An R Package for the inference in a multi-state illness-death model

October, 2011

III Jornadas de Usuarios de R

Madrid, Spain

Luís Meira-Machado, Javier Roca-Pardiñas

Overview

- Multi-state models
  - Generalities.
  - Progressive three-state models: empirical estimators.
  - Regression models.
- R based Software: p3state.msm
- Application

  Colon cancer data; Bladder cancer data.

Multi-state models

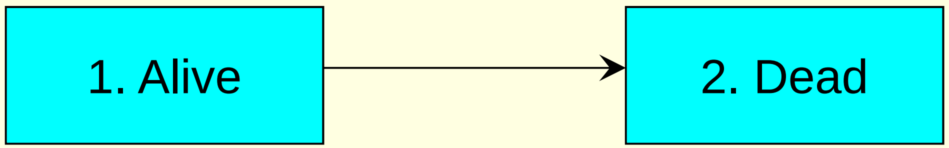
A multi-state model is a model for a stochastic process, which at any time point occupies one of a set of discrete states.

In biomedical applications, the states may be based:

- Clinical symptoms
- Biological markers
- Some scale of the disease
- A non-fatal complication in the course of the illness.

The mortality model

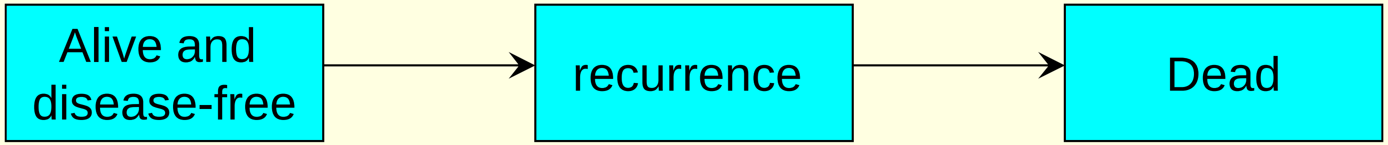
Multi-state models (MSMs) are very useful for describing complicated event history data. These models may be considered a generalization of survival analysis where survival is the ultimate outcome of interest but where intermediate (transient) states are identified.



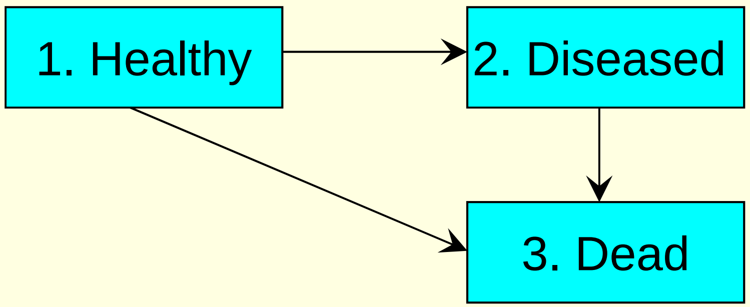
The mortality model (for survival analysis)

Progressive three-state model

The scope of multi-state models (Andersen et al., 1993) provides a rich framework to handle complex situations involving more than two states and a number of possible transitions among them.



The progressive three-state model for breast cancer data.



Three transition intensities:
- Incidence of the illness;
- Two death intensities (with and without the disease)

Notations

The multi-state process is fully characterized through:

Transition probabilities between states h and j,

$$\mathrm{P}\left(X(t)=j \middle| X(s)=h, \mathfrak{F}_{s-}\right)$$

being $\mathfrak{F}_{s-}$ the observed history of the process up to time t that is generated by $\{X(u), 0 \le u < s\}$. Or through transition intensities: $\alpha_{hj}(t \mid \mathfrak{F}_{t-}) = \lim_{dt \to 0} \dfrac{\mathrm{P}\left(X(t+dt)=j \middle| X(t)=h, \mathfrak{F}_{t-}\right)}{dt}$

Assumptions and goals

Assumptions

- *Time-Homogeneity*: the intensities are constant over time.
- *The Markov assumption*: future evolution only depends on the current state. That is, the transition intensities are independent of the history of the process.
- *The semi-Markov assumption*: future evolution does not depend on the current time, but only on the duration in the current state.

Goals

- Estimation of transition probabilities; Estimation of the bivariate distribution function.
- Multi-state regression (e.g., using Cox (semi-)Markov models)

Available R based software to implement Multi-state models

survival http://cran.r-project.org/web/packages/survival

msm http://cran.r-project.org/web/packages/msm

p3state.msm http://cran.r-project.org/web/packages/p3state.msm

survivalBIV http://cran.r-project.org/web/packages/survivalBIV

mstate http://cran.r-project.org/web/packages/mstate

etm http://cran.r-project.org/web/packages/etm

changeLOS http://cran.r-project.org/web/packages/changeLOS

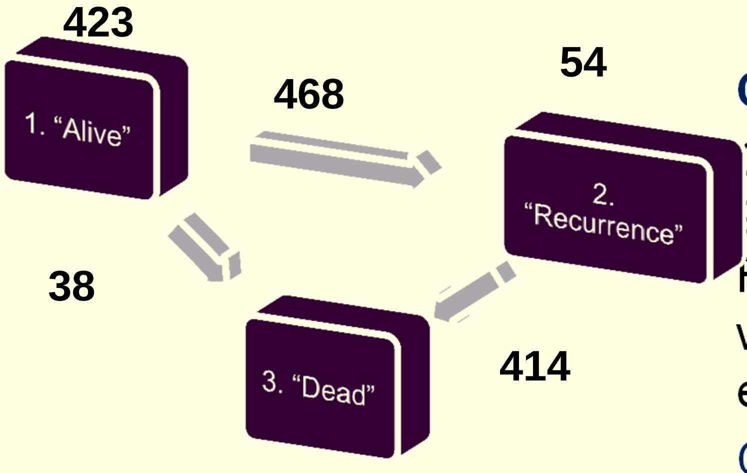
mvna http://cran.r-project.org/web/packages/mvna

timereg http://cran.r-project.org/web/packages/timereg

Epi http://cran.r-project.org/web/packages/Epi/index.html

New JSS Special issue: Competing Risks and Multi-State Models
http://www.jstatsoft.org/v38

Examples of Aplication



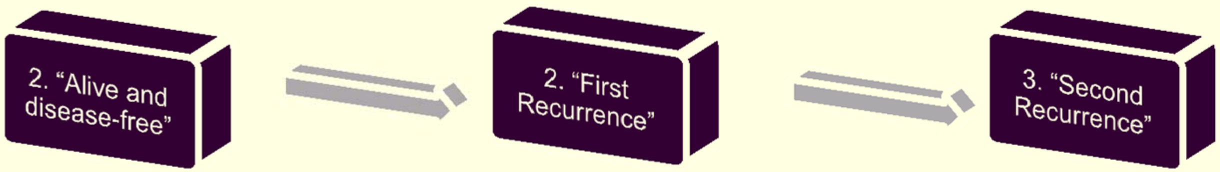
Colon Cancer Data

Clinical trial on Duke’s stage III patients with 929 incident cases of colon cancer. Recurrence is a time-dependent covariate which can be expressed as intermediate event and modeled as a multi-state model.

Covariates: rx, sex, age, etc.

Bladder Cancer Data

Available on the survival package of the R. The states are based on the occurrence of the first and second recurrence.



The **p3state.msm** Package

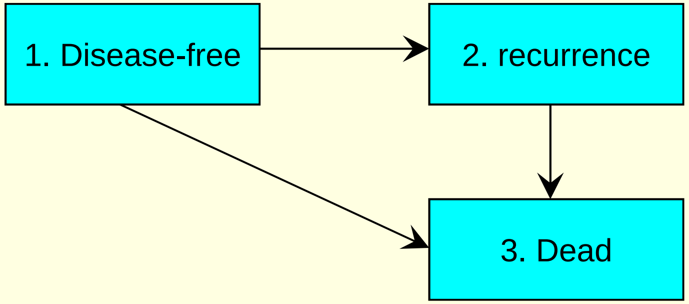
p3state.msm	
Numerical output:	Regression coefficients (TDCM, CMM, CSMM), bivariate distribution function, transition probabilities
Graphical output:	Transition probabilities, bivariate distribution function, marginal distribution
Models:	Progressive three-state and illness-death

Authors: Luis Meira-Machado and Javier Roca-Pardiñas.

- Performs multi-state regression
- Provides statistical methods for estimating quantities of interest such as transition probabilities.

Input data

times1	delta	times2	time	status	rx	sex	age
968	1	553	1521	1	3	1	43
3087	0	0	3087	0	3	1	63
542	1	421	963	1	1	0	71
245	1	48	293	1	3	0	66



times1 – sojourn time in state 1
delta – indicator of transition from state 1 to state 2
times2 – sojuourn time in state 2
time – times1+times2
status – final indicator status
covariates

p3state.msm

```
R> res.p3state<-p3state(colon2, formula = ~ factor(rx) + sex + age)
R> summary(res.p3state, model = "TDCM")
R> summary(res.p3state, model = "CMM")
R> summary(res.p3state, model = "CSMM")
```

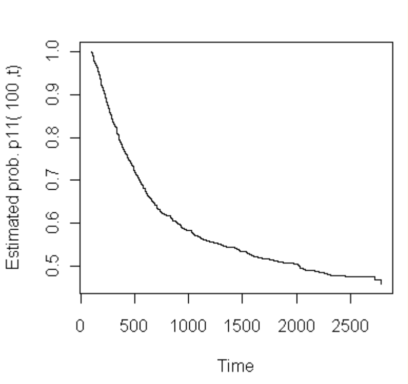
Cox Semi-Markov Model from state 1 -> 3				
	coef	**exp(coef)**	**95% CI**	**p-value**
factor(rx)2	-0.3353	0.7151	0.3132 – 1.6329	0.4261
factor(rx)3	-0.1670	0.8462	0.4011 – 1.7853	0.6611
sex	0.4238	1.5278	0.7922 – 2.9464	0.2059
age	0.0854	1.0892	1.0486 – 1.1313	1.0231e-05

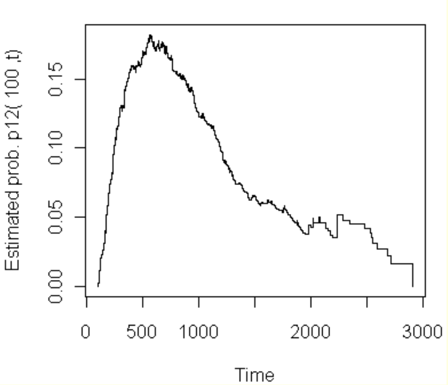
p3state.msm

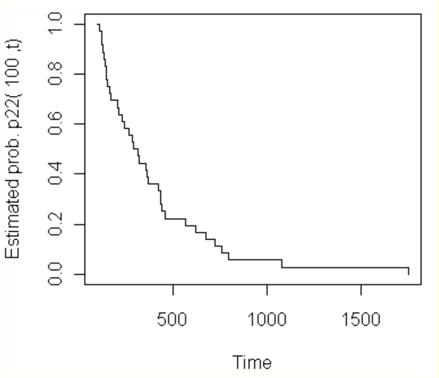
R> summary(res.p3state, time1 = 100, time2 = 800)

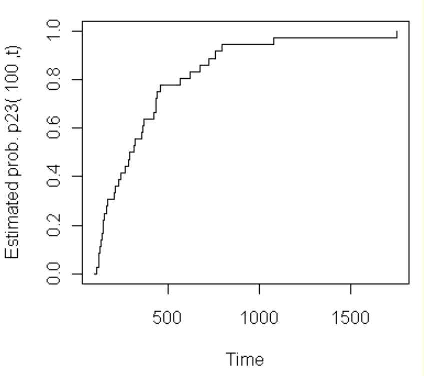
The estimate of the transition probability P11(100 , 800) is 0.6182574
The estimate of the transition probability P12(100 , 800) is 0.1553286
The estimate of the transition probability P13(100 , 800) is 0.226414
The estimate of the transition probability P22(100 , 800) is 0.05579566
The estimate of the transition probability P23(100 , 800) is 0.9442043

R> plot(res.p3state, plot.trans = "all", time1 = 100)









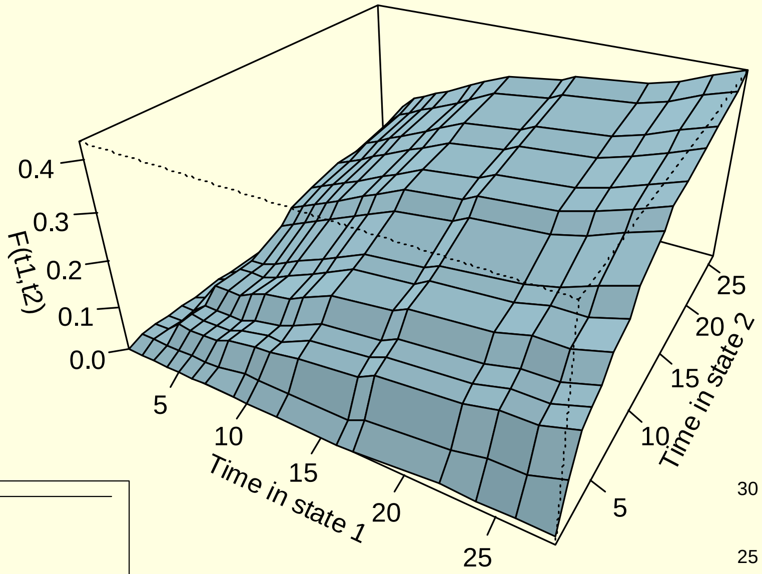
p3state.msm

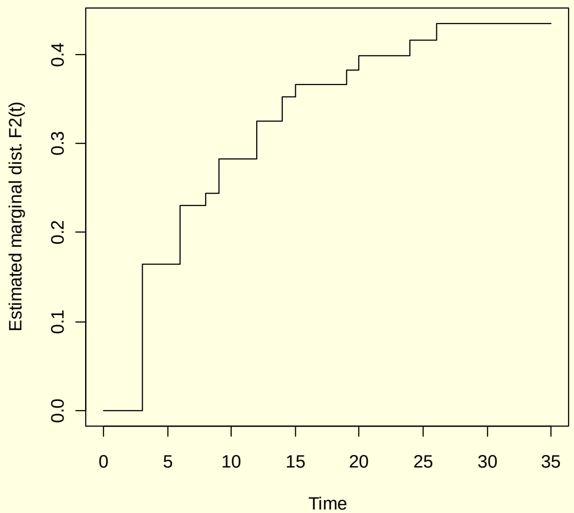

```
R> res.blad<-p3state(blad)
R> summary(res.blad, estimate = TRUE, time1 = 3, time2 = 12)
```

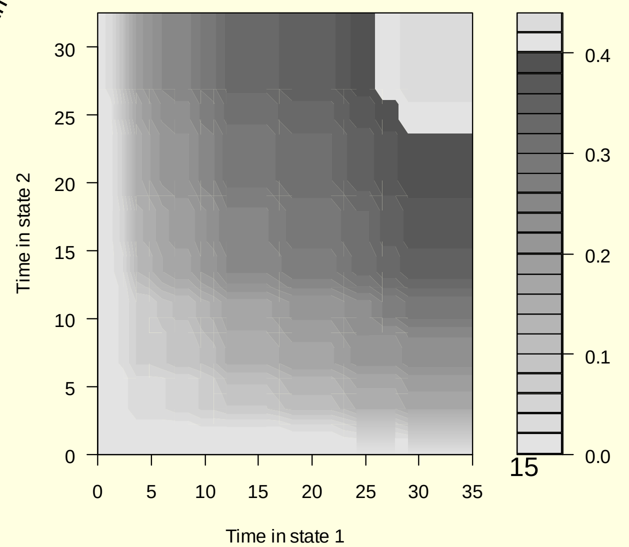
The estimate of the transition probability P11(3 , 12) is 0.7543621
The estimate of the transition probability P12(3 , 12) is 0.1300108
The estimate of the transition probability P13(3 , 12) is 0.1156272
The estimate of the transition probability P22(3 , 12) is 0.7074841
The estimate of the transition probability P23(3 , 12) is 0.2925159
The estimate of the bivariate distribution function F12(3 , 12) is 0.09060991
The estimate of the marginal distribution function of the second gap time, F2(12) is 0.3250242

p3state.msm

R> plot(res.blad, plot.marginal = TRUE, plot.bivariate = TRUE)







Two more packages

survivalBIV	
Numerical output:	Bivariate distribution function using several methods; marginal distribution of the second gap time.
Graphical output:	Bivariate distribution function, marginal distribution
Models:	Progressive three-state model
Authors	Ana Moreira, Artur Araujo and Luis Meira-Machado
Availability	Now on CRAN

TPmsm	
Numerical output:	Estimates for the transition probabilities with bootstrap confidence intervals.
Graphical output:	Transition probabilities
Models:	Progressive three-state model; illness-death model
Authors	Artur Araujo, Luis Meira-Machado and Javier Roca-Pardiñas
Availability	Soon on CRAN

Some References

- Andersen PK, Borgan O, Gill RD, Keiding N. (1993). *Statistical models based on counting processes*. New York, Springer.

- de Uña-Álvarez J, Meira-Machado LF (2008). A Simple Estimator of the Bivariate Distribution Function for Censored Gap Times. Statistics and Probability Letters, 78, 2440-2445.

- Meira-Machado LF, Roca Pardiñas J. (2011). p3state.msm: Analysing Survival Data from an Illness-Death Model. *Journal of Statistical Software*.

- Moreira A, Meira-Machado LF. survivalBIV: Estimation of the Bivariate Distribution Function for Sequentially Ordered Events Under Univariate Censoring. Submitted to *Journal of Statistical Software*.