Multilevel Discrete Time Survival Analysis

Tihomir Asparouhov and Bengt Muthén

March 11, 2021

1 Discrete Time Survival Analysis

Discrete time survival analysis is used to model survival variables T that take only integer values $1, 2, ..., L$. If C is a censoring indicator, then if $T_i = t$ and $C_i = 1$, individual i survived until time t but was not observed beyond that point. if $T_i = t$ and $C_i = 0$, individual i died at time t.

The data used for the analysis is represented by binary indicators U_{ij} where $i = 1, ..., N$ and $j = 1, ..., L$ (N is the sample size)

$$U_{ij} = \begin{cases} 1 \text{ if } T_i > j \text{ or } T_i = j, C_i = 1 \\ 0 \text{ if } T_i = j, C_i = 0 \\ * \text{ if } T_i < j \end{cases}$$

where $*$ represents the missing value symbol. The interpretation of U_{ij} is that when the binary indicator is 1, individual i survived period j; when it is 0, individual i died in period j; and when it is missing, individual i was not observed in period j. A basic survival model that estimates the survival probabilities in each period is then given by the equation

$$P(U_{ij} = 1) = \frac{1}{1 + Exp(\tau_j)}.$$

If we add a covariate X to this model, we can estimate the discrete time survival model where the effect of X varies with time

$$P(U_{ij} = 1) = \frac{1}{1 + Exp(\tau_j - \beta_j X_i)}.$$

Mplus user's guide example 6.19 shows another discrete time survival model.

Figure 1: Multilevel Discrete Time Survival Analysis

```
VARIABLE:  NAMES ARE u1-u4 x c;
   CATEGORICAL = u1-u4;
   MISSING = ALL (999);
   cluster=c;
   within=x;
ANALYSIS: ESTIMATOR = MLR; type=twolevel;
MODEL:
%within%
u1-u4 on x;
%between%
f BY u1-u4@1; f*1;
```

2 Multilevel Discrete Time Survival Analysis

When individuals are grouped in clusters, we can estimate the following multilevel discrete time survival analysis

$$P(U_{ijk} = 1) = \frac{1}{1 + Exp(\tau_j - f_k - \beta_j X_{ik})}$$

where U_{ijk} is the binary indicator for individual i in cluster k for period j and f_k is a cluster-specific frailty random effect. Such a model can be specified in Mplus as in Figure 1. The loadings on the between level can be freed as well. This will create a time-specific effect for the cluster-specific frailty f_k

$$P(U_{ijk} = 1) = \frac{1}{1 + Exp(\tau_j - \lambda_j f_k - \beta_j X_{ik})}.$$

Furthermore, the parameters τ_j can be estimated as cluster specific random parameters. To obtain such a model the variances of U_1,..., U_4 would be estimated on the between level.