

Flexible Additive Models for Multi-Event Survival Analysis

Johannes Piller¹, Helmut Küchenhoff¹, Andreas Bender^{1,2}

¹ Statistisches Beratungslabor (StaBLab), LMU Munich, Germany

² Munich Center for Machine Learning (MCML), Germany

E-mail for correspondence: johannes.piller@stat.uni-muenchen.de

Abstract: The Piecewise Exponential Additive Mixed Model (PAMM) has become a popular method for complex modelling of single-event survival data. Here, we extend the framework and the relating R package `pammtools` to event-history analysis, i.e. competing risks, recurrent events and multi-state settings.

Keywords: survival analysis; time-to-event; piecewise exponential; multi-state; competing risks

1 Introduction

Piecewise Exponential Additive Mixed Models (PAMMs) (Bender et al., 2018) have gained popularity in various domains due to their ability to tackle a wide variety of survival problems and their flexibility to model non-linear covariate effects, including time-varying effects and cumulative effects (Bender et al., 2019). One advantage of such reduction techniques is that they do not require any specialised models or software in order to be fit. Thus, in the case of the PAMM, they can be conveniently estimated using generalized additive mixed modeling methodology or, for example, respective boosting or deep learning based approaches (Bender et al., 2022). Nevertheless, their use in practice requires considerable pre-processing, which differs depending on the survival task at hand (e.g. left-truncation, competing risks, etc.) and post-processing (e.g. transforming estimated parameters to useful quantities like survival or transition probabilities). The R package `pammtools` facilitates the entire modeling process, so far, however, only for single-event data. Here we extend the framework and package capabilities to handle general multi-state models.

This paper was published as a part of the proceedings of the 38th International Workshop on Statistical Modelling (IWSM), Durham, UK, 14–19 July 2024. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

2 Methods

We consider the general multistate setting with observed data

$$(y_{i,k,e}^{entry}, y_{i,k,e}^{exit}, \delta_{i,k,e}, \mathbf{x}_{i,k,e}),$$

where $y_{i,k,e}^{entry}$ is the entry time of subject $i = 1, \dots, n$ into the risk set for transition $k = 1, \dots, q$ in episode $e = 1, \dots, m$ and $y_{i,k,e}^{exit}$ the respective exit time, either due to the transition occurring or censoring. Here, $\delta_{i,k,e}$ is the corresponding status indicator and $\mathbf{x}_{i,k,e}^\top = (x_{i,k,e,1} \cdots x_{i,k,e,p})$ the covariate row-vector, which, besides subject specific information, can contain information about the past (e.g. number and timing of past transitions and number and timing of previous episodes). Note that we need both indices, k and e , as, in the general case, some transitions could occur more often, for example in case of recurrent events (with competing risk) or multi-state models with back-transitions. The transitions are modeled via log-hazard rates

$$\log(h_{k,e}(t|\mathbf{x}_{i,k,e}, \ell_i)) = \beta_{0,k,e} + f_{0,k,e}(t) + \sum_{p=1}^P f_p(x_{i,k,e,p}, t) + b_{\ell_i}. \quad (1)$$

In Eq. (1) the log-baseline is given by $\log(h_{0,k,e}(t)) = \beta_{0,k,e} + f_{0,k,e}(t)$, $f_p(x_{i,k,e,p}, t)$ are potentially non-linear, potentially time-varying effects of covariates and b_{ℓ_i} potential random effects for cluster ℓ to which subject i belongs to. Eq. (1) could be further extended to include stratification of the baseline hazard according to subsets of subjects, interactions between different covariates (e.g. via tensor products), cumulative effects of time-dependent covariates, and more complex random-effect structures, but this is omitted here for simplicity.

Using PAMMs, we estimate Eq. (1) by splitting the follow-up into J intervals, with intervals $(\kappa_{j-1}, \kappa_j]$, $j = 1, \dots, J$, transforming the raw-data accordingly and estimating the interval-specific hazard rates $h_{k,e}(t|\mathbf{x}_{i,k,e}) = h_{k,e,j}(\mathbf{x}_{i,k,e}), \forall t \in (\kappa_{j-1}, \kappa_j]$. Non-linear functions are parameterized as $f_{p,k,e}(x_{i,k,e,p}) = \sum_{g=1}^G \gamma_{p,g} B_{p,g}(x_{i,k,e,p})$ with basis coefficients $\gamma_{p,g}$ and suitable basis functions $B_{p,g}$ (e.g. B-splines). The respective parameters are estimated by maximizing the corresponding penalized Poisson likelihood (Wood, 2020) or other suitable estimation techniques. Note that a separate smooth baseline hazard is estimated for each transition and episode, but could also be reduced to transition specific baseline hazard rates. Once the transition specific hazard rates in Eq. (1) are estimated, the transition probabilities are obtained via the empirical transition probability matrix (Beyersmann et al., 2011). To calculate the empirical probability matrix, first, we discretize the time component $t \in \mathcal{T}$, with $\mathcal{T} = \{T \in \mathbb{N} : t_1 < \dots < t_T\}$, i.e. t stems from a grid of ordered time points. Second, we integrate the specific hazard rates and obtain the cumulative transition hazards, i.e.

$H_{k,e}(t|\mathbf{x}) = \int_0^t h_{k,e}(u|\mathbf{x})du$. To define transition matrices, note that for each transition k , let Q be the set of all states and $Q_l \subseteq Q, l \in Q$ the set of all possible states after transitioning from l , then there exists a tuple $(l, o) \in Q \times Q_l$, describing the k -th transition from state l to state o , which, in the following, is denoted by $l \rightarrow o$. The transition matrix is then given by a finite matrix product over all event times t and matrices $\mathbf{I} + d\hat{\mathbf{H}}(t|\mathbf{x})$, with entries $d\hat{H}_{l \rightarrow o,e}(t|\mathbf{x}) = \hat{H}_{l \rightarrow o,e}(t|\mathbf{x}) - \hat{H}_{l \rightarrow o,e}(t-|\mathbf{x})$, i.e.

$$\prod_{t \in \mathcal{T}} (\mathbf{I} + d\hat{\mathbf{H}}(t|\mathbf{x})) \quad (2)$$

Since the rows of probability matrices must sum up to one, the diagonal elements of $\hat{\mathbf{H}}(t|\mathbf{x})$, are defined to be $\hat{H}_{l \rightarrow l,e}(t|\mathbf{x}) = -\sum_{o \in Q(l)} \hat{H}_{l \rightarrow o,e}(t|\mathbf{x})$. Using PAMMs, we estimate Eq. (2) by applying the same discretization and data transformation for the interval- and transition-specific hazard rates Eq. (1) and calculate

$$H_{k,e}(t|\mathbf{x}) = \sum_{l=1}^{j(t)-1} h_{k,e,j}(\mathbf{x})(\kappa_l - \kappa_{l-1}) \cdot h_{i,k,j(t)}(t - \kappa_{j(t)-1}) \quad (3)$$

where $j(t)$ the index of the interval for which $t \in (\kappa_{j(t)-1}, \kappa_{j(t)}]$.

3 Application

The presented PAMMs approach enables smooth effects to be used to estimate cumulative hazards and transition probabilities. To illustrate the flexibility, we used the `pamtools` package to analyze the `mgus2` data set in the `survival` package, see (Kyle et al., 2002). The data set contains a classical illness-death setup with possible transitions $0 \rightarrow 1$, i.e. progression to a plasma cell malignancy (pcm), $0 \rightarrow 2$, i.e. death, or $1 \rightarrow 2$, i.e. progression from pcm to death. We fitted a model with a linear effect and with a non-linear effect of hemoglobin (`hgb`) to estimate the hazard rates. Figure 1 visualizes the non-linear effect of `hgb` and the time `tend` on the logarithmic hazard rates for the transition $0 \rightarrow 2$ and the transition probabilities for the linear and non-linear model. As can be seen in the left facet of the figure, the change of the amount of `hgb` within $[11, 14]$ in $[g/dl]$ has a stronger decreasing effect on the hazard rates than for amounts within $[9, 11]$ c.p. The overall fit and the fact that the effective degrees of freedom are $4.65 > 1$ undermine the need of the more flexible smooth functions for modeling the effect of `hgb`. Figure 1 compares further the influence of the linearly (third facet) and the non-linearly (right facet) modeled hazard rates on the transition probabilities. In the linear case, the transition probabilities change equally over time and for different levels of `hgb`. In the non-linear case, the transition probabilities change according to the structure of the smooth functions in the first and second facet of figure 1, e.g.

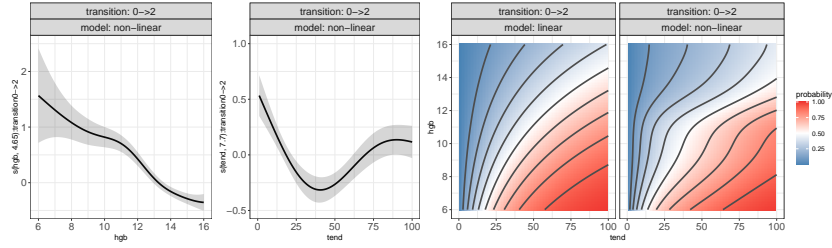


FIGURE 1. For transitions into the death stage, the first facet depicts the non-linear effect of hgb , the second depicts the non-linear effect of time, the third and forth facets depict the transition probabilities depending on the time and hgb , modelled with a linear effect (third facet) or a non-linear effect (fourth facet).

for hgb within $[11, 14]$, the difference in the transition probabilities is larger than for amounts within $[9, 11]$.

Summarizing, we have shown, that the flexible hazard modeling using PAMMs can be extended from single-event survival models to multi-stage settings. We provide an extension of our software package `pamtools` for an efficient application of these models.

References

- Bender, A, Groll, A., and Scheipl, F. (2018). A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, **18**, 299–321.
- Bender, A, Rügamer, D., Scheipl, F., and Bischl B. (2020). A General Machine Learning Framework for Survival Analysis In: *Hutter F, Kerting K, Lijffijt J, Valera I, editors, Machine learning and knowledge discovery in databases*, Lecture Notes in Computer Sciences, Springer International Publishing, 158–173.
- Bender, A, Scheipl, F., Hartl, W., Day, A.G., and Küchenhoff H. (2019). Penalized estimation of complex, non-linear exposure-lag-response associations. *Biostatistics*, **20(2)**, 315–331.
- Beyersmann J, Allignol A, and Schumacher M. (2011). *Competing risks and multistate models with R*. New York: Springer.
- Kyle, R., Therneau, T., Rajkumar, V., Offord, J., Larson, D., Plevak, M., and Melton III, L. J. (2002). A long-terms study of prognosis in monoclonal gammopathy of undertermined significance. *New Engl J Med*, **346**, 564–569
- Wood, S.N. (2020). Inference and computation with generalized additive models and their extensions. *TEST*, **29**, 307–339.