III Luso-Galician Meeting on Biometry
Aveiro, 28 - 30 June 2018

A COMPARATION OF PRESMOOTHING METHODS IN THE ESTIMATION OF TRANSITION PROBABILITIES

Gustavo Soutinho[1], Luís Meira-Machado[2] and Pedro Oliveira[3]

[1]University of Minho, Portugal.
[2]Centre of Molecular and Environmental Biology & Department of Mathematics and Applications, University of Minho, Portugal.
[3]EPIUnit, ICBADS, University of Porto, Portugal

ABSTRACT

One major goal in clinical applications of multi-state models is the estimation of transition probabilities. In a recent paper, landmark estimators were proposed to estimate these quantities, and their superiority with respect to the competing estimators has been proved in situations in which the Markov condition is violated. The idea behind their estimator is to use a procedure based on (differences between) Kaplan-Meier estimators derived from a subset of the data consisting of all subjects observed to be in the given state at the given time. Because of this, the computation of their estimator is performed in small sample sizes providing large standard errors in some circumstances. A valid approach is to consider a modification of the landmark estimator based on presmoothing. In this two presmoothing methods are compared. Simulation results indicate that both methods may be much more efficient than the unsmoothed estimator. Real data illustration is included.

Keywords and key sentences: Kaplan-Meier, Multi-state model, Nonparametric estimation, Presmoothing, Survival Analysis.

1. INTRODUCTION

Multi-state models can be successfully used to model the movement of patients among a set of several states. The so-called 'illness-death' model plays a central role in the theory and practice of these models. In these models one important goal is the estimation of the transition probabilities since they allow for long-term predictions of the process. These quantities have been traditionally estimated by the Aalen-Johansen estimator (Aalen and Johansen, 1978), which is consistent if the process is Markovian. Alternative landmark estimators which are consistent regardless the Markov conditions have been proposed in the recent literature (de Uña-Álvarez and Meira-Machado, 2015), and their superiority with respect to the competing estimators has been proved in situations in which the Markov condition is violated. The idea behind the proposed methods is to use specific subsamples or portions of data at hand (namely, those observed to be in a given state at a pre-specified time point) for which the

ordinary Kaplan-Meier (Kaplan and Meier, 1958) survival function leads to a consistent estimator of the target. A weakness of the new method emerges, in some circumstances, from the large estimated standard errors. To avoid this problem, a valid approach is to consider a modification of the landmark estimator based on presmoothing (Cao et al., 2005).

2. PRESMOOTHED ESTIMATORS OF THE TRANSITION PROBABILITIES

A multi-state model is a model for a time continuous stochastic process $(Y(t), t \in \mathcal{T})$ which at any time occupies one of a few possible states. In this paper we consider the progressive illness-death model depicted in Figure **??** and we assume that all subjects are in State 1 at time $t = 0$ (i.e., $Y(0) = 1$). This model is encountered in many medical studies for describing the progression of patients undergoing a given illness, particularly in cancer studies.

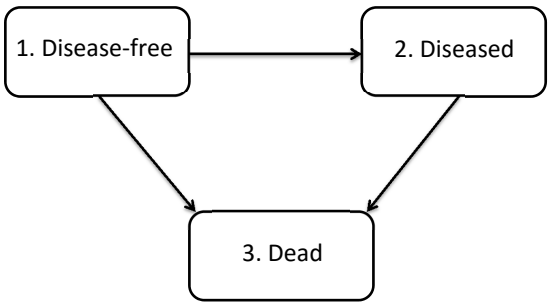

Figure 1: The progressive illness-death model.

For two states k, j and two time points $s < t$, introduce the so-called transition probabilities $p_{kj}(s,t) = P(Y(t) = j|Y(s) = k)$. In the illness-death model we have five different transition probabilities to estimate: $p_{11}(s,t)$, $p_{12}(s,t)$, $p_{13}(s,t)$, $p_{22}(s,t)$ and $p_{23}(s,t)$.
Recently, de Uña-Álvarez and Meira-Machado (2015) introduce new (landmark) estimators for non-Markov processes. The idea of the new methods is to use a procedure based on (differences between) Kaplan-Meier estimators derived from a subset of the data consisting of all subjects observed to be in the given state at the given time. In this paper, we use presmoothing to improve efficiency of the landmark estimators.
Several successful applications of presmoothed estimators have been used in recent literature. All these references concluded that the presmoothed estimators have improved variance when compared to purely nonparametric estimators. This 'presmoothing' is obtained by replacing the censoring indicator variables in the expression of the Kaplan-Meier weights by a smooth fit. This preliminary smoothing may be based on a certain parametric family such as the logistic, probit or cauchit, or on a nonparametric estimator of the binary regression curve. When the parametric family is the right one, parametric presmoothing leads to more efficient estimation than that associated to the unsmoothed estimator. Nonparametric presmoothing is useful when there is a clear risk of a miss-specification of the parametric model. The validity of a given parametric model for presmoothing can be checked graphically or formally, by applying a goodness-of-fit test such as the test proposed by Hosmer and Lemeshow (1989).

3. EXAMPLE OF APPLICATION

In this section we use data of 929 patients affected by colon cancer that underwent a curative surgery for colorectal cancer. In this study, 468 developed recurrence and among these 414 died. 38 patients died without recurrence. Recurrence can be considered as an intermediate transient state and modeled using an illness-death model with transient states 'alive and disease-free' and 'alive with recurrence', and an absorbing state 'dead'.

Figure **??** reports estimated transition probabilities for $p_{ij}(s,t)$, for a fixed value for $s = 365$ days along time t. Plots labbeled as 'Unsmoothed' correspond to the original unsmoothed landmark estimator proposed by de Uña-Álvarez and Meira-Machado (2015) which reveals higher variability on the right hand side. Remaining curves correspond to estimators with a preliminary presmoothing based on a parametric binomial family ('logit', 'probit' or 'cauchit'), on an additive logistic model ('logit.gam') or on a nonparametric regression model ('nonparametric') using the Nadaraya-Watson kernel estimator. We have applied the goodness-of-fit test proposed by Hosmer and Lemeshow (1989) which revealed that the test was able to reject the logistic model when used to presmoothed estimation of the transition probabilities $p_{1j}(365,t)$. Note that the choice of this parametric model is a common choice for a parametric presmoothing. Though one could consider a different parametric model, nonparametric presmoothing is a useful approach when there is a clear risk of a miss-specification of the parametric model.

Plots for the transition probabilities $p_{11}(365,t)$ and $p_{22}(365,t)$ reveal minor differences in the estimated curves based on different methods. Some differences are observed in the right tail of the curves obtained from the different methods when estimating the transition probabilities $p_{12}(365,t)$ and $p_{13}(365,t)$. Plots on bottom of the left-hand side allow for an inspection along time of the probability of being alive with recurrence for the individuals who are disease-free one year, one year after surgery. Since the recurrence state is transient, this curve is first increasing and then decreasing. Major differences are observed in the right tail when comparing the methods based on a parametric presmoothing with their counterparts. A similar behavior is observed when estimating the transition probability $p_{13}(365,t)$ reported in the right-hand side of Figure **??** (top). These plots report one minus the survival fraction along time, among the individuals in the recurrence state. Curves depicted in Figure **??** reveal that the nonparametric presmoothed landmark estimator provide in all cases reliable curves, similar to those obtained from the nonparametric estimators but with less variability. Since there is a clear risk of a miss-specification of the parametric model, the use of estimators based on nonparametric presmoothing as those labelled with 'nonparametric' are preferable.

3. CONCLUSIONS

There have been several recent contributions for the estimation of the transition probabilities in the context of non-Markov multi-state models. Recently, the problem of estimating the transition probabilities in a non-Markov illness-death model has been reviewed, and new estimators have been proposed which are built by considering specific subsets of individuals (namely, those observed to be in a given state at a prespecified time point s for which the ordinary Kaplan-Meier survival function leads to a consistent estimator of the target. As a weakness, it provides large standard errors for large values of s and higher censoring percentages. In this article we compare several approaches based on a presmoothed version of the Kaplan-Meier estimator that can be used to reduce the variability of the proposed estimator. Results obtained in simulations studies (not reported here) suggest that presmoothed approaches are preferable to the original nonparametric estimator, since they often have less variance while providing more reliable curves. Parametric presmoothing is a recommended approach if there is no clear risk of miss-specification of the parametric model. Otherwise the use of nonparametric pressmoothing or based on an additive model is recommended.

ACKNOWLEDGMENT

This research was financed by Portuguese Funds through FCT - "Fundação para a Ciência e a Tecnologia".

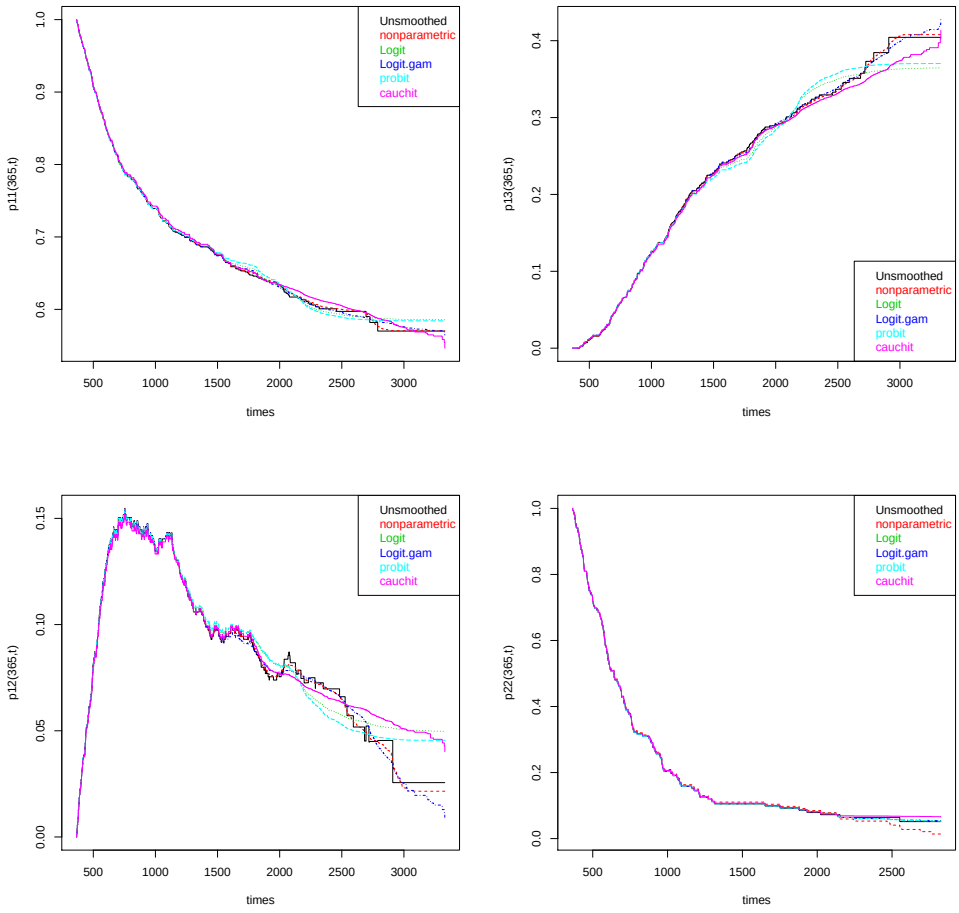

Figure 2: Estimated transition probabilities.

References

[1] Aalen, O.O. and Johansen, S. (1978). An empirical transition matrix for non homogeneous Markov and chains based on censored observations Matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* 5, 141–150.

[2] Cao, R., Lopez-de Ullibarri, I., Janssen, P., and Veraverbeke, N. (2005). Presmoothed Kaplan-Meier and Nelson-Aalen estimators. *Journal of Nonparametric Statistics* 17, 31–56.

[3] de Uña-Álvarez, J. and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study. *Biometrics* 71, 364–375.

[4] Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley & Sons, New York.

[5] Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457–481.