FITTING A CONTINUOUS-TIME MULTIPROCESS EVENT HISTORY MODEL FOR MARITAL SEPARATION AND MARITAL BIRTHS IN AML

Fiona Steele
May 2005

1. Introduction

In this document we describe how a continuous-time multiprocess event history model can be fitted in aML. Details of the data structure required by aML and syntax for fitting the model are given. However, anyone who plans to use aML is referred to the User Guide for fuller explanations and further examples. Readers may also find the software review by Steele (2004) useful.

2. Hierarchical data structure

Some women have more than one marriage, leading to a two-level structure with marriages nested within women. In aML the levels are numbered starting at the highest level, so we have women at level 1 and marriages at level 2. The birth data have a further level of nesting because some women have more than one birth in a marriage, so we have births at level 3.

3. Multiprocess event history model

Denote by $h_{ij}^{S}(t)$ the hazard of marital separation at duration *t* for marriage *j* of woman *i*, and by $h_{ijk}^{B}(t)$ the hazard of marital birth *k*.

The multiprocess model for marital separation and marital childbearing can be written:

$$\ln h_{ij}^{S}(t) = \alpha^{S} D_{ij}^{S}(t) + \beta^{S} X_{ij}^{S}(t) + u_i^{S}$$
$$\ln h_{ijk}^{B}(t) = \alpha^{B} D_{ijk}^{B}(t) + \beta^{B} X_{ijk}^{B}(t) + u_i^{B}$$

where the baseline log-hazard function is $\alpha^{S} D_{ij}^{S}(t)$ for marriage duration and $\alpha^{B} D_{ij}^{B}(t)$ for the birth interval. In aML these are specified as piecewise-linear splines. Continuously time-varying covariates such as age can be treated in the same way. Covariates for the separation and birth processes are denoted by $X_{ij}^{S}(t)$ and $X_{ijk}^{B}(t)$ respectively; these can be defined at any level in the hierarchy (woman, marriage or birth). In the case of the marital separation process, we have one time-varying covariate which is a count of the number of children; this can be thought of as a variable at the birth level. Woman-level random effects u_i^{S} and u_i^{B} allow for unobserved heterogeneity between women (and correlation between durations to events for the same woman). The separation and birth equations are linked by allowing these random effects to be correlated. We assume that the random effects follow a bivariate normal distribution with zero mean. In aML we estimate the standard deviation of each random effect and their correlation.

4. Data description

The data consist of women from the National Child Development Study (NCDS) who have been married at some point between the ages of 16 and 42. For this illustration, we use a subsample of 937 of the eligible women. These women contribute 1065 marriages and 1764 marital births. We consider five covariates: previous marriage, previous cohabitation, age at marriage, education at marriage and number of children. The two partnership history variables are defined at the level of the marriage. For the purposes of this exercise, we also treat age and education as marriage-level variables. We consider one time-varying covariate whose value can change during a marriage, number of children.

The dataset **amlex.raw** contains the following variables:

WOMAN	Woman identifier
NBINT	Number of birth intervals within marriage (equals dimension of arrays TBIRTH(i) etc.)
MNO	Marriage number within woman
MDURLO	Marriage duration in months (lower limit) – see below for explanation
MDURHI	Marriage duration in months (upper limit) – see below for explanation
SEP	Marital separation indicator (1=yes, 0=no)
PRM	Previously married (1=yes, 0=no)
PRC	Previously cohabited with any partner (1=yes, 0=no)
AGE	Age at start of current marriage (1=less than 20, 2=20-24, 3=25-29, 4=30+)
FTED	Number of years of post-16 education at start of marriage (1=none, 2=1-2 years, 3=3-5 years, 4=6+ years)
TBIRTH(i)	Duration of marriage at birth of child *i*
NKID(i)	Total number of children before birth of child *i* (including any from previous relationships)
BIRTH(i)	Birth indicator (1=birth, 0=censored)
BINTLO(i)	Birth interval in months; duration between birth *i* and *i*-1 or marriage if first birth (lower limit)
BINTHI(i)	Birth interval in months; duration between birth *i* and *i*-1 or marriage if first birth (upper limit)

5. aML data input

aML will read data files in ascii format only and the data must be structured in a particular way before they can be imported. Our data have a three-level structure with births (level 3) nested within marriages (level 2) within women (level 1). There are two ways that we could structure the data for aML. One option is to have one record per woman with variables relating to marriages, and births within them, written out as arrays. However, it is more convenient to organise the data so that there is one record per marriage, with variables relating to births stored as arrays. This is best illustrated with an example. We will consider the records for two women in the dataset.

Woman	Mno	Mardur	Sep	Nkid1	Birth1	Tbirth1	Nkid2	Birth2	Tbirth2	Nkid3	Birth3	Tbirth3
1	1	129	0	1	1	25	2	0	129	.	.	.
5	1	20	1	0	0	20	.	.	.	.	.	.
5	2	160	0	0	1	34	1	1	68	2	0	160

NKID1 is the number of children born before the start of the marriage, BIRTH1 indicates whether the first birth interval is censored or closed, TBIRTH1 is the duration of the marriage at the birth of the first child (or interview if BIRTH1=0). If BIRTH1=1, then NKID2=NKID1+1. BIRTH2 and TBIRTH2 are the censoring indicator and duration of marriage at the second birth. The birth interval durations are calculated from TBIRTH1, TBIRTH2 etc. For example, the first birth interval is equal to TBIRTH1, and the second birth interval is the difference between TBIRTH2 and TBIRTH1.

The first woman has only one marriage which is in progress at interview (SEP=0). She already had one child at the start of the marriage (NKID1=1), which could have been fathered by her husband before marriage or by a previous partner. Then in the 25th month of marriage (TBIRTH1=25) she had a second child (NKID2=2). She had no other children during the marriage, so the second birth interval is censored (BIRTH2=0) at interview.

The second woman in the example (WOMAN=5) has had two marriages. The first marriage ended at the 20th month (SEP=1) and the second is in progress at interview (SEP=0). At the start of the first marriage, she had no children (NKID1=0) and none were born during the marriage so the first birth interval is open (BIRTH1=0). In the second marriage, she had two children; the first was born in the 34th month of marriage, and the second was born in the 68th month. The third birth interval, the duration between the second and third birth, is censored by the interview (BIRTH3=0).

The ascii file **amlex.raw** has the above structure. An additional requirement is that the variables are in the following order. The first two variables must be the woman ID and a variable indicating the dimension of the birth variable arrays (2, 1 and 3 for the above records). These are called control variables and are not declared in the data entry syntax file. The remaining variables must be in order of the level at which they are defined, starting with level 1 (woman), and followed by level 2 (marriage) and level 3 (birth).

For hazards models, there is one further data requirement. A duration must be denoted by two variables, corresponding to the lower and upper limits of the interval in which an event occurred. For example, suppose we know that a marriage ended in May 1990. The marriage could have ended any time between 1 May and 31 May, so the difference between the upper and lower limits of the interval equals one month. For censored durations, the lower and upper limits must be equal.

The syntax file **amlex.r2a** contains the syntax for reading the ascii file **amlex.raw** into aML. It contains the following commands:

```
ascii data file = amlex.raw;

level 2 var = mno mdurlo mdurhi sep prm prc age fted;
level 3 var = tbirth nkid birth bintlo binthi;
```

The file **amlex.r2a** is executed using the raw2aml program, which produces the aML data file **amlex.dat**.

After running raw2aml, the following summary of the dataset is given:

```
Number of observations:     937
Maximum number of level 2 branches in any observation:      4
Maximum number of level 3 branches in any observation:     15
Maximum number of level 3 branches in any level 2 branch:  15

------------------------------------------------------------

LEVEL 1 VARIABLES:
Variable     N       Mean    Std Dev        Min        Max
_id        937      469.0   270.6329        1.0      937.0

LEVEL 2 VARIABLES:
Variable     N       Mean    Std Dev        Min        Max
mno       1065     1.1277   .3583495        1.0        4.0
mdurlo    1065   163.4967   82.45096        1.0      313.0
mdurhi    1065   163.7709   82.21991        1.0      313.0
sep       1065   .2741784   .4463089        0.0        1.0
prm       1065   .1201878   .3253338        0.0        1.0
prc       1065   1.538028   .8264369        1.0        4.0
age       1065   2.266667   .8464136        1.0        4.0
fted      1065   1.603756   .8750859        1.0        4.0

LEVEL 3 VARIABLES:
Variable     N       Mean    Std Dev        Min        Max
tbirth    2829   98.22411   79.21545        1.0      313.0
nkid      2829   1.243549    1.21971        0.0       14.0
birth     2829   .6235419   .4845827        0.0        1.0
bintlo    2829   62.27642   56.47104        1.0      282.0
binthi    2829   62.89996    56.1921        1.0      282.0

------------------------------------------------------------

NOTE: there is variation in all data variables.
```

The number of observations refers to the number of level 1 units (women), and the maximum number of level 2 branches in any observation is the maximum number of marriages per woman. The variable _id is the control variable WOMAN.

6. aML syntax for multiprocess model

The syntax in file **amlex.aml** specifies a multiprocess model. Readers are referred to the User Guide for full descriptions of the commands but, briefly, the following elements of the model need to be declared:

- Splines for continuously-varying variables, here the durations of marriages and birth intervals. We must also specify the nodes (or knows) for the splines. The parameters to be estimated are the slopes for each of linear piece of a spline.
- Covariates (regressor sets) for each process. Here we include the same covariates in the marital separation and birth equations.

- Random effects. We include a random effect in each equation and assume that they follow a bivariate normal distribution.
- Specify the hazards model for marital separation. Note that for the time-varying covariate, NKID, we must specify the durations ('timemarks') at which the covariate changes value. The 'draw' option in 'intres' tells aML that the random effects declared earlier are at the woman (_id) level.
- Specify the hazards model for marital births.
- Give starting values for each parameter, starting with the spline parameters and ending with the random effect parameters. The starting values given here are from fitting a separate hazards model for each process, i.e. constraining the random effect correlation to equal zero. It is likely that convergence problems will be encountered if poor starting values are provided. To avoid this, it is advisable to start with simple models, building up to the full model gradually.

```
dsn=amlex.dat;

define spline Durmar; node=24 48 72 120;
define spline Durbir; node=24 36 48 60;

define regressor set MarX;
  var = 1 fted==2 fted==3 fted==4 age==2 age==3 age==4 prm prc nkid;
define regressor set BirX;
  var = 1 fted==2 fted==3 fted==4 age==2 age==3 age==4 prm prc nkid;

define normal distribution; dim=2; number of integration points=4;
  name=u1;
  name=u2;

hazard model;
  censor=(sep==0);
  duration=mdurlo mdurhi;
  timemarks=tbirth;
  model=durspline(origin=1,ref=Durmar)+regset MarX+intres(draw=_id,ref=u1);

hazard model;
  censor=(birth==0);
  duration=bintlo binthi;
  model=durspline(origin=1,ref=Durbir)+regset BirX+intres(draw=_id,ref=u2);

starting values;
m1-24       T    .03668093301
m24-48      T    .01223780974
m48-72      T   -.00869942068
m72-120     T    .00969131168
m120+       T   -.00398903679
b1-24       T    .08646769103
b24-36      T    0.0042266794
b36-48      T   -.02910625354
b48-60      T   -.01195173519
b60+        T   -.02031847719
mcons       T   -6.9298459798
mfted2      T   -0.2084728966
mfted3      T   -.37209696238
mfted4      T   -0.6484922208
mage2       T   -.72050600795
mage3       T   -1.2613801056
mage4       T   -1.1991930977
mprm        T    .04668891847
```

```
mprc        T    .11228513718
mnkid       T   -.19245470217
bcons       T   -4.9181917462
bfted2      T   -.16145241951
bfted3      T   -.22223683694
bfted4      T    .03220286175
bage2       T   -.40821334642
bage3       T   -.36751018281
bage4       T   -.57781510314
bprm        T    .42482561255
bprc        T    .03581181215
bnkid       T   -.70901524659
Sigu1       T    1.0872609208
Sigu2       T    .41610642911
Rho         T     0
;
```

References

Lillard, Lee A. and Constantijn W.A. Panis (2003) *aML Multilevel Multiprocess Statistical Software, Version 2.0*. EconWare, Los Angeles, California.

Steele, F. (2004) A review of aML (Release 2.0). Centre for Multilevel Modelling. Downloadable from www.mlwin.com/***