\widehat{\lambda}(\alpha) = (d/V)^{1/\alpha}$$

with $V = \sum t_i^\alpha$ and $d$ the total number of deaths

- Since the MLE $\widehat{\lambda}(\alpha)$ can easily be obtained, we can define as

$$\ell^*(\alpha) = \ell(\widehat{\lambda}(\alpha), \alpha)$$

the Weibull profile likelihood

- Maximizing $\ell^*(\alpha)$ yield the MLE of $\alpha$ and the MLE for $\lambda(\alpha)$ is

$$\widehat{\lambda}(\widehat{\alpha}) = (d/V)^{1/\widehat{\alpha}}$$

# Example : the profile Weibull likelihood

- When applied to the smokers data, we obtained

$$\widehat{\sigma} = 1/\widehat{\alpha} = 2.041063$$

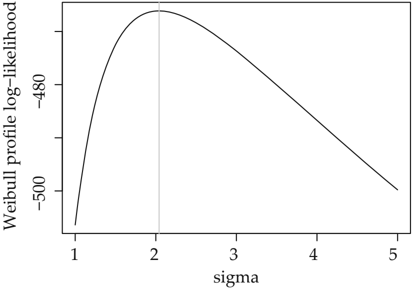
which is almost identical to $\widehat{\sigma}_{MLE}$

- Then, $\widehat{\alpha}$ is used to obtain $\widehat{\lambda}$ and finally

$$\widehat{\mu} = 4.656329$$

which is indistinguishable from $\widehat{\mu}_{MLE}$

⇒ As the MLE only relies on 1 parameter we can plot the profile likelihood



# The Accelerated Failure Time model

- When comparing patients of two groups (e.g. treatment and control)

  $$e^{\beta},$$

  i.e. the hazard ratio, was the quantity we used
- It was assumed to be time-invariant (proportional hazards hypothesis)

⇒ If the treatment group is effective in increasing survival

$$\beta < 0,$$

i.e. the log-hazard ratio, such that the hazard ratio is less than 1

- An alternative way of comparing two groups is called AFT
- We assume here that the survival time of the first group is a multiple

  $$\theta = e^{\gamma}$$

  of what the survival time would have been had if the patient was in the second group

## More intuition on AFT models

- The AFT approach assumes that the effect of a covariate is to accelerate or decelerate the life course of a disease by some constant

e.g. If $\theta = 2$ everything in the life history of patient happens twice as fast

⇒ If the model concerns the development of a tumor, this implies that

1 all of the stages progress twice as fast as for the unexposed individual

2 the expected time until the failure event is 0.5 of the baseline time

⇒ Formally, the survival distributions for the AFT models are given by

$$S_1 = S_0(e^{-\gamma}t)$$

and the hazards are given by

$$h_1(t) = e^{-\gamma}h_0(e^{-\gamma}t)$$

# The AFT model with Weibull distribution

- In the case of the Weibull distribution we have (see S37)

$$h_1(t) = e^{-\gamma} h_0(e^{-\gamma} t) = e^{-\gamma} \frac{1}{\sigma} e^{-\mu/\sigma} (e^{-\gamma} t)^{1/\sigma - 1}$$

- Rearranging, we have

$$h_1(t) = e^{-\gamma/\sigma} \frac{1}{\sigma} e^{-\mu/\sigma} t^{1/\sigma - 1} = e^{-\gamma/\sigma} h_0(t) = e^{\beta} h_0(t)$$

  that is, the AFT model is equivalent to the Cox model with $\beta = -\gamma/\sigma$

Note This equivalence only exists for the Weibull distribution

# Comparison of two groups with the parametric Weibull model

- Consider again the smokers data and the comparison of the triple therapy treatment group to the patch therapy group

- When estimating the AFP model we obtain the following results

| | coeffs | se | $z$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 5.286 | 0.3320 | 15.92 | 4.59e-57 |
| grppatchOnly | -1.251 | 0.4348 | -2.88 | 4.00e-03 |
| Log(scale) | 0.689 | 0.0911 | 7.56 | 3.97e-14 |

$\Rightarrow$ $\widehat{\gamma} = -1.251$ indicates that by a factor of

$$\widehat{\theta} = e^{\widehat{\gamma}} = 0.286$$

the patch therapy group has shorter times to relapse (life course to relapse decelerates for the triple therapy group)

$\Rightarrow$ The scale parameter estimate is $\widehat{\sigma} = \exp(0.689) = 1.992$, leading to

$$\widehat{\beta} = -\widehat{\gamma}/\widehat{\sigma} = 0.629$$

for the log proportional hazard in the Cox model

Note In comparison, a Cox-model-based estimation of $\beta$ gives $\widehat{\beta} = 0.6050$

## Interpreting the intercept in the AFT model

- The Cox-model fit provides only 1 estimate, $\widehat{\beta}$
- The AFT Weibull model provides 3 estimates, 2 of them being linked to the baseline Weibull distribution
- In particular, the intercept $\mu$, cannot be estimated in the Cox approach

⇒ it would cancel out of the partial likelihood (as the baseline hazard does)

- The AFT model allows for direct estimation of the baseline hazard as

$$\widehat{\mu} = 5.286 \text{ and } \widehat{\sigma} = 1.992$$

lead to $\widehat{\alpha} = 1/1.992 = 0.502$ and $\widehat{\lambda} = \exp(-5.286) = 0.00506$ and finally

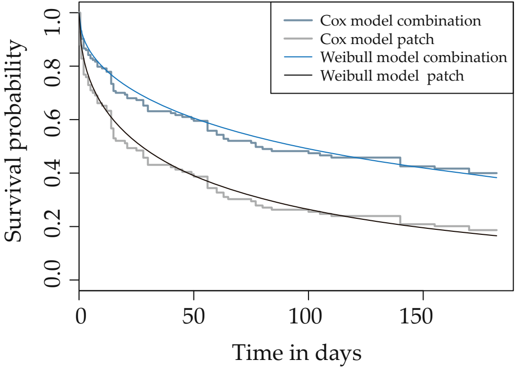
$$\widehat{S}_0(t) = e^{-(\widehat{\lambda}t)^{\widehat{\alpha}}}$$

# Comparison of two groups based on survival functions

- The survival function for the combination group is

$$S_1(t) = \big(S_0(t)\big)^{e^{-\gamma/\sigma}}$$

  and can be estimated by replacing all quantities by their estimates
- We can compare $\widehat{S}_0(t)$ and $\widehat{S}_1(t)$ with those obtained from the Cox model
- Notice that the parametric nature of the AFT produces smooth curves



## AFT-Weibull-based regression

- An alternative way of looking at Weibull AFT model is to define

  $$\log(T) = \mu + \gamma x + \sigma\varepsilon^*$$

  i.e. to model the log-survival time as a location-scale model where

  $$\varepsilon^* = \log \varepsilon$$

  with $\varepsilon$, a unit exponential distribution and $x$ a vector of covariates

- Then, the survival function is given by

  $$\begin{aligned} S(t) = \mathbb{P}(T > t) &= \mathbb{P}(\varepsilon^* > \frac{\log(t) - \mu - \gamma x}{\sigma}) \\ &= S_0(te^{-\gamma x}) \end{aligned}$$

- This formulation is quite general as different choices for

  $$\varepsilon \sim \mathcal{L}(\theta)$$

  can lead to other parametric survival models

# Example : AFT-Weibull-based regression

- Consider again the smokers data but with the covariates (see Ch. 2)

Recall For the Cox model we obtained the results below

Recall In this model, $\widehat{\beta}_{patch} = 0.608$ means that the hazard is higher for this treatment group by a constant factor of $\exp(0.608) = 1.83654$

| | coef | exp(coef) | se(coef) | $z$ | $p$ |
|---|---|---|---|---|---|
| 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 |

- For the AFT Weibull model we obtain

Note $\widehat{\gamma}_{patch} = -1.1902$ means that patients with the patch only have shorter times to relapse by a deceleration factor of $\exp(-1.1902) = 0.304$

| | coeffs | se | $z$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 2.4024 | 0.9653 | 2.490 | 1.28e-02 |
| grppatchOnly | -1.1902 | 0.4133 | -2.880 | 3.98e-03 |
| age | 0.0697 | 0.0203 | 3.430 | 6.02e-04 |
| employmentother | -1.3890 | 0.5029 | -2.760 | 5.74e-03 |
| employmentpt | -1.3143 | 0.6132 | -2.140 | 3.21e-02 |
| Log(scale) | 0.6313 | 0.0900 | 7.020 | 2.26e-12 |

# Exercise : AFT-Weibull-based regression

- Express the results of the AFT Weibull model in terms of proportional hazards coefficients
- Then, compare these coefficients we those obtained from the Cox model

# Exercise : AFT-Weibull-based regression

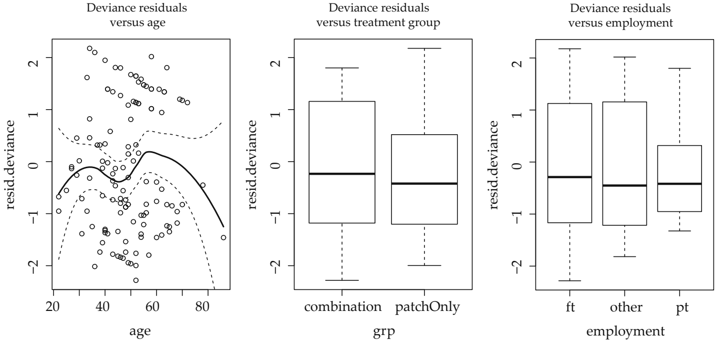
- Express the results of the AFT Weibull model in terms of proportional hazards coefficients
- Then, compare these coefficients we those obtained from the Cox model
- For each regression coefficient $\gamma_j$, we have $\beta_j = -\gamma_j/\sigma$

| | weib.coef.ph | coxph.coef |
|---|---|---|
| grppatchOnly | 0.63301278 | 0.60788405 |
| age | -0.03708786 | -0.03528934 |
| employmentother | 0.73878031 | 0.70347664 |
| employmentpt | 0.69903157 | 0.65369019 |

# Model selection and residual analysis

- Many of the facilities for model selection and residual analysis of Ch. 2 remain valid

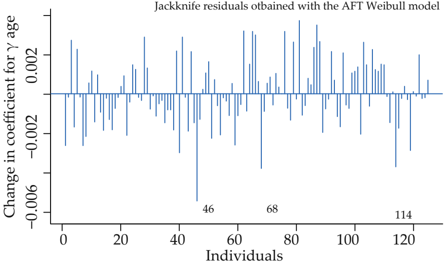
e.g. We plot below the deviance residuals from the previous Weibull model

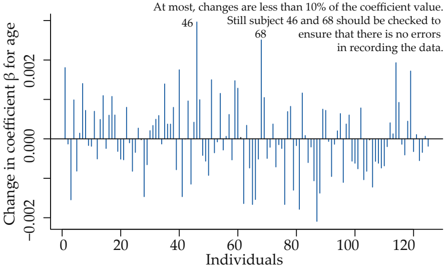


- The residual distributions of both "grp" and "employ" are reasonably comparable, indicating that these variables are modeled successfully
- For "age", the distribution may be consistent with a linear model, when one considers the width of the 95% confidence intervals

# Jackknife residuals

Recall These residuals are computed as the difference in the value of $\widehat{\gamma}$ when all data are used and when an individual is deleted from the data


Jackknife residuals otbained with the AFT Weibull model


At most, changes are less than 10% of the coefficient value. Still subject 46 and 68 should be checked to ensure that there is no errors in recording the data.

# AFT log-normal model

- Various distribution can be considered in the AFT framework
- For instance, when

$$\varepsilon \sim \mathcal{N}(0,1)$$

  $\varepsilon^*$ follows a log-normal distribution and we obtain the following results

| | coeffs | se | $z$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 1.6579 | 1.0084 | 1.64 | 1.00e-01 |
| grppatchOnly | -1.2623 | 0.4523 | -2.79 | 5.25e-03 |
| age | 0.0648 | 0.0203 | 3.20 | 1.39e-03 |
| employmentother | -1.1711 | 0.5316 | -2.20 | 2.76e-02 |
| employmentpt | -0.9543 | 0.7198 | -1.33 | 1.85e-01 |
| Log(scale) | 0.8754 | 0.0796 | 10.99 | 4.15e-28 |
| Scale | 2.4 | | | |

- All estimates are quite different from what we obtain with the Weibull model albeit with similar signs

# AFT log-logistic model

- If $\varepsilon$ has a logistic distribution, with survival distribution given by

$$S(u) = \frac{1}{1+e^u}$$

then, $T$ has a log-logistic distribution and the results are now

| | coeffs | se | $z$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 1.9150 | 0.9708 | 1.97 | 4.85e-02 |
| grppatchOnly | -1.3260 | 0.4588 | -2.89 | 3.85e-03 |
| age | 0.0617 | 0.0196 | 3.15 | 1.66e-03 |
| employmentother | -1.2605 | 0.5392 | -2.34 | 1.94e-02 |
| employmentpt | -1.0991 | 0.7050 | -1.56 | 1.19e-01 |
| Log(scale) | 0.3565 | 0.0884 | 4.03 | 5.47e-05 |
| Scale | 1.43 | | | |

- Again, the estimates are different from what we obtain with the two other models

# Plan

1 Competing Risks

2 Parametric Models

3 Lasso Methods for Selection

4 Survival Analysis with R

## Large set of covariates

- One of the purpose analysis is to understand how covariates contributes to survival times
- Sometimes we focus on specific covariates such as age, employment, etc.
- By contrast, we can focus on the predictive ability of a set of covariates

⇒ In many cases, dozens or thousands or predictors may be available

- In such a study, many of them are unrelated with survival

and those that are relevant may be strongly correlated amongst themselves

⇒ this multicollinearity is likely to complicate estimation and inference

⇒ Penalized methods such as the Lasso method are useful in such situation

# The Lasso method for survival models

- This approach maximizes the partial likelihood function but now with

... the additional stipulation that the $L_1$ norm of $\beta_j$ satisfies

$$\sum_{j=1}^{p} |\beta_j| \leq t$$

for a constant $t$ and with $p$ the number of parameters

⇒ This may be shown to be equivalent to maximizing the penalized likelihood

$$\ell_p(\beta) = \ell(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|$$

for $\lambda$ a pre-specified value of $\lambda$

Note 1 Adding this constraint on coefficients shrinks them toward zero (as compared to non-penalized MLE)

Note 2 A too large $\lambda$ will result in no covariates at all in the model

Note 3 A too small $\lambda$ will result in a large number of covariates in the model

# Optimization issues with the Lasso method

- A complication is that $\ell_p(\beta)$ may not be strictly concave (weakly concave or flat)

⇒ this causes convergence problems

- A crucial issue is hence to select $\lambda$

⇒ As in other econometric fields, cross validation procedures are helpful

1. we randomly divide the data set into 5 subsets of equal size
2. we select 1 subset to be the so-called “validation” set
3. we combine the remaining subsets in the so-called “training” set
4. we use the training set ($\approx 80\%$ of the data) to build the Lasso model
5. we use this model to predict the survival times in the validation set
6. we use a partial-likelihood-based measure of goodness-of-fit to this set
7. we repeat steps 1-6 with each of the remaining 4 subsets in turn playing the role of the validation set
8. we derive an average partial-likelihood goodness-of-fit
9. we repeat the whole process for a wide range of values of $\lambda$
10. we select the value of $\lambda$ that produces the optimum goodness-of-fit : $\lambda^*$

## Example : biomarkers data

- Consider 227 patients with hepatocellular carcinoma (cancer du foie)
- For each patients, a wide range of clinical and biomarker covariates is collected
- The dataset is composed of 48 clinical and biomarker measurements
- Of the 227 patients, 117 have levels of a variety of chemokines markers

⇒ some represent the levels in the tumor itself

Note In medical study, building a predictive model is a complex process that involves interplay between the known medical science and the optimal predictive model

⇒ as we are economists we omit this dimension and consider 26 biomarkers

5 chemokines markers for 3 patients as an example

|  | OS | Death | CD4T | CD4N | CD8T | CD8N | CD20T |
|---|---|---|---|---|---|---|---|
| 1 | 83 | 0 | 2.600000 | 0.000000 | 190.6000 | 126.80 | 20.950000 |
| 76 | 20 | 1 | 14.450000 | 2.758621 | 2.1500 | 38.95 | 26.100000 |
| 131 | 35 | 1 | 2.821133 | 8.294828 | 8.0064 | 62.64 | 2.821133 |

# Example : Lasso method for selecting biomarkers

- First, we select the 117 patients for which all biomarkers are available

- Before implement the Lasso, we standardize the covariates

⇒ as the biomarker ranges vary widely

- Then, we set $\lambda = 10$ and fit the Lasso model using 26 biomarkers

⇒ we see that 7 are retained and here are there coefficient estimates

| CD8N | CD68T | CD4TR | CD8TR | CD68TR | Ki67 | CD34 |
|---|---|---|---|---|---|---|
| 0.104 | 0.258 | -0.035 | -0.096 | 0.111 | 0.285 | -0.013 |

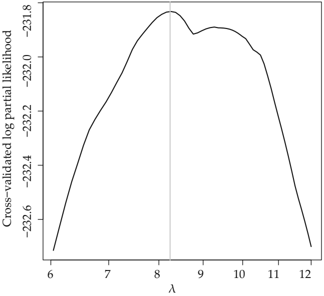
- As $\lambda = 10$ has been specified arbitrarily, we can question the results

⇒ To investigate this we implement the cross validation procedure

# Example : Cross validation procedure for the Lasso method

- The cross-validated partial log-likelihood can be plotted to visualize $\lambda^*$

$\Rightarrow$ we see that the global maximum is obtained for $\lambda^* = 8.24$



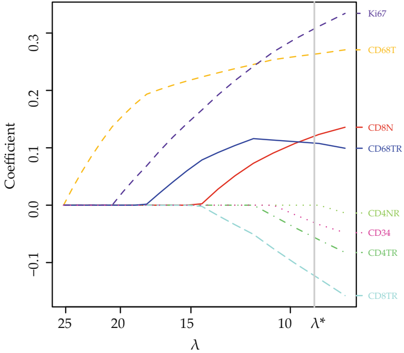
- The results where obtained for $\lambda \in [2, 12]$

| CD8N | CD68T | CD4NR | CD4TR | CD8TR | CD68TR | Ki67 | CD34 |
|---|---|---|---|---|---|---|---|
| 0.133 | 0.269 | -0.009 | -0.076 | -0.149 | 0.102 | 0.328 | -0.044 |

# Example : Cross validation procedure for the Lasso method

- One can also be interested in the impact of $\lambda$ on the estimates

$\Rightarrow$ we can plot the selected markers estimated coefficients for $\lambda \in [20, \lambda^*]$



- The 8 paths that intersect the $\lambda^*$ vertical line are
  - positive coefficients Ki67, CD68T, CD8N, and CD68TR
  - negative coefficients CD4NR, CD34, CD4TR, and CD8TR

# Interpretation of Lasso-based estimates

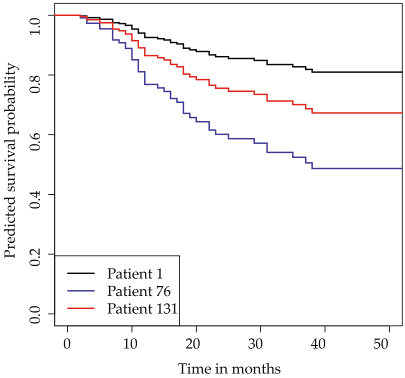
- All $\widehat{\beta}_j$ obtained with $\lambda^*$ are not interpretable in terms of hazard ratio

1 the lasso procedure has shrunken them

2 they are standardized to have standard deviation one

⇒ However, they can be use to predict the survival profile of patients

Note These patients are those who were in the sample Table of S60



# Plan

1. Competing Risks
2. Parametric Models
3. Lasso Methods for Selection
4. Survival Analysis with R

## Survival Analysis with R

- We will review the implementation of most of the examples, in R
- All packages we use are listed below (databases are packed in “asaur”)

Note This package is attached to the book of Dirk F. Moore (Springer, 2016) that I mainly use for this course

- “asaur” package
- “bshazard” package
- “cmprsk” package
- “coxme” package
- “forestplot” package
- “muhaz” package
- “numDeriv” package
- “Hmisc” package
- “kmconfband,” package
- “stats” package
- “penalize” package
- “survival” package

# Loading packages and visualizing data

- One of the first dataset we introduce is the aid to smokers to quit
- To visualize the data we need the “asaur” package
- Then we display for the 6 first subjects some columns (2 to 8)

```r
> library (asaur)
> pharmacoSmoking[1:6, 2:8]
```

- In the same package we also have, e.g., the pancreatic cancer data
- To quickly visualize the first observations of the database we use

```r
> head(pancreatic)
```

# Manipulating and visualizing parametric survival distributions

- In the second section we introduce some parametric distributions
- For example, we can plot the Weibull survival function as follows

```
weibSurv <- function(t, shape, scale) pweibull(t, shape=shape,
scale=scale, lower.tail=F)

curve(weibSurv(x, shape=1.5, scale=1/0.03), from=0, to=80,
ylim=c(0,1), ylab='Survival probability', xlab='Time')
```

- We can also plot the Weibull hazard function as follows

```
weibHaz <- function(x, shape, scale) dweibull(x, shape=shape,
scale=scale)/pweibull(x, shape=shape, scale=scale,
lower.tail=F)

curve(weibHaz(x, shape=1.5, scale=1/0.03), from=0, to=80,
ylab='Hazard', xlab='Time', col=''red'')
```

# Manipulating and visualizing parametric survival distributions

- If needed we can simulate data from Weibull distribution as

```r
tt.weib <- rweibull(1000, shape=1.5, scale=1/0.03)
```

- We can then check whether some empirical quantities converge to their theoretical values

```r
> mean(tt.weib)
[1] 31.35497
> median(tt.weib)
[1] 26.84281

> gamma(1 + 1/1.5)/0.03
[1] 30.09151
> (log(2)^(1/1.5))/0.03
[1] 26.10733
```

## Computation of the Survival function from the Hazard function

- As discussed in Ch. 1, one can use the hazard function to approximate the survival function
- Then, we can compute the empirical mean and median estimates

```
> library (survival)
> tm <- c(0,1/365,7/365,28/365,1:107)
> tm.diff <- diff(tm)
> survMale <- exp(-cumsum(hazMale*tm.diff)*365.24)
> survFemale <- exp(-cumsum(hazFemale*tm.diff)*365.24)
> sum(survMale*tm.diff)
 [1] 71.99964
> sum(survFemale*tm.diff)
 [1] 76.98838
```

- At this stage, to get an estimate of the hazard function we rely on

```
> tm <- c(0, birth
1/365, first day of life
7/365, seventh day of life
28/365,fourth week of life
1:106) subsequent years
> hazMale <- survexp.us[,"male","2004"]
> hazFemale <- survexp.us[,"female","2004"]
```

# The Kaplan-Meier estimator

- The Kaplan-Meier estimator is the most used non-parametric estimator of the survival function
- In the course we first apply it to artificial data

```
> library(survival)
> tt <- c(7,6,6,5,2,4)
> cens <- c(0,1,0,0,1,1)
> Surv(tt, cens)
 [1] 7+ 6 6+ 5+ 2 4
```

- Then, the KME rely on the following function of the survival library

```
result.km <- survfit(Surv(tt, cens) ~ 1, conf.type="log-log")
> summary(result.km)
> plot(result.km)
```

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. Biometrika, 66(1), 59-65.