ELECT: Estimation of life expectancies using continuous-time multi-state survival models[1]

Vignette ELECT version 0.2

Ardo van den Hout
Department of Statistical Science, University College London, UK

June 2016

Abstract

ELECT is a set of functions in R to compute state-specific and marginal life expectancies on the basis of a fitted continuous-time multi-state model that includes an absorbing dead state. Multi-state survival models can be used to describe, understand, and predict health-related processes over time. Life expectancy in a specified state is defined as the expected remaining number of years in that state and is conditional on current age. The multi-state survival model for panel data is estimated using the R package msm with age as the time-scale. Estimation of life expectancies is explained and illustrated using the functions in ELECT.

[1]If you use ELECT, please cite: Van den Hout, A. (2016) ELECT: Estimation of life expectancies using continuous-time multi-state survival models. ELECT version 0.2. Vignette.

1 Introduction

Multi-state models can be used to describe health-related stochastic processes over time. To formulate the model, a finite number of health states are defined and potential transitions between these states are distinguished. A multi-state model that includes a dead state is called a multi-state survival model. For such a model it is of interest how total life expectancy at a given age subdivides into life expectancies in the living states. As an example, consider for an older population the three-state illness-death model defined by a healthy state, an ill-health state, and the dead state. For an individual at a specified age, we can distinguish two residual life expectancies, namely expected remaining time spent in the healthy state and expected remaining time spent in the ill-health state. The sum of these expectancies make up the total residual life expectancy.

Important methodological work on continuous-time multi-state models for panel data is presented in Kalbfleish and Lawless (1985), Kay (1986), and Satten and Longini (1996). Continuous-time models are based on theory for continuous-time Markov chains as discussed in, for example, Cox and Miller (1965). Jackson (2011) presents the freely available R package msm that provides a flexible framework for fitting continuous-time multi-state models to panel data.

We will use msm to fit multi-state survival models and we will use a collection of functions in R called ELECT to estimate life expectancies. ELECT was written by the author and the acronym stands for Estimating Life Expectancies using Continuous Time. The aim of this note is to explain and illustrate ELECT. Additional information on multi-state survival models and the estimation of life expectancies can be found in Van den Hout (2016).

Section 2 illustrates estimating life expectancies using a multi-state survival model with two living states and a dead state. In this model, a transition from state 2 back to state 1 is possible. In Section 3, an illustration is given for a model without backwards transitions. Section 4 discusses an application with four states. These three sections should get the user started. Section 5 discusses the details of the estimation in ELECT and Section 6 provides the complete specification of the functions in ELECT.

The software environment R is free as is the package msm. Both can be downloaded from `www.r-project.org`. ELECT, the simulated data that are used in this report, and text files with the command lines can be downloaded from the author's website: `www.homepages.ucl.ac.uk/~ucakadl/`.

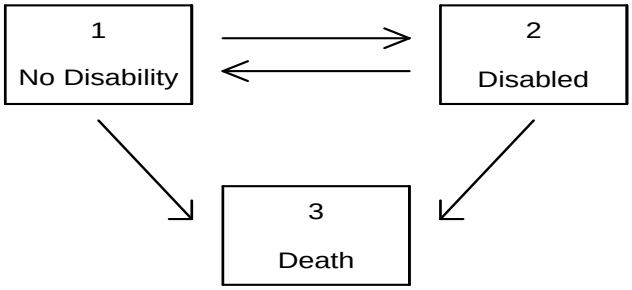

Figure 1: Model for disability in the MRC Cognitive Function and Ageing Study.

2 Estimating life expectancies: Example I

2.1 Data

A three-state model for disability and survival is considered. State 1 is defined as the disability-free state, state 2 is the disabled state, and state 3 is the dead state; see Figure 1. Simulated data are used, which mimic the observed trajectories of men in the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS, `www.cfas.ac.uk`, Brayne et al. 2006). CFAS data is not publicly available but can be requested via the CFAS website. Simulated data are used in what follows to protect the privacy of the CFAS respondents. Data were simulated as follows. Firstly, a multi-state model was fitted to CFAS data and, secondly, estimated model parameters were used to simulate 2000 individual trajectories with a follow-up of 15 years. All individuals are 65 years or older at the time of the first observation. These simulated data are provided for illustration purposes only.

The longitudinal panel data format for the three-state model is one row per observation. An example for an individual is given by

```
id state   age ybirth
 3     1 -0.50     19
 3     1  1.50     19
 3     1  3.50     19
 3     1  5.50     19
 3     2  7.50     19
 3     3  8.64     19
```

The identifier is `id`. The variable for state is `state`, where state 1 and 2 are living states, and state 3 is the dead state. Age (centred by minus 75) is time-dependent and denoted by `age`. For the fitting of the model it is essential that consecutive records for one individual do not contain the same age. This would imply that no time has passed between the two observations. For this reason rounding age to whole years is not recommended. Ideally, date of birth is available and age can be calculated to the required precision. Another method would be to fix rounded age at baseline of the study and add the study time (which is often given in more detail) to obtain time-dependent age. Year of birth (minus 1900) is `ybirth`. Year of birth is transformed for numerical reasons.

Another example of data for an individual is

```
 id state  age ybirth
  4     1 -2.3     21
  4     2 -0.3     21
  4     1  1.7     21
  4     1  3.7     21
  4     1  5.7     21
  4     1  7.7     21
  4    -2  9.7     21
```

Here the last state is right-censored at the end of the follow-up (denoted by the value -2). Right censoring here implies being alive but in an unknown living state.

For the three-state model with transitions $1 \to 2$, $1 \to 3$, $2 \to 1$, and $2 \to 3$, we specify the loglinear model for the transition intensities (hazards) by

$$\log[q_{rs}(\texttt{age})] = \beta_{rs.0} + \beta_{rs.1}\texttt{age} + \beta_{rs.2}\texttt{ybirth}, \tag{1}$$

where $(r, s) \in \{(1,2), (1,3), (2,1), (2,3)\}$.

2.2 Estimation

An example is given how to fit a three-state model in R using msm and how to estimate life expectancies using ELECT.

Change the working directory in R to the one with the data and ELECT. Load the data for the three-state model using

```
> load("dataExample1.RData")
```

Next load the library msm and print some descriptive statistics for the data:

```
> library(msm)
> cat("Sample size:")
> print(length(table(data$id)))
> cat("Frequencies observed state:")
> print(table(data$state))
> cat("State table:"); print(statetable.msm(state,id,data=data))
    Sample size:[1] 2000

    Frequencies observed state:
      -2    1    2    3
     582 6077 2094 1418

    State table:    to
    from   -2    1    2    3
       1  379 4153  938  607
       2  203  119  961  811
```

For the model fitting, we first specify the initial value of the $\mathbf{Q}$-matrix:

```
>  q <- 0.001
>  Q <- rbind(c(0,q,q), c(q,0,q),c(0,0,0))
```

This definition of $\mathbf{Q}$ defines the transitions $1 \to 2$, $1 \to 3$, $2 \to 1$, and $2 \to 3$, where state 3 is the dead state. The specification defines the transitions that are possible according to the model and also provides the starting values for the transition intensities in the maximum likelihood estimation in msm.

Next the three-state model is fitted using msm in R with the command

```
>   model <- msm(state~age, subject=id, data=data, center=FALSE,
          qmatrix=Q, death=TRUE, covariates=~age+ybirth,
          censor= -2, censor.states=c(1,2), method="BFGS",
          control=list(reltol=1e-8, maxit=1000, fnscale=100000))
```

The specification `censor=-2` defines the denotation for the censored state, and `censor.states= c(1,2)` specifies the underlying states of the censored states. The maximisation will take about 10 minutes.

To use ELECT, there are prerequisite elements in the `msm`-call: using names `state` and `age`, and using options `center=FALSE` and `death=TRUE`. If different names are used, or covariates are centred internally by msm, ELECT will give an error message. ELECT cannot deal with covariates that are encoded as a

factor; use dummy variables instead when fitting the model with msm. The choice `death=TRUE` is because ELECT is restricted to illness-death models where the final state is the dead state and times of death are known exactly.

The package msm has its own functions to summarise the fitted model; for example, `print(model)` or `summary(model)`. Alternatively, extract the β-parameters using the commands

```
>   qnames <- c("q12","q13","q21","q23")
>   p <- model$estimates
>   p.se <- sqrt(diag(model$covmat))
>   print(cbind(q=qnames,p=round(p,3),se=round(p.se,3)),quote=FALSE)
```

```
         q   p      se
qbase q12 -1.529 0.179
qbase q13 -2.874 0.538
qbase q21 -2.886 0.483
qbase q23 -0.71  0.173
qcov  q12 0.057  0.007
qcov  q13 0.036  0.021
qcov  q21 -0.075 0.018
qcov  q23 0.013  0.006
qcov  q12 -0.035 0.009
qcov  q13 -0.049 0.028
qcov  q21 0.015  0.025
qcov  q23 -0.037 0.009
```

Note that the estimated regression coefficients for age (the first four named `qcov`) are positive for transitions $1 \to 2$, $1 \to 3$, and $2 \to 3$, and negative for $2 \to 1$ confirming that increasing age is associated with an increased risk of a transition to disability and/or death, and a decreased likelihood of a recovery from disability.

Life expectancies can now be computed and displayed by using ELECT. First load ELECT in R and define the data used for the distribution of the living states:

```
>   source("ELECT.r")
>   sddata <- data[data$state%in%c(1,2),]
```

The first line loads the ELECT functions in R. The second line defines the data that are used to estimate the distribution of the living states conditional on a

specified age. This distribution is needed to compute marginal life expectancies. It is up to the user how to choose this data. Next, the command lines for age 75 (`age=0`) and year of birth 1920 (`ybirth=20`) are

```
>   LEs <- elect(model=model, b.covariates=list(age=0,ybirth=20),
              statedistdata=sddata, time.scale.msm="years", h=0.5,
              statedist.covariates="age", age.max=40, S=0)
>   summary.elect(LEs, digits=2)
-----------------------------
ELECT summary
-----------------------------
Covariates values in the multi-state model:
   age ybirth
     0     20
Covariates in the state-distribution model:
   age

Life expectancies:
Point estimates:
     pnt
e11 6.20
e12 3.15
e21 1.16
e22 3.68
e.1 5.32
e.2 3.25
e   8.56
-----------------------------
```

A few guidelines for the `elect`-call: `age` should be the first covariate in the list `b.covariates`, and the order of the covariates in `b.covariates` should correspond to the order in the `msm`-call. Data `statedistdata` should only contain records with observed living states. Character string `time.scale.msm` should correspond to the time scale of `age`. Value `h` ≥ 0 is the parameter for the grid in the integral that is used in the estimation of the life expectancies and its value should be provided in the same scale as `time.scale.msm`. The specification of `age.max` should take into account `time.scale.msm` and possible transformation

of age. In the example, specifying `age.max=40` corresponds with an assumed maximum age of $75 + 40 = 115$ years.

ELECT always estimates life expectancies in years, but `time.scale.msm` allows for a different scale in the multi-state model (`"years"`, `"months"`, or `"weeks"`).

If no estimation of the uncertainty is required, choose `S` $= 0$. In that case only point estimates of the life expectancies will be provided. If `S` is specified by a non-negative integer, `S` replications of estimated life expectancies will be undertaken such that estimated uncertainty can be reported in `summary.elect`.

Estimated state-specific life expectancies (LEs) in the output above are `e11`, `e12`, `e21`, `e22`, where `eij` is LE in state `j` for an individual who is in state `i` at age 75. Marginal LE `e.j` is expected years in state `j` when state at age 75 is not taken into account explicitly. Total LE at age 75 is `e`.

To estimate marginal LEs, and additional model is fitted by ELECT for the distribution of the living states. The default of ELECT is to use a multinomial logistic regression model with age as only covariate. This is the specification `statedist.covariates="age"` in the `elect`-call. It is possible to extend this logistic regression model using the other covariates that are in the fitted multi-state model. In the above example, this would imply using `statedist.covariates= c("age","ybirth")`.

Estimating LEs for a range of specified ages and plotting the result is easily done by calling `elect` repeatedly. Choosing `S` larger than 0 makes it possible to add 95% confidence bands to the plot. The text file on the ELECT website with the R commands includes lines of code that will produce Figure 2, where 95% confidence bands are based on `S` $= 500$ simulation per specified age.

Warning: When using age as a time-dependent covariate in the model above, some of the functions in msm do not produce the correct output since they do not take into account the piecewise-constant approximation to the transition intensities. As an example: `summary(model)` produces expected prevalence of state at a grid of time points. For the above model, this is not the correct output as it does not take changing age into account. There are solutions to this problem in msm by explicitly using the function `prevalence.msm` and specifying the argument `piecewise.times`.

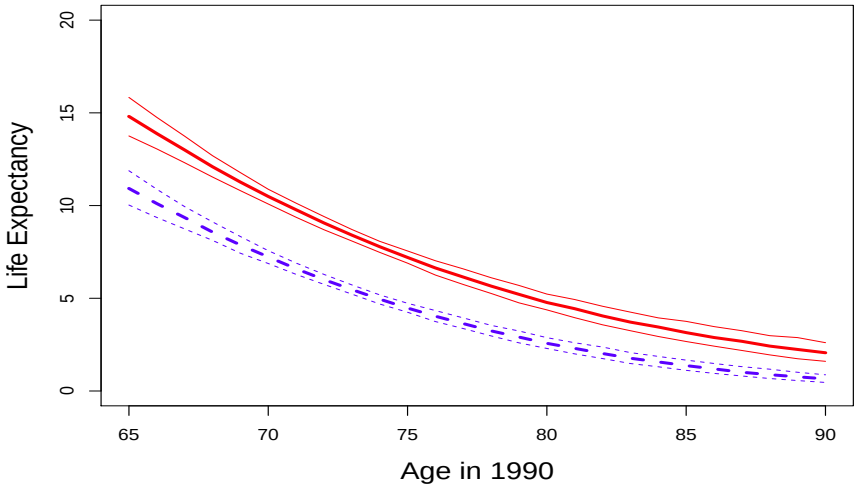

Figure 2: Marginal LEs for age specified in 1990, and 95% confidence bands. Solid line for total LEs, dashed line for LE in state 1.

3 Estimating life expectancies: Example II

This section presents an illness-death model for stroke. State 1 is the stroke-free state and state 2 indicates a history of one or more strokes; see Figure 3. The times of transitions $1 \to 2$ are interval-censored. State 3 is the dead state. Note that by definition there are no transitions from state 2 to state 1. The data stem from CFAS centre Newcastle and are not publicly available. The research was presented at REVES 2011 in Paris. Here we report the R commands, see the slides (downloadable from the ELECT website) for more information.

First the three-state model is fitted where the definition of $\mathbf{Q}$ allows only transitions $1 \to 2$, $1 \to 3$, and $2 \to 3$. The commands are

```
> # Data information:
> cat("Sample size:")
> print(length(table(data$id)))
> cat("Frequencies observed state:")
> print(table(data$state))
> cat("State table:")
> print(statetable.msm(state,id,data=data))
> # Define the Q matrix:
> q <- 0.05
```

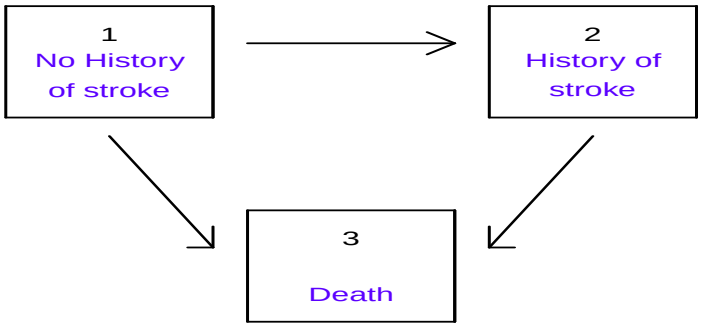

Figure 3: Three-state model for history of stroke in the MRC Cognitive Function and Ageing Study.

```
> Q <- rbind(c(0,q,q), c(0,0,q),c(0,0,0))
> covariates <- as.formula("~age+ybrth+sex+educ")
> # Fit the model:
> model <- msm(state~age, subject=id, data=data,
    center=FALSE, qmatrix=Q,death=TRUE,
    covariates=covariates, censor=c(-1,-2),
    censor.states=list(c(1,2),c(1,2)),method="BFGS",
    control=list(trace=0, REPORT=1,maxit=10000, fnscale=80000))
> # Organise output:
> qnames <- c("q12","q13","q23")
> p    <- model$estimates
> p.se <- sqrt(diag(model$covmat))
> print(cbind(q=qnames,p=round(p,3),se=round(p.se,3)),quote=FALSE)

Sample size:[1] 2321
Frequencies observed state:
  -2   -1    1    2    3
 766 2115 5121  583 1555

State table:    to
from   -2   -1    1    2    3
  -1  341 1200   24    8  542
  1   382  855 2942  105  837
  2    43   60    0  304  176

      q   p      se
qbase q12 -4.297 0.745
qbase q13 -2.867 0.184
```

```
qbase q23 -1.81  0.316
qcov  q12 0.111  0.052
qcov  q13 0.094  0.011
qcov  q23 0.05   0.018
qcov  q12 0.032  0.049
qcov  q13 0.004  0.011
qcov  q23 -0.012 0.018
qcov  q12 0.4    0.198
qcov  q13 0.355  0.077
qcov  q23 0.433  0.129
qcov  q12 -0.023 0.231
qcov  q13 -0.27  0.098
qcov  q23 0.163  0.155
```

A few remarks. The number of right-censored states and observed death (766 + 1555) is equal to the sample size 2321. Note that there are no backwards transitions in the state table: the entry for transition $2 \rightarrow 1$ is zero. The value -1 for `state` denotes an intermediate missing state; that is, there is a time of interview but without an observed state. All estimated regression coefficients for age (the first three named `qcov`) are positive confirming that increasing age is associated with an increased risk of a transition to a next state.

Next the life expectancies are estimated using the functions in ELECT for a specified age and specified covariates at baseline. Here we take into account that age is centred by minus 78.5 years.

```
> # Define the data for the logistic regression:
> sddata  <- data[data$state%in%c(1,2),]
> # Define the parameters for the LEs
> age     <- 70 - 78.5
> age.max <- 115 - 78.5
> ybrth   <- 1920 - 1900
> educ    <- 1
> # Life expectancies for women:
> LEsW <- elect(model=model,
                b.covariates=list(age=age,ybrth=ybrth,sex=0,educ=educ),
                statedistdata=sddata, h=0.5,time.scale.msm="years",
                age.max=age.max, S=1000)
> summary.elect(LEsW,probs=c(.025,.975),digits=2)
-----------------------------
ELECT summary
-----------------------------
Covariates values in the multi-state model:
  age ybrth   sex  educ
 -8.5  20.0   0.0   1.0
```

```
Covariates in the state-distribution model:
   age

Life expectancies:Using simulation with  1000 replications
Point estimates, and mean, SEs, and quantiles from simulation:
      pnt    mn   se 0.025q 0.975q
e11 11.93 11.79 0.68  10.40  13.00
e12  1.67  1.71 0.52   0.88   2.86
e21  0.00  0.00 0.00   0.00   0.00
e22  7.10  7.14 0.88   5.61   9.00
e.1 11.04 10.90 0.64   9.60  12.00
e.2  2.08  2.12 0.48   1.38   3.21
e   13.12 13.02 0.47  12.11  13.93
-----------------------------

> # Life expectancies for men:
> LEsM <- elect(model=model,
                b.covariates=list(age=age,ybrth=ybrth,sex=1,educ=educ),
                statedistdata=sddata, h=0.5, time.scale.msm="years",
                age.max=age.max, S=1000)
```

LEs are estimated and the uncertainty in the estimation is assessed by simulation. In the example, there are `S = 1000` repetitions used to quantity the uncertainty. Instead of providing quantiles of the generated distributions, the repetitions can also be used to plot the estimated distribution for each of the LEs using `plot.elect(LEsW)`. The following code produces similar plots, but with the LEs for men and women combined; see Figure 4.

```
> lwd <- 5; cex.lab <- 2; ylab <- c("Density"," "," ","Density"," ", " ")
> opar <- par(mfrow=c(2,3), mex=0.8,mar=c(5,5,2,1)+.1)
> index <- 1; tekst <- names(LEsW$pnt)
> for(i in c(1,2,4,5,6,7)){
     plot(c(0,14),c(0,1.5),main="",ylab=ylab[index],xlab="Years",type="n",
            cex=1,cex.axis=1.5,cex.lab=cex.lab)
     lines(density(LEsW$sim[,i]),col=1,lwd=lwd,cex.lab=cex.lab)
     lines(density(LEsM$sim[,i]),col=2,lwd=lwd,cex.lab=cex.lab)
     text(12,1.4,tekst[i],cex=3,col="blue")
     if(index==1){
       legend(0, 1.5, c("Women", "Men"), col = c(1,2),
       text.col = c(1,2), lty=c(1,1),lwd=c(3,3),merge=TRUE,cex=2, bg = 'white')
     }
     index <- index+1
 }
```

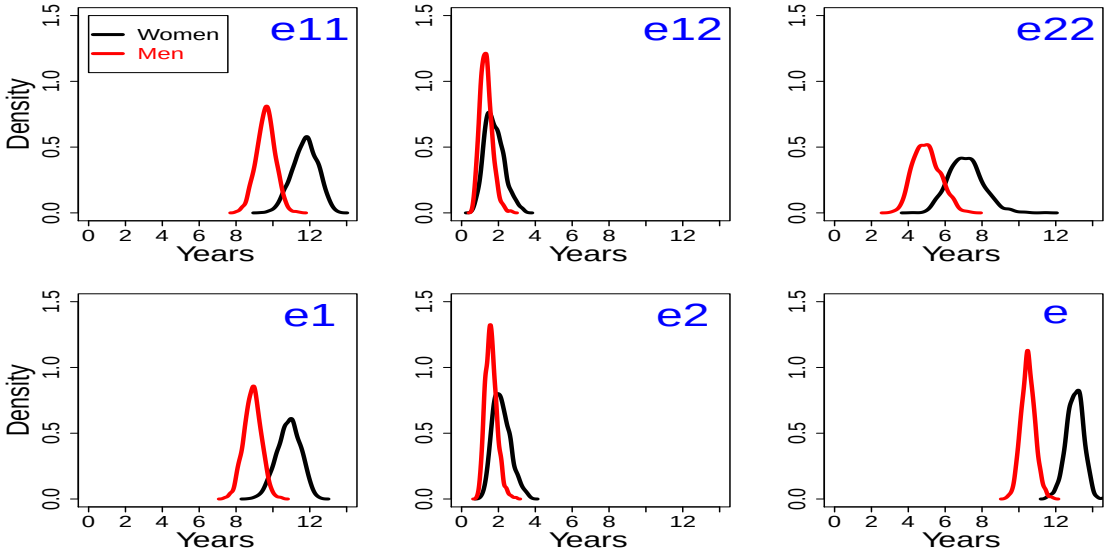

Figure 4: The distributions of the estimated life expectancies for men and women aged 70 in 1990.

4 Estimating life expectancies: Example III

Heart transplant monitoring data are taken from the msm package: "A series of approximately yearly angiographic examinations of heart transplant recipients. The state at each time is a grade of cardiac allograft vasculopathy (CAV), a deterioration of the arterial walls". After loading the msm package in R, type

```
> ?cav
```

for more information, see also Sharples et al. (2003). There are three living states and a dead state, and variables that will be used are patient number (`PTNUM`), `age`, age of donor (`dage`), `sex`, and an indicator for the baseline (`firstobs`). All patients start in CAV-free state 1. Further data descriptives are

```
Sample size = 622

Frequencies observed state:
   1    2    3    4
2039  351  205  251

State table:     to
from    1    2    3    4
   1 1367  204   44  148
```

```
       2   46  134   54   48
       3    4   13  107   55
```

A hidden Markov model is fitted inspired by the model presented in Sharples et al. (2003), but with age as a time-dependent covariate. The model is progressive (no transitions back), but the transitions in the data are not and potential misclassification of the living states is accounted for. The model formulation and the code for msm are given by

```
> q <-0.01
> Q <- rbind( c(0,q,0,q), c(0,0,q,q),c(0,0,0,q),c(0,0,0,0) )
> ematrix <- rbind( c(0,0.1,0,0), c(0.1,0,0.1,0), c(0,0.1,0,0),c(0,0,0,0) )
> model <- msm(state~age, subject=PTNUM, data=cav, center=FALSE,
             qmatrix=Q, death=TRUE, covariates=~age+dage+sex,
             ematrix=ematrix,fixedpars=c(6,8,9,10), method="BFGS",
             control=list(reltol=1e-8,maxit=1000,fnscale=10000))
```

Note that some of the coefficients for age are restricted to zero, and that the misclassification is only possible between states 1 and 2, or states 2 and 3.

By selecting the baseline data and using that as input for ELECT, the fact that all individuals start in state 1 is automatically taken into account. A possible call to ELECT is as follows.

```
> age0   <- round(mean(data[data$firstobs==1,]$age))
> dage0  <- round(mean(data[data$firstobs==1,]$dage))
> sex0   <- median(data[data$firstobs==1,]$sex)
> sddata <- data[data$firstobs==1,]
> LEs    <- elect(model=model,
                  b.covariates=list(age=age0,dage=dage0,sex=sex0),
                  statedistdata=sddata, time.scale="years",
                  h=0.5, age.max=100, S=500,setseed=12345)
> summary.elect(LEs, digits=2)

----------------------------
ELECT summary
----------------------------
Covariates values in the multi-state model:
 age dage  sex
  47   31    0
Life expectancies:Using simulation with  500 replications

Point estimates, and mean, SEs, and quantiles from simulation:
      pnt    mn   se 0.025q  0.5q 0.975q
e11  5.91  5.89 0.35   5.21  5.89   6.57
e12  2.92  2.80 0.36   2.09  2.80   3.58
e13  2.18  2.09 0.31   1.47  2.09   2.66
```

```
[...]

e.1  5.91  5.89 0.35   5.21  5.89   6.57
e.2  2.92  2.80 0.36   2.09  2.80   3.58
e.3  2.18  2.09 0.31   1.47  2.09   2.66
e   11.00 10.78 0.55   9.64 10.83  11.74
-----------------------------
```

Note that the marginal LEs `e.1`, `e.2`, and `e.3`, are the same as `e11`, `e12`, and `e13`, respectively, because all patients start in state 1 (a multinomial logistic regression model was not fitted).

This example is for illustration only. Given that patients are followed up after a transplant, time since transplant would be a better choice for the time scale in the model.

5 Estimation in ELECT

For a discussion of state-specific life expectancy (LE) for multi-state models see, for example, Izmirlian et al. (2000), Van den Hout and Matthews (2010), and Van den Hout (2016). LEs in a multi-state model is a generalisation of mean survival in a standard survival model where there is one living state and one dead state.

Let the finite state space be given by $\{1, 2, ..., D\}$ where D is the dead state. Assume that the deterministic covariate process is given by $\mathcal{Z} = \{\boldsymbol{z}(t) |\ t \geq t_0\}$ for specified age t_0. LE in living state s given state r at t_0, for $r, s \in \{1, 2, ..., D-1\}$, is defined by

$$e_{rs}(t_0) = e_{rs}(t_0, \mathcal{Z}) = \int_0^\infty \mathbb{P}(X_{t_0+t} = s |\ X_{t_0} = r, \mathcal{Z}) dt, \tag{2}$$

where $\mathbb{P}(X_{t_0+t} = s |\ X_{t_0} = r, \mathcal{Z})$ is the transition probability of being in state s at age $t_0 + t$, given starting state r at age t_0 and covariate process $\mathcal{Z}$. Marginal LE is given by

$$e_{\bullet s}(t_0) = \sum_{r \neq D} \mathbb{P}(X_{t_0} = r | \mathcal{Z}) e_{rs}(t_0), \tag{3}$$

which is the LE in state s irrespective of the initial state at age t_0. To derive this quantity we need the distribution of the living states at age t_0, that is, we need

$\mathbb{P}(X_{t_0} = r|\mathcal{Z})$ for all $r \in \{1, 2, ..., D-1\}$. Total LE at age t_0 is defined as

$$e(t_0) = \sum_{s \neq D} e_{\bullet s}(t_0). \tag{4}$$

Quantities (2), (3), and (4) can be derived from the fitted multi-state model and the data. The specification of $\mathcal{Z}$ is of course crucial, and within $\mathcal{Z}$ the age t_0 will in most cases be the most influential specification. To approximate the integral (2), a maximum age has to specified such that we may safely assume that the integrand $\mathbb{P}(X_{t_0+t} = s|\ X_{t_0} = r, \mathcal{Z})$ is zero beyond a certain age.

With regard to the estimation of (2), (3), and (4), we will only discuss the basic elements without going into detail. Note that with a fitted multi-state model and specified $\mathcal{Z}$, the integrand in (2) can be computed for any t. In the computation of this integrand, a piecewise-constant approximation is used to account for changing age over time. Computationally it is convenient to use the same grid for the piecewise-constant approximation *and* the numerical approximation of the integral.

For the marginal LE in (3), multinomial logistic regression is used to estimate the distribution of the living states at age t_0. This distribution is necessary to estimate $\mathbb{P}(X_{t_0} = r|\mathcal{Z})$ in (3). The default of ELECT uses `age` as a covariate in the multinomial logistic model. This model can be extended by using additional covariates (as long as they are also used in the multi-state model). For the logistic model to be estimated, `statedistdata` has to be provided in the `elect`-call.

The above will provide a point estimate of LEs and total LE. To estimate the uncertainty (standard errors and/or confidence intervals) we make use of the asymptotic properties of the maximum likelihood estimator of the parameters for the multi-state model. Given a fitted model, consider the multivariate normal distribution with expectation equal to the maximum likelihood estimate of the parameter vector and the covariance matrix equal to the estimated covariance matrix at the optimum. The sample variation in the estimation of the life expectancies is evaluated by drawing parameters values from this multivariate distribution and computing the life expectancies for each of the drawn values (cf. Aalen et al. 1997).

6 Functions in ELECT

Function

`elect`

Description

Estimates state-specific and marginal life expectancies given a fitted illness-death model in msm.

Usage

```
elect(model, b.covariates, statedistdata, time.scale.msm="years",
      h, age.max, S=0, setseed=NULL, RestrAndConst=NULL,
      statedist.covariates="age", method="step")
```

Arguments

`model`	fitted msm model.
`b.covariates`	list with specified covariates values (ignore intercept).
`statedistdata`	data to derive univariate distribution of living states.
`time.scale.msm`	time scale in multi-state model either in the set {"years", "months","weeks"} or a value in (0,1].
`h`	grid parameter for integration where scale is `time.scale.msm`.
`max.age`	assumed maximum age in same time scale as in model.
`S`	number of replications for estimation of uncertainty (`S`=0 for no estimation).
`setseed`	seed for the random number generation in the simulations.
`RestrAndConst`	vector which indexes the independent model parameters. Only needed when `constraint` is used in `msm` call.
`statedist.covariates`	names of covariates for model for univariate distribution of living states.
`method`	approximation of integral: `"step"` for simple step function `"MiddleRiemann"` or `"Simpson"`.

Details

In the msm-call for model fit use `center=FALSE` and `death=TRUE`, and names `state` and `age`. Do not use variables encoded as factor by R. Covariate `age` should be the first entry in `b.covariates`. The other covariates in `b.covariates` should follow the order in the msm call. The life expectancies are computed by approximating the integral numerically with a grid defined by `h`. The specification of `statedist.covariates` should be a subset of `b.covariates`.

Value

`pnt`	life expectancies derived from MLE of model parameters.
`sim`	simulated life expectancies using the MLE of model parameters.
`h`	`h` as specified in `elect`-call.
`covars`	covariates as specified in `elect`-call.
`S`	`S` as specified in `elect`-call.
`sd.model`	fitted model for the univariate distribution of living states.

Function

`summary.elect`

Description

Summarises the life expectancies as estimated by `elect`.

Usage

```
summary.elect(LEs,probs=c(.025,0.5,.975),digits=3,
      print=TRUE, StartStateTotals=FALSE, sd.model=FALSE)
```

Arguments

`LEs`	life expectancies estimated by `elect`.
`probs`	numeric vector of probabilities with values in [0,1] for quantiles.
`digits`	number of decimal places in output.
`print`	TRUE for printing output to screen, FALSE otherwise.
`StartStateTotals`	TRUE for output on start-state totals $e_{r.}$ (for $S > 0$).
`sd.model`	TRUE for output on fitted logistic regression model.

Details

Reports state-specific and marginal life expectancies derived from the maximum likelihood point estimate of the model parameters. In addition, quantiles of the simulated distribution derived from the maximum likelihood estimation.

Function

`plot.elect`

Description

Graphical representation by smoothed densities of the distribution of the life expectancies as estimated by `elect`.

Usage

```
plot.elect(LEs, kernel="gaussian", col="red", lwd=2, cex.lab=1)
```

Arguments

`LEs`	life expectancies estimated by `elect`.
`kernel`	character string for smoothing kernel ("gaussian", "rectangular", "triangular", "epanechnikov", "biweight", "cosine" or "optcosine").
`col`	colour of curve.
`lwd`	line width of curve.
`cex.lab`	magnification to be used for axis-labels.

Details

Presents distribution of the estimated life expectancies derived from the maximum likelihood estimate of the model parameters. The smoothing is undertaken using the R function `density`.

Function

`check.RestrAndConst.elect`

Description

Function to check definition of `RestrAndConst` in `elect` call.

Usage

```
check.RestrAndConst.elect(model, RestrAndConst, PRINT=FALSE)
```

Arguments

`model`	fitted msm model.
`RestrAndConst`	vector which indexes the independent parameters in `model$opt$par` w.r.t. to the model parameters.
`PRINT`	print the comparison, or not.

Details

Function to help defining the parameter constraints in the `elect` call when the `constraint` option is used in `msm`.

Value

`OK` when `RestrAndConst` is well-defined `TRUE`, else `FALSE`

Acknowledgements

Work on ELECT started at the MRC Biostatistics Unit, Cambridge, under supervision of Fiona Matthews. The work was partly funded by the Medical Research Council grant US UC A030 0031. I also would like to thank some early users of ELECT for feedback and suggestions.

References

Aalen OO, Farewell VT, De Angelis D, Day NE, Gill ON (1997). A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Statistics in Medicine* **16**, 2191–2210.

Brayne C, McCracken C, Matthews FE (2006). Cohort profile: the Medical Research Council Cognitive Function and Ageing Study (CFAS). *International Journal of Epidemiology* **35**, 1140-1145.

Cox DR, Miller HD (1965). *The Theory of Stochastic Processes*. London: Chapman and Hall.

Izmirlian G, Brock D, Ferrucci L, and Phillips C (2000). Active Life Expectancy from Annual Follow-Up Data with Missing Responses. *Biometrics* **56**, 244–248.

Jackson C (2011). Multi-state models for panel data: The msm Package for R. *Journal of Statistical Software* **38**.

Kalbfleisch J, Lawless JF (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863–871.

Kay R (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics* **42**, 855–865.

Satten GA, Longini IM (1996). Markov chains with measurement error: estimating the 'true' course of a marker of the progression of human immunodeficiency virus disease (with discussion). *Applied Statistics* **45**, 275–309.

Sharples LD, Jackson CH, Parameshwar J, Wallwork J, Large SR (2003). Diagnostic accuracy of coronary angiography and risk factors for post-heart-transplant cardiac allograft vasculopathy. *Transplantation* **76**, 679–682.

Van den Hout A, Matthews FE (2010). Estimating stroke-free and total life expectancy in the presence of non-ignorable missing values. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **173**, 331–349.

Van den Hout A (2016). *Multi-State Survival Models for Interval-Censored Data*. Boca Raton: CRC/Chapman & Hall.