ANALYSING LIFE HISTORY CALENDAR DATA: A METHODOLOGICAL COMPARISON

Mervi Eerola[1], Satu Helske[2]

[1] Department of Mathematics and Statistics,
FIN-20014 University of Turku, FINLAND, mervi.eerola@utu.fi,
[2] Methodology Centre for Human Sciences/Department of Mathematics and Statistics,
P.O.Box 35, FIN-40014 University of Jyväskylä, FINLAND, satu.helske@jyu.fi

ABSTRACT

The life history calendar, also called an event-history calendar, is a data-collection tool for obtaining reliable retrospective data about life events. The advantage of a life history calendar is that the order and proximity of important transitions in multiple life domains can be studied at the same time.

To illustrate the analysis of such data, we compare the model-based probabilistic event history analysis and a more recent type of approach of model-free data-mining, sequence analysis. The latter is well known in bioinformatics in the analysis of protein or DNA sequences. In life course analysis it is less familiar but has provided novel insight to the diversity of life trajectories and their relationship to life satisfaction. We emphasize the differences, but also the complementary advantages of the methods.

In event history analysis, we consider the data generated by a marked point process $(T_n, X_n)_{n \geq 1}$, a time-ordered sequence of points or events, characterised by pairs of random variables, the occurrence times $T_1, T_2, ...$ and marks $X_1, X_2, ...$ describing what happens at a particular T. Instead of transition hazards, we estimate the cumulative prediction probabilities of a particular life event in the entire observed trajectory, given the history of the marked point process. This way of combining information in multi-state event history models has been called 'survival synthesis'. The innovation gain from observing a life event at a particular age, related to the prediction of another life event, can be quantified and monitored visually.

In sequence analysis, we compare several dissimilarity measures between the life sequences, either assuming independence or using some *ad hoc* definition of dependence between the sequence elements. We also contrast data-driven (estimated) and user-defined costs of substituting one sequence element with another.

As an example, we study young adults' transition to adulthood as a sequence of events in three life domains (partnership, parenthood and employment). The events define the multi-state event history model and the parallel life domains in the multidimensional sequence analysis.

We conclude that the two approaches complement each other in life course analysis; sequence analysis can effectively find typical and atypical life patterns while event history analysis is needed for causal inquiries.

1. REFERENCES

[1] Avshalom Caspi, Terrie E. Moffitt, Arland Thornton, Deborah Freedman, et al., "The life history calendar: A research and clinical assessment method for collecting retrospec-

tive event-history data.," *International Journal of Methods in Psychiatric Research*, vol. 6, no. 2, pp. 101–114, 1996.

[2] M. Eerola, *Probabilistic causality in longitudinal studies*, vol. 92 of *Lecture Notes in Statistics*, Springer-Verlag, 1994.

[3] Jacques-Antoine Gauthier, Eric D. Widmer, Philipp Bucher, and Cédric Notredame, "Multichannel sequence analysis applied to social science data," *Sociological Methodology*, vol. 40, no. 1, pp. 1–38, 2010.

[4] I. Tabus J. Helske, M. Eerola, "Minimun description length based hidden markov model clustering for life sequence analysis," in *Proc. Third WITMSE Conf.*, Tampere, Finland, Aug. 2010.

[5] G. Pollock, "Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis," *J. R. Statist. Soc. A*, vol. 170, pp. 167–183, Dec. 2007.