The Journal of Open Source Software

smmR: A Semi-Markov R package

Vlad Stefan Barbu[1], **Florian Lecocq**[1], **Corentin Lothodé**[1], **and Nicolas Vergne**[1]

1 Laboratory of Mathematics Raphaël Salem (LMRS), UMR CNRS 6085, University of Rouen Normandy, France

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 25 March 2022
Published: unpublished

Summary

This package performs parametric and non-parametric estimation and simulation for multi-state discrete-time semi-Markov processes (Barbu et al., submitted). For the parametric estimation, several discrete distributions are considered for the sojourn times: Uniform, Geometric, Poisson, Discrete Weibull of type 1 and Negative Binomial. The non-parametric estimation concerns the sojourn time distributions, where no assumptions are done on the shape of distributions. Moreover, the estimation can be done on the basis of one or several sample paths, with or without censoring at the beginning or/and at the end of the sample paths. Estimation and simulation of discrete-time k-th order Markov chains are also considered.

Semi-Markov models are specified by using the functions `smmparametric()` and `smmnonparametric()` for parametric and non-parametric specifications respectively. These functions return objects of S3 class (`smm`, `smmparametric`) and (`smm`, `smmnonparametric`) respectively (`smm` class inherits from S3 classes `smmparametric` or `smmnonparametric`). Thus, `smm` is like a wrapper class for semi-Markov model specifications.

Based on a model specification (an object of class `smm`), it is possible to:

- simulate one or several sequences with the method `simulate.smm()`;
- plot conditional sojourn time distributions (method `plot.smm()`);
- compute log-likelihood, AIC and BIC criteria (methods `loglik()`, `aic()`, `bic()`);
- compute reliability, maintainability, availability, failure rates (methods `reliability()`, `maintainability()`, `availability()`, `failureRate()`).

Estimations of parametric and non-parametric semi-Markov models can be done by using the function `fitsmm()`. This function returns an object of S3 class `smmfit`. The class `smmfit` inherits from classes (`smm`, `smmparametric`) or (`smm`, `smmnonparametric`).

Based on a fitted/estimated semi-Markov model (an object of class `smmfit`), it is possible to:

- simulate one or several sequences with the method `simulate.smmfit()`;
- plot estimated conditional sojourn time distributions (method `plot.smmfit()`);
- compute log-likelihood, AIC and BIC criteria (methods `loglik()`, `aic()`, `bic()`);
- compute estimated reliability, maintainability, availability, failure rates and their confidence intervals (methods `reliability()`, `maintainability()`, `availability()`, `failureRate()`).

The implemented methods are described in:

- Barbu & Limnios (2008)
- Barbu & Limnios (2006)
- Trevezas & Limnios (2011)

The Journal of Open Source Software

Statement of need

The semi-Markov processes represent a versatile tool that is applied in many fields of science like reliability, survival analysis, bioinformatics, engineering, finance, etc. Few R packages have been developed to handle semi-Markov models or hidden semi-Markov models. For semi-Markov models we have the recent `semiMarkov` R package (Listwon & Saint-Pierre, 2015) that performs maximum likelihood estimation for parametric continuous-time semi-Markov processes, where the distribution can be chosen between Exponential, Weibull or exponentiated Weibull. That package computes associated hazard rates; covariates can also be taken into account through the Cox proportional hazard model. Few R packages are also dedicated to hidden semi-Markov models, implementing estimation and prediction methods. Among them, we can cite the `hsmm` R package (Bulla et al., 2010) and the `mhsmm` R package (O'Connell et al., 2011). The package `SMM` (Barbu et al., 2018) deals with discrete-time multi-state semi-Markov models but does not compute reliability, maintainability, availability and failure rates and was not object oriented.

Acknowledgements

We acknowledge the project AStERiCs *Apprentissage Statistique à l'Echelle pour la Représentation et la Classification non-supervisées* (RIN project funded by the Normandy Region), DAISI on Biomedical Data Classification (co-financed by the European Union with the European Regional Development Fund (ERDF) and by the Normandy Region).

We also acknowledge Mathilde Sautreuil, Caroline Bérard and Dominique Cellier for the help they provided in creating the first working package `SMM` (Barbu et al., 2018).

References

Barbu, V. S., Bérard, C., Cellier, D., Sautreuil, M., & Vergne, N. (2018). SMM: An R package for estimation and simulation of discrete-time semi-Markov models. *R Journal*, *10*(2). https://doi.org/10.32614/rj-2018-050

Barbu, V. S., Lecocq, F., Lothodé, C., & Vergne, N. (submitted). *smmR: An R package for estimation and simulation of discrete-time semi-Markov models.*

Barbu, V. S., & Limnios, N. (2006). Empirical estimation for discrete time semi-Markov processes with applications in reliability. *Journal of Nonparametric Statistics*, *18*(4), 483–498. https://doi.org/10.1080/10485250701261913

Barbu, V. S., & Limnios, N. (2008). *Semi-Markov chains and hidden semi-Markov models toward applications* (Vol. 191). Springer New York. https://doi.org/10.1007/978-0-387-73173-5

Bulla, J., Bulla, I., & Nenadić, O. (2010). hsmm - an R package for analyzing hidden semi-Markov models. *Computational Statistics & Data Analysis*, *54*(3), 611–619. https://doi.org/10.1016/j.csda.2008.08.025

Listwon, A., & Saint-Pierre, P. (2015). SemiMarkov: An R package for parametric estimation in multi-state semi-Markov models. *Journal of Statistical Software*, *66*(6), 784. https://doi.org/10.18637/jss.v066.i06

O'Connell, J., Højsgaard, S., & others. (2011). Hidden semi Markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software*, *39*(4), 1–22.

Trevezas, S., & Limnios, N. (2011). Exact MLE and asymptotic properties for nonparametric semi-Markov models. *Journal of Nonparametric Statistics*, *23*(3), 719–739. https://doi.

org/10.1080/10485252.2011.555543