DEAKIN UNIVERSITY AUSTRALIA

A Bayesian Nonparametric Approach to Multilevel Regression

Vu Nguyen (1), Dinh Phung (1), Svetha Venkatesh (1), Hung Bui (2)

(1) Deakin University, Australia

(2) Adobe Research, USA

PAKDD 2015

DEAKIN UNIVERSITY AUSTRALIA

Content

Multilevel Data

Multilevel Regression

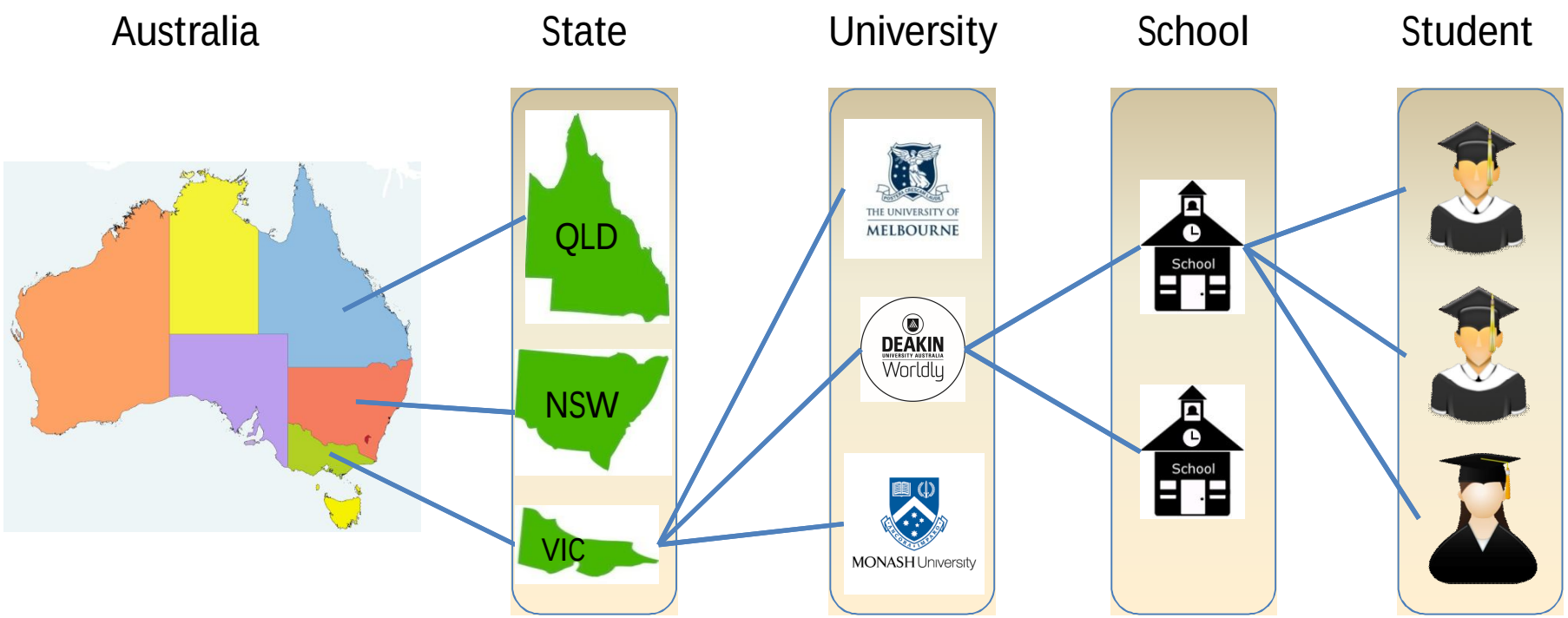
Dirichlet Process Mixture

Bayesian Nonparametric Multilevel Regression

Experiments

Question & Answer

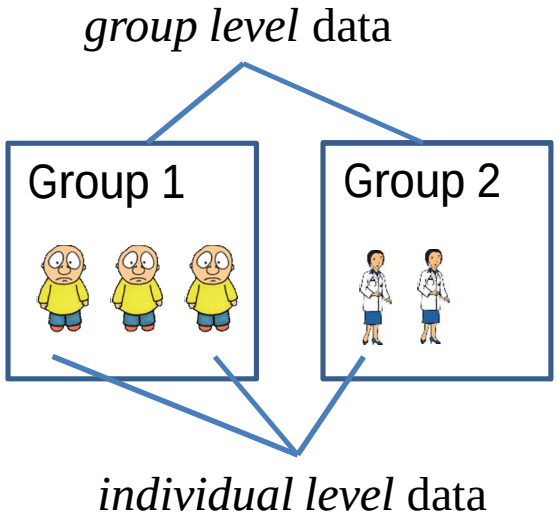
Real world data is often in multilevel.



Examples of Multilevel Data

Data naturally present themselves in groups.

- Students are grouped into classes.
- Words are grouped into documents.
- Pixels are grouped into images.

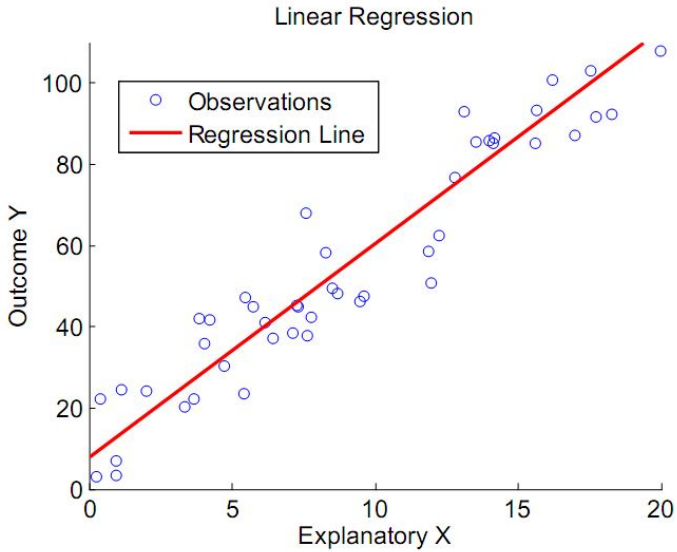


Individuals in the same group are likely to be more similar than individuals in different groups.

DEAKIN UNIVERSITY AUSTRALIA

Linear Regression

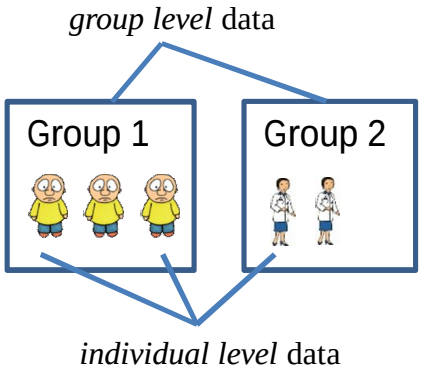
Modelling and analyzing the linear relationship between explanatory X and outcome Y.



Linear regression is suitable for flat data setting, not for multilevel data.

Multilevel Regression

Multilevel regression is a type of regression analysis for multilevel data where explanatory and outcome variables are organized in groups.



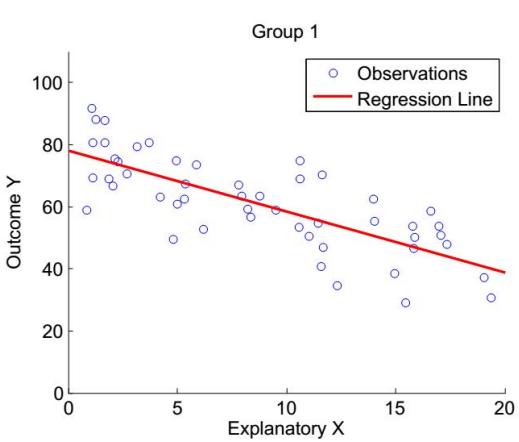
The task is named as *multilevel regression* for two reasons:

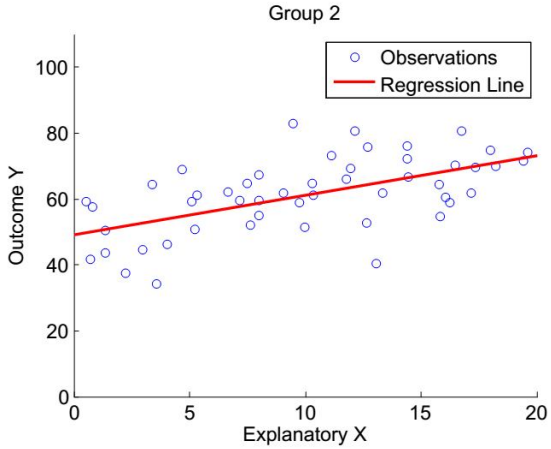
1. The model is designed to perform regression for multilevel data.
2. The model parameters are in multilevel structure.

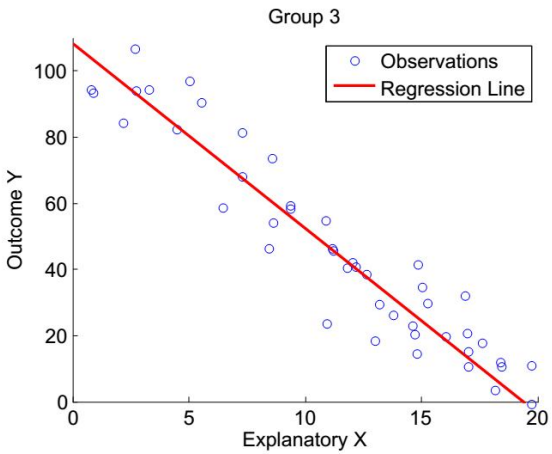
Multilevel Regression

Individuals in the same group are likely to be more similar than individuals in different groups.

Different groups may have dissimilar behaviours in regression.







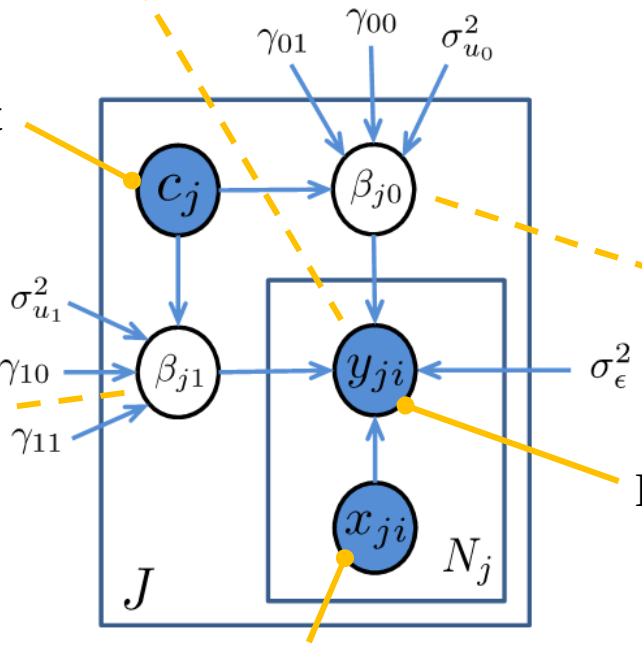
Standard single-level regression may fail due to poor fitting for multilevel data.

Multilevel Regression

Linear Mixed Effects model

$$y_{ji} = \beta_{j0} + x_{ji}^T \beta_{j1} + \epsilon_{ji}$$

$$\epsilon_{ji} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$$



Group-level context

$$\beta_{j0} = \gamma_{00} + \gamma_{01} c_j + u_{j0}$$

$$u_{j0} \sim \mathcal{N}\left(0, \sigma_{u0}^2\right)$$

$$\beta_{j1} = \gamma_{01} + \gamma_{11} c_j + u_{j1}$$

$$u_{j1} \sim \mathcal{N}\left(0, \sigma_{u1}^2\right)$$

Individual-level outcome

Individual-level explanatory

Multilevel Regression

Linear Mixed Effects model

$$y_{ji} = \underbrace{\gamma_{00} + \gamma_{01} c_j + \gamma_{10} x_{ji} + \gamma_{11} c_j x_{ji}}_{\text{fixed effects}} + \underbrace{u_{j0} + u_{j1} x_{ji} + \epsilon_{ji}}_{\text{random effects}} .$$

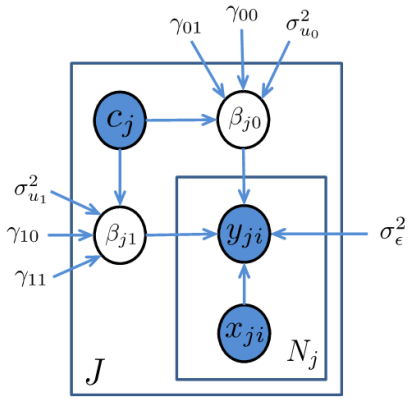
Mixed Effects

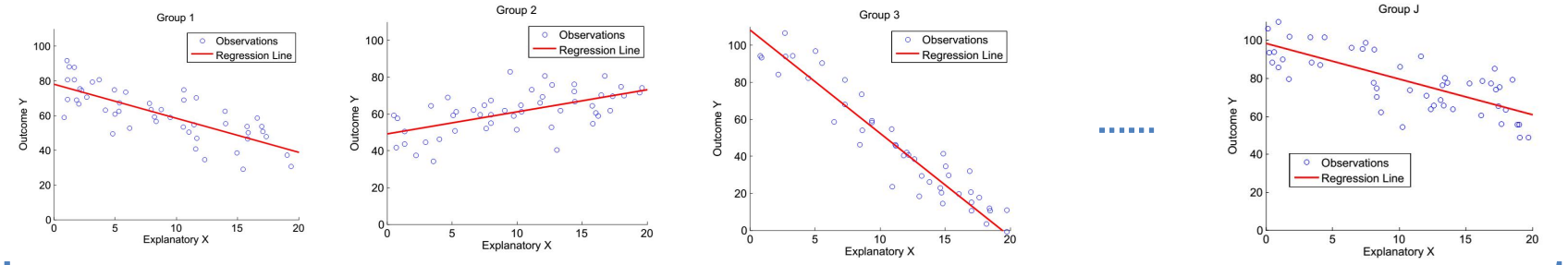
The random effects u_{j0} and u_{j1} varied across groups, but ϵ_{ji} does not.

Multilevel Regression

Linear Mixed Effects model

Parameter estimation using Expectation Maximization.





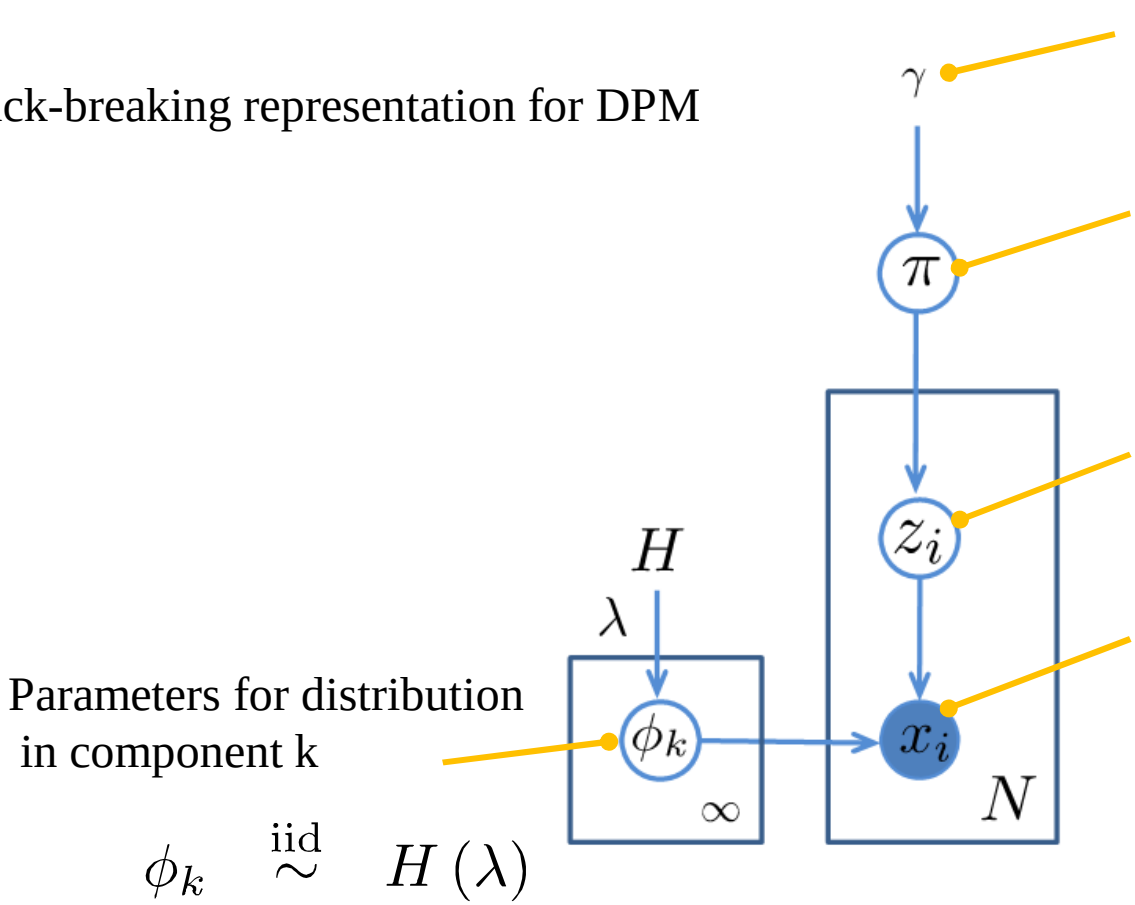
Output of LME: J regression patterns β_j for J groups

Problem:

- Predicting unseen groups (e.g., J+1) in multilevel regression is hard.
- How many regression patterns should be used for modelling?

Dirichlet Process Mixture

Stick-breaking representation for DPM



Concentration parameter

Mixing proportion

$$\pi \sim \mathrm{GEM}(\gamma)$$

Latent assignment

$$z_i \overset{\text{iid}}{\sim} \pi$$

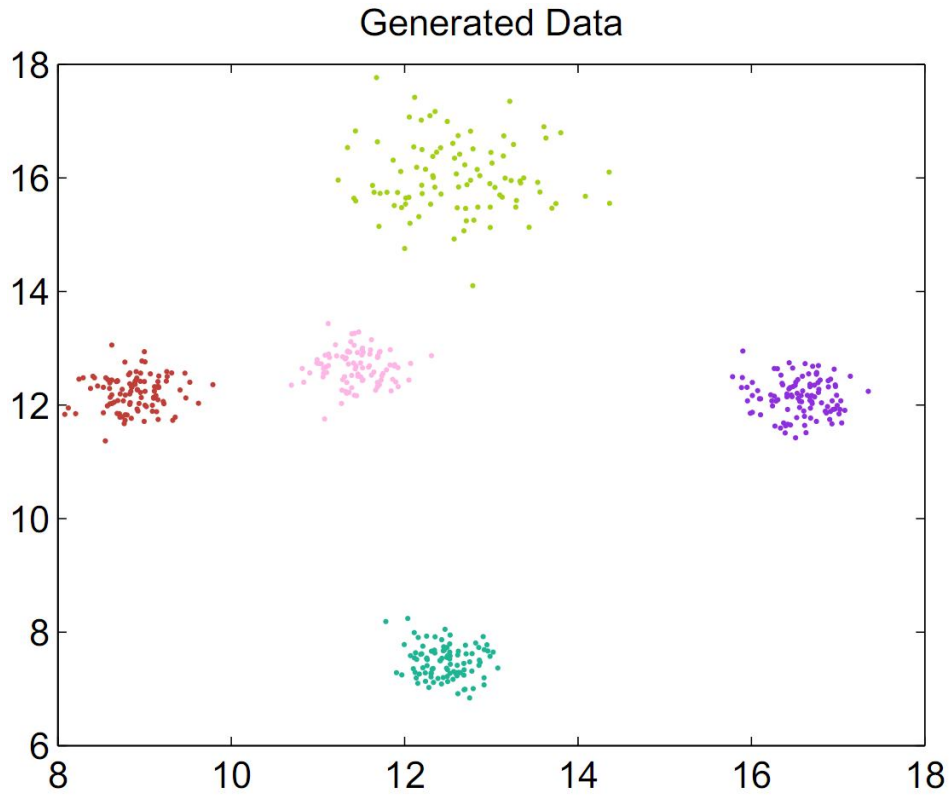
Data observation

$$x_i \sim F(\phi_{z_i})$$

Parameters for distribution in component k

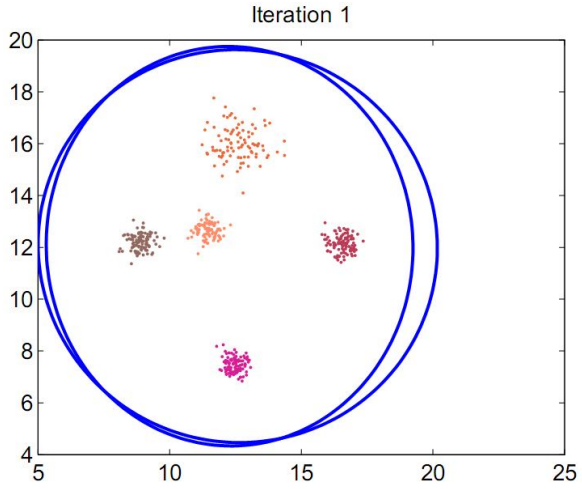
$$\phi_k \overset{\text{iid}}{\sim} H(\lambda)$$

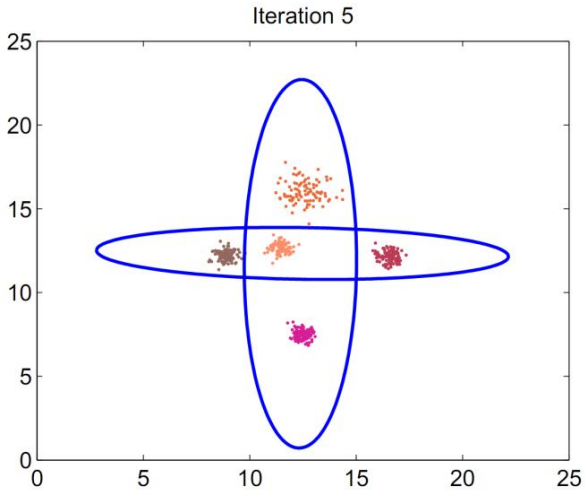
Demo DPM for nonparametric clustering

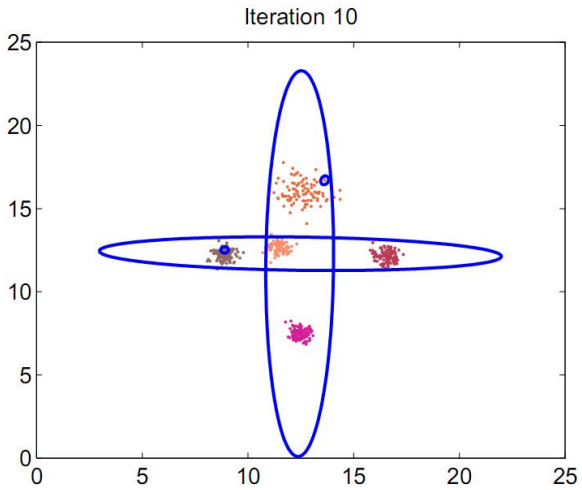


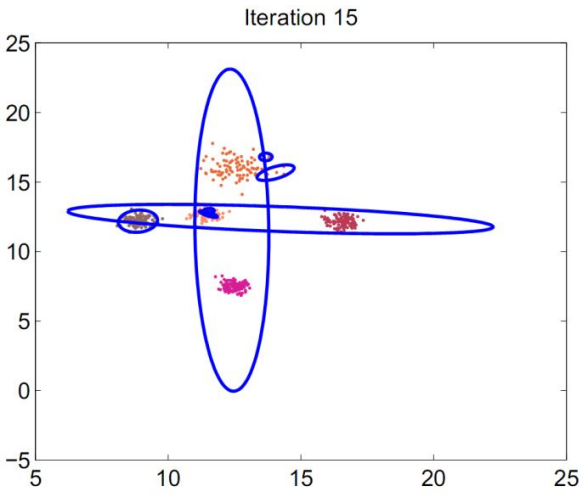
There are K=5 clusters
which different means and variances.

Demo Dirichlet Process Mixture

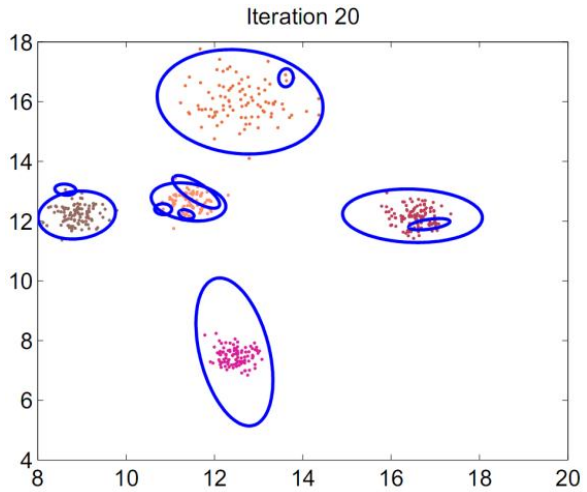


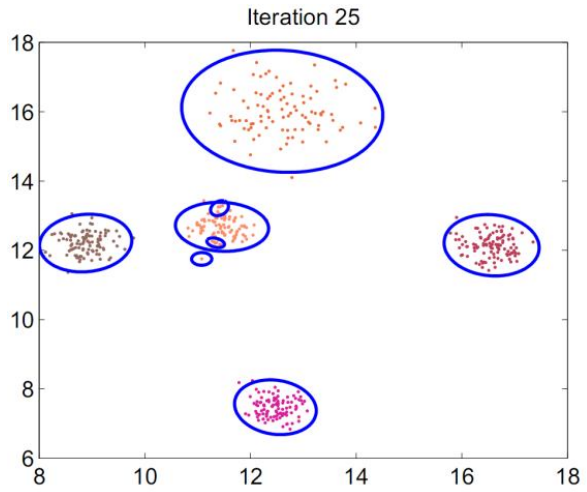


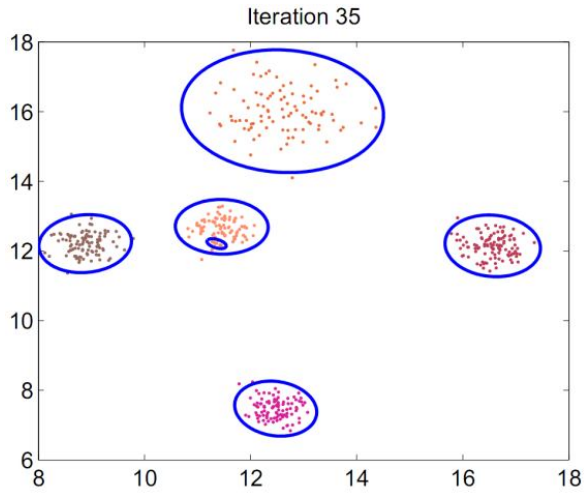


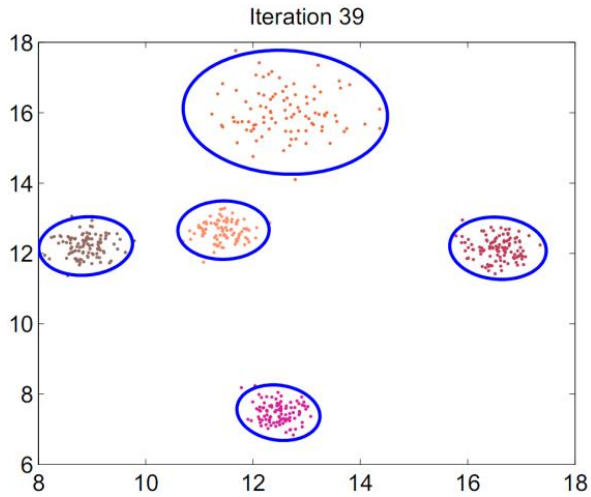


Demo Dirichlet Process Mixture









DEAKIN UNIVERSITY AUSTRALIA

Content

DEAKIN UNIVERSITY AUSTRALIA

Bayesian Nonparametric Multilevel Regression

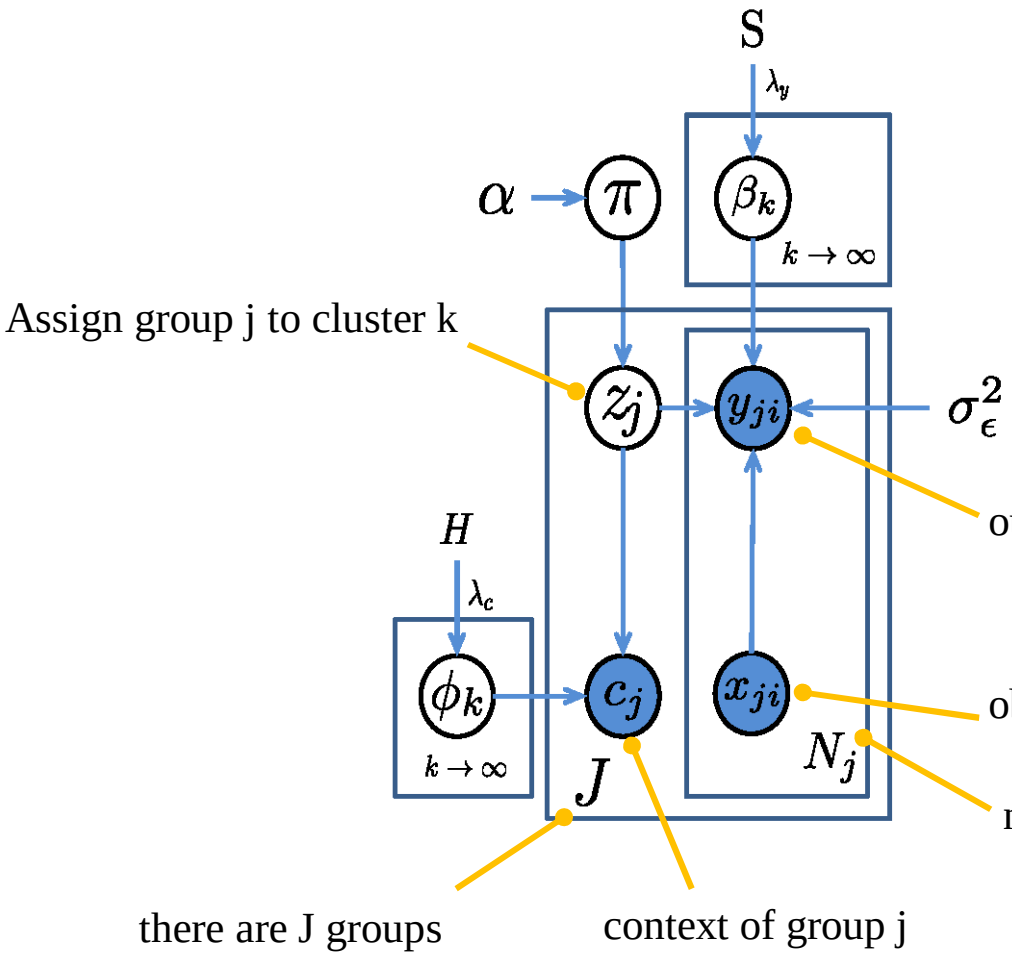
Problems in multilevel regression:

- Predicting unseen groups (e.g., J+1) in multilevel regression is hard.
- How many model should be used for regression?

Our solution is to propose a model that can:

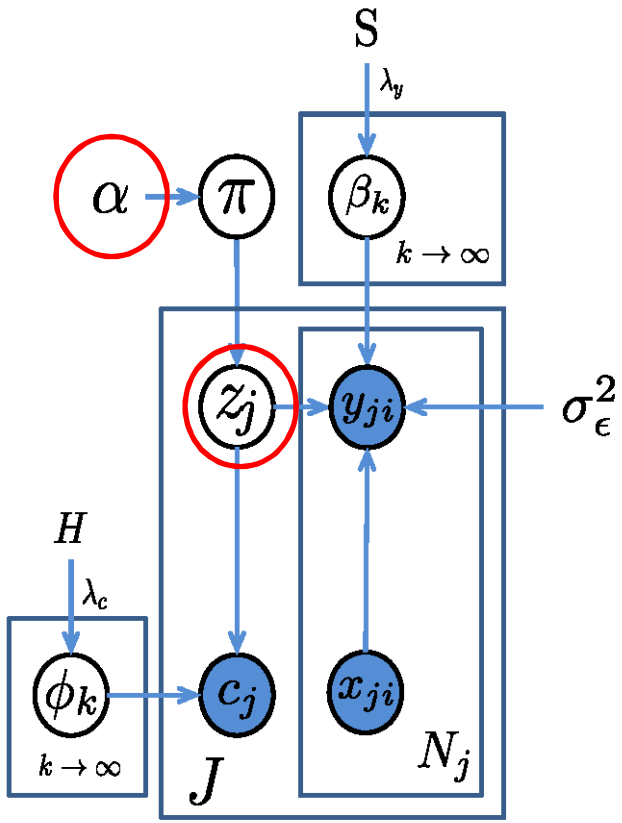
- Model multilevel data for regression.
- Utilize group-level context observation for improving performance.
- Exploit the sharing property of individuals within groups.
- Predict individuals from unseen groups.
- Identify the suitable number of regression patterns underlying inside the data.

Bayesian Nonparametric Multilevel Regression



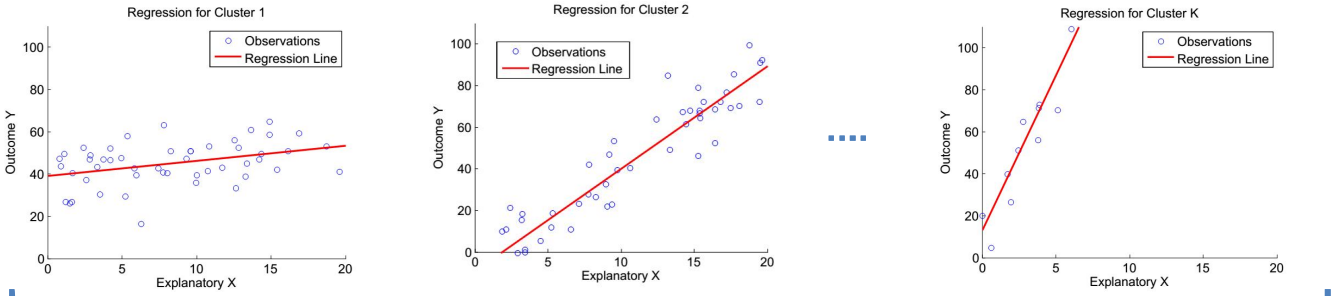
Bayesian Nonparametric Multilevel Regression

Using collapsed Gibbs sampler to learn latent parameters z_j and α



Bayesian Nonparametric Multilevel Regression

BNMR identifies a set of *core* regression patterns which can representing for all groups. Each regression pattern β_k associated with a context atom ϕ_k

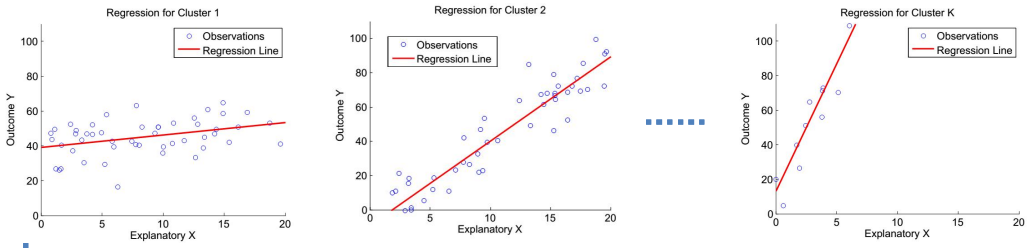


Estimating K regression patterns for J groups

- K is unknown and discovered by Bayesian nonparametric setting.
- K ranges from 1 to J.
- When K=1, BNMR becomes standard single-level regression.
- When K=J, BNMR is similar to LME.

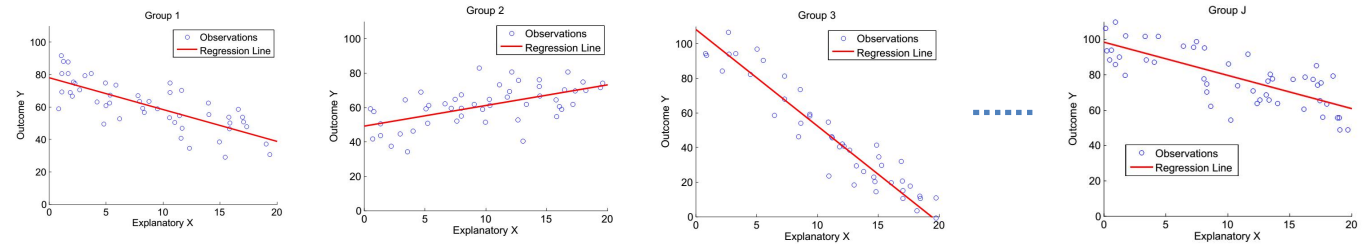
BNMR compare to LME

- BNMR identifies a set of *core* regression patterns for all groups.



$K = [1...J]$ regression patterns for J groups

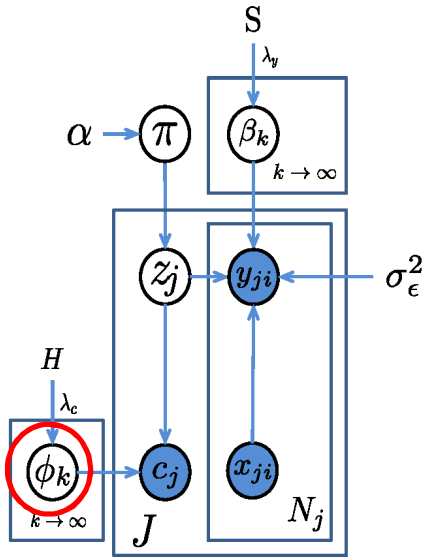
- LME estimates a regression pattern for each group.

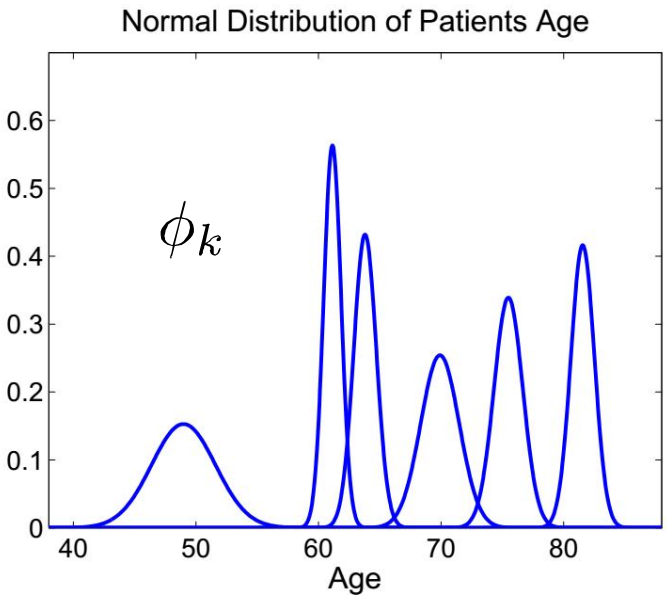


J regression patterns, each of which is for a group

BNMR compare to LME (continue)

- BNMR also discovers the context distributions ϕ_k





- LME utilizes context observation but does not result in distributions over context.

Bayesian Nonparametric Multilevel Regression

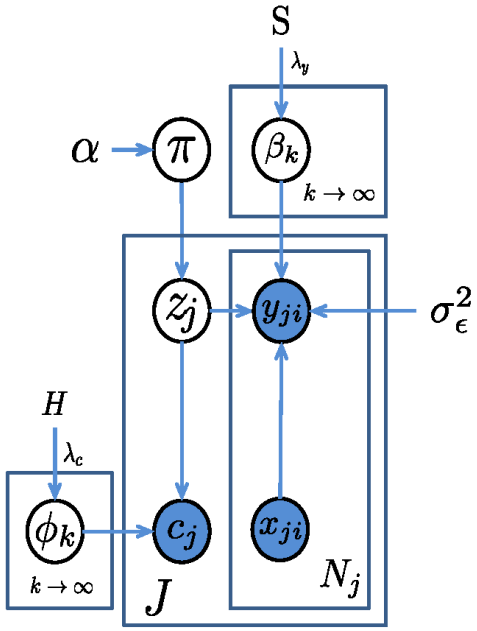
Predicting outcome for the test data

$$\hat{y}_{ji}^{\text{Test}} \propto \sum_{z_j^{\text{Test}}=1}^{K} \left[\beta_{z_j}^T x_{ji}^{\text{Test}} \right] \times p\left(z_j^{\text{Test}} \mid c_j^{\text{Test}} \right)$$

where $p\left(z_j^{\text{Test}} \mid c_j^{\text{Test}} \right) \propto p\left(z_j^{\text{Test}} \mid \pi \right) \times p\left(c_j^{\text{Test}} \mid \phi_{z_j^{\text{Test}}} \right)$

BNMR utilizes context information available in unseen groups (J+1), then perform prediction.

It is applicable for predicting individuals from unseen groups of data.



DEAKIN UNIVERSITY AUSTRALIA

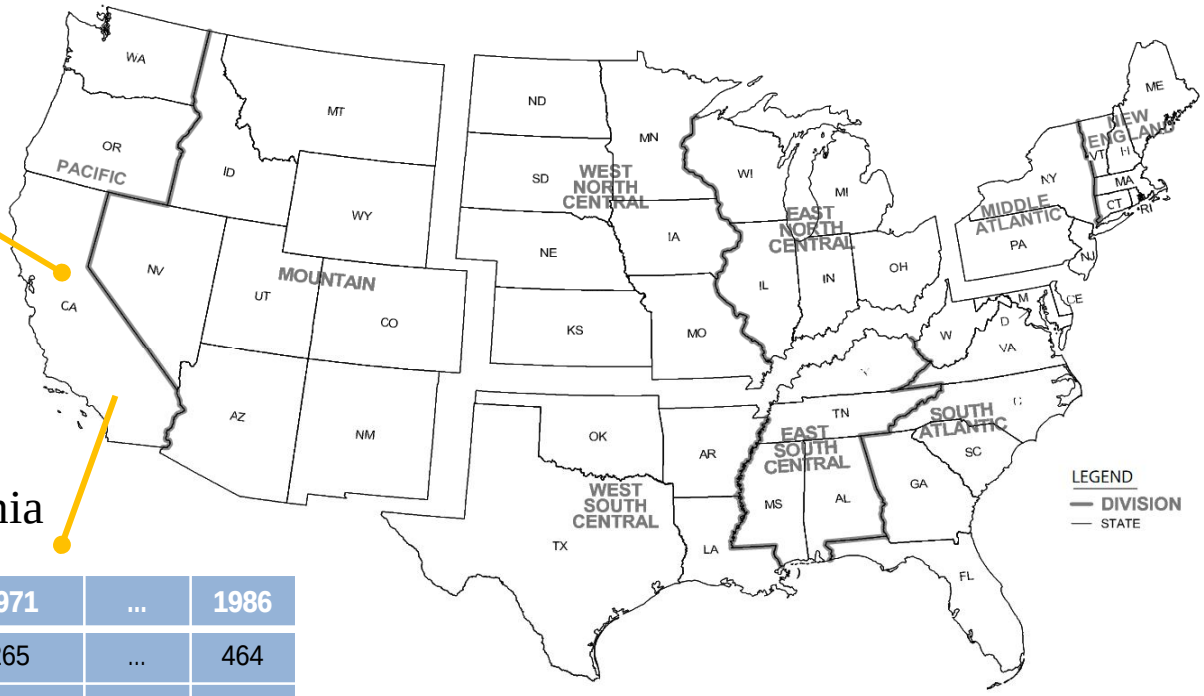
Experiments

- We emphasize on predicting outcomes of individuals from unseen groups of data.
- Unseen groups (J+1) are not available during training (1,2,…J).

Experiments

Econometric Panel Data

state = group



Individuals of group as California

	Year	1970	1971	...	1986
Y	GPD (billions $)	263	265	...	464
X	Capital Stock (billions $)	128	132	...	139
	Unemployment Rate	7	8	...	6
	...	.	.	.	.

Group level context is state population

Predicting GDP of each years in new states.

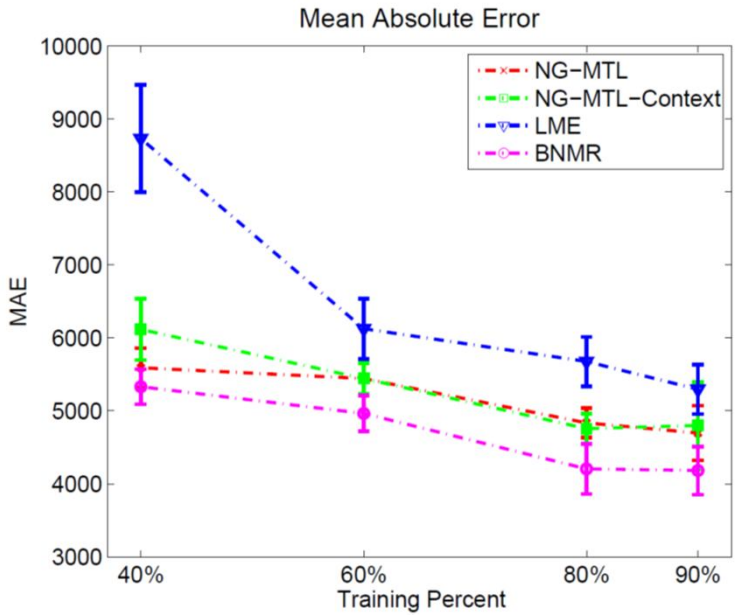
Experiments

Econometric Panel Data

We identify there are three clusters of states:

- High population (e.g., Cali, Texas, New York)
- Mid population (e.g., Arizona, Indiana, Missouri)
- Low population (e.g., Vermont, Montana)

BNMR achieves the best performance.



Predicting GDP of new states.

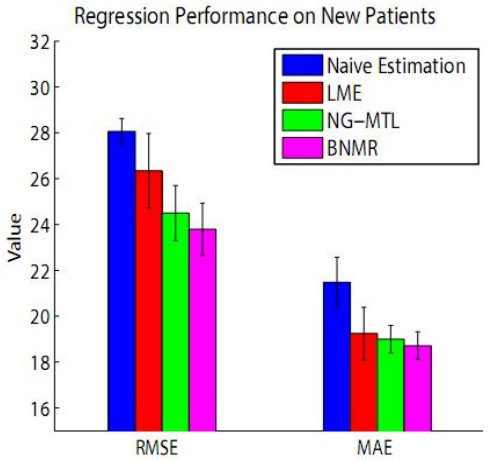
Experiments

Healthcare Longitudinal Data

Polyvascular Disease cohort

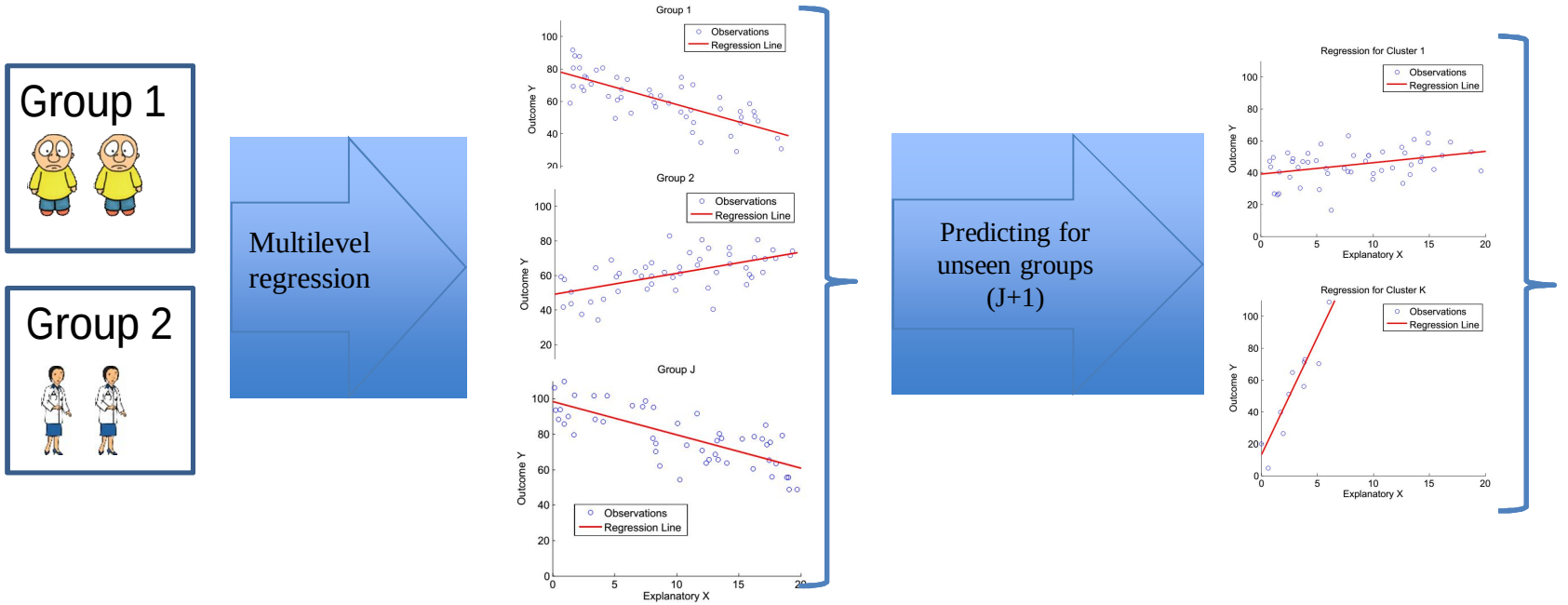
- 209 patients, each of whom has multiple admissions into hospital
- 3207 admissions in total
- x_{ji} includes *External Factor Code* and *Diagnosis Code*
- y_{ji} readmission gap (days)
- c_j is patient's age.

Predicting patient's readmission interval for new patients.



Summary

Individuals in the same group are likely to be more similar than individuals in different groups.



Multilevel data

Mixed Effects model

Bayesian Nonparametric Multilevel Regression

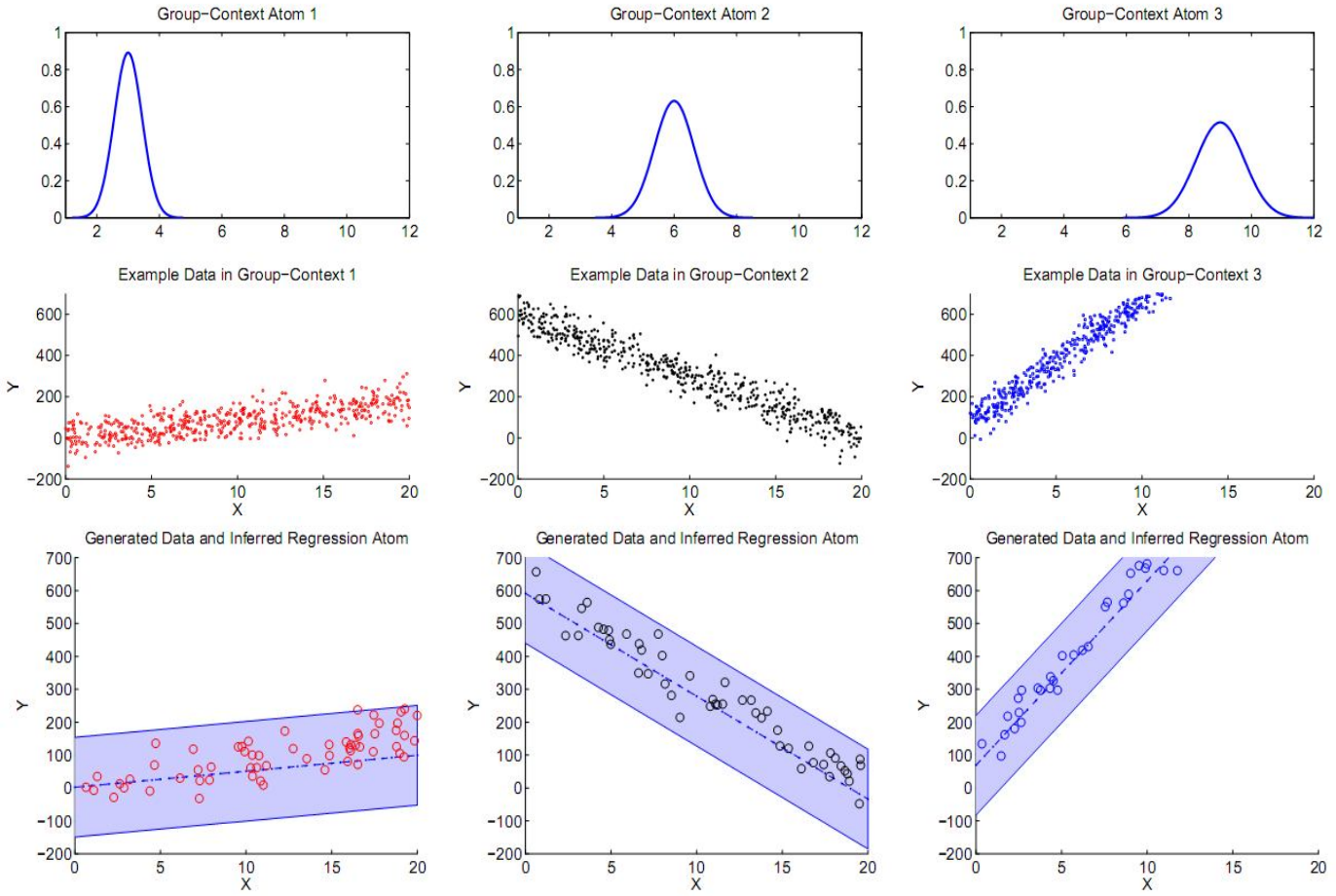
Simulation Study

We generate J=200 groups, each group includes Nj=20 individuals.

Group level context distributions

Generated observations

Estimated regression patterns

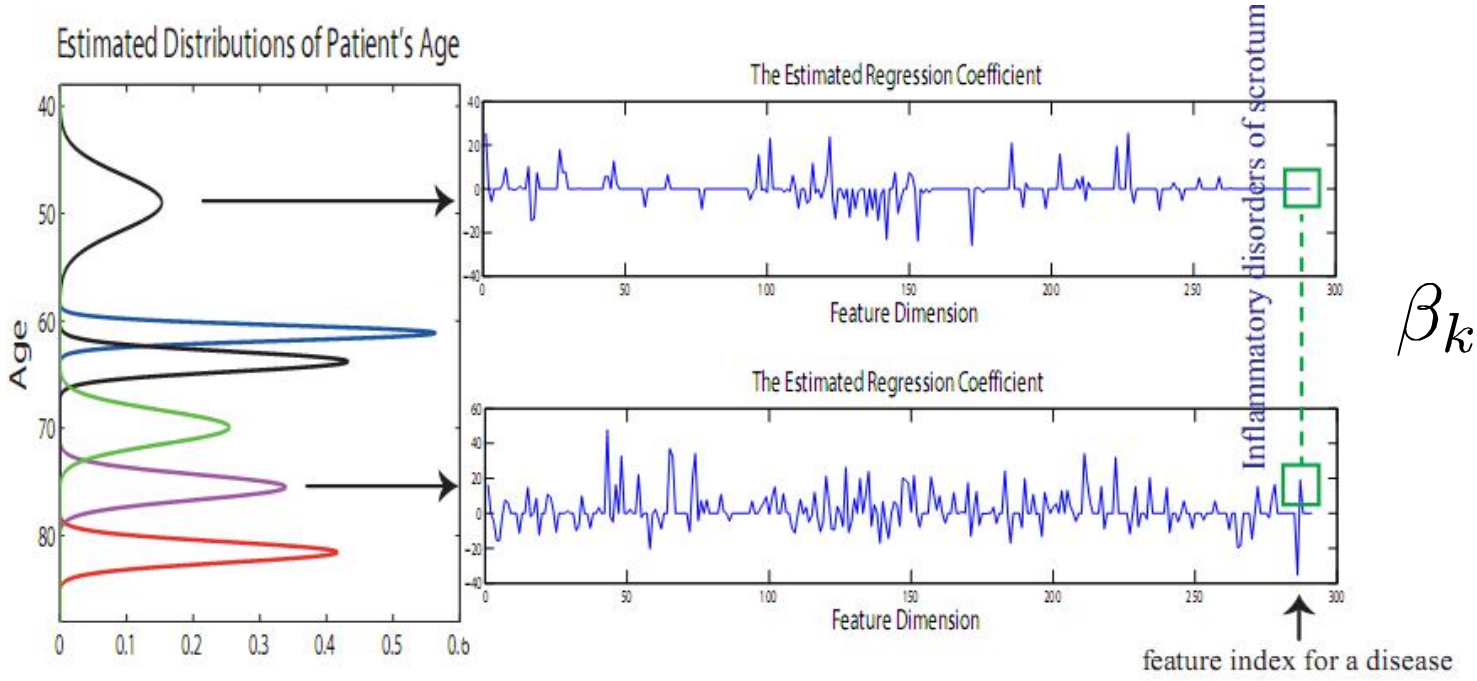


Experiments

Further analysis using estimated regression coefficient β_k and context atom ϕ_k

Zero element in regression coefficient index indicates there is no influence from this feature.

ϕ_k



β_k