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

\[
\text{fit=chngptm (formula.1=V3_{BioV3B} ~ 1, formula.2=\sim NAb_score, dat.mtct.2, type="segmented", family="gaussian", est.method="fastgrid", var.type="bootstrap", save.boot=TRUE)}
\]

- 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:

<table>
<thead>
<tr>
<th></th>
<th>Est</th>
<th>p.value</th>
<th>(lower)</th>
<th>(upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-22.33152</td>
<td>NA</td>
<td>-30.07675</td>
<td>-14.58628</td>
</tr>
<tr>
<td>NAb_score</td>
<td>67.23925</td>
<td>NA</td>
<td>49.98398</td>
<td>84.49452</td>
</tr>
<tr>
<td>(NAb_score-chngpt)+</td>
<td>-64.83129</td>
<td>NA</td>
<td>-81.61413</td>
<td>-48.04845</td>
</tr>
</tbody>
</table>

Threshold:

31.8%  (lower  upper)

0.4653923 0.4535000 0.4772845
To get an estimate of the slope after threshold, we call

\[
est = \text{lincomb}(\text{fit.mtct}, \text{comb} = c(0, 1, 1), \text{alpha} = 0.05); \text{print(est)}\]

and get

\[
\begin{align*}
95\% & \quad 95\% \\
2.40795883 & -0.06780353 & 4.88372120
\end{align*}
\]

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](image)

**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 \textit{Girth} for the \textit{trees} dataset, we call

\begin{verbatim}
fit=chngptm(formula.1=Volume~1, formula.2=~Girth, data=trees, type="segmented", family="gaussian", var.type="bootstrap", weights=NULL)
\end{verbatim}

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

Calling \texttt{summary(fit)}, we get

\textbf{Change point model type: segmented}

\begin{tabular}{lcccc}
  \hline
  & Est & p.value & (lower.(Intercept) & upper.(Intercept) \\
  \text{(Intercept)} & -24.614440 & NA & -37.580354 & -11.648527 \\
  \text{Girth} & 3.993966 & NA & 2.785558 & 5.202373 \\
  \text{(Girth-chngpt)+} & 4.266618 & NA & 1.765770 & 6.767467 \\
  \hline
\end{tabular}

\textbf{Threshold:}

74.2\% (lower upper)

16.0 12.9 19.1

Calling \texttt{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

```r
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

```r
library(splines)
fit=chngptm(formula.1=y\sim\text{birth}, formula.2=~NAb\_SF162LS, dat.mtct,
  type="hinge", family="binomial",
est.method="smoothapprox", var.type="robust",
aux.fit=glm(y\sim\text{birth} + \text{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:**

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>p.value</th>
<th>(lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.7026</td>
<td>0.3414</td>
<td>0.3388 366</td>
<td>1.4571 44</td>
</tr>
<tr>
<td>birthVaginal</td>
<td>1.2398</td>
<td>0.5232</td>
<td>0.64 3632</td>
<td>2.4039 81</td>
</tr>
<tr>
<td>( NAb_SF162LS-chngpt )+</td>
<td>0.6712</td>
<td>0.0013</td>
<td>0.5270 327</td>
<td>0.8548 327</td>
</tr>
</tbody>
</table>

**Threshold:**

- 26.3\% (lower 7.373374, upper 8.186464)

To test whether there is a change point, we call

```r
test=chngpt.test(formula.null=y\sim\text{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


