Understanding product integration

Jan Beyersmann[1,2], Arthur Allignol[1,2] and Martin Schumacher[2]

[1] Freiburg Centre for Data Analysis and Modelling, University of Freiburg, Eckerstraße 1, 79104 Freiburg, Germany, `jan.beyersmann@fdm.uni-freiburg.de`

[2] Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Stefan-Meier-Straße 26, 79104 Freiburg, Germany

Abstract. Product integration is a very powerful, but somewhat neglected topic in applied survival analysis: Survival data are usually incompletely observed, the most important example being independent right-censoring. This leads to survival analysis being based on hazards, because the hazard of seeing an event is undisturbed by censoring. The Kaplan-Meier estimator of the survival function is a finite product over one minus empirical hazards, and it approaches e to the negative true cumulative hazard. This result is not very intuitive, but it is much better understood using product integration: A product integral is a 'continuous time product', like a usual integral is a 'continuous time sum'. The product integral over one minus the true hazard is a 'product' over infinitesimal conditional survival probabilities, and therefore equal to the survival probability. The product integral over one minus the empirical hazard equals the Kaplan-Meier estimator. The 'e to the negative true cumulative hazard'-formula is then seen to simply be the solution of a product integral. We explore these connections in R, where one function `prodint` both approximates the true survival function arbitrarily close and results in the Kaplan-Meier estimate when applied to data based empirical hazards.

Both theory and the R implementation generalize to the matrix-valued case that is important for multistate models: Here one individual may experience a possibly random number of events, e.g. transitions between 'healthy' and 'ill' before dying. Closed formulae for the transition probabilities will in general not be available anymore, but a matrix-valued version of `prodint` may still be used for numerical approximation. It also results in the Aalen-Johansen estimator of the matrix of transition probabilities, a generalization of the Kaplan-Meier estimator, when applied to empirical transition hazards. Empirical transition hazards can be obtained in R using the `mvna`-package.

References

Gill, R and Johansen, S (1990): A survey of product-integration with a view towards application in survival analysis. *Ann Stat, 18, 1501–1555.*