

Methods for checking the Markov condition in multi-state survival data

Gustavo Soutinho¹, Luís Meira-Machado², Pedro Oliveira¹

¹ EPIUnit, ICBADS, University of Porto, Portugal.

² Department of Mathematics and Centre of Molecular and Environmental Biology (CBMA), University of Minho, Portugal.

E-mail for correspondence: lmachado@math.uminho.pt

Abstract: The inference in multi-state models is traditionally performed under a Markov assumption. This assumption claims that given the present state, the future evolution of the process is independent of the states previously visited and the transition times among them. Usually, this assumption is checked including covariates depending on the history. However, since the landmark methods of the transition probabilities proposed by de Uña-Álvarez and Meira-Machado (2015), and by Putter and Spitoni (2018) are free of the Markov assumption, they can also be used to introduce such tests (at least in the scope of the progressive multi-state models) by measuring their discrepancy to Markovian estimators. In this paper, we introduce two local tests for the Markov assumption and compare them with the usual approach based on local future-past association through simulations. An application to a clinical trial on colon cancer is included.

Keywords: Censoring; Markov assumption; Multi-state models; Transition probabilities.

1 Introduction

Multi-state models are the most suitable models for the description of complex longitudinal survival data involving several events of interest. A multi-state model is a model for a stochastic process, which is characterized by a finite number of states and the possible transitions among them. In general, the multi-state analysis deals with inference for transition intensities and transition probabilities. The inference for transition intensities often includes regression analysis which usually involves the modelling of each transition intensity separately. A popular choice is to model each transition intensity using a proportional hazards model assuming the process to

This paper was published as a part of the proceedings of the 34th International Workshop on Statistical Modelling (IWSM), Guimarães, Portugal, 7–12 July 2019. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

be Markovian. This assumption claims that given the present state, the future evolution of the process is independent of the states previously visited and the transition times among them; in other words, the history of the process is summarized by the state occupied at time t . However, it has been quoted that the Markov assumption is violated in some applications (Andersen 2000, 2002). In such cases, if interest is on multi-state regression, one alternative approach is to use a semi-Markov model in which the future of the process does not depend on the current time but rather on the duration in the current state. Semi-Markov models are also called “clock reset” models, because each time the patient enters a new state time is reset to 0. This Markov assumption allows also the construction of simple estimators for the transition probabilities, since individuals with different past histories become comparable. Unfortunately, when this assumption is violated, the use of the Aalen-Johansen estimators for transition probabilities can induce bias, and thus may not be recommended. Substitute estimators for the Aalen-Johansen estimator for a non-Markov process were recently introduced by de Uña-Álvarez and Meira-Machado (2015). The proposed estimators are based on subsampling, also referred to as landmarking, and may behave more efficiently than the Aalen-Johansen estimators when the Markov assumption does not hold. Putter and Spitoni (2018) recover the work by Meira-Machado and Meira-Machado (2015) to propose alternative non-Markovian estimation methods which are based on the landmark methodology combined with the Aalen-Johansen estimate of the state occupation probabilities derived from specific subsamples. Therefore, when the aim is to perform inference for transition intensities or for the transition probabilities it is essential to check if the Markov assumption is tenable.

1.1 Markov assumption

The Markov assumption can be checked by including covariates depending on the history. In the illness-death model, for example, we can examine whether the time spent in the initial state is important on the transition from the disease state (the intermediate state) to death (the absorbing state) or not. Then, one only needs to fit the regression model with that covariate and test the null hypothesis $H_0 : \beta = 0$, against the general alternative $H_0 : \beta \neq 0$. This would assess the assumption that the transition rate from the disease state into death is unaffected by the time spent in the previous state.

Since the landmark methods (LM) of the transition probabilities proposed by de Uña-Álvarez and Meira-Machado (2015), and by Putter and Spitoni (2018) are free of the Markov assumption, they can also be used to introduce such tests (at least in the scope of the progressive multi-state models) by measuring their discrepancy to Markovian Aalen-Johansen estimators (AJ). A log-rank test, for example, can be used to introduce a local test for Markovianity by measuring the discrepancy of the landmark

estimators to the Aalen-Johansen Markovian estimators, for a fixed value $s > 0$. A test based on the areas between the two curves is also introduced to compare the two curves. We propose the use of the following test statistic based on direct nonparametric estimates of the transition probabilities $T = \int_s^\tau (p_{ij}^{LM}(s, u) - p_{ij}^{AJ}(s, u) du)$, where τ is the upper bound of the support of T , and p_{ij}^{LM} and p_{ij}^{AJ} denote the transition probabilities for the landmark and Aalen-Johansen estimators, respectively. Note that if the null hypothesis of Markovianity holds, the value of T should be close to zero. To approximate the distributions of the test statistic, bootstrap methods with a large number of resamples, M , are used. According to large sample asymptotic distribution theory, when M , the number of replicates goes to infinity, we have the following statistic distributed approximately as a standard normal distribution with a mean of 0 and variance of 1: $Z = (\mu_{(T^*)} - 0)/\sigma_{(T^*)} \sim N(0, 1)$. The null hypothesis will be rejected if $Z > Z_{(1-\alpha/2)}$ or $Z < Z_{(\alpha/2)}$.

2 Simulation study

In this section we report results of a simulation study, where the aim is to compare the finite sample performance of three methods to test the Markov assumption in an illness-death model. To simulate the data in the illness-death model, we assume that all the individuals are in the initial state 1 (State 1), at time $t = 0$, and the course of the history is given by two groups of subjects. The individuals who pass through the intermediate state (State 2), at some specific time and those who directly go to the absorbing state (State 3). Transition times from the initial state are generated from the cause-specific hazards given by $h_{12}(t) = 0.29/(t + 1)$ and $h_{13}(t) = 0.024t$ where $t > 0$, denotes the time since the start point. To study the Markov assumption, three different hazards were considered to generate death times for individuals passing through the intermediate state: $h_{23}^1(t) = 0.05$, $h_{23}^2(t) = 0.25(t_{12} + 1)^{-0.8}$ and $h_{23}^3(t) = 0.04 \times \log(t + 1)$, where t_{12} is the transition time to the intermediate event. The first scenario is Markov since the hazard is independent of time, whereas the second is semi-Markov and the third is non-Markov. Censoring times were generated from uniform distributions. Two samples size were considered for each scenario ($n = 250$ and $n = 500$).

Table 1 reports the rejection proportions of the proposed local tests (comparing the transition probabilities $p_{23}(s, t)$) for the fixed values $s = 1$, $s = 2$, $s = 4$ and $s = 8$ with sample sizes $n = 250$ and $n = 500$. These quantities were estimated by the empirical rejection proportions from 1000 trials at the significant level 0.05. This table also reports the global test which is obtained through the fit of a Cox model with covariates depending on the history (time in the previous state).

For all three scenarios the results show that the power of the local tests increase with the sample sizes. The log-rank test revealed to be much con-

TABLE 1. Rejection proportions of the local tests log-rank and the test based on the area for the fixed values $s = 1$, $s = 2$, $s = 4$ and $s = 8$. Rejection proportions for the global test based on the Cox model is also shown.

Model	Markov									
Sample size	250	250	250	250	250	500	500	500	500	500
s	1	2	4	6	8	1	2	4	6	8
Log-rank	0.006	0	0	0	0	0.001	0.001	0	0	0
Area	0.304	0.303	0.308	0.255	0.228	0.096	0.094	0.110	0.107	0.078
Cox	0.052	0.052	0.052	0.052	0.052	0.052	0.042	0.042	0.042	0.042
Model	semi-Markov									
Log-rank	0.452	0.233	0.023	0	0	0.903	0.767	0.179	0.014	0
Area	0.966	0.977	0.884	0.771	0.600	0.993	0.990	0.962	0.830	0.643
Cox	0.894	0.894	0.894	0.894	0.894	0.998	0.998	0.998	0.998	0.998
Model	non-Markov									
Log-rank	0.059	0.03	0.005	0.003	0.001	0.135	0.086	0.035	0.006	0.001
Area	0.705	0.801	0.797	0.760	0.721	0.594	0.743	0.812	0.785	0.703
Cox	0.392	0.392	0.392	0.392	0.392	0.727	0.727	0.727	0.727	0.727

servative obtaining low power for small sample sizes. In general, this method is unable to distinguish the inequality between AJ and LM curves in semi-Markov and non-Markov scenarios for sample sizes of $n = 100$ (results not shown) and $n = 250$. This feature is present even in the case of lower values of s in the semi-Markov scenario. In this scenario, departures between the two curves (obtained for the transition probabilities $p_{23}(s, t)$ from AJ and LM methods) are expected to increase as the difference $t - s$ increases. On the other hand, the bootstrap test based on the areas, is able to identify the differences between survival curves in the semi-Markov and non-Markov scenarios. However this local test also obtains high rejection proportions when the data is generated from a Markov scenario. The results for both local tests greatly improve when considering a sample size of $n = 500$. For completeness purposes, Table 1 also shows the results from the global test based on the Cox proportional hazard model. This approach seems to have better results in the Markov scenario, while obtaining reasonable proportions of rejections in the semi-Markov and non-Markov scenarios.

3 Real data analysis

Our methodology is motivated by the re-analysis of colon cancer from a large clinical trial on Duke's stage III patients. In this study, 929 patients affected by colon cancer were followed from the date of a curative surgery for colorectal cancer until censoring or death from colon cancer. From this total, 468 developed a recurrence and among these 414 died. 38 patients died without recurrence. The rest of the patients (423) remained alive and disease-free up to the end of the follow-up. These data can be viewed as arising from a progressive illness-death model with states 'Alive and disease-free', 'Alive with Recurrence' and 'Dead'.

Figure 1 reports estimated transition probabilities for fixed value of $s = 365$ and $s = 1825$ days (1 and 5 years respectively), along time. Plots shown in the top of the figure (for s equal to one year) show departures between the two Markov-free estimators and the Aalen-Johansen estimator. On the

other hand, the corresponding plot for s equal to 5 years show that all proposed methods behave quite similar.

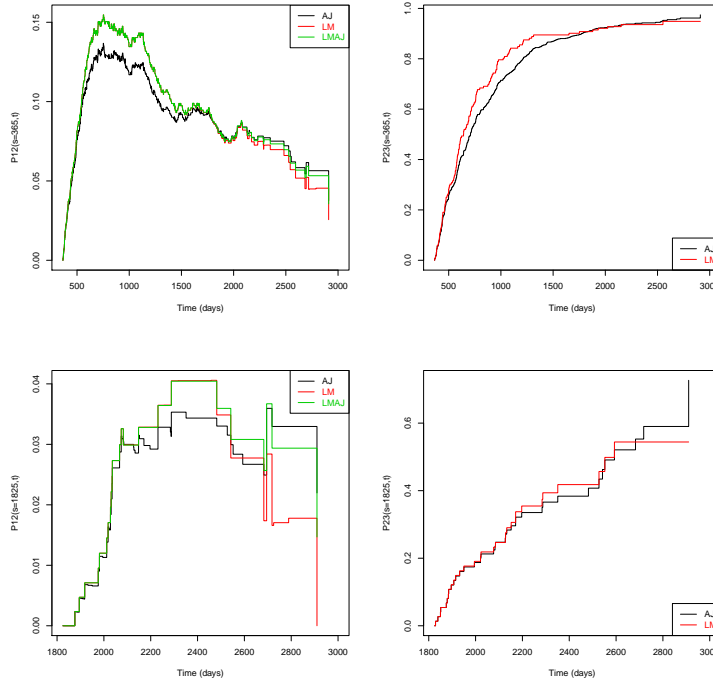


FIGURE 1. Estimates of the transition probabilities for the Aalen-Johansen (AJ) and Markov-free estimators (landmark and landmark Aalen-Johansen). Colon cancer data.

A p-value of 0.154 was obtained from the global test (based on the Cox model) proving no evidence against the Markov model for the colon data. However, a graphical test (not shown here) displaying the two curves with the respective bootstrap confidence intervals reveal differences in the two curves for $s = 365$ (one year). These findings were not confirmed by the local tests based on the log-rank test. Results from this local test confirm its low power to distinguish the differences in the survival curves. For a fixed value of $s = 365$ the log-rank test provided a p-value of 0.279. For a fixed value of $s = 1825$ (five years) the corresponding p-value was 0.996. The bootstrap test based on the area of the two curves lead a p-value of 0.032 for $s = 365$ and 0.326 for $s = 1825$.

Acknowledgments: This research was financed by Portuguese Funds through FCT - “Fundação para a Ciência e a Tecnologia”, within the research grants PTDC/MAT-STA/28248/2017 and PD/BD/142887/2018.

References

- Aalen, O. and Johansen, S. (1978). An Empirical transition matrix for non homogeneous Markov and chains based on censored observations. *Scandinavian Journal of Statistics*, **5**, 141–150.
- Andersen, P.K., Esbjerg, S. and Sorensen, TIA. (2000). Multistate models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine*, **19**, 587–599.
- Andersen, P.K. and Keiding N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, **11**, 91–115.
- Meira-Machado, L., de Uña-Álvarez, J. and Cadarso-Suárez, C. (2006). Non-parametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis*, **12**, 325–344.
- Moreira, A., de Uña-Álvarez, J. and Meira-Machado, L. (2013). Presmoothing the Aalen-Johansen estimator in the illness-death model. *Electronic Journal of Statistics*, **7**, 1491–1516.
- 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.
- Putter, H. and Spitoni, C. (2018). Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator. *Statistical Methods in Medical Research*, **27**:7, 2081–2092.
- Meira-Machado, L. (2016). Smoothed landmark estimators of the transition probabilities. *SORT-Statistics and Operations Research Transactions*, **40**, 375–398.
- Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C., and Andersen, P.K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, **18**, 195–222.