

Tutorials on using the R package chngpt

Youyi Fong

October 9, 2017

This tutorial is best read in combination with Fong et al. (2017).

1 Continuous threshold regression models

1.1 Logistic regression example

To estimate a logistic regression model with a hinge-type change point in *NAb_SF162L* for the MTCT dataset, we call

```
> fit=chngptm(formula.1=y~birth, formula.2=~NAb_SF162LS, dat.mtct,
type="hinge", family="binomial", est.method="smoothapprox", var.type="robust",
aux.fit=glm(y ~birth + ns(NAb_SF162LS,3), dat.mtct, family="binomial"))
```

The argument *formula.2* gives the threshold variable and the argument *formula.1* gives the rest of the model. The argument *type* conveys the type of threshold effects one wishes to fit, and the argument *est.method* controls the type of search algorithm to use. If the estimation method is grid search and the argument *verbose* is set to 2, the program will output a plot of log likelihoods versus candidate change points (Figure 2a, Fong et al., 2017). The returned *fit* is an object of type *chngptm*. Calling summary function on the *fit* object produces the following information:

```
Coefficients:
              OR      p.value    (lower  upper)
(Intercept)  0.7026523  0.341429662  0.3388366  1.4571044
birthVaginal  1.2397649  0.523159883  0.6393632  2.4039809
(NAb_SF162LS-chngpt)+ 0.6712371  0.001332547  0.5270730  0.8548327

Threshold:
  26.3%    (lower  upper)
 7.373374  5.472271  8.186464
```

To test whether there is a change point, we call

```
> test=chngpt.test(formula.null=y~birth, formula.chngpt=~NAb_SF162LS, dat.mtct,
type="hinge", family="binomial", main.method="score")
```

The returned *test* is an object of type *htest* and type *chngp.test*. When printed, `print.htest` is called and generates a standard output that reports the estimated change point, the test statistic and :

```
Maximal Score Test
data:  dat.mtct
Maximal statistic = 3.3209, change point = 7.0347, p-value = 0.00284
alternative hypothesis:  two-sided
```

The first line gives the type of test carried out, and it may be maximal likelihood ratio test. In addition, a plot function can be called on the test object to show the score or likelihood ratio statistic as a function of candidate change points.

1.2 Linear regression example

To estimate a linear regression model with a segmented-type change point in *Girth* for the *trees* dataset, we call

```
> fit=chngp(mformula.1=Volume~1, formula.2=~Girth, trees,
type="segmented", family="gaussian", est.method="smoothapprox", var.type="bootstrap",
aux.fit=glm(Volume~ns(Girth,df=3), trees, family="gaussian"))
```

The argument *formula.2* gives the threshold variable and the argument *formula.1* gives the rest of the model. The argument *type* conveys the type of threshold effects one wishes to fit, and the argument *est.method* controls the type of search algorithm to use. When the estimation method is grid search and the argument *verbose* is set to 2, the program will output a plot of log likelihoods versus candidate change points (Figure B.1a, Fong et al., 2017). The returned *fit* is an object of type *chngp.m*. Calling summary function on the *fit* object produces the following information:

```
Coefficients:
              Est      p.value    (lower    upper)
(Intercept) -24.614440 NA -39.336944 -17.530558
Girth        3.993966 NA  3.249892  5.370237
(Girth-chngp)+ 4.266618 NA  2.952945 10.083153

Threshold:
 74.2% (lower upper)
 16.0  12.9  18.0
```

To test whether there is a change point, we call

```
> test=chngp.test(formula.null=Volume~1, formula.chngp=~Girth, trees,
type="segmented", family="gaussian")
```

The returned *test* is an object of type *htest* and type *chngpt.test*. When printed, `print.htest` is called and generates a standard output that reports the estimated change point, the test statistic and :

```
Maximum of Likelihood Ratio Statistics
data:  trees
Maximal statistic = 17.694, change point = 15.388, p-value = 0.00014
alternative hypothesis:  two-sided
```

The first line gives the type of test carried out, and it is maximal likelihood ratio test here, which is the default. In addition, a plot function can be called on the test object to show the score or likelihood ratio statistic as a function of candidate change points.

2 Discontinuous threshold regression models

To be added.

References

Fong, Y., Huang, Y., Gilbert, P. and Permar, S. (2017), “chngpt: threshold regression model estimation and inference,” *BMC Bioinformatics*, in press.