Stochastic Process Model (SPM)

Ilya Y. Zhbannikov

2015-10-08

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 some parameters. Specifically, developed R-package spm allows (i) one-dimensional SPM; (ii) multiple dimensional SPM; (iii) data simulation for one- and multiple dimensions.

Data description

Data represents a typical clinical trait data and must be presented in form of two datasets: longitudinal dataset, in which one record represents a single observation, and vital statistics, where one record represents a patient and other related information. Longitudinal dataset contains person id, status (dead/alive), time and measurements across the covariates. The spm can handle infinite number of covariates but usually 5-7 covariates are enough in practice.

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

```
##   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 <- sim_discrete(N=10, ystart=c(80), k=1)
head(data)
```

```
##      id xi t1 t2   par1_1   par1_2
## [1,]  1  0 30 31 80.00000 75.39909
## [2,]  1  0 31 32 75.39909 70.50492
## [3,]  1  0 32 33 70.50492 70.87460
## [4,]  1  0 33 34 70.87460 79.30900
## [5,]  1  0 34 35 79.30900 81.45933
## [6,]  1  0 35 36 81.45933 81.67704
```

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(N=5,ystart = c(50))
head(data)
```

```
##   id xi       t1       t2        y   y.next
## 1  1  0 75.57755 77.23481 52.41572 54.79142
## 2  1  0 77.23481 78.38687 54.79142 57.78377
## 3  1  0 78.38687 79.24429 57.78377 57.11729
## 4  1  0 79.24429 80.14292 57.11729 60.69150
## 5  1  0 80.14292 80.55143 60.69150 60.56253
## 6  1  0 80.55143 81.18079 60.56253 64.98930
```

Discrete case

In discrete case, we use the following assumptions:

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

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

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:

```r
library(spm)
data <- sim_discrete(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 70.20654 94.00000 92.21295
## [2,]  1  0 31 32 70.20654 61.26589 92.21295 91.94005
## [3,]  1  0 32 33 61.26589 52.42369 91.94005 89.90354
## [4,]  1  0 33 34 52.42369 50.78972 89.90354 82.90639
## [5,]  1  0 34 35 50.78972 54.48376 82.90639 86.99052
## [6,]  1  0 35 36 54.48376 45.13650 86.99052 83.30642
```

Continuous

For continuous case:

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

```
##   id xi       t1       t2        y   y.next
## 1  1  0 47.08371 48.21060 74.53240 71.47093
## 2  1  0 48.21060 49.91341 71.47093 72.69258
## 3  1  0 49.91341 51.77316 72.69258 77.22813
## 4  1  0 51.77316 52.16391 77.22813 75.48182
## 5  1  0 52.16391 52.46469 75.48182 77.54071
## 6  1  0 52.46469 53.71344 77.54071 78.19149
```

More Examples