Variable Selection and Model Choice in Survival Models with Time-Varying Effects

Benjamin Hofner [1]
Institut für Medizininformatik, Biometrie und Epidemiologie (IMBE)
Friedrich-Alexander-Universität Erlangen-Nürnberg

joint work with Torsten Hothorn and Thomas Kneib
Institut für Statistik
Ludwig-Maximilians-Universität München

55. Biometrisches Kolloquium
Hannover - 2009

[1] benjamin.hofner@imbe.med.uni-erlangen.de

Introduction

Cox PH model:

$$\lambda_i(t) = \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta})$$

Problem: restrictive model, not allowing for

- non-proportional hazards (e.g., time-varying effects)
- non-linear effects

Notation:

- $\lambda_i(t)$ hazard rate of observation i $[i = 1, \ldots, n]$
- $\lambda_0(t)$ baseline hazard rate
- $\mathbf{x}_i$ vector of covariates for observation i $[i = 1, \ldots, n]$
- $\boldsymbol{\beta}$ vector of regression coefficients

Generalization: Structured Survival Models

(Kneib & Fahrmeir, 2007)

$$\lambda_i(t) = \exp(\eta_i(t))$$

with additive predictor

$$\eta_i(t) = \sum_{l=1}^{L} f_l(\mathbf{x_i}(t)),$$

Generic representation of covariate effects $f_l(\mathbf{x}_i)$

a) linear effects: $f_l(\mathbf{x}_i(t)) = f_{l,\text{linear}}(\tilde{x}_i) = \tilde{x}_i\beta$

b) smooth effects: $f_l(\mathbf{x}_i(t)) = f_{l,\text{smooth}}(\tilde{x}_i)$

c) time-varying effects: $f_l(\mathbf{x}_i(t)) = f_{l,\text{smooth}}(t) \cdot \tilde{x}_i$

where $\tilde{x}_i$ is a covariate from $\mathbf{x}_i(t)$.

Note:

c) includes log-baseline ($\tilde{x}_i \equiv 1$)

Estimation

- Flexible terms can be represented using P-splines (Eilers & Marx, 1996)
- This leads to:

Penalized Likelihood Criterion:

$$\mathcal{L}_{\text{pen}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\delta_i \eta_i(t_i) - \int_0^{t_i} \exp(\eta_i(t))\, dt \right] - \sum_{l=0}^{L} \text{pen}_l(\boldsymbol{\beta}_l)$$

- NB: this is the **full** log-likelihood

Problem:

Estimation and in particular model choice

- t_i observed survival time
- δ_i indicator for non-censoring
- $\text{pen}_l(\boldsymbol{\beta}_l)$ P-spline penalty for smooth effects

CoxFlexBoost

Aim:

Maximization of a (potentially) high-dimensional log-likelihood with different modeling alternatives

Thus, we use:

- Iterative algorithm
- Likelihood-based boosting algorithm
- Component-wise base-learners

Therefore:

- Use one base-learner $g_j(\cdot)$ for each covariate (or each model component) $[\ j \in \{1, \ldots, J\}\ ]$

Component-Wise Boosting

as a means of estimation and variable selection combined with model choice.

CoxFlexBoost

Aim:

Maximization of a (potentially) high-dimensional log-likelihood with different modeling alternatives

Thus, we use:

- Iterative algorithm
- Likelihood-based boosting algorithm
- Component-wise base-learners

Therefore:

- Use one base-learner $g_j(\cdot)$ for each covariate (or each model component) $[\ j \in \{1, \ldots, J\}\ ]$

Component-Wise Boosting

as a means of estimation and variable selection combined with model choice.

CoxFlexBoost

Aim:

Maximization of a (potentially) high-dimensional log-likelihood with different modeling alternatives

Thus, we use:

- Iterative algorithm
- Likelihood-based boosting algorithm
- Component-wise base-learners

Therefore:

- Use one base-learner $g_j(\cdot)$ for each covariate (or each model component) $[\ j \in \{1, \dots, J\}\ ]$

Component-Wise Boosting

as a means of estimation and variable selection combined with model choice.

Some Details on CoxFlexBoost

In each boosting iteration m (until $m = m_{\text{stop}}$):

- All base-learners $g_j(\cdot)$ (i.e., modeling possibility) are fitted separately (based on penalized MLE).
- Choose best fitting base-learner $\hat{g}_{j^*}$ (i.e., the base-learner that maximizes the unpenalized LH)
- Add ...

  ... fraction ν of the fit ($\hat{g}_{j^*}$) to the model
  ... fraction ν of the parameter estimate ($\boldsymbol{\beta}_{\mathbf{j}^*}$) to the estimation

  ($\nu = 0.1$ in our case)

What happens then?

(parameters of) previously selected base-learners are treated as a constant in the next iteration

We stated that we use . . .

. . . component-wise boosting as a means of estimation and variable selection combined with model choice.

But how?

. . . is achieved by

- selection of base-learner, i.e., component-wise boosting

and

- early stopping,
  i.e., choose $\widehat{m}_{\text{stop,opt}}$ via cross validation, out-of-bag sample, . . .

We stated that we use ...

...component-wise boosting as a means of estimation and variable selection combined with model choice.

Variable Selection and Model Choice

... is achieved by

- selection of base-learner, i.e., component-wise boosting

and

- early stopping,
  i.e., choose $\widehat{m}_{\text{stop,opt}}$ via cross validation, out-of-bag sample, ...

- For model choice we apply the decomposition

$$f_{\text{smooth}}(x) = \underbrace{\beta_0 \quad + \beta_1 x}_{\text{unpenalized, parametric part}} + \underbrace{f_{\text{smooth,centered}}(x)}_{\text{deviation from polynomial}}$$

(based on Kneib, Hothorn, & Tutz, 2008)

- Add unpenalized part as separate, parametric base-learners
- Assign df $= 1$ to the centered effect (and add as P-spline base-learner)
- Analogously for time-varying effects

Thus:

- Modeling components are comparable (w.r.t. df)
  Important as otherwise more flexible base-learners are preferred
- Model choice between:
  - linear effects
  - smooth effects
  - linear time-varying effects
  - smooth time-varying effects

- For model choice we apply the decomposition

$$f_{\text{smooth}}(x) \cdot t = \underbrace{\beta_0 \cdot t + \beta_1 x \cdot t}_{\text{unpenalized, parametric part}} + \underbrace{f_{\text{smooth,centered}}(x) \cdot t}_{\text{deviation from polynomial}}$$

(based on Kneib et al., 2008)

- Add unpenalized part as separate, parametric base-learners
- Assign df $= 1$ to the centered effect (and add as P-spline base-learner)
- Analogously for time-varying effects

Thus:

- Modeling components are comparable (w.r.t. df)
  Important as otherwise more flexible base-learners are preferred
- Model choice between:
  - linear effects
  - smooth effects
  - linear time-varying effects
  - smooth time-varying effects

- For model choice we apply the decomposition

$$f_{\text{smooth}}(x) = \underbrace{\beta_0 + \beta_1 x}_{\text{unpenalized, parametric part}} + \underbrace{f_{\text{smooth,centered}}(x)}_{\text{deviation from polynomial}}$$

(based on Kneib et al., 2008)

- Add unpenalized part as separate, parametric base-learners
- Assign df $= 1$ to the centered effect (and add as P-spline base-learner)
- Analogously for time-varying effects

Thus:

- Modeling components are comparable (w.r.t. df)
  Important as otherwise more flexible base-learners are preferred
- Model choice between:
  - linear effects
  - smooth effects
  - linear time-varying effects
  - smooth time-varying effects

Application - Intensive Care Patients with Severe Sepsis

- **Response:** 90-day survival
- **Predictors:** 14 categorical predictors (sex, fungal infection (y/n), . . .)
  6 continuous predictors (age, Apache II Score, . . .)

Aim:

- flexible survival model for patients suffering from severe sepsis
- identify prognostic factors (at appropriate complexity)

Further Details of the Data-Set:

- **Origin:** Department of Surgery, Campus Großhadern, LMU Munich
- **Period of observation:** March 1993 – February 2005 (12 years)
- **N:** 462 septic patients (180 observations right-censored)

Application - Intensive Care Patients with Severe Sepsis (II)

CoxFlexBoost

- selected 10 out of 20 variables + baseline hazard
- used 15 different base-learners (out of 68)

⇒ sparse model

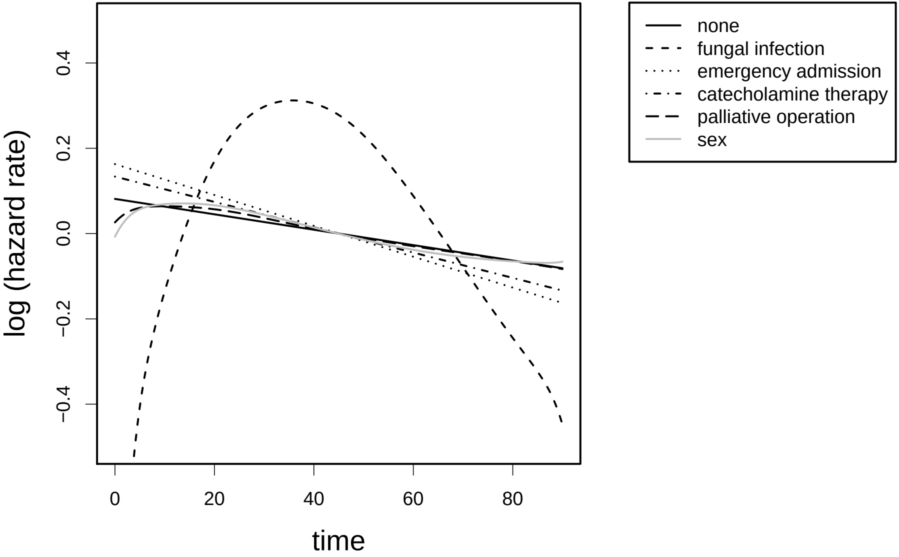
Out of 14 categorical covariates:

- 7 were selected
  - 2 were selected as linear effects
  - 4 were selected as time-varying effects
  - 1 was selected as linear and time-varying effect

Out of 6 continuous covariates:

- 3 were selected
  - 1 with linear effect
  - 2 with linear and time-varying effects

Application - Intensive Care Patients with Severe Sepsis (III)



Summary & Outlook

R-package **CoxFlexBoost** available (Hofner, 2008)

CoxFlexBoost . . .

- . . . allows for variable selection and model choice.
- . . . allows for flexible modeling
  - flexible, non-linear effects
  - time-varying effects (i.e., non-proportional hazards)
- . . . provides convenient functions to manipulate and show results (`summary()`, `plot()`, `subset()`, . . .)

More Details:
Tomorrow, 11:05; Room: "Kika"

Literature

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, *11*, 89–121.

Hofner, B. (2008). *CoxFlexBoost: Boosting Flexible Cox Models (with Time-Varying Effects).* (R package version 0.6-0)

Hofner, B., Hothorn, T., & Kneib, T. (2008). *Variable selection and model choice in structured survival models* (Tech. Rep. No. 43). Department of Statistics, Ludwig-Maximilans-Universität München.

Kneib, T., & Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, *34*, 207–228.

Kneib, T., Hothorn, T., & Tutz, G. (2008). Variable selection and model choice in geoadditive regression models. *Biometrics*. (accepted)

Find out more: `http://benjaminhofner.de/`