Using R for Step Detection
Original 2012-05-25

Srikesh G. Arunajadai

June 2, 2013

Abstract

The codes were run in Mac OSX and Windows XP without problems. We cannot guarantee its performance in other platforms/ systems. Please do report any bugs to the author but unfortunately we cannot guarantee a timely solution.
In the Folder :

1. Manual
2. GLS-STEP.R
3. MyCode.R
4. MyCode2.R
5. *msProcess*_1.0.7.*tar.gz*

Installing the package msProcess

As the package *msProcess* is no longer actively maintained by its author as of this writing, the archived version of the package is provided with this code. The package is found in the file *msProcess*_1.0.7.*tar.gz*. This will be a one time process to install the package (or whenever you update R).

1. Install the latest version of R from http://www.r-project.org/
2. You are provided *msProcess*_1.0.7.*tar.gz* along with this code.
3. Note down the path where you have saved *msProcess*_1.0.7.*tar.gz*
4. In R type the following command : *msProcess.install()*
5. Next in R type the following : *install.packages("path/filename",repos=NULL,type="source")* where path is the path you noted down earlier and filename is *msProcess*_1.0.7.*tar.gz*
6. The above steps should have installed msProcess in R.

1 Step Detection

1. Assume all relevant codes, datasets are in the same directory.
2. Change the working directory to where you have stored the code **GLS-STEP.R** using *setwd()*.
3. Source the code in by

```
source("GLS-STEP.R")
```

You can check if the codes are sourced in by

```
ls()
```

4. Input the sample data set by

   ```
   RNA<-read.table("RNAtrace1.txt",sep="\t",header=TRUE)
   ```

5. You can see the first few lines of the data by

   ```
   head(RNA)
   ```

6. Set window sizes by

   ```
   w<-c(seq(10,90,by=10),seq(100,1000,by=25))
   ```

7. Set

   ```
   y<-RNA[,1]
   ```

8. Set

   ```
   times<-RNA[,2]
   ```

9. Plot your trace and check

   ```
   plot(y,type="l")
   ```

10. Run the command

    ```
    check.packages()
    ```

    to install/ load required R packages.

11. Getting the η statistics:

    ```
    chpt0<-get.zstat(y,times,w)
    ```

12. Doing the step Fit :

    ```
    chpt1<-ChPt(chpt0)
    ```

13. Plot Step :

    ```
    plot.step(chpt0,chpt1,type="bic")
    ```

    You can choose $type =$ "aic" for an AIC fit.

14. To get the results:

    ```
    RNA.RESULT1<-get.results(chpt1,type="bic")
    ```

    You can choose $type =$ "aic" for an AIC fit. Check results by

```
RNA.RESULT1
```

15. If you want to plot to a pdf file:

```
plot.step(chpt0,chpt1,type="bic",pdfname="MyRNAPlot.pdf")
```

16. The command

```
myModel<-get.model(chpt1,type="bic")

names(myModel)

[1] "betas"      "betas.vcov" "tstat"      "fit"        "res"        "sigma2"
     "pval"       "aic"        "bic"        "AR-order"   "AR-Coefs"
```

As the names imply these are - Step size estimates, the variance-covariance matrix of the step estimates, the t-statistic for the steps for hypothesis testing, the fitted step, the residuals from the fit, the variance of the noise, the p-value of the steps, the aic value, the bic value, The AR-order of the noise and the AR coefficients.

17. One can look at these value as follows. So to see the AR order of the noise one types

```
MyModel$AR-order
```

and to view the BIC value

```
myModel$bic
```

18. To look at the ACF and PACF of residuals

```
acf(myModel$res$)
```

and

```
pacf(myModel$res$)
```

2 Step Detection - For Large Traces

It can be very time consuming. We suggest that you run the code in the background in the UNIX shell. Save the following code in say *MyCode.R*

```
source("GLS-STEP.R")
check.packages()
RNA<-read.table("RNAtrace1.txt",sep="\t",header=TRUE)
w<-c(seq(10,90,by=10),seq(100,1000,by=25))
Y<-RNA[,1]
times<-RNA[,2]
chpt0<-get.zstat(y,times,w)
chpt1<-ChPt(chpt0)
plot.step(chpt0,chpt1,pdfname="RNAFIG1.pdf")
RNA.RESULT1<-get.results(chpt1)
write.table(RNA.RESULT1,file="RNA-RESULT1.txt",sep="\t",header=TRUE)
myModel<-get.model(chpt1,type="bic")
save(myModel,file="All-results-1.Rdata")
#load("All-results-1.Rdata")
```

1. Open X11
2. Change the working directory to where the data, codes are located
3. Type in

```
nohup R CMD BATCH MyCode.R myLOG.txt &
```

4. For Windows try the following. Make sure you have the corrects paths.

```
"PATH\R.exe" CMD BATCH --vanilla --slave "PATH\MyCode.R"
```

5. Once the analysis is done check *RNA-RESULT1.txt* for the results and *RNAFIG1.pdf* for the plot.

3 To Run Multiple Traces

To run more than one trace at a time - create a folder say MyTraces and place all your traces in this folder.
Save the following code in say *MyCode2.R*

```
source("GLS-STEP.R")
check.packages()
setwd("MyTraces FOLDER")  # Set the My traces Folder Directory path
file.names<-system("ls",intern=TRUE)   # This should get the names of the trace files in the folder.
for(i in 1:length(file.names)){
 RNA<-read.table(file.names[i],sep="\t",header=TRUE)
 w<-c(seq(10,90,by=10),seq(100,1000,by=25))
 Y<-RNA[,1]
 times<-RNA[,2]
 chpt0<-get.zstat(y,times,w)
 chpt1<-ChPt(chpt0)
 plot.step(chpt0,chpt1,pdfname=paste("RNAFIG-",i,".pdf",sep=""))
 RNA.RESULT1<-get.results(chpt1)
 write.table(RNA.RESULT1,file=paste("RNA-RESULT-",i,".txt",sep=""),sep="\t",header=TRUE)
 myModel<-get.model(chpt1,type="bic")
 save(myModel,file=paste("myModel-",i,".Rdata",sep=""))
 # This is saves as a R file .Rdata, you can load it into R by   load("myModel-1.Rdata")
}
```

1. Open X11
2. Change the working directory to where the data, codes are located
3. Type in

```
nohup R CMD BATCH MyCode2.R myLOG.txt &
```

4. For Windows try the following. Make sure you have the corrects paths.

```
"PATH\R.exe" CMD BATCH --vanilla --slave "PATH\MyCode2.R"
```

5. Once the analysis is done check *RNA-RESULT1.txt* for the results and *RNAFIG1.pdf* for the plot and other files.