Handling sampling design in Bayesian multistate life table estimation

Scott M. Lynch, Princeton University
J. Scott Brown, Miami University

Over the last several years, we have developed two Bayesian approaches to multistate life table estimation that allow the construction of interval estimates of life table quantities. One method requires panel data; the other requires independent cross-sectional data for mortality and health. Both methods involve Gibbs sampling, and both implicitly assume the sample data are from a simple random sample, which is not the case with most panels or cross-sections. Here, we investigate the implications of ignoring sample design. We ask (1) whether sample design influences interval estimates to a substantively meaningful extent, and (2) whether the Bayesian approach can be adapted to handle design. We find that sample design increases interval estimates only slightly in most cases. We also show how the Bayesian approach can be easily adapted to incorporate most design effects via a bootstrap. We describe the procedure and give an applied presentation of its implementation.

Slightly Extended abstract

Over the last 5 years, we have developed two Bayesian approaches to multistate life tables estimation which allow the construction of interval estimates of life table quantities for any desired covariate profile. Our original method (see Lynch and Brown, *Sociological Methodology*, 2005) requires panel data and involves using Gibbs sampling of parameters from a discrete time bivariate dichotomous probit model capturing transitions between states between waves of the panel study. The new method (currently under review) is an extension of Sullivan's method and requires independent cross-sectional data sets for mortality and health. The mortality data can be vital statistics rate data, while the health data should come from a cross-sectional survey. This method also involves Gibbs sampling of parameters, but from a bivariate dichotomous probit model that has been altered somewhat to handle (1) known probabilities for mortality rather than dichotomous indicators, and (2) covariates measured at different levels of disaggregation for health and mortality data.

Under both approaches, once the Gibbs sampling is complete, the parameter samples are then combined with a desired set of covariate values, and age-specific predicted transition probability matrices are produced for each Gibbs sample using integral calculus. The result is a collection of age-specific transition matrices that can be used to generate a distribution of multistate life tables. Quantities from these life tables can then be summarized by constructing basic summary measures, including confidence intervals.

Both methods implicitly assume that the sample data are from a simple random sample, which is not the case with the vast majority of public use panels or cross-sections. In this paper, we investigate the implications of ignoring sample design issues. Specifically, we ask (1) whether sample design influences interval estimates at a substantively meaningful level, and (2) whether the Bayesian approach can be adapted to compensate for design.

In order to answer (1), we use bootstrapped parameter estimates from probit models in Stata that adjust for design (e.g., the svy approach) and compare multistate life table results from that approach to results from our original approaches. We find that sample design increases interval estimates only slightly, especially when the estimate of interest is the proportion of remaining life to be spent in different states. When years are the measure of interest, interval estimates are affected to a slightly greater degree, but they are still not substantially larger than unadjusted intervals.

We also show how the Bayesian approach can be easily adapted to incorporate most design issues via the use of a weighted bootstrap. Under the weighted bootstrap, at each step of the Gibbs sampler, the data are resampled using sample weights as inverse probabilities of inclusion into the bootstrap sample. After a bootstrap sample is obtained, the Gibbs sampler proceeds as usual. All in all, this approach requires only a slight adjustment to our original approaches.

We describe the procedure and give a very applied presentation of its implementation via freely available software.