Tutorials for the R package chngpt

Youyi Fong

March 20, 2018

This tutorial illustrates the use of *chngpt* (Fong et al., 2017a) with code examples. Estimation methods are described in Fong (2018); Fong et al. (2017b), hypothesis testing methods are described in Fong et al. (2015).

1 Continuous threshold regression models

1.1 Threshold linear regression example 1

To estimate a threshold linear regression model with a segmented-type change point for the relationship between *V3_BioV3B* and *NAb_score* in the *MTCT* dataset, we call

```
fit=chngptm (formula.1=V3_BioV3B~1, formula.2=~NAb_score, dat.mtct.2,
  type="segmented", family="gaussian",
  est.method="fastgrid", var.type="bootstrap", save.boot=TRUE)
```

- `formula.2` and `formula.1`: threshold variable and the rest of the model
- `type`: type of threshold model to fit
- `est.method` defaults to *fastgrid* and is recommended
- `var.type`: *bootstrap* method is recommended here
- `save.boot`: saves bootstrap samples for plotting bootstrap distributions

Calling `summary(fit)`, we get

```
Change point model type:  segmented

Coefficients:
                      Est p.value          (lower          upper)
(Intercept)      -22.33152      NA       -30.07675       -14.58628
NAb_score         67.23925      NA        49.98398        84.49452
(NAb_score-chngpt)+ -64.83129    NA       -81.61413       -48.04845

Threshold:
    31.8%    (lower    upper)
0.4653923 0.4535000 0.4772845
```

To get an estimate of the slope after threshold, we call

```
est=lincomb(fit.mtct, comb=c(0,1,1), alpha=0.05); print(est)
```

and get

```
                         95%            95%
 2.40795883 -0.06780353   4.88372120
```

Calling `plot(fit, which=1)` and `plot(fit, which=3)`, we get the two plots on the left-hand side of Figure 1. Changing `est.method` to `smoothapprox` in the model fit led us to the two plots on the right-hand side.
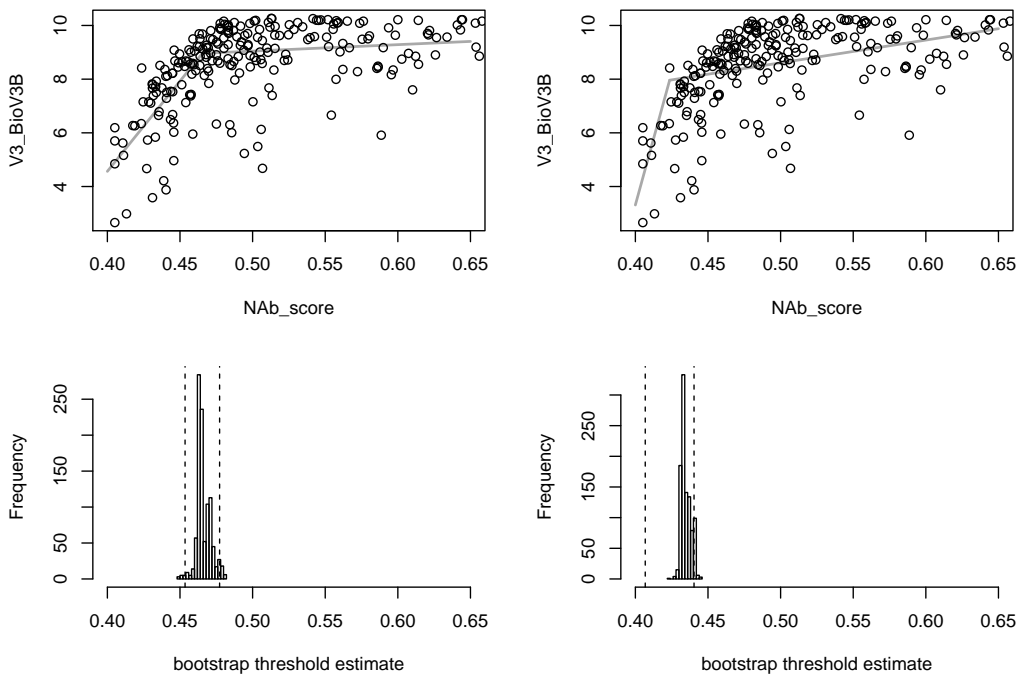

Figure 1: This is a replicate of Fong (2018) Figure 1. Left: results by fast grid search; right: results by smooth approximation search. Top: scatterplots with fitted models (gray lines); bottom: bootstrap distributions of the threshold estimate from 10^3 replicates. The dashed lines correspond to the 95% symmetric bootstrap confidence interval.

1.2 Threshold linear regression example 2

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

```
fit=chngptm(formula.1=Volume~1, formula.2=~Girth, data=trees,
type="segmented", family="gaussian",
 var.type="bootstrap", weights=NULL)
```

- `formula.2` and `formula.1`: threshold variable and the rest of the model
- `type`: type of threshold model to fit
- `var.type`: *bootstrap* method is recommended for confidence interval
- `weights` can be supplied

Calling `summary(fit)`, we get

```
Change point model type:  segmented

Coefficients:
                  Est p.value (lower.(Intercept) upper).(Intercept)
(Intercept)     -24.614440      NA         -37.580354         -11.648527
Girth             3.993966      NA           2.785558           5.202373
(Girth-chngpt)+   4.266618      NA           1.765770           6.767467

Threshold:
 74.2% (lower upper)
  16.0   12.9   19.1
```
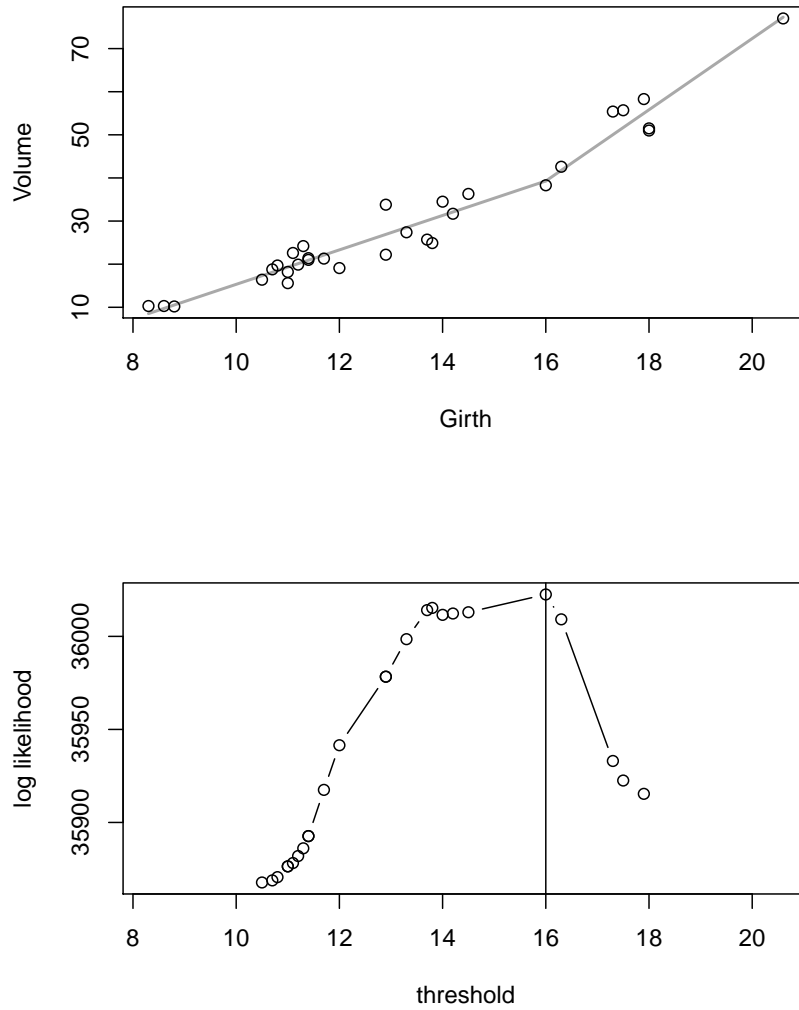
Calling `plot(fit)`, we get Figure 2.

Figure 2: (a) Scatterplot of timber volume vs girth. The gray line shows the fitted segmented model. (b) Log likelihood of the submodel versus threshold parameter.

To test whether there is a change point, we call

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

When printed, we get

```
        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.

1.3   Threshold logistic regression example 1

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

```
library(splines)
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"))
```

- `formula.2` and `formula.1`: threshold variable and the rest of the model
- `type`: type of threshold model to fit
- `est.method`: *smoothapprox* is recommended
- `var.type`: *robust* is recommended for confidence interval
- `aux.fit`: required for *robust* variance estimation

Calling `summary(fit)`, we get

```
Change point model type:  hinge


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")
```

When printed, we get

```
        Maximum of Score Statistics


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.

2 Discontinuous threshold regression models

The package also provides some support for estimation and hypothesis testing under discontinuous threshold regression models. What is missing, though, is confidence intervals for parameter estimates.

References

Fong, Y. (2018), "Fast Bootstrap Confidence Intervals for Continuous Threshold Linear Regression," *Journal of Computational and Graphical Statistics*, in press.

Fong, Y., Di, C. and Permar, S. (2015), "Change point testing in logistic regression models with interaction term," *Statistics in medicine*, 34, 1483–1494.

Fong, Y., Huang, Y., Gilbert, P. and Permar, S. (2017a), "chngpt: threshold regression model estimation and inference," *BMC Bioinformatics*, 18, 454–460.

Fong, Y., Chong, D., Huang, Y. and Gilbert, P. (2017b), "Model-robust Inference for Continuous Threshold Regression Models," *Biometrics*, 73, 452–462.