Model-based Recursive Partitioning for Subgroup Analyses

Torsten Hothorn; joint work with Heidi Seibold and Achim Zeileis
Subgroup analyses

Identifying groups of patients for whom the treatment has a different effect than for others.

Effect is:
- Stronger
- Lower
- Contrary

than the average treatment effect.

Suitable models promise better prediction of treatment effect and thus individualised treatments.
Situations of interest for subgroup analyses

EMA (2014, 1):

1. “The clinical data presented are overall statistically persuasive with therapeutic efficacy demonstrated globally. It is of interest to verify that the conclusions of therapeutic efficacy (and safety) apply consistently across subgroups of the clinical trial population.”

2. “The clinical data presented are overall statistically persuasive but with therapeutic efficacy or benefit/risk which is borderline or unconvincing and it is of interest to identify post-hoc a subgroup, where efficacy and risk-benefit is convincing.”

3. “The clinical data presented fail to establish statistically persuasive evidence but there is interest in identifying a subgroup, where a relevant treatment effect and compelling evidence of a favourable risk-benefit profile can be assessed.”
Subgroup analyses

Goal: Find predictive factors

\[ \wedge \text{ covariate } \times \text{ treatment interactions} \]
Covariate $\times$ treatment interactions

- Subgroups known: incorporate subgroup $\times$ treatment interaction in linear predictor.
- (Few) Categorical covariates at few levels: incorporate covariate $\times$ treatment interaction in linear predictor.
- Many, potentially numeric covariates: Interactions hard to interpret, difficult to derive subgroups.
- Common approach: Automated interaction detection.
- BUT: Ordinary trees don’t know treatment effect parameters.
- We need both parametric models AND trees.

Solution: Model-based recursive partitioning (MOB).
MOB Basics

MOB: Model-based recursive partitioning (Zeileis, Hothorn, Hornik, JCGS, 2008, 3)

Start with model $\mathcal{M}((Y, X), \vartheta)$ with

$$
\vartheta = \begin{pmatrix}
\alpha \\
\beta \\
\gamma \\
\nu
\end{pmatrix}
$$

intercept(s) 

treatment effect 

other parameter(s) of interest 

nuisance parameter(s),

which fits data $(Y, X) \in (Y, X)$. 

MOB Basics

Estimation

\[
\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^{N} \Psi((y, x)_i, \theta)
\]

or equivalently solving the score equation

\[
\sum_{i=1}^{N} \psi((y, x)_i, \theta) = 0
\]

\(\Psi\): objective function
\(\psi\): score function; gradient of \(\Psi\)
MOB Basics

Maybe the treatment effect is not the same for all patients, but depends on their characteristics (covariates).

⇒ Find partitions $\{B_b\}$ $(b = 1, \ldots, B)$ based on patient characteristics $Z = (Z_1, \ldots, Z_J) \in \mathcal{Z}$

⇒ Fit separate models $\mathcal{M}((Y, X), \vartheta(b))$ in partitions.

$$\vartheta(b) = (\alpha(b), \beta(b), \gamma, \nu)\top$$
If partition \( \{ \mathcal{B}_b \} \) is known, the partitioned model parameters \( \vartheta(b) \) could be estimated by minimizing the segmented objective function:

\[
(\hat{\vartheta}(b))_{b=1,\ldots,B} = \arg\min_{\vartheta(b)} \sum_{i=1}^{N} \sum_{b=1}^{B} 1(z_i \in \mathcal{B}_b) \psi((y, x)_i, \vartheta(b))
\]

Subgroup-specific intercept and treatment parameters can be written as functions of the partitioning variables

\[
\alpha(z) = \sum_{b=1}^{B} 1(z \in \mathcal{B}_b) \cdot \alpha(b) \quad \text{and} \quad \beta(z) = \sum_{b=1}^{B} 1(z \in \mathcal{B}_b) \cdot \beta(b).
\]
Partitioning

How to find the partitions?

⇒ Test:

\[ H_0^{\alpha,j} : \psi_\alpha((Y, X), \hat{\theta}) \perp Z_j \]
\[ H_0^{\beta,j} : \psi_\beta((Y, X), \hat{\theta}) \perp Z_j, \quad j = 1, \ldots, J \]

\( \psi_\alpha, \psi_\beta \) partial derivatives of \( \Psi \) with respect to \( \alpha/\beta \).

– Partition if global permutation p-value smaller than significance level

– Use as split variable the one with the smallest p-value
Example: Linear model

Two-arm trial comparing active (A) to control (C):

\[ Y|X = x \sim \mathcal{N}(\alpha + \beta x_A + \gamma x_{\text{stratum}}, \sigma^2). \]

\[ Y|X = x, Z = z \sim \mathcal{N}(\alpha(z) + \beta(z)x_A + \gamma x_{\text{stratum}}, \sigma^2), \]

\[
\begin{align*}
\psi((y, x), \hat{\theta}) &= \left( \begin{array}{c}
\frac{\partial \psi((y, x), \theta)}{\partial \alpha} \\ \frac{\partial \psi((y, x), \theta)}{\partial \beta} \\
\end{array} \right)_{\theta = \hat{\theta}}^\top \\
&= \frac{1}{\sigma^2} \left( \begin{array}{c}
y - (\hat{\alpha} + \hat{\beta}x_A + \hat{\gamma}x_{\text{stratum}}) \\
(y - (\hat{\alpha} + \hat{\beta}x_A + \hat{\gamma}x_{\text{stratum}})) \cdot x_A \\
\end{array} \right)^\top
\end{align*}
\]
Example: Linear model

Prognostic
\[ \alpha(1) \neq \alpha(2) ; \beta(1) = \beta(2) \]

Predictive
\[ \alpha(1) = \alpha(2) ; \beta(1) \neq \beta(2) \]

Both
\[ \alpha(1) \neq \alpha(2) ; \beta(1) \neq \beta(2) \]
Example: Linear model

Linear model:
\[ Y|X = x \sim \mathcal{N}(\alpha + \beta x_B, \sigma^2) \]

Data generating process (Loh, He, Man, 2014, 4):
\[ Y|X = x, Z = z \sim \mathcal{N}(1.9 + 0.2 \cdot x_A + 1.8 \cdot 1(z_1 < 0) + 3.6 \cdot 1(z_1 > 0) \cdot x_A, 0.7) \]
Example: Linear model

Linear model:
\[ Y|X = x \sim \mathcal{N}(\alpha + \beta x_B, \sigma^2) \]

Data generating process:
\[ Y|X = x, Z = z \sim \mathcal{N}(1.9 + 0.2 \cdot x_A + 1.8 \cdot 1(z_1 < 0) + 3.6 \cdot 1(z_1 < 0) \cdot x_A, 0.7) \]
Example: Linear model

Linear model:
\[ Y | X = x \sim \mathcal{N}(\alpha + \beta x_B, \sigma^2) \]

Data generating process:
\[ Y | X = x, Z = z \sim \mathcal{N}(2 \cdot x_A + \mathbb{1}(z_1 > 0), 0.7) \]
Partitioning effects of Riluzole on ALS patients

PRO-ACT database (2)

- Amyotrophic lateral sclerosis (ALS) patients
- Data of several clinical trials
- Treatment of interest: Riluzole
- Primary endpoints of interest:
  - ALS Functional Rating Scale (ALSFRS)
  - ALSFRS items
  - Survival time
Question

Riluzole modestly prolongs life expectancy

But: Are there any groups of patients for whom it is better or worse?

⇒ Subgroup analysis
ALS functional rating scale: Measure of functional status of ALS patients

Sum-score of ten items (0 < 1 < 2 < 3 < 4):

- speech
- salivation
- swallowing
- handwriting
- cutting food and handling utensils,
- dressing and hygiene
- turning in bed and adjusting bed clothes
- walking
- climbing stairs
- breathing
\[ \mathbb{E} \left( \frac{\text{ALSFRS}_6}{\text{ALSFRS}_0} \middle| X = x \right) = \frac{\mathbb{E}(\text{ALSFRS}_6 \middle| X = x)}{\text{ALSFRS}_0} = \exp\{\alpha + \beta x_R\} \]

or equivalently

\[ \mathbb{E}(\text{ALSFRS}_6 \middle| X = x) = \exp\{\alpha + \beta x_R\} \cdot \exp\{\log(\text{ALSFRS}_0)\} \]

GLM with log-link and offset
ALSFRS

Results

- Time onset treatment: 
  - $p < 0.001$
  - $\leq 467 > 467$

- FVC: 
  - $p < 0.001$
  - $\leq 3.99 > 3.99$

- Phosphorus: 
  - $p < 0.001$
  - $\leq 1 > 1$

Values and confidence intervals:

- ALSFRS
  - $\alpha = -0.2397$ ($-0.2680, -0.2122$)
  - $\beta = 0.0276$ ($-0.0077, 0.0632$)
  - $n = 625$

- ALSFRS
  - $\alpha = -0.1344$ ($-0.1599, -0.1096$)
  - $\beta = -0.0396$ ($-0.0726, -0.0064$)
  - $n = 378$

- ALSFRS
  - $\alpha = -0.1105$ ($-0.1410, -0.0809$)
  - $\beta = 0.0284$ ($-0.0055, 0.0630$)
  - $n = 292$

- ALSFRS
  - $\alpha = -0.1266$ ($-0.1435, -0.1101$)
  - $\beta = -0.0046$ ($-0.0242, 0.0151$)
  - $n = 1239$
ALSFRS items

Unidimensionality of ALSFRS is questionable.

Two components to consider:

1. Baseline adjustment
   Adjust for item score at beginning of treatment
   ⇒ compute separate models

2. Multivariate primary endpoint
   Look at 10 items simultaneously
   ⇒ compute 10 item-models in every node

Score matrix of dimension $n \times p \cdot 10$
ALSFRS items

Each item ordinal with values 0, ..., 4
⇒ proportional odds model adjusted for baseline:

\[
P(Y_6 \leq r | Y_0 = k, X = x) = \frac{\exp(\alpha_{rk} - \beta_k x_R)}{1 + \exp(\alpha_{rk} - \beta_k x_R)}
\]

for \( k = 0, ..., 4 \)

Compute stratified permutation tests treating baseline item values as blocks.
ALSFRS items

Results

1. time_onset_treatment
   \( p < 0.001 \)
   - \( \leq 584 \)
   - \( > 584 \)

2. FVC
   \( p < 0.001 \)
   - \( \leq 2.73 \)
   - \( > 2.73 \)

3. \( n = 348 \)
4. \( n = 1005 \)

5. lymphocytes
   \( p = 0.006 \)
   - \( \leq 21.4 \)
   - \( > 21.4 \)

6. \( n = 407 \)
7. \( n = 774 \)
## ALSFRS items

### Results

<table>
<thead>
<tr>
<th>Item</th>
<th>No.</th>
<th>Start</th>
<th>Node 3</th>
<th>Node 4</th>
<th>Node 6</th>
<th>Node 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>-0.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.33</td>
<td>0.04</td>
<td>-0.27</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.84</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salivation</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>0.22</td>
<td>0.43</td>
<td>1.36</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.15</td>
<td>-0.26</td>
<td>-0.24</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.49</td>
<td>-0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swallowing</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>0.35</td>
<td>-0.89</td>
<td>1.51</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.57</td>
<td>-0.36</td>
<td>-0.40</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.62</td>
<td></td>
<td></td>
<td>0.15</td>
<td>0.28</td>
</tr>
<tr>
<td>Handwriting</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>-1.45</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>-1.15</td>
<td>-0.36</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>-0.54</td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.13</td>
<td>0.14</td>
<td>-0.08</td>
<td>-0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-0.10</td>
<td>0.04</td>
<td></td>
<td></td>
<td>-0.14</td>
</tr>
<tr>
<td>Cutting</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>-0.01</td>
<td>-0.79</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>0.15</td>
<td></td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.03</td>
<td>-0.07</td>
<td>0.10</td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.13</td>
<td>-0.09</td>
<td>-0.14</td>
<td></td>
<td>-0.21</td>
</tr>
</tbody>
</table>
### ALSFRS items

#### Results

<table>
<thead>
<tr>
<th>Item</th>
<th>No.</th>
<th>Start</th>
<th>Node 3</th>
<th>Node 4</th>
<th>Node 6</th>
<th>Node 7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hygiene</strong></td>
<td>6</td>
<td>0</td>
<td>-0.11</td>
<td>-0.37</td>
<td>0.14</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>-0.22</td>
<td>-0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>0.26</td>
<td>0.01</td>
<td>0.14</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bed</strong></td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>0.15</td>
<td>-0.32</td>
<td>-0.03</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>-0.21</td>
<td>-0.10</td>
<td>-0.12</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>-0.11</td>
<td>-0.35</td>
</tr>
<tr>
<td><strong>Walking</strong></td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>0.48</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>0.11</td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>0.51</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td><strong>Stairs</strong></td>
<td>9</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>-0.02</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>-0.80</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.26</td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1.01</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>Respiratory</strong></td>
<td>10</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.58</td>
<td>-0.08</td>
</tr>
</tbody>
</table>
Survival time

We present two ways of modeling:

- Weibull model
  (parametric survival model)
- Cox model
  (semiparametric survival model)
Weibull model

\[ P(Y \leq y|X = x) = F \left( \frac{\log(y) - \alpha_1 - \beta x_R}{\alpha_2} \right) \]

with \( F \) cumulative distribution function of Gompertz distribution

and \( \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \) intercept, scale parameter

\( \alpha \) defines the shape of the baseline hazard \( \Rightarrow \) use as "intercept"
Weibull model

Results

<table>
<thead>
<tr>
<th>Group</th>
<th>Value (CI)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riluzole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Riluzole</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age ≤ 55.7</td>
<td>α₁ 7.017 (6.784, 7.251)</td>
<td>570</td>
</tr>
<tr>
<td></td>
<td>β₀ -0.031 (-0.228, 0.166)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log(α₂) -1.031 (-1.268, -0.794)</td>
<td></td>
</tr>
<tr>
<td>time_onset_treatment ≤ 456</td>
<td>α₁ 6.3939 (6.2970, 6.4909)</td>
<td>732</td>
</tr>
<tr>
<td></td>
<td>β₀ 0.1162 (-0.0096, 0.2421)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log(α₂) -0.5774 (-0.6696, -0.4852)</td>
<td></td>
</tr>
<tr>
<td>time_onset_treatment &gt; 456</td>
<td>α₁ 6.671 (6.565, 6.777)</td>
<td>1020</td>
</tr>
<tr>
<td></td>
<td>β₀ 0.043 (-0.090, 0.175)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log(α₂) -0.533 (-0.615, -0.451)</td>
<td></td>
</tr>
</tbody>
</table>
Cox model

\[ \lambda(y|x) = \lambda_0(y) \exp(\beta x_R) \]

Objective function: negative partial likelihood (without baseline hazard)

- No classical score function
- Use surrogate score function
  - Martingale residuals (as score with respect to the baseline hazard/"intercept")
  - Score residuals (as score with respect to \( \beta \))
Cox model

Results

- **Age**
  - \( p < 0.001 \)
  - \( \leq 55.7 \):
    - Value (CI): 0.11 (-0.46, 0.68)
    - \( n = 572 \)
  - \( > 55.7 \):
    - Value (CI): -0.260 (-0.537, 0.017)
    - \( n = 982 \)

- **Time on onset treatment**
  - \( p < 0.001 \)
  - \( \leq 456 \):
    - Value (CI): -0.164 (-0.388, 0.061)
    - \( n = 732 \)
  - \( > 456 \):
    - Value (CI): -0.159 (-0.382, 0.065)
    - \( n = 1020 \)

- **Riluzole vs. No Riluzole**
  - **Age**
    - \( p < 0.001 \)
    - \( \leq 55.7 \):
      - Value (CI): 0.11 (-0.46, 0.68)
      - \( n = 572 \)
    - \( > 55.7 \):
      - Value (CI): -0.260 (-0.537, 0.017)
      - \( n = 982 \)
Computational details

PRO-ACT data are available at https://nctu.partners.org/ProACT/.

The source code for reading and cleaning the database is provided in the TH.data package.

All computations were conducted using partykit (version 0.8-2) in the R system for statistical computing (version 3.1.2).
library("partykit")

## Function to compute Weibull model and return score matrix
mywb <- function(data, weights, parm) {
  mod <- survreg(Surv(survival.time, cens) ~ Riluzole,
                 data = data, subset = weights > 0,
                 dist = "weibull")
  ef <- as.matrix(estfun(mod)[,parm])
  ret <- matrix(0, nrow = nrow(data), ncol = ncol(ef))
  ret[weights > 0,] <- ef
  ret
}

## Compute tree
tree <- ctree(fm, data = data, ytrafo = my.wb,
              control = ctree_control(maxdepth = 2,
                                      testtype = "Bonferroni"))
Summary

- MOB partitions a large class of models suitable for the treatment effect on the primary endpoint of interest.
- Score functions capture instabilities and thus help to identify predictive and prognostic variables.
- It is hard to differentiate between predictive and prognostic variables.
- Permutation tests suitable for all models (distribution free) with good small sample properties and error control.
- Multiplicity adjustment for subgroup-specific treatment effects unclear.
# Literature

**European Medicines Agency**  *EMA Guideline on the investigation of subgroups in confirmatory clinical trials (draft)*  

**Massachusetts General Hospital, Neurological Clinical Research Institute**  *Pooled Resource Open-Access ALS Clinical Trials Database*  

**A. Zeileis, T. Hothorn, K. Hornik**  *Model-Based Recursive Partitioning*  

**W. Loh, X. He, M. Man**  *A regression tree approach to identifying subgroups with differential treatment effects*  