Flexible Additive Models for Multi-Event Survival Analysis

Johannes Piller[1,2], Helmut Küchenhoff[1], Andreas Bender[1,2]

[1] Statistiches Beratungslabor (StaBLab), LMU Munich, Germany
[2] Munich Center for Machine Learning (MCML), Munich, 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 software for the estimation of the model parameters. 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 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.

2 Methods

We consider the general multistate setting with observed data

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

where $y^{entry}_{i,k,e}$ is the entry time of subject $i = 1, \ldots, n$ into the risk set for transition $k = 1, \ldots, q$ in episode $e = 1, \ldots, m$ and $y^{exit}_{i,k,e}$ the respective exit time, either due to the transition occurring or censoring. Here, $\delta_{i,k,e}$ is the corresponding status indicator and $\mathbf{x}^{\top}_{i,k,e} = (x_{i,k,e,1} \cdots x_{i,k,e,p})$ 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 or back-transitions. The transitions are modeled via log-hazard rates

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

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 $(\kappa_{j-1}, \kappa_j]$, $j = 1, \ldots, J$, transforming the raw-data accodringly 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 $\mathcal{T} = \{T \in \mathbb{N} : t_1 < \ldots < t_T\}$, 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. $H_{k,e}(t|\mathbf{x}) = \int_0^t h_{k,e}(u|\mathbf{x})du$. Next, we define transition matrices, which

contain the transition probabilities. Let Q be the set of all states for each transition k 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 $l \to o$. 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 $d\hat{H}_{l\to o,e}(t|\mathbf{x}) = \hat{H}_{l\to o,e}(t|\mathbf{x}) - \hat{H}_{l\to o,e}(t - |\mathbf{x})$, i.e.

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

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 $\hat{H}_{l\to l,e}(t|\mathbf{x}) = -\sum_{o\in Q(l)} \hat{H}_{l\to o,e}(t|\mathbf{x})$. 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

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

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

3 Simulation

In this section, we illustrate the capabilities of the model via simulation studies. While we focus on simple settings (e.g. to illustrate equivalence to Aalen-Johanson estimator), the model class can be used in more complex settings.

In the first simulation study, we sample an illness-death multi-state setting with 200 subjects and constant transition-specific log-hazards $h_{0\to1} = 0.3, h_{0\to2} = 0.6$, and $h_{1\to2} = 0.5$. Using the functionalities of the `mvga` package, see (Allignol et al., 2008), we calculate the Aalen-Johanson estimator for the simulated multi-state model. Using PAMMs, we estimate the log-hazards and calculate the transition probabilites. Figure 1 shows equivalence of the two approaches.

In the second simulation study, we sample an illness-death multi-state setting with non-linear covariate effects and time dependent transition-specific log-hazards given by $h_{0\to1}(t|(x)) = 0.5 + 0.25x^3$, $h_{0\to2}(t|(x)) = \frac{1}{2^8\cdot\Gamma(8)} t^7 e^{-t/2}$, and $h_{1\to2}(t|(x)) = 0.4 - x^2$. Note that the hazards are by construction highly non-linear and also time-dependent. The simulation is built on a data set with 1500 subjects, a covariate x, sampled uniformly from $[-3, 3]$, and 100 repetitions. Using PAMMs, for each repetition, we fitted a model with non-linear effects in `x` and `t`. Finally, we calculated the average and compared the true curve with the average fit. Figure 2 shows that the average fit of the non-linear log-hazards (red) is close to the pre-defined transition-specific log-hazards of the simulation setup.

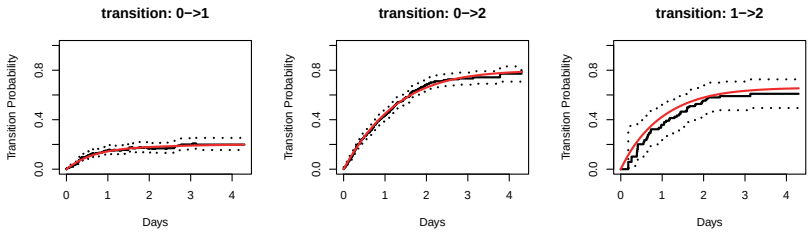

FIGURE 1. Shown are the transition probabilities based on the Aalen-Johanson estimator (solid, black) with confidence bands (dotted, black), and the estimation using `pammtools` (solid, red).

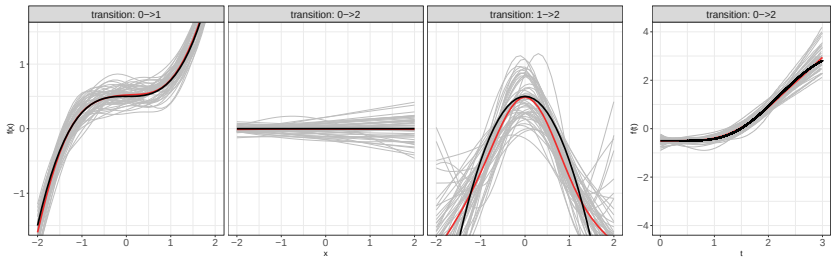

FIGURE 2. The first three facets depict the non-linear effect of `x` for each transition. The last facet depicts the non-linear effect of `t` for $0 \to 2$. Results from each iteration are grey, the average is red, and the true function is black.

4 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 `pammtools` 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 $0 \to 1$, i.e. progression to a plasma cell malignancy (pcm), $0 \to 2$, i.e. death, or $1 \to 2$, i.e. progression from pcm to death. We estimated the hazards with a linear and a non-linear effect of hemoglobin (`hgb`). Figure 3 visualizes the non-linear effect of `hgb` and the time `tend` on the logarithmic hazard rates for the transition $0 \to 2$ 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 3 compares further the influence of the linearly (third facet) and the non-linearly (right facet) modeled hazard rates on the transition

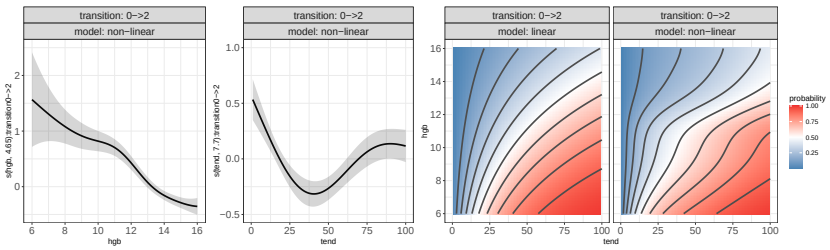

FIGURE 3. For transitions into the death stage, the first facet depicts the non–linear effect of `hgb`, the second depicts the non-linear effect of time, the third and forth facets depict the transition probabilities depending on the time and `hgb`, modelled with a linear effect (third facet) or a non-linear effect (fourth facet).

probabilities. In the linear case, the transition probabilities change equally over time and 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 3.

5 Discussion

This article illustrates a workflow for reducing complex multi-state tasks to more standard regression tasks using the R package `pammtools`. All transition hazards and probabilities, including their dependencies on covariates and time, can be estimated in one single model. The additive predictor used to define the transition hazards can be very flexible, including non-linear terms, spatial effects, random effects and their interactions. While this approach was illustrated mainly on non-linearity with a focus on extending single-event survival models to multi-state settings in Sections 3 and 4, the approach will be particularly useful when the Markov assumption may be violated, i.e., when the hazard at time t depends on the past. In this approach, one could model such dependencies by introducing time-varying covariates like number of past transitions or time spent in previous states into the restructured data set, which natively supports dependencies on multiple timescales (Iacobelli et al., 2013). One disadvantage of the data transformation, in particular in the multi-state setting, is the increase of the data set. However, since the estimation problem is reduced to a Poisson regression task, available techniques for efficient estimation in the big data context (Wood et. al., 2017, Reulen et. al., 2015, Sennhenn-Reulen et. al., 2016) are applicable, such that even models with millions of rows and complex predictors are estimable within reasonable time and memory requirements. Nevertheless, further reduction of the transformed data set size is desirable and could be addressed in future iterations of the implementation.

References

Allignol, A., Beyersman, J., and Schumacher, M. (2008). mvna: An R Package for the Nelson–Aalen Estimator in Multistate Models. *R News*, **8(2)**, 48 – 50.

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, Kersting 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.

Iacobelli, S., and Carstensen, B. (2013). Multiple Time Scales in Multi-State Models. *Statistics in Medicine*, **32(30)**, 5315 – 5327.

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.

Reulen, H., and Kneib, T. (2015). Boosting Multi-State Models. *Lifetime Data Analysis*, 1 – 22.

Sennhenn-Reulen, H., and Kneib, T. (2016). Structured Fusion Lasso Penalized Multi-State Models. *Statistics in Medicine*, **35(25)**, 4637 – 4659.

Wood, S.N. (2020). Inference and computation with generalized additive models and their extensions. *TEST*, **29**, 307 – 339.

Wood, S.N., Li, Z., Shaddick, G., and Augustin, N.H. (2017). Generalized Additive Models for Gigadata: Modelling the U.K. Black Smoke Network Daily Data. *Journal of the American Statistical Association*, **112(519)**, 1199 – 1210.