Some Applications of Marginally Interpretable Linear Transformation Models for Clustered Observations

Torsten Hothorn
Universität Zürich

Abstract

Owing to their generality, transformation models can be used to set-up and compute many interesting regression models for discrete and continuous responses. This document focuses on the analysis of clustered observations. Marginal predictive distributions are defined by transformation models and their joint normal distribution depends on a structured covariance matrix. Applications with skewed, bounded, and survival continuous outcomes as well as binary and ordered categorical responses are presented. Data is analysed by a proof-of-concept implementation of parametric linear transformation models for clustered observations available in the tram add-on package to the \texttt{R} system for statistical computing.

\textit{Keywords:} conditional mixed models, marginal models, marginal predictive distributions, survival analysis, categorical data analysis.

1. Introduction

The purpose of this document is to compare marginally interpretable linear transformation models for clustered observations (Hothorn 2019) to conventional conditional formulations of mixed-effects models where such an overlap exists. In addition, novel transformation models going beyond the capabilities of conventional mixed-effects models are estimated and interpreted. A proof-of-concept implementation (meaning: the algorithms work but need optimisation, the user interface is very rough) available in package tram (Hothorn 2020) is applied. The results presented in this document can be reproduced from the mtram package vignette

\begin{verbatim}
R> install.packages("tram")
R> vignette("mtram", package = "tram")
\end{verbatim}

2. Normal and Non-normal Mixed-effects Models

First we consider mixed-effects models for reaction times in the sleep deprivation study (Be-
Figure 1: Sleep deprivation: Average reaction times to a specific task over several days of sleep deprivation are given for 18 subjects in Figure 1.

The classical normal linear random-intercept/random-slope model, treating the study participants as independent observations, is fitted by maximum likelihood to the data using the \texttt{lmer()} function from the \texttt{lme4} add-on package (Bates et al. 2015):

\begin{verbatim}
R> library("lme4")
R> sleep_lmer <- lmer(Reaction ~ Days + (Days | Subject),
+ data = sleepstudy, REML = FALSE)
\end{verbatim}

The corresponding conditional model for subject $i$ reads

$$
\mathbb{P}(\text{Reaction} \leq y \mid \text{day}, i) = \Phi \left( \frac{y - \alpha - \beta \text{day} - \alpha_i - \beta_i \text{day}}{\sigma} \right), \quad (\alpha_i, \beta_i) \sim \mathcal{N}_2(0, G(\gamma))
$$

with $\sigma^{-2} G = \Lambda(\gamma) \Lambda(\gamma)^\top$ and

$$
\Lambda(\gamma) = \begin{pmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_3
\end{pmatrix}, \quad \gamma = (\gamma_1, \gamma_2, \gamma_3)^\top.
$$

The same model, however using the alternative parameterisation and an independent (of \texttt{lme4}, only the \texttt{update()} method for Cholesky factors is reused) gradient-based maximisation of the log-likelihood, is estimated in a two-step approach as

\begin{verbatim}
R> sleep_LM <- Lm(Reaction ~ Days, data = sleepstudy)
R> sleep_LMmer <- mtram(sleep_LM, ~ (Days | Subject), data = sleepstudy)
\end{verbatim}
The first call to \texttt{Lm()} computes the equivalent of a normal linear regression model parameterised as a linear transformation model \textit{ignoring} the longitudinal nature of the observations. The purpose is to set-up the necessary model infrastructure (model matrices, inverse link functions, etc.) and to compute reasonable starting values for the fixed effects. The second call to \texttt{mtram()} specifies the random effects structure (here a correlated pair of random intercept for subject and random slope for days) and optimises the likelihood for all model parameters $\vartheta_1, \tilde{\alpha}, \tilde{\beta}$, and $\gamma$ in the model (here also looking at the conditional model for subject $i$)

$$
\mathbb{P}(\text{Reaction} \leq y \mid \text{day}, i) = \Phi \left( \vartheta_1 y + \tilde{\alpha} - \tilde{\beta}_{\text{day}} - \tilde{\alpha}_i - \tilde{\beta}_i \text{day} \right), \quad (\tilde{\alpha}_i, \tilde{\beta}_i) \sim N_2(0, \Lambda(\gamma)\Lambda(\gamma))
$$

that is, all fixed and random effect parameters are divided by the residual standard deviation $\sigma$ (this is the reparameterisation applied by \texttt{Lm()}). Of course, the parameter $\vartheta_1$, the inverse residual standard deviation, is ensured to be positive via an additional constraint in the optimiser maximising the log-likelihood.

The log-likelihoods of the two models fitted by \texttt{lmer()} and \texttt{mtram()} are very close

\begin{verbatim}
R> logLik(sleep_lmer)
'log Lik.' -875.9697 (df=6)
R> logLik(sleep_LMmer)
'log Lik.' -875.9697 (df=6)
\end{verbatim}

Looking at the model coefficients, the two procedures lead to almost identical inverse residual standard deviations

\begin{verbatim}
R> (sdinv <- 1 / summary(sleep_lmer)$sigma)
[1] 0.03907485
R> coef(sleep_LMmer)['Reaction']
Reaction
0.03907741
\end{verbatim}

and fixed effects (the slope can be interpreted as inverse coefficient of variation)

\begin{verbatim}
R> fixef(sleep_lmer) * c(-1, 1) * sdinv

(Intercept)        Days
-9.8236175  0.4090077

R> coef(sleep_LMmer)[c("(Intercept)", "Days")]

(Intercept)        Days
-9.8243917  0.4089265
\end{verbatim}
The random-effect parameters $\gamma$ are also reasonably close

```r
R> sleep_lmer@theta
[1] 0.92919061 0.01816575 0.22264321
```

```r
R> coef(sleep_LMmer)[-c(1:3)]
      gamma1       gamma2       gamma3
0.92901066 0.01843056 0.22280431
```

Consequently, the variance-covariance and correlation matrices

```r
R> sleep_LMmer$G * (1 / sdinv)^2
```

2 x 2 sparse Matrix of class "dsCMatrix"

```r
[1,]  565.2580  11.21410
[2,]  11.2141  32.73513
```

```r
R> cov2cor(sleep_LMmer$G * (1 / sdinv)^2)
```

2 x 2 sparse Matrix of class "dsCMatrix"

```r
[1,] 1.00000000 0.08243925
[2,] 0.08243925 1.00000000
```

```r
R> unclass(VarCorr(sleep_lmer))$Subject
    (Intercept) Days
(Intercept)  565.47697 11.05512
Days  11.05512 32.68179
```

```r
attr("stddev")
(Intercept) Days
23.779760 5.716799
```

```r
attr("correlation")
(Intercept) Days
0.08132109 1.00000000
```

are practically equivalent. This result indicates the correctness of the alternative implementation of normal linear mixed-effects models in the transformation model framework: `mtram()` reuses some infrastructure from `lme4` and `Matrix`, most importantly fast update methods for Cholesky factors, but the likelihood and corresponding optimisation relies on an independent implementation. So why are we doing this? Because `mtram()` is able to deal with models or likelihoods not available in `lme4`, for example the likelihood for interval-censored observations.

Let’s assume that the timing of the reaction times was less accurate than suggested by the numerical representation of the results. The following code
R> library("survival")
R> sleepstudy$Reaction_I <- with(sleepstudy, Surv(Reaction - 20, Reaction + 20, +
+ type = "interval2"))
R> sleepstudy$Reaction_I[1:5]

[1] [229.5600, 269.5600] [238.7047, 278.7047] [230.8006, 270.8006]
[4] [301.4398, 341.4398] [336.8519, 376.8519]

converts the outcome to interval-censored values, where each interval has length 40. The above mixed model can now be estimated by maximising the likelihood corresponding to interval-censored observations:

R> sleep_LM_I <- Lm(Reaction_I ~ Days, data = sleepstudy)
R> sleep_LMmer_I <- mtram(sleep_LM_I, ~ (Days | Subject), data = sleepstudy)

Of course, the log-likelihood changes (because this is a log-probability and not a log-density of a continuous distribution) but the parameter estimates are reasonably close

R> logLik(sleep_LMmer_I)

'log Lik.' -214.9663 (df=6)

R> coef(sleep_LMmer_I)

(Intercept) Reaction_I Days gamma1 gamma2 gamma3
-9.7878878 0.03900261 0.41620117 0.83409026 0.07594894 0.19048189

R> coef(sleep_LMmer)

(Intercept) Reaction Days gamma1 gamma2 gamma3
-9.82439168 0.03907741 0.40892652 0.92901066 0.01843056 0.22280431

The next question is if the normal assumption for reaction times is appropriate. In the transformation world, this assumption is simple to assess because we can easily (theoretically and in-silico) switch to the non-normal linear mixed-effects transformation model

$$P(\text{Reaction} \leq y | \text{day}, i) = \Phi \left( h(y) - \hat{\beta}_{\text{day}} - \hat{\alpha}_i - \hat{\beta}_i \text{day} \right), \quad (\hat{\alpha}_i, \hat{\beta}_i) \sim N_2(\mathbf{0}, \Lambda(\gamma)\Lambda(\gamma))$$

where $h(y) = \mathbf{a}(y)^\top \vartheta$ represents a monotone non-decreasing transformation function. The function implementing such a more flexible model in named in honor of the first paper on the analysis of transformed responses by Box and Cox (1964) but it does not simply apply what is known as a Box-Cox transformation. Bernstein polynomials $h(y) = \mathbf{a}(y)^\top \vartheta$ under suitable constraints are applied instead by

R> sleep_BC <- BoxCox(Reaction ~ Days, data = sleepstudy)
R> sleep_BCmer <- mtram(sleep_BC, ~ (Days | Subject), data = sleepstudy, +
+ Hessian = TRUE)
R> logLik(sleep_BCmer)
Marginally Interpretable Transformation Models

Average reaction time (in ms)

<table>
<thead>
<tr>
<th>h(y)</th>
<th>-4</th>
<th>-2</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>250</td>
<td>300</td>
<td>350</td>
<td>400</td>
<td>450</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Sleep deprivation: Data-driven transformation $\hat{h}$ of average reaction times to sleep deprivation. The non-linearity induces a non-normal marginal distribution function of reaction times.

$log$ Lik. $-859.5455$ (df=11)

The increase in the log-likelihood compared to the normal model is not a big surprise. Plotting the transformation function $h(y) = a(y)^T \vartheta$ as a function of reaction time can help to assess deviations from normality because the latter assumption implies a linear transformation function. Figure 2 clearly indicates that models allowing a certain skewness of reaction times will provide a better fit to the data. This might also not come as a big surprise to experienced data analysts.

However, what does this finding mean in terms of a direct comparison of the model and the data? Looking at the marginal cumulative distribution functions of average reaction time conditional on days of sleep deprivation in Figure 3 one finds that the non-normal marginal transformation models provided a better fit to the marginal empirical cumulative distribution functions than the normal marginal models. Especially for short reaction times in the first week of sleep deprivation, the yellowish marginal cumulative distribution is much closer to the empirical cumulative distribution function representing the marginal distribution of reaction times at each single day of study participation.

It should be noted that the small positive correlation between random intercept and random slope observed in the normal linear mixed-effects model turned into a negative correlation in this non-normal model.
Figure 3: Sleep deprivation: Marginal distribution of reaction times, separately for each day of study participation. The grey step-function corresponds to the empirical cumulative distribution function, the blue line to the marginal cumulative distribution of a normal linear mixed-effects model, and the yellowish line to a non-normal linear mixed-effects transformation model.

R> cov2cor(sleep_BCmer$G)

2 x 2 sparse Matrix of class "dsCMatrix"

[1,] 1.0000000 -0.1946629
[2,] -0.1946629 1.0000000

What is the uncertainty associated with this parameter? The correlation is a non-linear function of \( \gamma \) and therefore the direct computation of confidence intervals questionable. However, we can extract an estimate of the covariance of the estimated model parameters from the model and, relying on the asymptotic normality of the maximum likelihood estimators, we can sample from the asymptotic distribution of the variance of the random intercept \( \tilde{\alpha} \), the random slope \( \tilde{\beta} \), and their correlation

R> VC <- solve(sleep_BCmer$Hessian)
R> idx <- (nrow(VC) - 2):nrow(VC)
R> Rcoef <- rmvnorm(1000, mean = coef(sleep_BCmer), sigma = VC)[,idx]
R> ret <- apply(Rcoef, 1, function(gamma) {
+   L <- matrix(c(gamma[1:2], 0, gamma[3]), nrow = 2)
+   V <- tcrossprod(L)
+   c(diag(V), cov2cor(V)[1,2])
+ })
The 95% confidence intervals

```r
R> ### variance random intercept
R> quantile(ret[1,], c(.025, .5, .975))
   2.5%      50%     97.5%
0.9127821 2.5713595 5.2493469
```

```r
R> ### variance random slope
R> quantile(ret[2,], c(.025, .5, .975))
   2.5%      50%     97.5%
0.01890987 0.05348231 0.10594879
```

```r
R> ### correlation random intercept / random slope
R> quantile(ret[3,], c(.025, .5, .975))
   2.5%      50%     97.5%
-0.6193527 -0.1883314  0.4689778
```

indicate rather strong unobserved heterogeneity affecting the intercept and less pronounced variability in the slope. There is only weak information about the correlation of the two random effects contained in the data.

The downside of this approach is that, although the model is nicely interpretable on the scale of marginal or conditional distribution functions, the direct interpretation of the fixed effect $\beta$ is not very straightforward because it corresponds to the conditional mean after transforming the outcome. This interpretability issue can be addressed by exchanging the probit link to a logit link in Section 4.

### 3. Binary Probit Mixed-effects Models

Here we compare different implementations of binary probit mixed models for the notoriously difficult toenail data (Backer et al. 1998). The outcome was categorised to two levels (this being probably the root of all troubles) and a conditional density plot (Figure 4) suggests an improvement in both treatment groups over time, however with a more rapid advance in patients treated with terbinafine.

The random intercept probit model fitted by Laplace and Adaptive Gauss-Hermite Quadrature (AGQ) approximations to the likelihood give quite different results:

```r
R> ### Laplace
toenail_glmer_RI_1 <-
  + glmer(outcome ~ treatment * time + (1 | patientID),
  +     data = toenail, family = binomial(link = "probit"),
  +     nAGQ = 1)
R> summary(toenail_glmer_RI_1)
```
Figure 4: Toe nail data: Conditional density plot of two outcome classes (none or mild vs. moderate or severe) under two treatments.

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]
Family: binomial  ( probit )
Formula: outcome ~ treatment * time + (1 | patientID)
Data: toenail

AIC BIC logLik deviance df.resid
1279.0 1306.8 -634.5 1269.0 1898

Scaled residuals:
Min 1Q Median 3Q Max
-3.507 -0.017 -0.004 0.000 54.046

Random effects:
Groups   Name        Variance Std.Dev.
patientID (Intercept) 20.68     4.548
Number of obs: 1903, groups: patientID, 289

Fixed effects:
(Intercept) -3.39650 0.22091 -15.375 <2e-16 ***
treatmentterbinafine -0.01532 0.25359 -0.060 0.9518
time -0.21749 0.02256 -9.639 <2e-16 ***
treatmentterbinafine:time -0.07155 0.03425 -2.089 0.0367 *
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
   (Intr) trtmnt  time
trtmnttrbnf  -0.593
  time  -0.009   0.102
trtmnttrbn:  0.093  -0.143 -0.629

R> toenail_glmer_RI_1@theta

[1] 4.547891

R> ### Adaptive Gaussian Quadrature
R> toenail_glmer_RI_2 <-
+   glmer(outcome ~ treatment * time + (1 / patientID),
+          data = toenail, family = binomial(link = "probit"),
+          nAGQ = 20)
R> summary(toenail_glmer_RI_2)

Generalized linear mixed model fit by maximum likelihood (Adaptive
Gauss-Hermite Quadrature, nAGQ = 20) [glmerMod]
Family:  binomial   (  probit )
Formula:  outcome ~ treatment * time + (1 | patientID)
Data: toenail  

   AIC   BIC logLik deviance df.resid
1277.8 1305.6  -633.9 1267.8       1898

Scaled residuals:
     Min      1Q  Median      3Q     Max
-2.847  -0.189  -0.078  -0.001  33.997

Random effects:
   Groups    Name     Variance  Std.Dev.
   patientID (Intercept)  4.485   2.118
Number of obs: 1903, groups: patientID, 289

Fixed effects:
                           Estimate Std. Error z value Pr(>|z|)
   (Intercept)            -0.93061   0.23176  -4.015  5.93e-05  ***
treatmentterbinafine     -0.07609   0.30921  -0.246    0.8056
   time                   -0.19074   0.02059  -9.263  < 2e-16  ***
treatmentterbinafine:time -0.06419   0.03099  -2.071    0.0383 *

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
The sequential reduction (SR) algorithm (Ogden 2015) gives results close to AGQ

```
R> library("glmmsr")
R> toenail_glmm_RI_3 <-
+   glmm(outcome ~ treatment * time + (1 | patientID),
+   data = toenail, family = binomial(link = "probit"),
+   method = "SR", control = list(nSL = 3))
```

Fitting the model......................... done.

```
R> summary(toenail_glmm_RI_3)
```

Generalized linear mixed model fit by maximum likelihood [glmmFit]
Likelihood approximation: Sequential reduction at level 3

Family: binomial (probit)
Formula: outcome ~ treatment * time + (1 | patientID)

Random effects:
  Groups   Name            Estimate  Std.Error
  patientID (Intercept)    2.1190   0.1954

Number of obs: 1903, groups: patientID, 289;

Fixed effects:
             Estimate  Std.Error z value  Pr(>|z|)
(Intercept)  -0.93105   0.23217  4.0102 6.066e-05

Because of the probit link, this binary generalised linear model is equivalent to a linear transformation model and we can thus use the exact likelihood implemented for the latter model in mtram() for parameter estimation (it is still a bit nasty to set-up a constant transformation function $h(y) = \alpha$, I plan to add a more convenient interface later)

```
R> m <- ctm(as.basis(~ outcome, data = toenail),
+         shifting = ~ treatment * time,
+         data = toenail, todistr = "Normal")
```
Marginally Interpretable Transformation Models

R> toenail_probit <- mlt(m, data = toenail,
+     fixed = c("outcomemoderate or severe" = 0))
R> toenail_mtram RI <-
+     mtram(toenail_probit, ~ (1 | patientID),
+     data = toenail, Hessian = TRUE)
R> logLik(toenail_mtram RI)

'log Lik.' -633.9638 (df=5)

R> coef(toenail_mtram RI)

(Intercept) treatmentterbinafine
0.92946575 0.07699772
time treatmentterbinafine:time
0.19056617 0.06355426
gamma1
2.11447751

For this random intercept model, the exact likelihood is defined as a one-dimensional integral over the unit interval. We use sparse grids (Heiss and Winschel 2008; Ypma 2013) to approximate this integral. The integrand is defined by products of normal probabilities, which are approximated as described by Matić et al. (2018). It is important to note that this likelihood can be computed as accurately as necessary whereas Laplace, AGQ, and SR are approximations of limited accuracy.

The results are very close to SR and AGQ, indicating a very good quality of the AGQ and SR approximations. We can also compare the corresponding covariances

R> vcov(toenail_glmer RI_2)

4 x 4 Matrix of class "dpoMatrix"

(Intercept) treatmentterbinafine time
(Intercept) 0.0537125190 -0.046953461 -0.0008877046
treatmentterbinafine -0.0469534610 0.095609401 0.0013522326
time -0.0008877046 0.001352233 0.0004239968
treatmentterbinafine:time 0.0013871422 -0.002754670 -0.0003896725

treatmentterbinafine:time
(Intercept) 0.0013871422
treatmentterbinafine -0.0027546700
time -0.0003896725
treatmentterbinafine:time 0.0009603451

R> solve(toenail_mtram RI$Hessian)[1:4, 1:4]

[1,] 0.0535092145 -0.046828492 -0.0008863063 0.0013728319
[2,] -0.0468284922 0.095401203 0.0013522326 -0.0027201601
[3,] -0.0008863063 0.001352233 0.0004239968 -0.0003896725
[4,] 0.0013728319 -0.002720160 -0.0003896727 0.0009479429
Things get a bit less straightforward when a random slope is added to the model. The two implementations of the Laplace approximation in packages `lme4`

```r
R> toenail_glmer_RS <-
    +   glmer(outcome ~ treatment * time + (1 + time | patientID),
    +     data = toenail, family = binomial(link = "probit"))
R> summary(toenail_glmer_RS)
```

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]
Family: binomial ( probit )
Formula: outcome ~ treatment * time + (1 + time | patientID)
Data: toenail

AIC  BIC   logLik deviance df.resid
985.8 1024.7  -485.9  971.8  1896

Scaled residuals:
  Min 1Q Median 3Q Max
-1.85421 -0.00210 -0.00037 0.00000 2.35828

Random effects:
  Groups     Name      Variance Std.Dev. Corr
  patientID (Intercept) 118.433  10.883
  time         3.305   1.818  -0.90
Number of obs: 1903, groups: patientID, 289

Fixed effects:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          -4.30119   0.26361 -16.316  <2e-16 ***
treatmentterbinafine  0.05419   0.34652  0.156   0.8757
  time                -0.06791   0.07846  -0.866   0.3867
treatmentterbinafine:time -0.23478   0.13885  -1.691   0.0909 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
   (Intr) trtmnt time
trtmnttrbnf  -0.662
  time   -0.453  0.342
trtmnttrbn:  0.270 -0.438 -0.335

R> toenail_glmer_RS@theta
[1] 10.8826888 -1.6359617  0.7930867

and `glmmsr`
Fitting the model..... done.

R> toenail_glmm_RS_1$estim[1:3]
[1]  4.9992215 -0.5644628  0.4110986
R> toenail_glmm_RS_1$estim[-(1:3)]
[1] -3.49232631  0.04197115 -0.06769437 -0.12100940
do not quite agree. Note that the standard deviation of the random intercept is twice as large in the glmer() output.
The optimisation of the exact discrete likelihood in the transformation framework gives

R> toenail_mtram_RS <-
+   mtram(toenail_probit, ~(1 + time / patientID),
+          data = toenail)
R> logLik(toenail_mtram_RS)

'log Lik.' -545.1164 (df=7)

R> coef(toenail_mtram_RS)

(Intercept) treatmentterbinafine
  1.5765323        -0.2666840
   time treatmentterbinafine:time
  0.5323918          0.1842506
gamma1 gamma2
  5.2172376       -0.3723898
gamma3
  0.5285640

The variance parameters are not too far off the results reported by glmm(), but the fixed effects differ quite a bit.

At least in biostatistics, the probit model is less popular than the logit model owing to the better interpretability of the fixed effects as conditional log-odds ratios in the latter. Using a logit link, we can use the transformation approach to compute marginally interpretable time-dependent log-odds ratios from random intercept transformation logit models (standardise = TRUE() computes model (M2) instead of the default (M1), see Hothorn 2019):
It is important to note that this model is not a logistic mixed-effects model and thus we can’t expect to obtain identical results from \texttt{glmer()} as it was (partially) the case for the probit model.

From the standardised model, we can compute marginally interpretable probabilities and odds ratios over time.

We can also sample from the distribution of the maximum likelihood estimators to obtain an idea about the uncertainty (Figure 5).

From the unstandardised logit and probit models, we can also obtain marginally interpretable probabilities as (probit)

and (logit)
Figure 5: Toe nail data: Marginal odds ratio over time (from a logistic random intercept model). The blue line represents the maximum likelihood estimator, the grey lines are samples from the corresponding distribution.

\[
\begin{align*}
R> & \text{tmp <- toenail_probit} \\
R> & \text{cf <- coef(tmp)} \\
R> & \text{cf <- cf[\{names(cf) != "outcome moderate or severe"\}]} \\
R> & \text{sdrf <- rev(coef(toenail_mtram RI))[1]} \\
R> & \text{cf <- coef(toenail_mtram RI)[\{names(cf)\}] / sqrt(sdrf^2 + 1)} \\
R> & \text{cf <- c(cf[1], "outcome moderate or severe" = 0, cf[-1])} \\
R> & \text{coef(tmp) <- cf} \\
R> & \text{nd$prob_probit <- predict(tmp, newdata = nd, type = "distribution")[1,]} \\
\end{align*}
\]

The marginal time-dependent probabilities obtained from all three models are very similar as shown in Figure 6.

4. Proportional Odds Models for Bounded Responses

Manuguerra and Heller (2010) proposed a mixed-effects model for bounded responses whose fixed effects can be interpreted as log-odds ratios. We fit a transformation model to data from a randomised controlled trial on chronic neck pain treatment (Chow et al. 2006). The data are visualised in Figure 7. Subjective neck pain levels were assessed on a visual analog scale, that is, on a bounded interval.

Manuguerra and Heller (2010) suggested the conditional model

\[
\text{logit}(P(\text{pain} \leq y \mid \text{treatment}, \text{time}, i)) = h(y) + \beta_{\text{Active}} + \beta_{7 \text{ weeks}} + \beta_{12 \text{ weeks}} + \beta_{7 \text{ weeks}} \text{ Active} + \beta_{12 \text{ weeks}} \text{ Active} + \alpha_i
\]
with random intercepts $\tilde{\alpha}_i$ such that the odds at baseline, for example, are given by
\[
\frac{P(\text{pain} \leq y \mid \text{Active, baseline, } i)}{P(\text{pain} > y \mid \text{Active, baseline, } i)} = \exp(\beta_{\text{Active}}) \frac{P(\text{pain} \leq y \mid \text{Placebo, baseline, } i)}{P(\text{pain} > y \mid \text{Placebo, baseline, } i)}
\]

```r
R> library("ordinalCont")
R> neck_ocm <- ocm(vas ~ laser * time + (1 | id), data = pain_df,
+ scale = c(0, 1))
```

The results

```r
R> summary(neck_ocm)
```

Call:
```
ocm(formula = vas ~ laser * time + (1 | id), data = pain_df,
   scale = c(0, 1))
```

Random effects:
```
   Name   Variance Std.Dev.
Intercept|id  5.755    2.399
```

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>StdErr</th>
<th>t.value</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>laserActive</td>
<td>-2.07922</td>
<td>0.65055</td>
<td>-3.1961</td>
<td>0.001918**</td>
</tr>
<tr>
<td>time7 weeks</td>
<td>-0.60366</td>
<td>0.35744</td>
<td>-1.6889</td>
<td>0.094689</td>
</tr>
</tbody>
</table>
Figure 7: Neck pain: Trajectories of neck pain assessed on a visual analog scale with and without low-level laser therapy.

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>7 weeks</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>12 weeks</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>7 weeks</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>12 weeks</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Signif. codes:** 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

suggest that there is a difference at baseline; the pain distribution of subjects in the placebo group on the odds scale is only 13% of the odds in the active group for any cut-off y:

```r
R> exp(cbind(coef(neck_ocm)[2:6], confint(neck_ocm)[2:6,]))
```

<table>
<thead>
<tr>
<th></th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>laserActive</td>
<td>0.125028</td>
<td>0.4474608</td>
</tr>
<tr>
<td>time7 weeks</td>
<td>0.5468040</td>
<td>1.1017581</td>
</tr>
<tr>
<td>time12 weeks</td>
<td>0.7881704</td>
<td>1.6075391</td>
</tr>
<tr>
<td>laserActive:time7 weeks</td>
<td>82.1194073</td>
<td>246.4577275</td>
</tr>
<tr>
<td>laserActive:time12 weeks</td>
<td>29.5454666</td>
<td>85.0162253</td>
</tr>
</tbody>
</table>

In contrast, there seems to be a very large treatment effect (at week 7, the odds in the placebo group is 1 times larger than in the active group. This levels off after 12 weeks, but the effect is still significant at the 5% level.

The corresponding transformation model with a different parameterisation of the transformation function $h$ and a completely different optimisation procedure for maximising the log-likelihood, can be estimated by

```r
R> neck_Colr <- Colr(vas ~ laser * time, data = pain_df, +                     bounds = c(0, 1), support = c(0, 1),
```
Neck pain (on visual analog scale)
Marginal distribution

0.0 0.2 0.4 0.6 0.8 1.0
Baseline
7 weeks
12 weeks
Placebo
Active

Figure 8: Neck pain: Marginal distribution functions of chronic neck pain evaluated at three different time points under placebo or active low-level laser therapy.

+ extrapolate = TRUE)
R> neck_Colrmer <- mtram(neck_Colr, ~ (1 | id), data = pain_df,
+ Hessian = TRUE)
R> logLik(neck_Colrmer)

'log_Lik.' 74.35908 (df=13)

Based on this model of form (M1), it is possible to derive the marginal distribution functions in the two groups, see Figure 8.

5. Marginally Interpretable Weibull and Cox Models

The CAO/ARO/AIO-04 randomised clinical trial (Rödel et al. 2015) compared Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy to the same therapy using fluorouracil only for rectal cancer patients. Patients were randomised in the two treatment arms by block randomisation taking the study center, the lymph node involvement (negative, positive), and tumour grading (T1-3 vs. T4) into account. The primary endpoint was disease-free survival, defined as the time between randomisation and non-radical surgery of the primary tumour (R2 resection), locoregional recurrence after R0/1 resection, metastatic disease or progression, or death from any cause, whichever occurred first. The observed outcomes are a mix of exact dates (time to death or incomplete removal of the primary tumour), right-censoring (end of follow-up or drop-out), and interval-censoring (local or distant metastases). We are interested in a clustered Cox or Weibull model for interval-censored survival times. The survivor functions, estimated separately for each of the four strata defined by lymph node involvement and tumour grading, are given in Figure 9.

The implementation of mixed transformation models is currently not able to deal with mixed exact and censored outcomes in the same cluster. We therefore recode exact event times as
Figure 9: Rectal cancer: Distribution of disease-free survival times for two treatments in the four strata defined by lymph node involvement (negative or positive) and tumour grading (T1-3 or T4).

being interval-censored by adding a 4-day window to each exact event time (variable iDFS2).

R> ## convert "exact" event dates to interval-censoring (+/- one day)
R> exact <- CAOsurv$iDFS
R> tmp <- CAOsurv$iDFS2
R> exact <- tmp[,3] == 1
R> tmp[exact,2] <- tmp[exact,1] + 2
R> tmp[exact,1] <- pmax(tmp[exact,1] - 2, 0)
R> tmp[exact,3] <- 3
R> CAOsurv$iDFS2 <- tmp

We start with the random intercept model

\[
\Pr(Y > y \mid \text{treatment}) = \exp\left(-\exp\left(\varphi_1 + \varphi_2 \log(y) - \beta_{5-FU + Ox}\sqrt{\gamma_1^2 + 1}\right)\right)
\]

assuming a marginal Weibull model whose effects are scaled depending on the variance \(\gamma_1^2\) of a block-specific (interaction of lymph node involvement, tumour grading, and study center) random intercept:

R> CAO_SR <- Survreg(iDFS2 ~ randarm, data = CAOsurv)
R> CAO_SR_mtram_s <- mtram(CAO_SR, ~ (1 | Block), data = CAOsurv,
+               standardise = TRUE, Hessian = TRUE)
R> logLik(CAO_SR_mtram_s)

'log Lik.' -2081.542 (df=4)
\[(\text{Intercept}) \quad \log(\text{idFS2}) \]
\[
-6.2990054 \quad 0.7412855
\]
\[
\text{randarm5-FU + Oxaliplatin} \quad \gamma_1
\]
\[
0.2328600 \quad 0.1683613
\]

\[R> (\text{OR} \left< \exp\left(-\text{cf}["\text{randarm5-FU + Oxaliplatin}"] / \sqrt{\text{cf}["\gamma_1"]^2 + 1}\right)\right)\]

\[
\text{randarm5-FU + Oxaliplatin} \quad 0.794829
\]

We are, of course, interested in the marginal treatment effect, that is, the odds ratio

\[
\exp(-\beta_{\text{5-FU + Ox}} / \sqrt{\gamma_1^2 + 1}).
\]

We simply sample from the joint normal distribution of the maximum likelihood estimators \(\hat{\varphi}_1, \hat{\varphi}_2, \hat{\beta}_{\text{5-FU + Ox}}, \hat{\gamma}_1\) and compute confidence intervals for the marginal treatment effect 0.79 as

\[R> S \left< \text{solve(CAO_SR_mtram_s$Hessian)}\right>\]
\[R> \text{sqrt(diag(S))}\]

\[
[1] \quad 0.29019989 \quad 0.03872268 \quad 0.10722445 \quad 0.12433065
\]

\[R> \text{rbeta} \left< \text{rmvnorm(10000, mean = coef(CAO_SR_mtram_s),}\right>\]
\[
+ \text{sigma = S)}\]
\[R> s \left< \text{rbeta[,ncol(rbeta)]}\right>\]
\[R> \text{rbeta} \left< \text{rbeta[,ncol(rbeta)]} / \sqrt{\text{s}^2 + 1}\right>\]
\[R> \text{quantile(exp(-rbeta[, ncol(rbeta)]), prob = c(.025, .5, .975))}\]

\[
2.5\% \quad 50\% \quad 97.5\%
\]
\[
0.6478684 \quad 0.7953985 \quad 0.9747696
\]

In a next step, we stratify with respect to lymph node involvement and tumour grading: For each of the four strata, the parameters \(\varphi_1\) and \(\varphi_2\) are estimated separately:

\[R> \text{CAO_SR_2} \left< \text{Survreg(idFS2} \mid 0 + \text{strat_n: strat_t} \sim \text{randarm, data = CAOsurv)}\right>\]
\[R> \text{CAO_SR_2_mtram_s} \left< \text{mtram(CAO_SR_2,} \sim (1 \mid \text{Block), data = CAOsurv,}\right>\]
\[
+ \text{standardise = TRUE, Hessian = TRUE)}\]
\[R> \text{logLik(CAQ_SR_2_mtram_s)}\]

'log Lik.' -2067.797 (df=10)

\[R> (\text{cf} \left< \text{coef(CAO_SR_2_mtram_s)}\right>\]
(Intercept):strat_ncN0:strat_tcT1-3 log(idFS2):strat_ncN0:strat_tcT1-3
-7.8833653 0.9584499
(Intercept):strat_ncN+:strat_tcT1-3 log(idFS2):strat_ncN+:strat_tcT1-3
-6.2225174 0.7198965
(Intercept):strat_ncN0:strat_tcT4 log(idFS2):strat_ncN0:strat_tcT4
-3.0467542 0.3711277
(Intercept):strat_ncN+:strat_tcT4 log(idFS2):strat_ncN+:strat_tcT4
-4.8207090 0.6214653
randarm5-FU + Oxaliplatin
0.2240023 0.1474685

R> (OR_2 <- exp(-cf["randarm5-FU + Oxaliplatin"] / sqrt(cf["gamma1"]^2 + 1)))

randarm5-FU + Oxaliplatin
0.8012313

The corresponding confidence interval for the marginal treatment effect is then

[1] 0.68882415 0.09402924 0.34618756 0.04633063 1.01843933 0.13874107
[7] 0.68657931 0.09534206 0.10731700 0.13455556

2.5% 50% 97.5%
0.6528915 0.8044045 0.9856882

We now relax the Weibull assumption in the Cox model

\[
P(Y > y | treatment, i) = \exp \left( - \exp \left( \alpha (\log(y))^T \theta + \beta_{5-FU + Ox} \gamma^1 + 1 \right) \right)
\]

(note the positive sign of the treatment effect).

R> CAO_Cox_2 <- Coxph(idFS2 ~ 0 + strat_n:strat_t ~ randarm, data = CAOsurv,+
+ support = c(1, 1700), log_first = TRUE, order = 4)
R> logLik(CAO_Cox_2)

'log Lik.' -2021.875 (df=21)

R> CAO_Cox_2_mtram_s <- mtram(CAO_Cox_2, ~ (1 | Block), data = CAOsurv,+
+ standardise = TRUE, Hessian = TRUE)
R> logLik(CAO_Cox_2_mtram_s)

'log Lik.' -2031.051 (df=22)

R> coef(CAO_Cox_2_mtram_s)
Because the estimated variance parameter $\gamma_1$ is not very large, we would expect to see similar results in a conditional Cox model with normal frailty term.

R> library("coxme")
R> m <- coxme(DFS ~ randarm + (1 | Block), data = CAOsurv)
R> summary(m)

Cox mixed-effects model fit by maximum likelihood
Data: CAOsurv
events, n = 357, 1236
Iterations= 22 91

NULL Integrated Fitted
Log-likelihood -2432.971 -2430.475 -2414.482

Chisq   df  p  AIC  BIC
Integrated loglik  4.99  2.00  0.0823850  0.99  -6.76
Penalized loglik 36.98 16.72  0.0029833  3.53 -61.33

Model: DFS ~ randarm + (1 | Block)
Fixed coefficients

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>randarm5-FU + Oxaliplatin</td>
<td>-0.2310483</td>
<td>0.7937011</td>
<td>0.1067215</td>
<td>-2.16</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Random effects

<table>
<thead>
<tr>
<th>Group Variable</th>
<th>Std Dev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block Intercept</td>
<td>0.21741797</td>
<td>0.04727057</td>
</tr>
</tbody>
</table>

```R
R> sd <- sqrt(diag(vcov(m)))
R> exp(coef(m) + c(-1, 0, 1) * qnorm(.975) * sd)
[1] 0.6438957 0.7937011 0.9783596
```

6. Assessment of Unexplained Variability

Pollet and Nettle (2009) reported on an association between partner wealth and female self-reported orgasm frequency. It was later (Herberich et al. 2010) pointed out that the finding was due to an incorrectly implemented variable selection procedure based on a proportional odds (cumulative logit) model for the ordinal variable corresponding to the question “When having sex with your current partner, how often did you have orgasm?” with possible answer categories $y_1 = \text{Always}$, $y_2 = \text{Often}$, $y_3 = \text{Sometimes}$, $y_4 = \text{Rarely}$, or $y_5 = \text{Never}$. The full model explains the conditional distribution of orgasm frequency by $\mathbf{x} = \text{partner income, partner height, the duration of the relationship, the respondents age, the difference between both partners regarding education and wealth, the respondents education, health, happiness, and place of living (regions in China)}$ of the form

$$
P(\text{orgasm} \leq y_k | \mathbf{x}) = \expit(\vartheta_k + \mathbf{x}^\top \mathbf{\beta})
$$

for $i = 1, \ldots, N = 1531$ independent heterosexual couples. In this model, the regression coefficients $\mathbf{\beta}$ can be interpreted as log-odds ratios and we question the appropriateness of this model here by including a subject-specific random intercept with standard deviation $\gamma_1$. This changes the model to

$$
P(\text{orgasm} \leq y_k | \mathbf{x}) = \Phi\left(\Phi^{-1}(\expit(\vartheta_k + \mathbf{x}^\top \mathbf{\beta})) \sqrt{\gamma_1^2 + 1}\right)
$$

A value of $\gamma_1$ close to zero corresponds to marginal distributions very similar to the proportional odds model and, consequently, it is appropriate to interpret $\mathbf{\beta}$ as log-odds ratios. Larger values of $\gamma_1$ indicate a more variable distribution and thus the choice $F = \expit$ might be questionable.

We obtain

```R
R> CHFLS_Polr <- Polr(orgasm ~ AincomeSD + AheightSD + RADurationSD +
+ RageSD + edudiffSD + wealthdiffSD + Redu +
+ Rhealth + Rhappy + Region, data = orgAcc)
R> logLik(CHFLS_Polr)
```
'log Lik.' -1852.615 (df=27)

R> orgAcc$ID <- factor(1:nrow(orgAcc))
R> CHFLS_mtram <- mtram(CHFLS_Polr, ~ (1 | ID),
+     data = orgAcc)
R> logLik(CHFLS_mtram)

'log Lik.' -1852.829 (df=28)

R> coef(CHFLS_mtram)

<table>
<thead>
<tr>
<th></th>
<th>orgasm1</th>
<th>orgasm2</th>
<th>orgasm3</th>
<th>orgasm4</th>
</tr>
</thead>
<tbody>
<tr>
<td>AincomeSD</td>
<td>0.02603981</td>
<td>-0.02217060</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>AheightSD</td>
<td>-0.17413324</td>
<td>-0.03428024</td>
<td>0.13682888</td>
<td>0.17668815</td>
</tr>
<tr>
<td>RADurationSD</td>
<td>0.02603981</td>
<td>-0.02217060</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>RageSD</td>
<td>0.13682888</td>
<td>0.17668815</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>Reducol</td>
<td>0.13682888</td>
<td>0.17668815</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>Reduupmid</td>
<td>0.17668815</td>
<td>0.13682888</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>Redulowmid</td>
<td>0.17668815</td>
<td>0.13682888</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>Reduprimary</td>
<td>0.17668815</td>
<td>0.13682888</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>Redunoschool</td>
<td>0.17668815</td>
<td>0.13682888</td>
<td>0.07529263</td>
<td>-0.35310380</td>
</tr>
<tr>
<td>Rhealthfair</td>
<td>1.69707889</td>
<td>1.83068385</td>
<td>1.86386980</td>
<td>0.29162606</td>
</tr>
<tr>
<td>Rhealthgood</td>
<td>1.83068385</td>
<td>1.86386980</td>
<td>0.29162606</td>
<td>0.20369845</td>
</tr>
<tr>
<td>Rhealthexcellent</td>
<td>1.86386980</td>
<td>0.29162606</td>
<td>0.20369845</td>
<td>0.20369845</td>
</tr>
<tr>
<td>Rhealthnot good</td>
<td>-1.017</td>
<td>0.20369845</td>
<td>0.20369845</td>
<td>0.20369845</td>
</tr>
<tr>
<td>Rhappyrelatively</td>
<td>1.02374623</td>
<td>0.40651849</td>
<td>0.20369845</td>
<td>0.20369845</td>
</tr>
<tr>
<td>Rhappyvery</td>
<td>0.20369845</td>
<td>0.20369845</td>
<td>0.20369845</td>
<td>0.20369845</td>
</tr>
<tr>
<td>RegionNortheas t</td>
<td>0.75865626</td>
<td>1.2374623</td>
<td>0.40651849</td>
<td>0.20369845</td>
</tr>
<tr>
<td>RegionCoastalS</td>
<td>0.40651849</td>
<td>0.20369845</td>
<td>0.20369845</td>
<td>0.20369845</td>
</tr>
<tr>
<td>RegionCoastalE</td>
<td>0.75865626</td>
<td>1.2374623</td>
<td>0.40651849</td>
<td>0.20369845</td>
</tr>
<tr>
<td>RegionInlandS</td>
<td>0.40651849</td>
<td>0.20369845</td>
<td>0.20369845</td>
<td>0.20369845</td>
</tr>
<tr>
<td>gamma1</td>
<td>0.49050565</td>
<td>0.20059228</td>
<td>0.58338272</td>
<td>0.01678467</td>
</tr>
</tbody>
</table>

and from $\hat{\gamma}_1 = 0.017$ can conclude that the proportional odds model is appropriate here.

References


Affiliation:
Torsten Hothorn
Institut für Epidemiologie, Biostatistik und Prävention
Universität Zürich
Hirschengraben 84, CH-8001 Zürich, Switzerland
Torsten.Hothorn@uzh.ch
http://tiny.uzh.ch/bh