

Test PMLSeg

Ninh

2024-08-09

This document present result of several test of the PMLSeg package consisting:

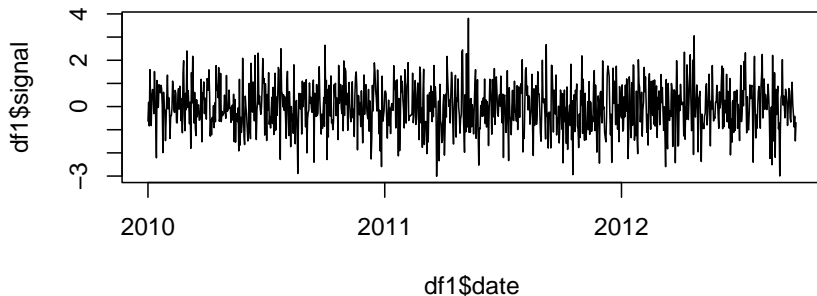
- ▶ Test of the `Segmentation` function in case with/ without offsets for the following examples :
 - ▶ Ex1 : zero mean + IID noise
 - ▶ Ex2 : periodic mean + IID noise
 - ▶ Ex3 : periodic mean + monthly variance
- ▶ Test of other functions such as:
 - ▶ `PlotSeg` to visualize segmentation results
 - ▶ `Cluster_screening` to detect the group of close change-points (usually due to the outliers) and check if it is needed to keep or remove the cluster.
 - ▶ `Validation` to validate the detected changepoints with the help of metadata.

Generate example data to test

Ex1 time series

```
set.seed(1)
length_series = 1000
df1 = data.frame(date = seq.Date(from = as.Date("2010-01-01"),
                                to = as.Date("2010-01-01")+(length_series-1),
                                by = "day"),
                 signal = rnorm(n = length_series, mean = 0, sd = 1))

plot(df1$date, df1$signal, type = "l")
```



```
head(df1, 3)
#>      date      signal
#> 1 2010-01-01 -0.6264538
#> 2 2010-01-02  0.1836433
#> 3 2010-01-03 -0.8356286
```

Generate example data to test

Ex2 time series : add the functional with 4 Fourier series with coefficient = 1

```
library(dplyr)
```

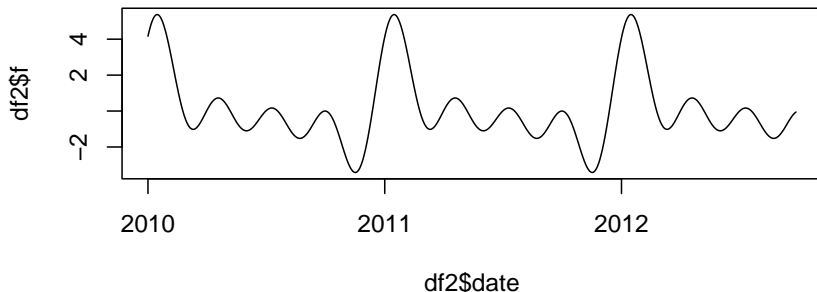
```
T <- 365.25
```

```
df2 <- df1 %>%
```

```
  mutate(t = as.numeric(date - date[1])+1,  
         f = rowSums(sapply(1:4, function(i) cos(i*t*(2*pi)/T) + sin(i*t*(2*pi)/T))),  
         signal = signal + f)
```

```
head(df2, 3)
```

```
#>      date      signal t      f  
#> 1 2010-01-01 3.541048 1 4.167502  
#> 2 2010-01-02 4.509279 2 4.325636  
#> 3 2010-01-03 3.638312 3 4.473941  
plot(df2$date, df2$f, type = "l")
```



Generate example data to test

Ex3 time series : add the functional with 4 Fourier series with coefficient = 1

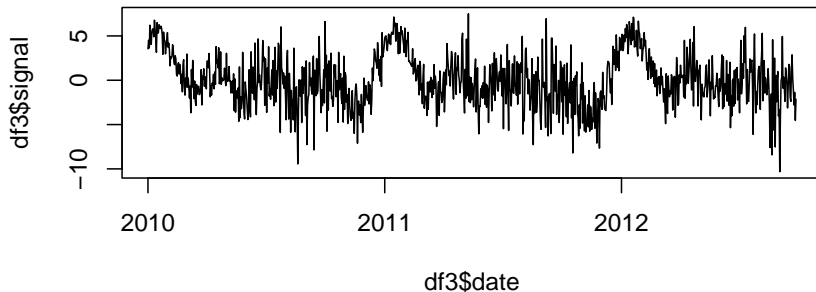
```
std = c(1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 2.5, 2, 1.5)
```

```
df3 <- df1 %>%  
  mutate(sd = std[as.numeric(format(date, "%m"))],  
         signal = signal * sd ) %>%  
  mutate(signal = signal + df2$f)
```

```
head(df3, 3)
```

```
#>      date    signal sd  
#> 1 2010-01-01 3.541048 1  
#> 2 2010-01-02 4.509279 1  
#> 3 2010-01-03 3.638312 1
```

```
plot(df3$date, df3$signal, type = "l")
```



Preliminary setting

Harmonize format of 3 dataframes to test:

```
df2 <- df2 %>% select(date, signal)
df3 <- df3 %>% select(date, signal)
```

```
names(df1)
#> [1] "date"    "signal"
names(df2)
#> [1] "date"    "signal"
names(df3)
#> [1] "date"    "signal"
```

Randomly select two position of changepoints :

```
jump_ind = c(200, 600)
jump_ind
#> [1] 200 600

jump_series = rep(0, length_series)

for (i in jump_ind) {
  jump_series[1:i] <- jump_series[1:i] + 1
}
plot(jump_series, type = "l")
```

