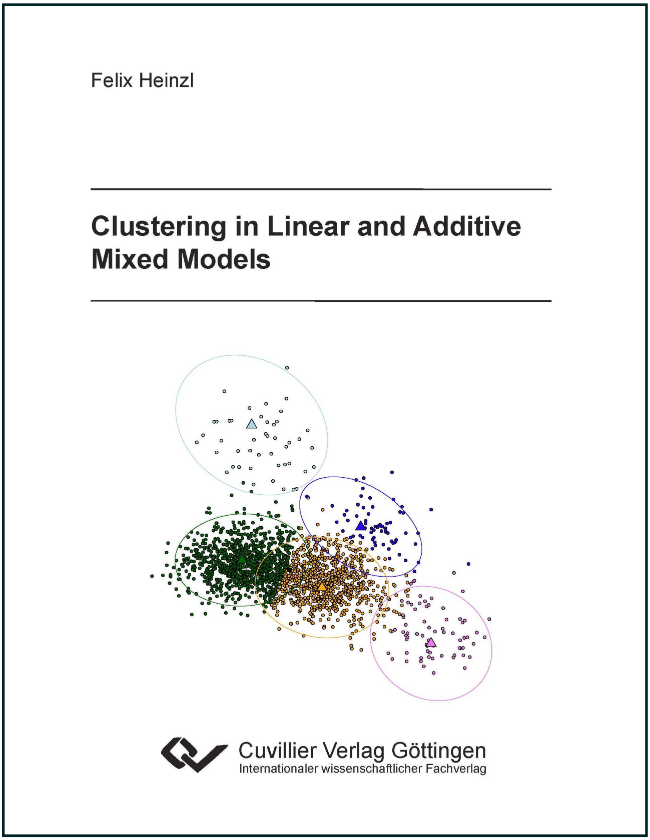
Felix Heinzl (Autor)

# Clustering in Linear and Additive Mixed Models




https://cuvillier.de/de/shop/publications/6371

# 1. Introduction

## Mixed Models

The analysis of longitudinal data is a popular task in statistics (Diggle et al., 2002; Fitzmaurice et al., 2004). This kind of data is given when statistical units like individuals are examined at several observation times with regard to some variables of interest. When regression models are applied for investigating the influence of different covariates on a response variable, one has to incorporate the dependence structure in repeated measurements that arises from the fact that measurements belonging to the same individual are typically correlated. This can either be achieved by considering *mixed models* also known as *random effects models* (Laird and Ware, 1982) or by using the *generalized estimation equation* approach proposed by Liang and Zeger (1986). While in the framework of generalized estimation equations the response values are modeled marginally by using only population-specific effects, mixed models contain population-specific fixed effects as well as individual-specific random effects and focus on the conditional distribution of each response value conditional on the corresponding random effect. Mixed models, which were introduced by Fisher (1918), assign each subject $i$ its own random effect $\boldsymbol{b}_i$. For longitudinal data random effects facilitate the modeling of individual deviations from the population trend of the response variable over time. In contrast to the fixed effects, for the random effects a distribution assumption is specified that is typically given by a normal distribution. A more flexible approach has been proposed by Verbeke and Lesaffre (1996). They consider a mixture of normal distributions as random effects distribution:

$$\boldsymbol{b}_i \sim \sum_{h=1}^{N} \pi_h \, N(\boldsymbol{\mu}_h, \boldsymbol{D}), \quad i = 1, \ldots, n.$$

This offers a possibility for clustering individuals due to their time-dependent trend curves of the response variable: If the number of clusters $N$ is smaller than the number of subjects $n$, several subjects share the same cluster center $\boldsymbol{\mu}_h$ and form a cluster. The covariance matrix $\boldsymbol{D}$ indicates the dispersion of the random effects around their cluster centers. However, this raises the question how to choose the number of clusters. In this dissertation, two penalization approaches are proposed to determine the number of clusters in a data driven way. One is based on the fusion of cluster centers: If the differences of cluster centers are penalized by an appropriate penalty term, some differences are shrunk to zero. Consequently some clusters are fused and the effective number of clusters is reduced. An alternative possibility consists in the penalization of the amounts of the weights $\pi_h$. If some weights are shrunk to zero, the corresponding clusters drop out.

## Discussion of Penalization Ideas

In regression models regularization approaches are widely used that aim at penalization of the fixed effects of predictors on a response variable. The fundamental papers of Hoerl and Kennard (1970) and Tibshirani (1996) introduced the penalized regression techniques *ridge* respectively *lasso* based on a $L_2$-norm respectively $L_1$-norm penalty. The latter one is particulary characterized by the possibility to shrink parameters *and* to set some of them to zero if the corresponding covariates have no impact on the response variable. In the following, it will be shortly discussed in which extent the lasso method could be used for the two penalization goals mentioned in the previous section: On the one hand, we want to shrink differences of cluster centers to zero. However, for fusion of parameters the *fused lasso* idea of Tibshirani et al. (2005) is much more helpful than the direct lasso approach. Furthermore, for incorporating multivariate random effects the fusion concept has to be combined with the *group lasso* approach by Yuan and Lin (2006), which also is an extension of the lasso idea. On the other hand, at first sight the lasso approach seems to be appropriate to shrink weights to zero. But note that probabilities with the range $[0, 1]$ and the restriction that the sum of all probabilities is one cannot be handled in the same way as usual regression coefficients. Thus, we prefer a completely different approach that is based on a Dirichlet process. In this approach, all restrictions are fulfilled automatically and we get rather a shift than a penalization of the weights: High weights become higher and small weights become nearly zero.

## Guideline through the Thesis

The main part of this dissertation consists of four chapters, which show different possibilities of clustering in linear and additive mixed models. In Chapters 3 and 4 the two different methods for penalizing the number of clusters introduced in the previous sections are elaborated and applied within the framework of linear mixed models. An Expectation-Maximization (EM) algorithm is used for inference in each case. A comparison of both methods with regard to simulation results and applications can be found in Sections 4.3.3 and 4.4. Chapters 5 and 6 deal with additive mixed models using an approximate Dirichlet process mixture (DPM) as random effects distribution. While in Chapter 5 the model is fitted by using Markov chain Monte Carlo (MCMC) methods, in Chapter 6 the EM algorithm of Chapter 4 is extended to additive mixed models and compared to the MCMC approach of Chapter 5. Chapter 2 takes a special role in the thesis. Here, the theoretical concepts of Dirichlet processes are explained for a better understanding of the methods in chapters using Dirichlet processes. Nevertheless, the single chapters can be read independently of each other. Just for background knowledge or comparisons to other approaches cross references are helpful. Short summaries of the chapters are given in the following:

### Chapter 2: Dirichlet Processes

In this chapter we want to depict the idea as well as the highly praised cluster property of the Dirichlet process. The stick breaking representation of the Dirichlet process and the concept of DPMs play a central role in thesis and are also outlined in this chapter.

### Chapter 3: Linear Mixed Models with a Group Fused Lasso Penalty

A method is proposed that aims at identifying clusters of individuals that show similar patterns when observed repeatedly. We consider linear mixed models, which are widely used for the modeling of longitudinal data. In contrast to the classical assumption of a normal distribution for the random effects a finite mixture of normal distributions is assumed. Typically, the number of mixture components is unknown and has to be chosen, ideally by data driven tools. For this purpose an EM algorithm-based approach is considered, that uses a penalized normal mixture as random effects distribution. The penalty term shrinks the pairwise distances of cluster centers based on the group lasso and the fused lasso method with the effect that individuals with similar time trends are merged into the same cluster. The strength of regularization is determined by one penalization parameter. For finding the optimal penalization parameter, a new model choice criterion is proposed. The usefulness of this method is illustrated in three applications and in a simulation study.

### Chapter 4: Linear Mixed Models with DPMs using EM Algorithm

For the same goal as in the previous chapter an alternative clustering approach is considered. Note that in linear mixed models the assumption of normally distributed random effects is often inappropriate and unnecessary restrictive. The proposed approximate DPM assumes a hierarchical Gaussian mixture that is based on the truncated version of the stick breaking presentation of the Dirichlet process. In addition to the weakening of distributional assumptions, the specification allows to identify clusters of observations with a similar random effects structure. An EM algorithm is given, that solves the estimation problem and that, in certain respects, may exhibit advantages over Markov chain Monte Carlo approaches when modeling with Dirichlet processes. The method is evaluated in a simulation study and applied to the dynamics of unemployment in Germany as well as lung function growth data.

### Chapter 5: Additive Mixed Models with DPMs using MCMC methods

When the population time trend is nonlinear, the methods of Chapters 3 and 4 cannot be used, and more flexible approaches like additive mixed models are necessary. For the handling of nonlinearity and heterogeneity in the data, a combination of flexible time trends and individual-specific random effects is required. We propose a fully Bayesian approach based on MCMC simulation techniques that allows for the semiparametric specification of both the trend function and the random effects distribution. Bayesian penalized splines (P-splines) are considered for the former while a DPM specification allows for an adaptive amount of deviations from normality for the latter. The advantages of such DPM prior structures for random effects are investigated in terms of a simulation study to improve understanding of the model specification before analyzing childhood obesity data.

**Chapter 6: Additive Mixed Models with DPMs using EM Algorithm**

As in the previous chapter, additive mixed models with a DPM as random effects distribution are considered, that are based on the truncated version of the stick breaking presentation of the Dirichlet process. In contrast to Chapter 5 an EM algorithm is given, that solves the estimation problem and that exhibits advantages over MCMC approaches, which are typically used when modeling with Dirichlet processes. For handling the trend curve the mixed model representation of P-splines is used. The method is evaluated in a simulation study and applied to theophylline data and childhood obesity data.

An important technical fact concerning regression models in general should be mentioned at this stage. Regression models are among other things specified by an assumption for the conditional distribution of the response variable given all covariates. Formally, we abstain from conditioning on the covariates in the model equations of this thesis for a clearer notation. Nevertheless, this condition is implied.

## Publications

As research is a dynamic process, parts of this dissertation have already been published in peer reviewed journals or as technical reports and have been done in cooperation with supervising coauthors. Parts of this thesis can be found in

- Heinzl, F. and G. Tutz (2012). Clustering in linear mixed models with a group fused lasso penalty. Technical Report 123, Ludwig-Maximilians-University Munich. (resubmitted). (Chapter 3)

- Heinzl, F. and G. Tutz (2013). Clustering in linear mixed models with approximate Dirichlet process mixtures using EM algorithm. *Statistical Modelling 13*, 41-67. (Chapter 4)

- Heinzl, F., L. Fahrmeir, and T. Kneib (2012). Additive mixed models with Dirichlet process mixture and P-spline priors. *Advances in Statistical Analysis 96*, 47–68. (Chapter 5)

See the corresponding chapters for more details.

## Software

For most of the computations in this thesis the programming language C++ (Stroustrup, 1997) and the statistical software R (R Development Core Team, 2012) were used. All new proposed methods are implemented in C++ for a computing time as low as possible. These C++ functions make use of the libraries `ASA047` (Burkhardt, 2008) and `Newmat` (Davies, 2008) and are embedded in R wrapper functions, that are made available by the self-implemented R add-on package `clustmixed` (Heinzl, 2012), which will presumably be made publicly accessible via CRAN (see `http://www.r-project.org`). A test version of the package can be downloaded from `http://www.statistik.lmu.de/~heinzl/research.html`. This package imports the packages `Matrix` (Bates and Maechler, 2012), `lme4` (Bates et al., 2012), `splines` (Bates and Venables, 2011), `ellipse` (Murdoch and Chow, 2012), and `coda` (Plummer et al., 2012). For comparison to other approaches in the simulation studies the R package `lme4` of Bates et al. (2012) and the software BayesX (Belitz et al., 2012) were used. More information can be found in the corresponding sections.