mvna: An R Package for the Nelson–Aalen Estimator in Multistate Models

by A. Allignol, J. Beyersmann and M. Schumacher

Introduction

The multivariate Nelson–Aalen estimator of cumulative transition hazards is the fundamental nonparametric estimator in event history analysis. However, there has not been a multivariate Nelson–Aalen R–package available, and the same appears to be true for other statistical softwares, like SAS or Stata. As the estimator is algebraically simple, its univariate entries are easily computed in R using the `survfit` function of the package **survival** (Therneau and Grambsch, 2001; Lumley, 2004), but other values returned by `survfit` may be meaningless in the multivariate setting and consequently misleading (Kalbfleisch and Prentice, 2002, Chapter VIII.2.4). In addition, the computations become cumbersome for even a moderate number of transition types, like 4 or 5.

For instance, time–dependent covariates may easily be incorporated into a proportional hazards analysis, but, in general, this approach will not result in a model for event probabilities any more (Kalbfleisch and Prentice, 2002, Chapter VI.3). The multivariate Nelson–Aalen estimate straightforwardly illustrates the analysis, while popular stratified Kaplan-Meier plots lack a probability interpretation (*e.g.,* Feuer et al. (1992); Beyersmann et al. (2006b)).

The article is organised as follows: First we describe the main features of the **mvna** package. Next we use the package to illustrate the analysis of a time–dependent covariate. We briefly discuss the relative merits of the available variance estimators, and conclude with a summary.

Package description

The **mvna** contains the following functions:

- `mvna`
- `xyplot.mvna`
- `print.mvna`
- `predict.mvna`

The main function, `mvna`, computes the Nelson–Aalen estimates at each of the observed event times, and the two variance estimators described in eq. (4.1.6) and (4.1.7) of Andersen et al. (1993). `mvna` takes as arguments a transition–oriented data.frame, where each row represents a transition:

```
 id from   to time
  1    0    2    4
  2    1    2   10
  3    1    2    2
  4    1    2   49
  5    1    0   36
  5    0    2   47
749    1    0   11
749    0 cens   22
```

`id` is the patient identification number, `from` is the state from which a transition occurs, `to` is the state to which a transition occurs and `time` is the transition time. Left–truncation, *e.g.,* due to delayed entry into the study cohort, can be handled replacing `time` with an `entry` and `exit` time within a state. One need to specify the state names, the possible transitions, and the censoring code. For example, patient 749 is right–censored at time 22.

`xyplot.mvna` plots the cumulative hazard estimates in a lattice plot (Sarkar, 2002), along with pointwise confidence intervals, with possible log or arcsin transformations to improve the approximation to the asymptotic distribution. The `predict.mvna` function gives Nelson–Aalen estimates at time points given by the user.

Two random samples of intensive care unit data are also included within the **mvna** package (Beyersmann et al., 2006a). One of these data sets focuses on the effect of pneumonia on admission on the hazard of discharge and death, respectively. The other one contains data to gauge the influence of ventilation (a time–dependent covariate) on the hazard of end of hospital stay (combined endpoint death/discharge).

Illustration

The aim of the present section is to analyse the influence of a time-dependent covariate, namely ventilation, which is binary and reversible. We use the model of Figure 1 to analyse the data, where ventilation is considered as an intermediate event in an illness–death–model (Andersen et al., 1993, Chapter I.3.3).

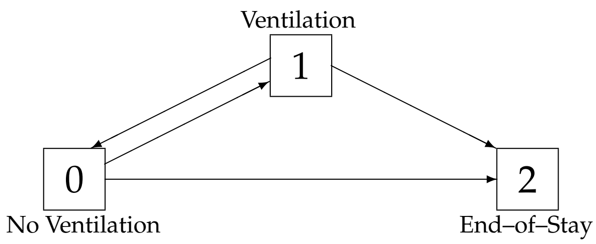

Figure 1: Model to assess the influence of ventilation.

The following call computes the estimated cumulative hazards along with the estimated variance

```
> data(sir.cont)
> na.cont <- mvna(sir.cont,c("0","1","2"),
+                    tra,"cens")
```

tra being a quadratic matrix of logical specifying the possible transitions:

```
> tra
      0     1     2
0  FALSE  TRUE  TRUE
1   TRUE FALSE  TRUE
2  FALSE FALSE FALSE
```

To assess the influence of the time–dependent covariate ventilation on the hazard of end–of–stay, we are specifically interested in the transitions from state "No ventilation" to "End–of–stay" and from state "Ventilation" to "End–of–stay". We specify to plot only these cumulative hazard estimates with the `tr.choice` option of the `xyplot.mvna` function.

```
> xyplot(na.cont,tr.choice=c("0 2","1 2"),
+ aspect=1,strip=strip.custom(bg="white",
+ factor.levels=
+    c("No ventilation -- Discharge/Death",
+      "Ventilation -- Discharge/Death"),
+ par.strip.text=list(cex=0.9)),
+ scales=list(alternating=1),xlab="Days",
+ ylab="Nelson-Aalen estimates")
```

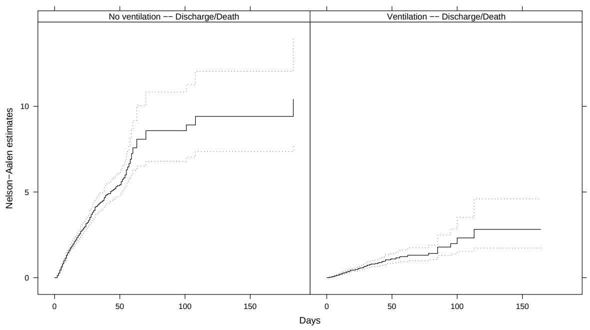

Figure 2: Nelson–Aalen estimates for the transitions "No ventilation" to "End–of–Stay", and "Ventilation" to "End–of–Stay".

We conclude from Figure 2 that ventilation prolongs the length of ICU stay since it reduces the hazard of end–of–stay.

Fig. 2 is well suited to illustrate a Cox analysis of ventilation as a time–dependent covariate (Therneau and Grambsch, 2001). Taking no ventilation as the baseline, we find a hazard ratio of 0.159 (95% CI, 0.132–0.193), meaning that ventilation prolongs hospital stay. Note that, in general, probability plots are not available to illustrate the analysis.

Variance estimation

`mvna` computes two variance estimators based on the optional variation process (Andersen et al., 1993, eq. (4.1.6)) and the predictable variation process (Andersen et al., 1993, eq. (4.1.7)) 'attached' to the Nelson-Aalen estimator, respectively. For standard survival data, Klein (1991) found that the (4.1.6) estimator tended to overestimate the true variance, but that the bias was of no practical importance for risk sets ≥ 5. He found that the (4.1.7) estimator tended to underestimate the true variance and had a smaller mean squared error. The two estimators appear to coincide from a practical point of view for risk sets ≥ 10. Small risk sets are a concern in multistate modelling because of multiple transient states. In a preliminary investigation, we found comparable results in a multistate model simulation. As in Klein (1991), we recommend the use of the (4.1.6) estimator, which was found to have a smaller absolute bias.

Summary

The **mvna** package provides a way to easily estimate and display the cumulative transition hazards from a time–inhomogeneous Markov multistate model. The estimator may remain valid under even more general assumptions; we refer to Andersen et al. (1993, Chapter III.1.3) and Glidden (2002) for mathematically detailed accounts. We hope that the **mvna** package will help to promote a computationally simple estimator that is extremely useful in illustrating and understanding complex event history processes, but that is underused in applications. We have illustrated the usefulness of the package for visualising the impact of a time–dependent covariate on survival. We also wish to mention that the package similarly illustrates standard Cox competing risks analyses in a straightforward way, whereas the usual cumulative incidence function plots don't; this issue is pursued in Beyersmann and Schumacher (2008). In closing, we also wish to mention the very useful R-package **muhaz** (Gentleman) for producing a smooth estimate of the survival hazard function, which also allows to estimate the univariate Nelson–Aalen estimator subject to right-censoring, but not to left-truncation.

Bibliography

P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer-Verlag, New-York, 1993.

J. Beyersmann and M. Schumacher. Time-dependent covariates in the proportional subdistribution hazards model for competing risks. *Biostatistics*, 2008.

J. Beyersmann, P. Gastmeier, H. Grundmann, S. Bärwolff, C. Geffers, M. Behnke, H. Rüden, and M. Schumacher. Use of Multistate Models to Assess Prolongation of Intensive Care Unit Stay Due to Nosocomial Infection. *Infection Control and Hospital Epidemiology*, 27:493–499, 2006a.

J. Beyersmann, T. Gerds, and M. Schumacher. Letter to the editor: comment on 'Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator' by Steven Snapinn, Qi Jiang, and Boris Iglewicz in the November 2005 issue of The American Statistician. *The American Statistician*, 60(30):295–296, 2006b.

E. J. Feuer, B. F. Hankey, J. J. Gaynor, M. N. Wesley, S. G. Baker, and J. S. Meyer. Graphical representation of survival curves associated with a binary non-reversible time dependent covariate. *Statistics in Medicine*, 11:455–474, 1992.

R. Gentleman. *muhaz: Hazard Function Estimation in Survival Analysis*. R package version 1.2.3.

D. Glidden. Robust inference for event probabilities with non-Markov data. *Biometrics*, 58:361–368, 2002.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, 2002.

J. P. Klein. Small sample moments of some estimators of the variance of the Kaplan–Meier and Nelson–Aalen estimators. *Scandinavian Journal of Statistics*, 18:333–340, 1991.

T. Lumley. The survival package. *R News*, 4(1):26–28, June 2004. URL `http://CRAN.R-project.org/doc/Rnews/`.

D. Sarkar. Lattice. *R News*, 2(2):19–23, June 2002. URL `http://CRAN.R-project.org/doc/Rnews/`.

T. Therneau and P. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Statistics for biology and health. Springer, New York, 2001.

Arthur Allignol
Jan Beyersmann
Martin Schumacher

Freiburg Center for Data Analysis and Modelling, University of Freiburg, Germany.
Institute of Medical Biometry and Medical Informatics, University Medical Centre Freiburg, Germany.

`arthur.allignol@fdm.uni-freiburg.de`