Longitudinal Data Analysis
8. (Bayesian) Growth Curve Mixture Models

Prof. Dr. Holger Brandt

University of Tübingen
Methods Center

December 11, 2023

Mixture modeling/latent class models

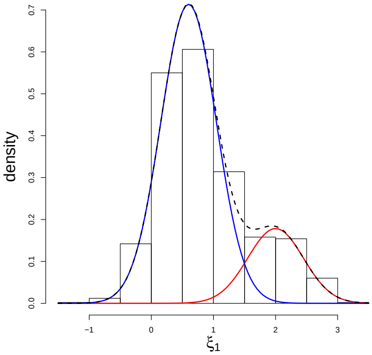


Figure: Mixture modeling allows to model nonnormality by approximating distributions with a mixture of (normal) distributions

Mixture modeling in astronomy

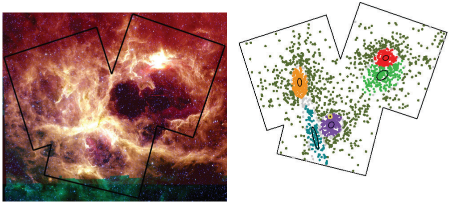


FIGURE 19.2
Left: The mid-infrared view of NGC 6357 seen by the *Spitzer Space Telescope*. The molecular clouds, forming several bubbles, are seen prominently in these images but the star clusters are not immediately evident. Right: The cluster members identified from the X-ray/infrared MYStIX study, which are color-coded by group from the mixture model. Light grey stars have ambiguous cluster memberships, and dark green stars are members of a distributed population in the model. The core-regions of the various mixture components are shown as black ellipses. (Feigelson *et al.* , 2013; Broos *et al.* , 2013; Kuhn *et al.* , 2013, 2014)

Figure: From Kuhn & Feigelson 2017, `https://arxiv.org/pdf/1711.11101.pdf`

Mixture modeling: How does it work?

- Assume a number of latent classes (e.g., $c = 2$).
- Specify a model for each class. The classes need to be different in at least one (meaningful) parameter; in principle all parameters can be different:

$$Y_{itc} = \eta_{0ic} + \lambda_{tc}\eta_{1ic} + \epsilon_{itc} \tag{1}$$

with

$$\boldsymbol{\eta}_c \sim N(\boldsymbol{\kappa}_c, \boldsymbol{\Psi}_c) \tag{2}$$
$$\boldsymbol{\epsilon}_c \sim N(\mathbf{0}, \boldsymbol{\Theta}_c) \tag{3}$$

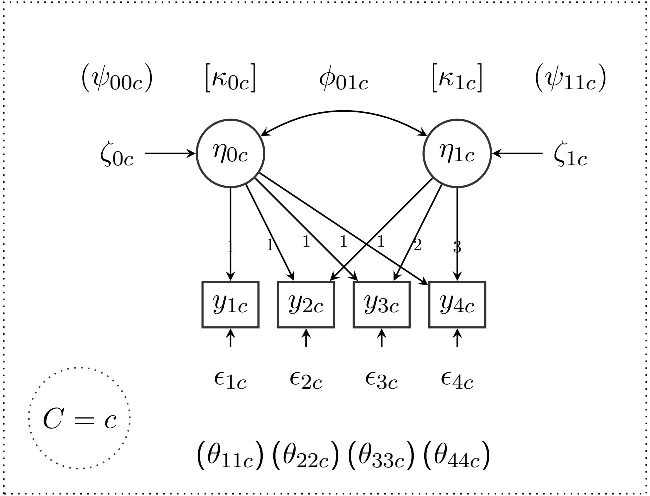
- Class membership is unknown a priori.
- Both class membership and class-specific models are estimated simultaneously.

Some examples

Mixture models are very flexible. They can account for a wide range of heterogeneous growth patterns.

- Class-specific factor loadings: They represent different growth curves for (unknown) subgroups, like a linear vs. quadratic
- Class-specific means:
    - The average trajectories are different (e.g., persons with a change vs. no change)
    - The variances of the trajectories are different (e.g., a homogeneous vs. a heterogeneous growth pattern)

Simple mixture cfa



Mixture model: Parameters to be estimated

- Parameters in the model: $\kappa_{0c}, \kappa_{1c}, \psi_{00c}, \psi_{11c}, \psi_{01c}, \theta_{11c}, \theta_{22c}, \theta_{33c}, \theta_{44c}$, i.e. 9 per class
- For the latent class model: $P(C = 1)$, . . . , $P(C = c - 1)$ for c classes (i.e. number of classes -1 parameters)
- All parameters can be class-specific.
- Typically, only some are actually class-specific and all else are constraint across classes.

Identification and interpretation

- If no differences in the classes exist, models sometimes do not converge (i.e. the data is normally distributed and homogeneous) or produce classes with very small class sizes.
- The mixture model only distinguishes groups by their distributions. It is sometimes complicated to interpret the classes as meaningful groups
  - *Direct* applications: Interpretation of classes as meaningful subpopulations
  - *Indirect* applications: Account for nonnormality, heterogeneity in growth or nonlinear relationships between growth factors/covariates
- For direct applications additional knowledge/theories about meaning of groups is important.

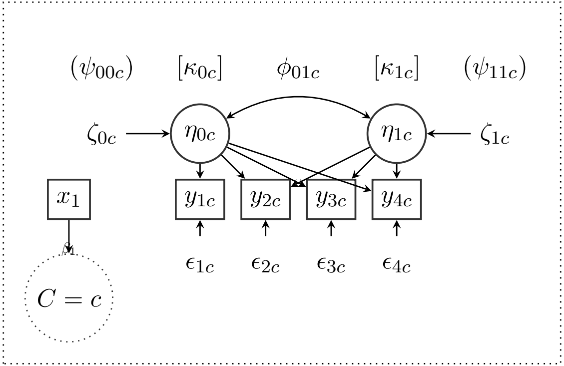
Explanation of class membership

- Increase the plausibility of class membership interpretation by including a (logistic) model to predict the classes

$$P(C = c|\mathbf{w}) = logit(\mathbf{w}\boldsymbol{\beta}) \tag{4}$$

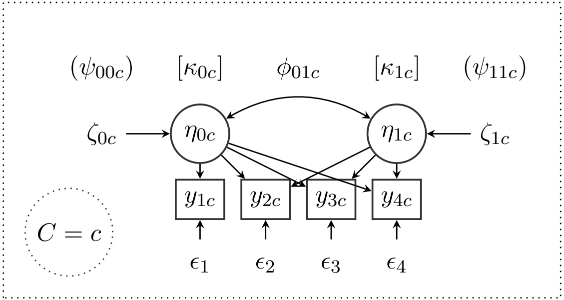
where $\mathbf{w}$ is a (baseline) covariate vector.

Mixture model: prediction of class membership



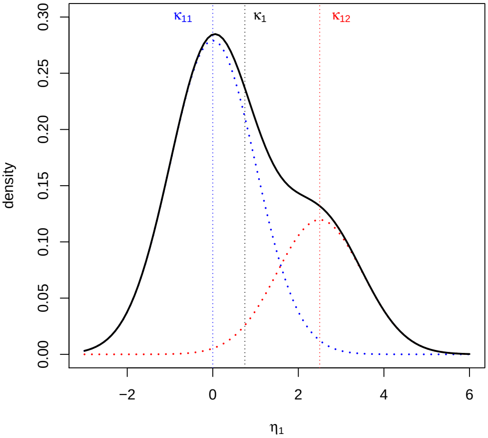
$$P(C = c|w) = \frac{\exp(\beta_{0c} + \beta_{1c}w)}{\sum_{j=1}^{C} \exp(\beta_{0j} + \beta_{1j}w)} \tag{5}$$

Indirect application of a GMM



- Constraint: $\theta_{ttc} = \theta_{tt}$ across classes
- Modeling of nonnormal distribution of factors.

Marginal mixture distribution of η_1



Marginal distribution in mixture models

- Probability for class membership

$$\pi_c := P(C = c) \in [0; 1] \text{ with } \sum_{c=1}^{C} \pi_c = 1 \tag{6}$$

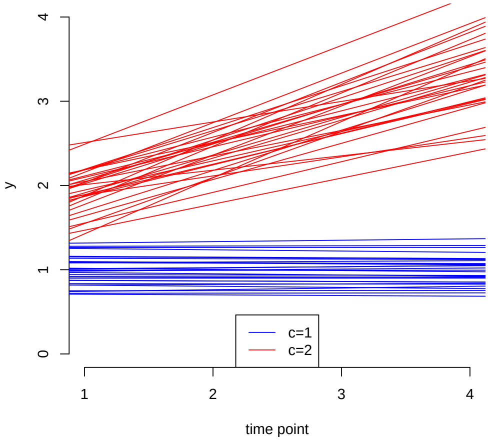
- Mean vector for latent growth factors

$$\boldsymbol{\kappa} = \sum_{c=1}^{C} \pi_c \boldsymbol{\kappa}_c \tag{7}$$

- Covariance matrix for latent growth factors

$$\boldsymbol{\Phi} = \sum_{c=1}^{C} \pi_c (\boldsymbol{\Phi}_c + \boldsymbol{\kappa}_c \boldsymbol{\kappa}_c') - \boldsymbol{\kappa}\boldsymbol{\kappa}' \tag{8}$$

Mixture distribution of trajectories



Frequentist Estimation

- Models are estimated using a so called "Expectation Maximization" (EM) algorithm (?, ?) that cycles through two steps until convergence:
  1. Expectation: Predict/(re)-assign class membership
  2. Maximization: Maximize the likelihood with the given class memberships

Bayesian Estimation

- Bayesian LGMs can be extended using an additional grouping variable (C). This variable is a parameter in the model that can be sampled with respective (conjugate) priors.
  1. In models without additional prediction model $C \sim dcat(P(C))$ with $P(C) \sim Beta(a, b)$ or $P(C) = logit(\mu_c), \mu_c \sim N(a, b)$, and $dcat$ as multinomial (categorical) distribution in jags (with a range of 1 to C)
  2. With additional predictors, μ_c can be modeled to be person-specific, i.e. a logistic regression for C.
  3. Extensions to more than $C = 2$ classes use a Dirichlet distribution for $P(C)$, which is an multivariate extension of the Beta distribution.
- Label switching problem: Across chains and even within chains (across iterations), labels can switch and groups switch their meaning. This needs to be avoided with respective modeling (e.g., constrain $\kappa_1 > \kappa_2$)

Model fit

- The GMM does not provide model fit statistics such as χ^2 tests or CFI, RMSEA etc. because the necessary saturated model ("S" for linear models) does not exist.
- However, models provide a likelihood value and information criteria (AIC, BIC).
- Similarly, Bayesian model fit criteria can be used (WAIC, loo).
- Typical strategy for fitting GMM's:
  - Start with a single class model, then increase the number of classes (use a meaningful model based on theory)
  - Choose the model with the smallest AIC/BIC
- In addition: Extraction of too many classes leads to very small class sizes (or even empty classes) with problems of identification.

Assessing mixture models

- Run models with different number of classes
- Compare the models using the information criteria. Models are generally not nested in general (so you cannot use a likelihood ratio test in frequentist models).
- Select the most parsimonious model
- Note: Model building is complex. A well-developed theory for the model is essential.