spm : an R-package that implements a Stochastic Process Model (SPM)

Ilya Y. Zhbannikov

2015-11-24

Overview

The R-package `spm` (https://github.com/izhbannikov/spm) is developed for modeling aging-related changes and it allows (1) data simulation and (2) estimating the process parameters using maximum likelihood estimation by optimizing parameters used in the model. Specifically, developed R-package spm allows (i) one-dimensional SPM; (ii) multiple dimensional SPM; (iii) data simulation for one- and multiple dimensions.

Introduction

General description of age-related changes.

Model description

What we propose

Data description

Data represents a typical clinical trait data and must be presented in form of two datasets: longitudinal dataset (follow-up studies), in which one record represents a single observation, and vital statistics, where one record represents all patient information. Longitudinal dataset must contain person ID (Identification number), status (dead(1)/alive(0)), time and measurements across the covariates. The `spm` can handle an infinite number of covariates but in practice, 5-7 covariates is enough.

Below there is an example of clinical data that can be used in `spm` and we will discuss the field later. Longitudinal studies:

```
##   X ID IndicatorDeath Age AgeNext      DBP      BMI
## 1 1  1              0  30      32 80.00000 25.00000
## 2 2  1              0  32      34 80.51659 26.61245
## 3 3  1              0  34      36 77.78412 29.16790
## 4 4  1              0  36      38 77.86665 32.40359
## 5 5  1              0  38      40 96.55673 31.92014
## 6 6  1              0  40      42 94.48616 32.89139
```

Vital statistics:

```
##   X ID IsDead   LSmort
## 1 1  1      1 85.34578
## 2 2  2      1 80.55053
## 3 3  3      1 98.07315
## 4 4  4      1 81.29779
## 5 5  5      1 89.89829
## 6 6  6      1 72.47687
```

Data fields description

Longitude studies

- ID - subject unique identificatin number.
- IndicatorDeath - 0/1, indicates death of a subject.
- Age - current age of subjects.
- AgeNext - next age of subject he will attend to the survey/exam.
- DBP, BMI - covariates, here "DBP" represents a diastolic blood pressure, "BMI" a body-mass index.

Vital statistics

- ID - subject's unique ID.
- IsDead - death indicator, 0 - alive, 1 - dead.
- LSmort - age at death of stopping observations.

Discrete and Continuous cases

There are two main SPM types in the package: discrete model and continuous model. Discrete model assumes equal intervals between follow-up observations. The example of discrete dataset is given below.

```
library(spm)
data <- simdata_discr_MD(N=10, ystart=c(80), k=1)
head(data)
```

```
##      id xi t1 t2   par1_1   par1_2
## [1,]  1  0 30 31 80.00000 66.66658
## [2,]  1  0 31 32 66.66658 63.38874
## [3,]  1  0 32 33 63.38874 64.77483
## [4,]  1  0 33 34 64.77483 67.15203
## [5,]  1  0 34 35 67.15203 74.06955
## [6,]  1  0 35 36 74.06955 75.32344
```

In this case there are equal intervals between t1 and t2 (Age and Age.next).

The opposite is continuous case, in which intervals between observations are not equal. The example of continuous case dataset is shown below:

```
library(spm)
data <- simdata_cont_MD(N=5,ystart = c(50))
head(data)
```

```
##   id xi       t1       t2       y1  y1.next
## 1  1  0 72.63791 74.24545 50.12166 52.37498
## 2  1  0 74.24545 75.47583 52.37498 53.01293
## 3  1  0 75.47583 76.65728 53.01293 51.81519
## 4  1  0 76.65728 77.84210 51.81519 49.32389
## 5  1  0 77.84210 78.18091 49.32389 50.86074
## 6  1  0 78.18091 79.89277 50.86074 50.53029
```

Discrete case

In discrete case, we use the following assumptions:

$$\bar{y}(t+1) = \bar{u} + \bar{R} \times \bar{y}(t) + \bar{\epsilon}$$

$$\mu(t) = \mu_0(t) + \bar{b}(t) \times \bar{y}(t) + \bar{Q} \times \bar{y}(t)^2$$

Where:

$$\mu_0(t) = \mu_0 e^{\theta t}$$
$$\bar{b}(t) = \bar{b} e^{\theta t}$$
$$\bar{Q}(t) = \bar{Q} e^{\theta t}$$

Continuous case

$$\mu(u) = \mu_0(u) + (\bar{m}(u) - \bar{f}(u)^* \times \bar{Q}(u) \times (\bar{m}(u) - \bar{f}(u)) + Tr(\bar{Q}(u) \times \bar{\gamma}(u))$$

$$dm(t)/dt = \bar{a}(t) \times (\bar{m}(t) - \bar{f}_1(t)) - 2\bar{\gamma}(t) \times \bar{Q}(t) \times (\bar{m}(t) - \bar{f}(t))$$
$$d\bar{\gamma}(t)/dt = \bar{a}(t) \times \bar{\gamma}(t) + \bar{\gamma}(t) \times \bar{a}(t)^* + \bar{b}(t) \times \bar{b}(t)^* - 2\bar{\gamma}t \times \bar{Q}(t) \times \bar{\gamma}(t)$$

Coefficient conversion between continuous and discrete cases

$$Q = Q$$
$$\bar{a} = \bar{R} - diag(k)$$
$$\bar{b} = \bar{\epsilon}$$
$$\bar{f}1 = -1 \times \bar{u} \times a^{\bar{-}1}$$
$$\bar{f} = -0.5 \times \bar{b} \times Q^{\bar{-}1}$$
$$mu_0 = mu_0 - \bar{f} \times \bar{Q} \times t(\bar{f})$$
$$\theta = \theta$$

Case with time-dependent coefficients

In two previous cases, we assumed that coefficients is sort of time-dependant: we multiplied them on to

$$e^{\theta t}$$

. In general, this may not be the case. We extend this to a general case, i.e. (we consider one-dimensional case):

$$a\bar{(t)} = par_1 t + par_2$$

- linear function.

The corresponding equations will be equivalent to one-dimensional continuous case described above.

Simulation

We added one- and multi- dimensional simulation to be able to generate test data for hyphotesis testing. Data, which can be simulated can be discrete (equal intervals between observations) and continuous (with arbitrary intervals).

Discrete

For discrete case:

```
library(spm)
data <- simdata_discr_MD(N=100, ystart=c(75, 94), k=2)
head(data)
```

```
##      id xi t1 t2   par1_1   par1_2    par2_1    par2_2
## [1,]  1  0 30 31 75.00000 69.29038  94.00000  96.27888
## [2,]  1  0 31 32 69.29038 63.42717  96.27888  94.22123
## [3,]  1  0 32 33 63.42717 54.07290  94.22123  95.23642
## [4,]  1  0 33 34 54.07290 49.58226  95.23642 100.91389
## [5,]  1  0 34 35 49.58226 47.11312 100.91389  98.87588
## [6,]  1  0 35 36 47.11312 40.63211  98.87588  86.45995
```

Continuous

For continuous case:

```
library(spm)
data <- simdata_cont_MD(N=100)
head(data)
```

```
##   id xi       t1       t2       y1  y1.next
## 1  1  0 30.97811 31.42534 85.74200 87.40471
## 2  1  0 31.42534 31.95067 87.40471 87.31503
## 3  1  0 31.95067 32.52584 87.31503 87.43195
## 4  1  0 32.52584 33.68741 87.43195 90.09190
## 5  1  0 33.68741 33.97101 90.09190 87.47674
## 6  1  0 33.97101 35.74413 87.47674 84.42130
```

More Examples

[TODO]