Frans Willekens
NIDI

Bridging the divide between sequence analysis and event history analysis

Aart helped us better understand human lives. The life course has been central to his research. Aart argued that the life events people experience, the life choices they make, the resources they accummulate, the opportunities and other benefits they receive from social networks and the welfare state, and the social inequalities in society should be studied in a life-course perspective and as outcomes of interactions between agency and structure.

The diversity of life courses and the determinants of life paths are important areas of study. Aart's research on the de-standardization of the life course is widely cited. Sequence analysis is the method of choice for analysing life-course trajectories. Ritschard and Studer (2018, p. 4) state that sequence analysis essentially remains exploratory and needs to be complemented with models that generate the sequences. In his introduction to a special issue of Advances in Life Course Research on methods of life course analysis, Aart calls for bridging the divide between sequence analysis (SA) and event-history analysis (EHA) (Liefbroer 2019, p. 2). Liao et al. (2022, p. 16), reflecting on four decades of sequence analysis in the social sciences, call for probabilistic models that are able to generate event sequences and build complete synthetic life trajectories. These probabilistic models are the subject of this contribution. I cover basic design principles and omit technical detail.

The life course is conceptualized as a sequence of states and transitions between states. The transitions are conveniently organized in an origin-destination matrix. States can be observable or latent. Aart published a model with latent states (Han et al. 2020). In this short contribution, states are observable. The central idea is that state and event sequences are outcomes of multistate stochastic processes. Several multistate processes can be specified depending on the factors and the dynamics impacting the trajectory. A widely used stochastic process is the Markov process. It assumes that the future of the process is independent of its past provided that the present state is known and the past is summarized in the present. In a Markov process, time is continuous and the sojourn time (duration) in a state follows an exponential distribution. The parameter of the exponential distribution, the rate of transition, usually differs by state of origin and state of destination, and may also differ by age (and other relevant factors). The transition rates are estimated from observed event sequences, which may be incomplete. Once the transition rates are known, event sequences can be generated using microsimulation. The simulated event sequences are synthetic because they are based on data from different individuals. If the number of simulated sequences is sufficiently large, some will be exact copies of observed sequences.

The estimation of transition rates requires tracking (a) the transitions and (b) the population at risk. The rate of transition from state i to state j is estimated by dividing the observed number of transitions ("occurrences", "counts") during a time interval by the total exposure time during that interval. The exposure time is the total time individuals at risk spend in state i during the interval. The estimate of the transition rate is known as occurrence-exposure rate. Tracking individual exposure times can be tricky because an individual may enter and leave state i at any time and may visit the state multiple times during an interval. The tracking of individual exposure times is illustrated by Willekens (2014), using the subsample of the German Life History Study used by Blossfeld and Rohwer (2002).

An alternative is to reduce the length of the time interval. When the length tends to zero, time becomes continuous. Suppose a transition from i to j is observed in the small interval from t to t+dt and no other transition occurs in that interval. Let Y(t) denote the number of individuals at risk at t. The estimator of the transition (hazard) rate at t is 1/Y(t). Summing the estimated rates of (i,j)-transition since the start of the observation is the non-parametric estimator of the cumulative hazard and is known as the Nelson-Aalen estimator. Its variance is the sum of the squared rates of (i,j)-transitions since the start of the observation (Aalen et al. 2008, p. 72). The estimator is widely used in the modelling of life histories. For the statistical theory, see Aalen et al. (2008) and Cook and Lawless (2018).

Several software packages exist for the estimation of transition rates in multistate models (Willekens and Putter 2014).

Bridging the divide between SA and EHA is a worthy challenge and contributes to a better understanding of human lives.

References

Aalen, O., Borgan, O., & Gjessing, H. (2008). Survival and event history analysis: A process point of view. *Statistics for Biology and Health.* New York: Springer.

Blossfeld, H.-P., & Rohwer, G. (2002). Techniques of event history modeling: New approaches to causal analysis (Second Edition.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Cook, R. J., & Lawless, J. F. (2018). Multistate models for the analysis of life history data. Boca Raton, Florida: Chapman and Hall / CRC Press.

Han, S. Y., Liefbroer, A. C., & Elzinga, C. H. (2020). Mechanisms of family formation: An application of Hidden Markov Models to a life course process. *Advances in Life Course Research, 43.*

Liao, T. F., Bolano, D., Brzinsky-Fay, C., Cornwell, B., Fasang, A. E., Helske, S., et al. (2022). Sequence analysis: Its past, present, and future*. Social Science Research*, *107*.

Liefbroer, A. C. (2019). Methodological diversity in life course research: Blessing or curse? *Advances in Life Course Research*, *41*, 1–3.

Ritschard, G., & Studer, M. (2018). Sequence analysis: Where are we, where are we going? In G. Ritschard & M. Studer (Eds.), Sequence analysis and related approaches: Innovative methods and applications, 1-11. Cham: Springer.

Willekens, F. (2014). Multistate analysis of life histories with R. Cham: Springer.

Willekens, F., & Putter, H. (2014). Software for multistate analysis*. Demographic Research*, *31*, 381–420.