Mixtures and Random Effects

John Hinde

Statistics Group,
School of Mathematics, Statistics and Applied Mathematics
National University of Ireland, Galway

john.hinde@nuigalway.ie

With thanks to:
Norma Coffey, Clarice Demétrio, Silvia de Freitas, Andrew Simpkin,
Georgios Papageorgio, Mariana Ragassi Urbano

Supported by SFI Awards 07/MI/012 & 07/RFP:MATF448

Seminar, Barcelona, Spain

15 March 2012

Summary

Normal Models

$$\mathbf{y} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\epsilon}$$

- single error term includes
  - individual observation/measurement error
  - *experimental* unit variability
  - unobserved covariates
- for simplest data structures/designs use **normal linear model**
- more complex situations
  - structure in *experimental* unit variability
  - repeated measures/longitudinal observations
  - ...

Normal Mixed Model

$$\mathbf{y} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{z} + \boldsymbol{\epsilon}$$

- $\mathbf{z}$ unobserved **random effects**
- shared random effects
  - multi-level/variance components models
  - longitudinal observations
  - spatial structure
- $\mathbf{z}$ normal
  - normal model with **structured covariance matrix**
- standard mixed model analyses – ML, REML
- widely available in standard software

Generalized Linear Models

Models for counts, proportions, times, ...

$$\mathbf{y} \sim F(\boldsymbol{\mu}) \qquad g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x}$$

- distributional assumption relates to the observation/measurement process
- how does this model incorporate
  - experimental/individual unit variability?
  - unobserved covariates?

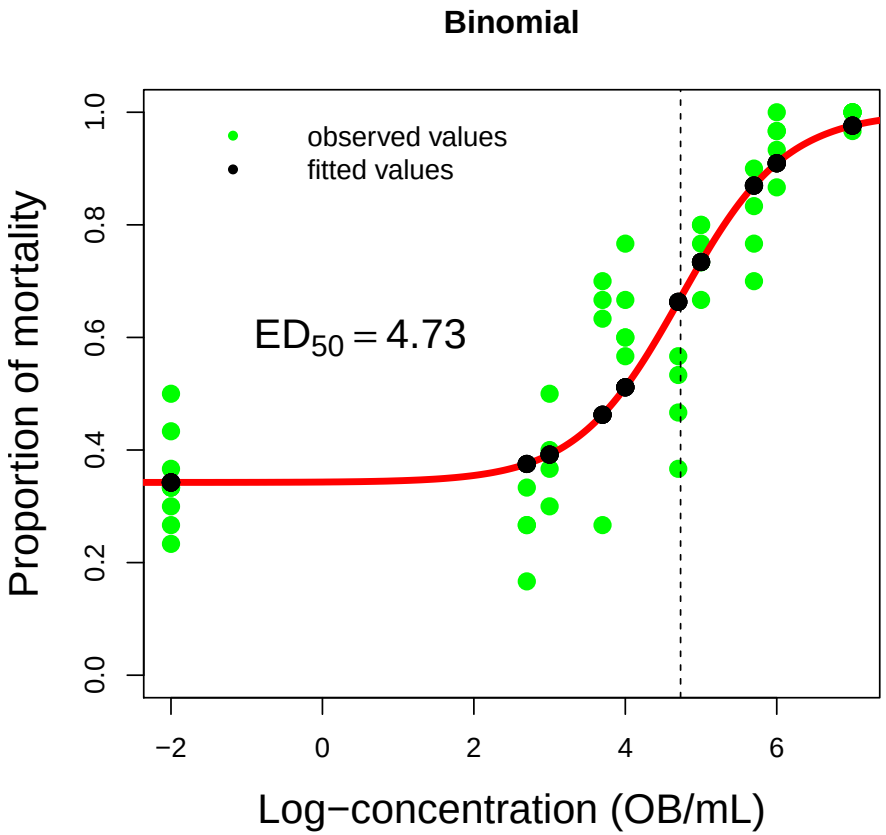
It doesn't!

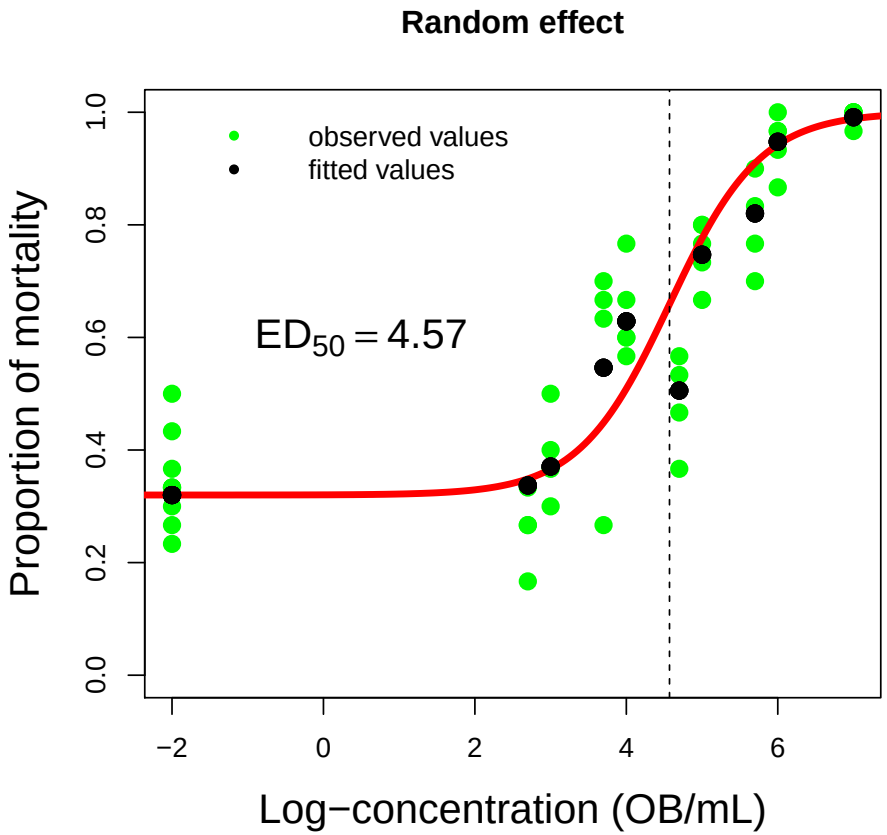
hence overdispersion, etc

Potato Larvae Bioassay

- Samples of potatoes were each infected with 30 larvae
- Different concentrations of an insecticide applied to potato samples
- Control sample (no insecticide) with 9 potatoes
- Experiment conducted at 18^{o}C, and after 60 days the numbers of dead larvae were counted
- Standard binomial model with log conc allowing for *natural mortality*
- Random effect at individual potato level — overdispersion
- Random effect at concentration level — measurement error

Potato Larvae Bioassay: Fitted Models





Random Effect Models

Include random effect(s) in the linear predictor

$$\boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{z}$$

- single conjugate random effect at individual level – standard overdispersion models
  - negative binomial for count data
  - beta-binomial for proportions
- $\mathbf{z}$ normal $\longrightarrow$ **generalized linear mixed models**
- $\mathbf{z}$ unspecified $\longrightarrow$ **nonparametric maximum likelihood**

Normal Random Effect

$$L(\boldsymbol{\beta}, \sigma) = \prod_{i=1}^{n} \int f(y_i|\boldsymbol{\beta}, \sigma, z_i)\phi(z_i)dz_i$$

where $\phi(z)$ is the normal density, and f the response density.

No analytic form for integral – approximate using K-point Gaussian Quadrature (mass points z_k with weights π_k)

$$L(\boldsymbol{\beta}, \sigma) \approx \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f(y_i|\boldsymbol{\beta}, \sigma, z_k)$$

Likelihood for K component mixture of response distribution with linear predictor for k-th component

$$\eta_{ik} = \boldsymbol{\beta}^T \mathbf{x}_i + \sigma z_k$$

EM Algorithm

Estimation for finite mixture conveniently viewed as EM algorithm.

E-Step: Calculate component weights w_{ik} – the posterior probability that observation y_i comes from component k:

$$w_{ik} = \frac{\pi_k f_{ik}}{\sum_\ell \pi_\ell f_{i\ell}}$$

M-step: Estimate $\widehat{\beta}$ and $\widehat{\sigma}$ by

- fitting response model to expanded data (K-copies)
- y-variate $\mathbf{y}^* = (\mathbf{y}', \mathbf{y}', \ldots, \mathbf{y}')'$
- explanatory variables for y^*_{ik}: $\mathbf{x}_i$ and z_k
- weights w_{ik}

Multi-Centre Beta-blocker Trial

- trial of beta-blockers to reduce mortality after myocardial infarction
- 22 centres
- single treatment – treatment and control groups
- patients *within* centres
- response r deaths out of n for each group
- centres very different sizes – 38 to 1916

require generalizability

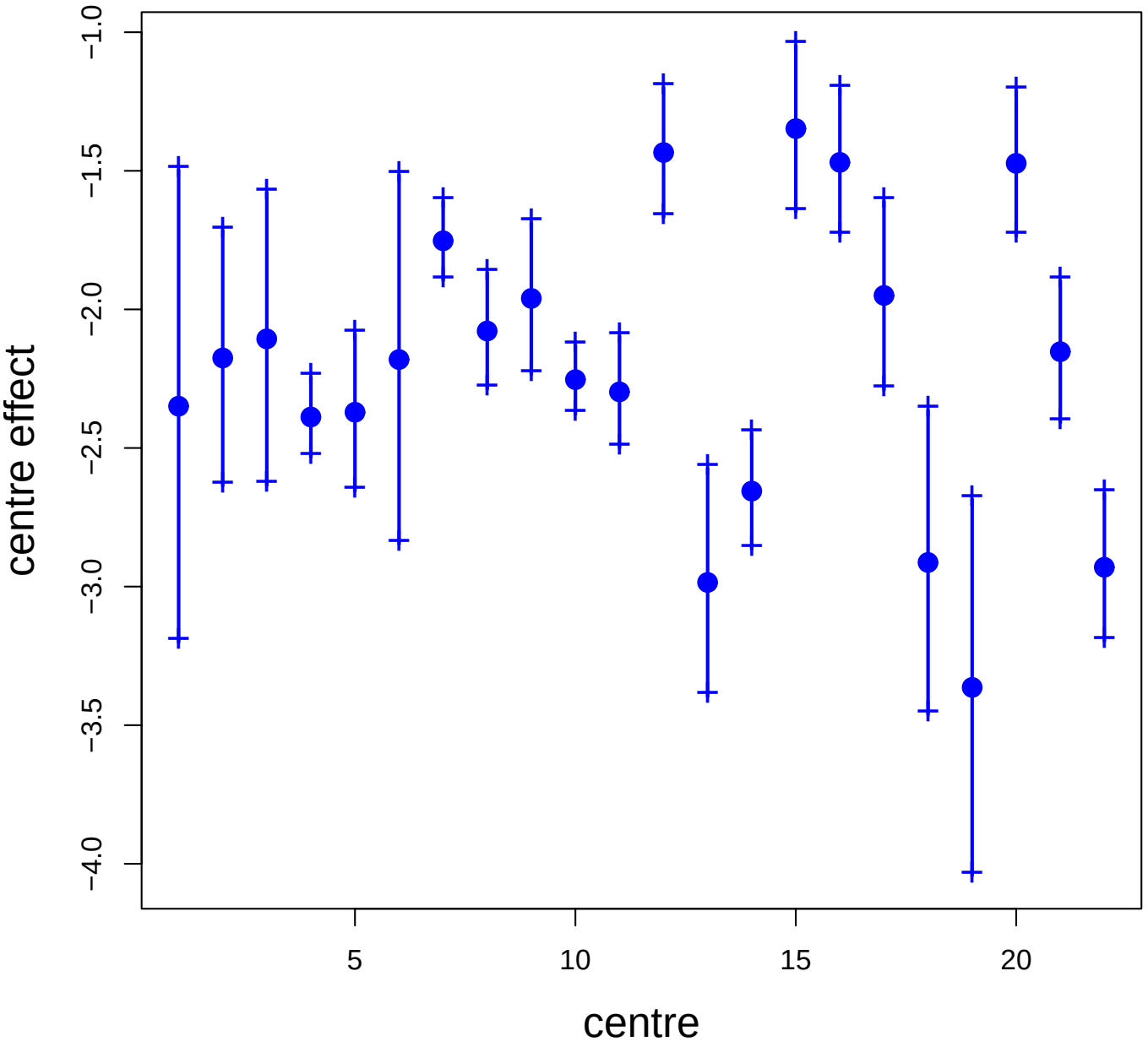
- Simple fixed effect model, ignoring among-centre variation

  Using binomial logit model, on log-odds scale

$$\text{treatment effect} = -0.257 \quad (\text{s.e.} = 0.049)$$

  Residual deviance: 305.76 on 42 df

- considerable among-centre variation . . .


centre effect
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
5
10
15
20
centre

Normal Random Effect – Variance Component

$$\text{logit}(p_{ij}) = \alpha + \beta\,\text{treat}_{ij} + \sigma\, Z_i \quad ; \quad Z_i \sim N(0,1)$$

Gaussian quadrature for betablockers

K	α	se	β	se	σ	$-2\log L$
1	−2.197	0.034	−0.257	0.049	0.000	523.2
2	−2.034	0.035	−0.257	0.050	0.366	365.5
3	−2.239	0.034	−0.258	0.050	0.360	321.0
5	−2.238	0.034	−0.258	0.050	0.455	319.8
10	−2.087	0.034	−0.262	0.050	0.454	318.6
20	−2.180	0.034	−0.261	0.050	0.432	316.7

Arbitrary Random Effects

$$L(\beta, g) = \prod_{i=1}^{n} \int f(y_i|\beta, z_i) g(z_i) dz_i$$

where $g(z)$ is the unspecified mixing distribution.
Use non-parametric maximum likelihood (NPML) estimate of g

a finite discrete distribution on K mass points

$$\begin{pmatrix} z_1, & z_2, & \cdots, & z_K \\ \pi_1, & \pi_2, & \cdots, & \pi_K \end{pmatrix}$$

The joint likelihood is

$$L(\beta, K, \pi_1, ..., \pi_{K-1}, z_1, ...z_K) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} f(y_i|z_k, \beta)\pi_k \right\}$$

Non-parametric Maximum Likelihood

EM technique is easily extended to incorporate estimation of a discrete mixing distribution for z with K mass points.

E-step: as before using current estimate of mixing distribution in place of quadrature points and weights.

M-step: Estimate β and $\{z_k\}$ by

- fitting response model to expanded data (K-copies)
- explanatory variables $\mathbf{x}_i$ and **a K-level factor**
- weights w_{ik}
- $\widehat{\pi}_k = \sum_{i=1}^{n} \frac{w_{ik}}{n}$

Fit models for different values of K until joint likelihood is maximized.

Beta-Blocker: NPML Variance Component

$$\text{logit}(p_{ij}) = \alpha + \beta\,\text{treat}_{ij} + Z_i \quad ; \quad Z_i \text{ unspecified}$$

K	β	se	σ	$-2\log L$	# Z-pars
1	−0.257	0.049	0.000	523.2	0
3	−0.258	0.050	*0.428*	318.7	4
5	−0.258	0.050	*0.488*	310.4	8
10	−0.258	0.050	*0.489*	308.6	18
Normal	−0.261	0.050	0.432	316.7	1

No evidence against normality

R-code and Computation

Simple model specification

```
allvc(cbind(r,(n-r))~treat,data=betablok,
        family=binomial,random=~1|center,
          k=5,random.distribution='np',tol=0.25))
```

- fixed and random models
- Gaussian or NPML random effects
- # of mass-points
- control over start for mass points – `tol`

Beta-blocker: Random Treatment Model

$$\text{logit}(p_{ij}) = \alpha + \beta\,\text{treat}_{ij} + Z_i + U_i\,\text{treat}_{ij} \quad ; \quad Z_i,\ U_i \text{ unspecified}$$

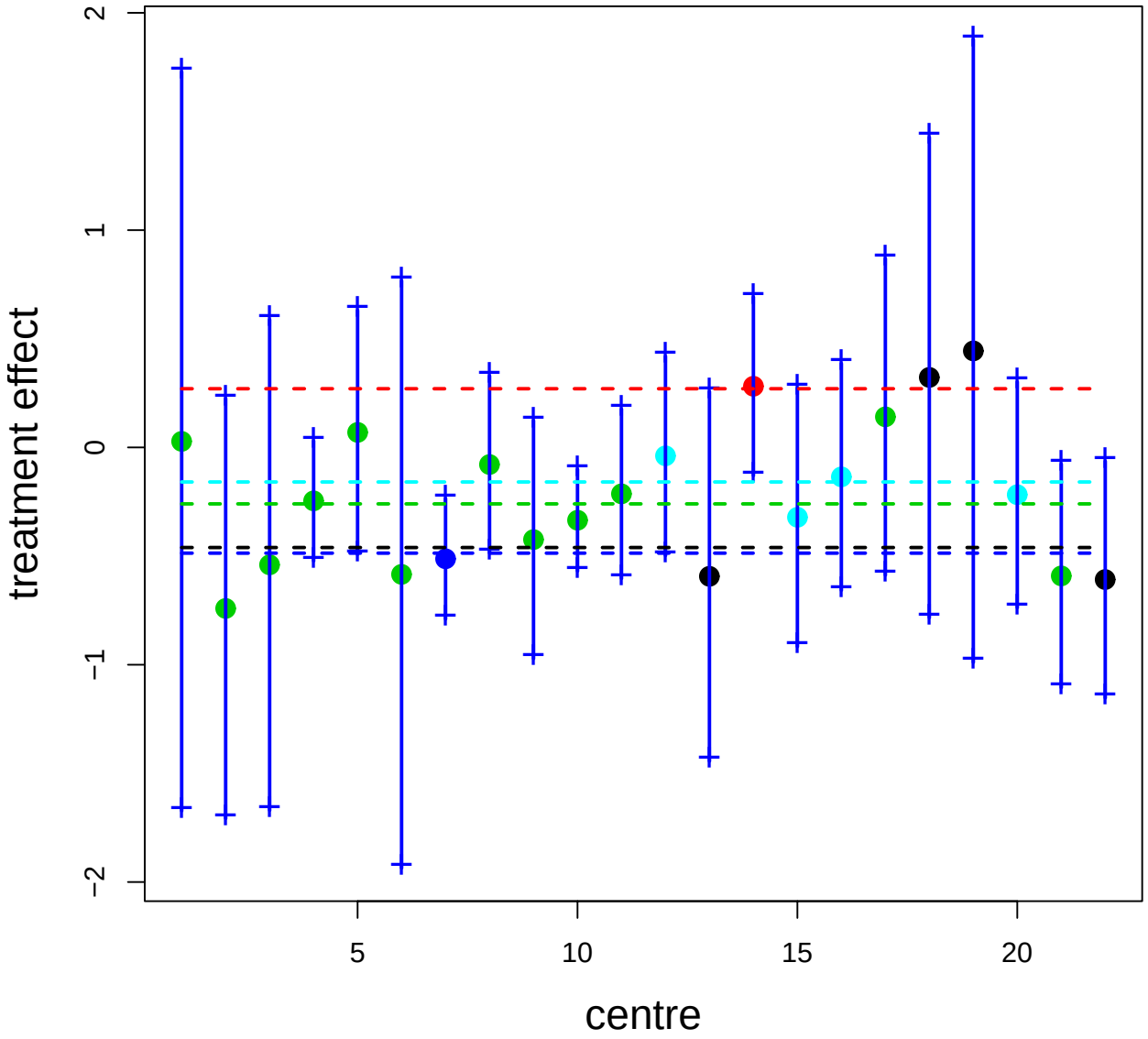
Finite mass-point distribution for (Z, U)

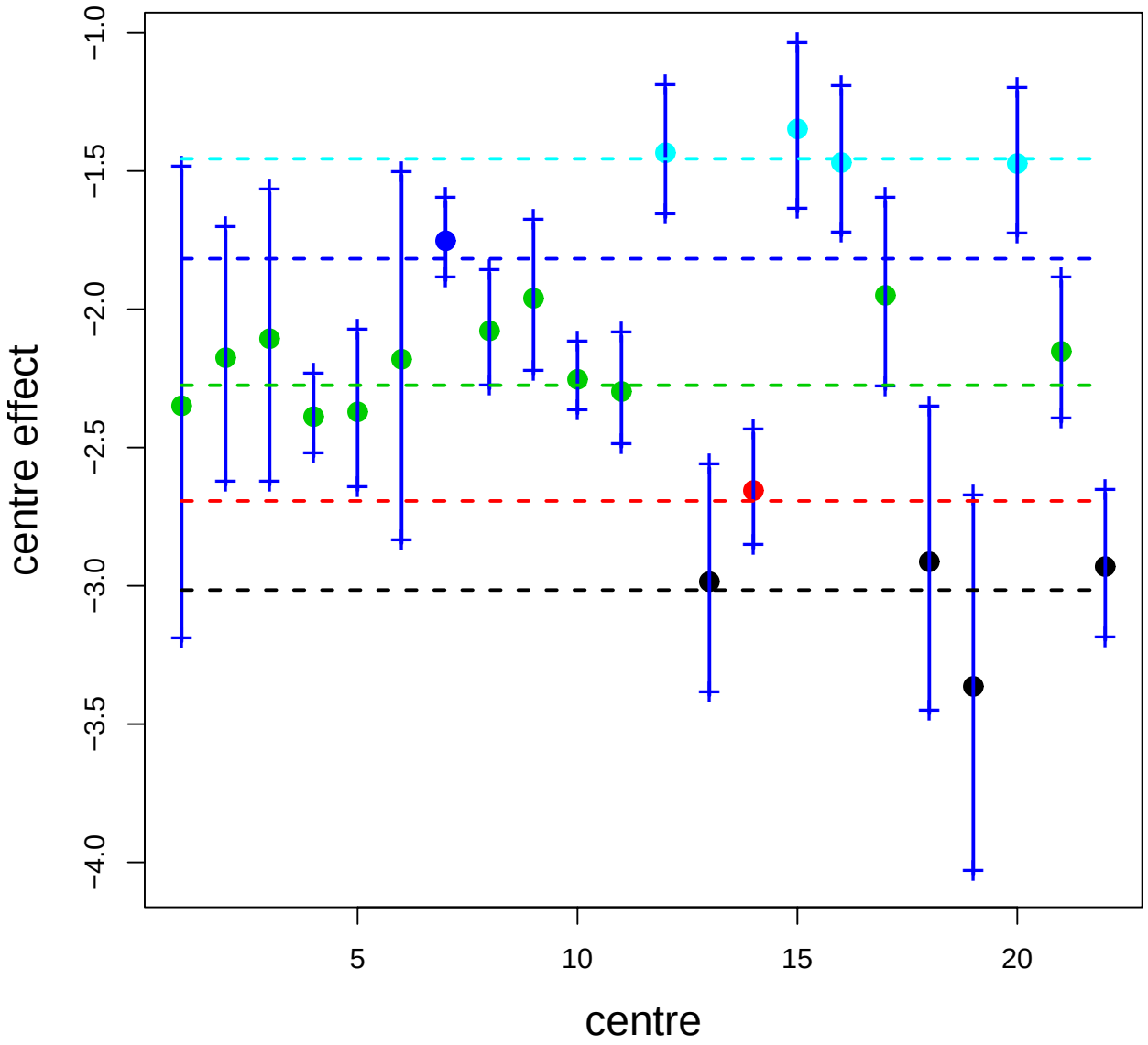
K	$-2\log L$	# (Z, U)-pars
1	523.2	0
3	316.6	6
5	299.0	12
VC-5	310.4	8
Normal	316.7	1

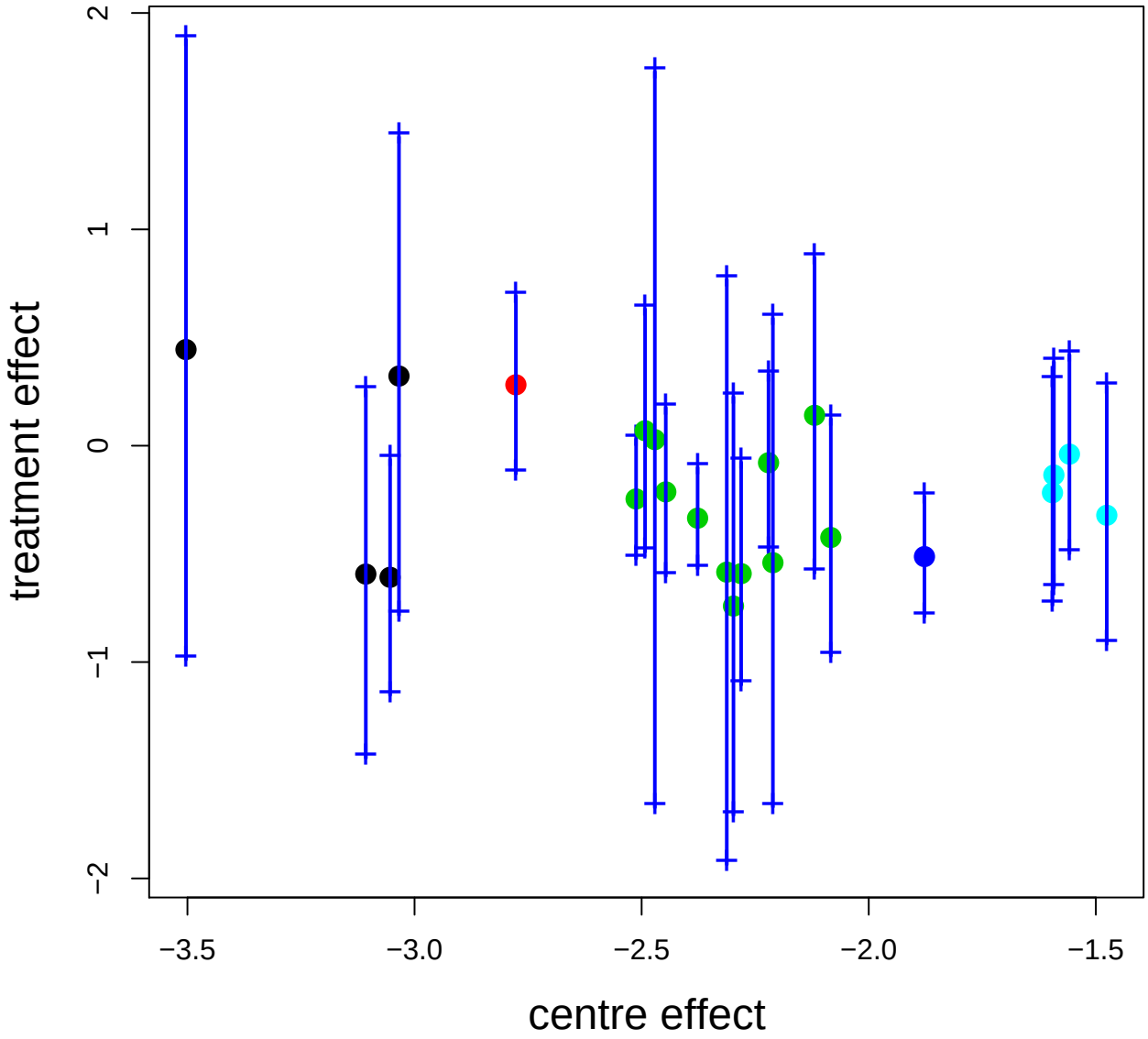
Regressions in each component

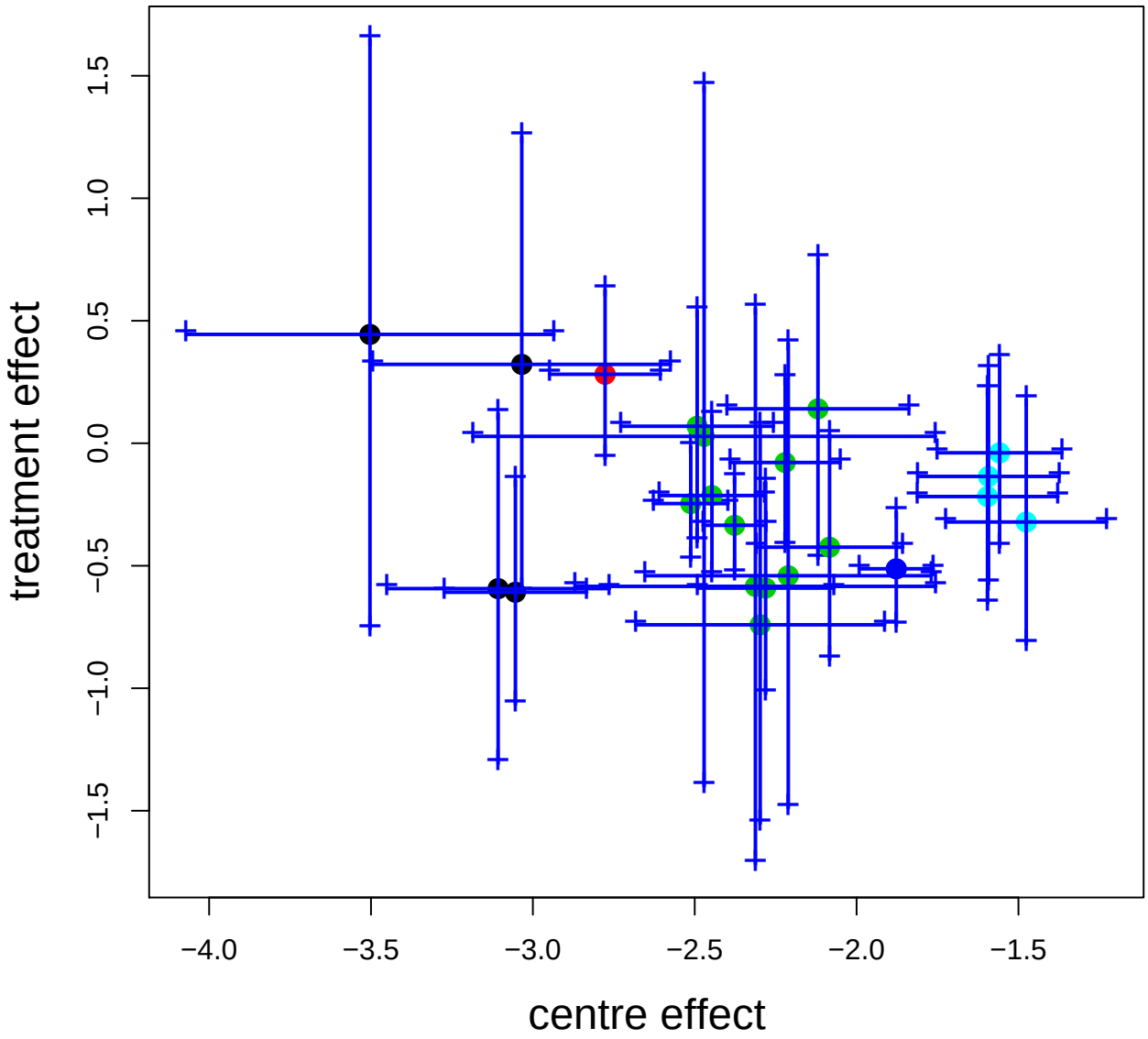
k	β_{0k}	β_{1k}	π_k
1	-1.486	-0.159	0.1816
2	-2.255	-0.260	0.4937
3	-2.895	-0.460	0.1581
4	-1.684	-0.487	0.0869
5	-2.937	$+0.270$	0.0798

- average treatment effect $= -0.250$
- significant treatment variation across centres
- component 5 — single large centre with **increased** death risk under treatment


treatment effect
2
1
0
−1
−2
5
10
15
20
centre


centre effect
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
5
10
15
20
centre

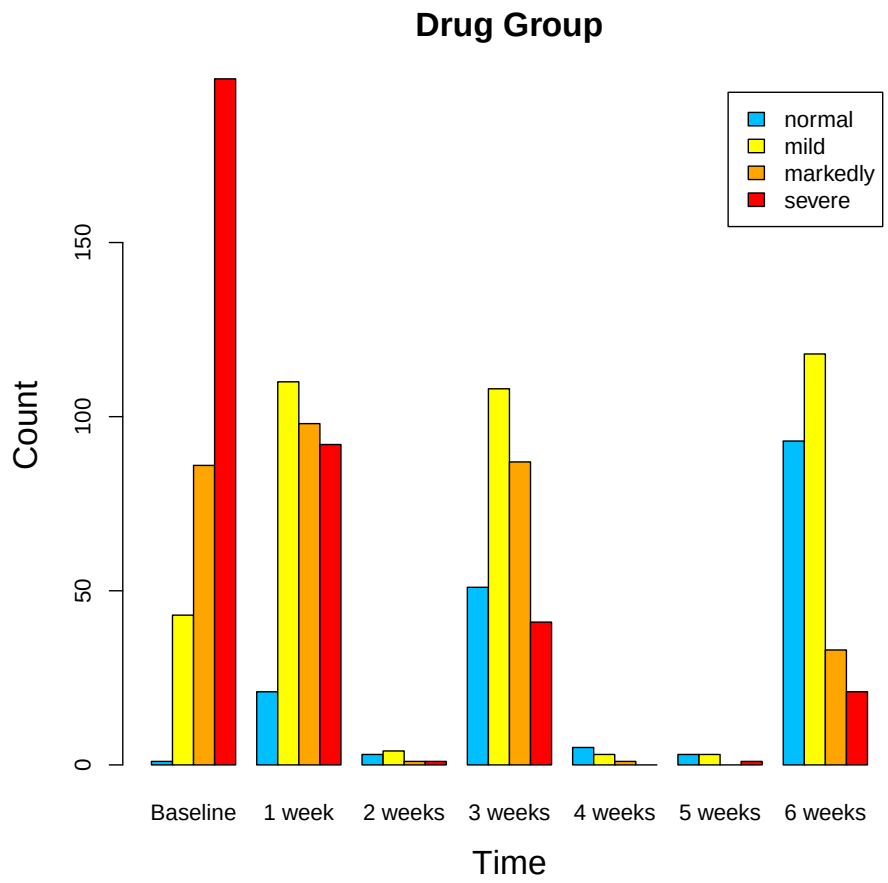

treatment effect
centre effect
2
1
0
−1
−2
−3.5
−3.0
−2.5
−2.0
−1.5

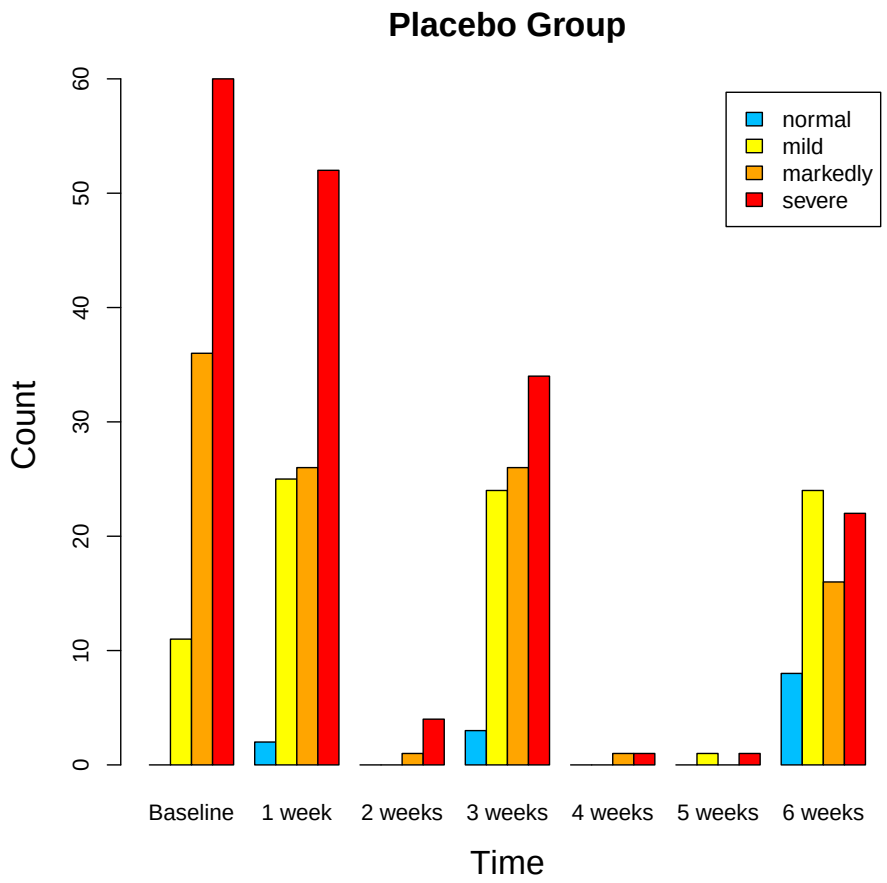

treatment effect
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−4.0
−3.5
−3.0
−2.5
−2.0
−1.5
centre effect

NIMH Schizophrenia Study

- Response variable measures 'Severity of Illness':
  - 1: normal or borderline mentally ill
  - 2: mildly or moderately ill
  - 3: markedly ill
  - 4: severely or among the most extremely ill
- 329 patients assigned to the drug group
- 108 patients assigned to the placebo group
- Responses measured at weeks $j = 0, 1, \ldots, 6$

NIMH Schizophrenia Study: Data





NIMH Schizophrenia Study: Model

Longitudinal cumulative logit model

Y_{ij} the jth R-category response from individual i at occasion j

$\mathbf{x}_{ij}$ vector of observed covariates

$\mathbf{z}_{ij}$ vector of random effect covariates

$$\log\left(\frac{P(Y_{ij} \leq r)}{1 - P(Y_{ij} \leq r)}\right) = \theta_r + \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{u}_i$$

where $r = 1, \ldots, R-1$.

$\mathbf{u}_i$ is an individual (cluster) level random effect

NIMH Schizophrenia Study: Random Effects Models

- **Normal**: u normally distributed
- **SNP** K: semi-nonparametric smooth density (Gallant and Nychka, 1987)
  - truncated series expansion to approximate these densities

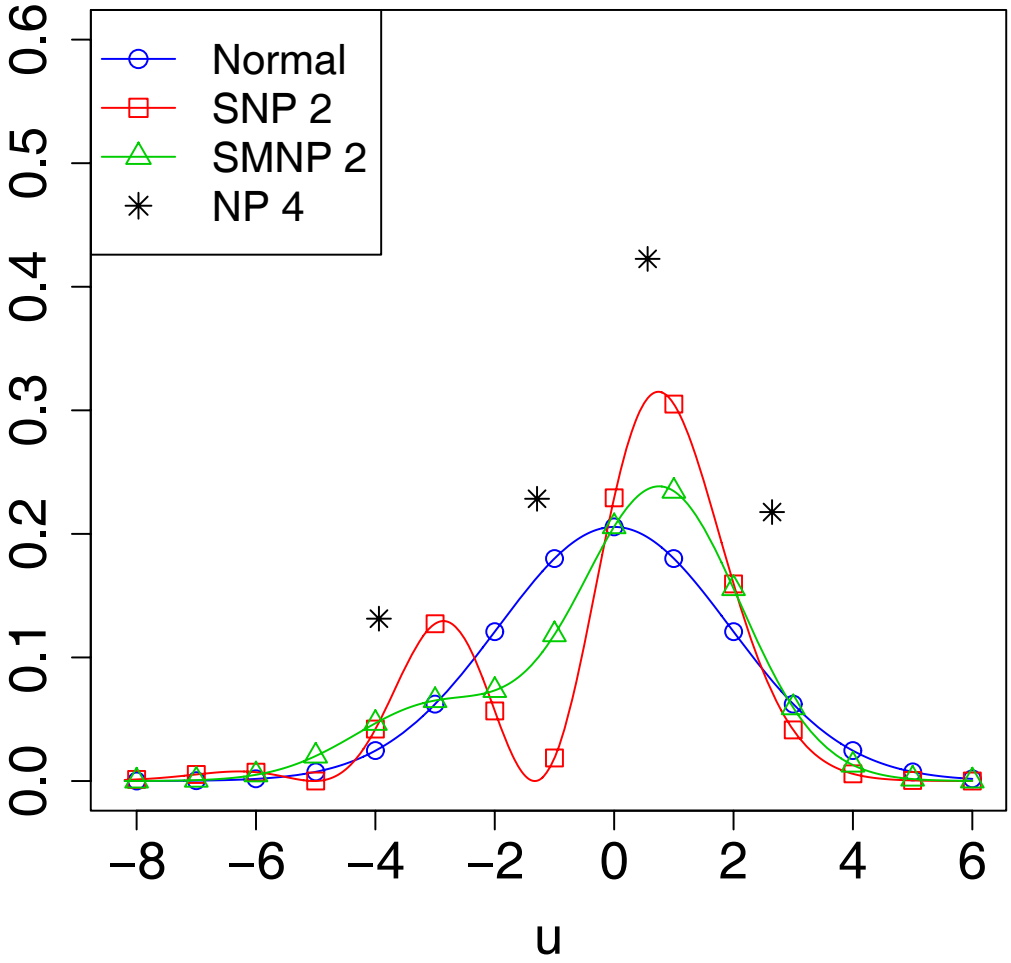
$$g_K(u) = \left\{ \sum_{i=0}^{K} a_i u^i \right\}^2 \phi(u)$$

  - tuning parameter K controls the flexibility:
    $K = 0$, normal density
    $K = 2$, allows up to 3 modes, skewed and thick tailed distributions
- **SMNP** K: smooth non-parametric densities
  mixtures of K normal distributions (Verbeke and Lesaffre, 1996)
- **NP** K: finite mass point distribution

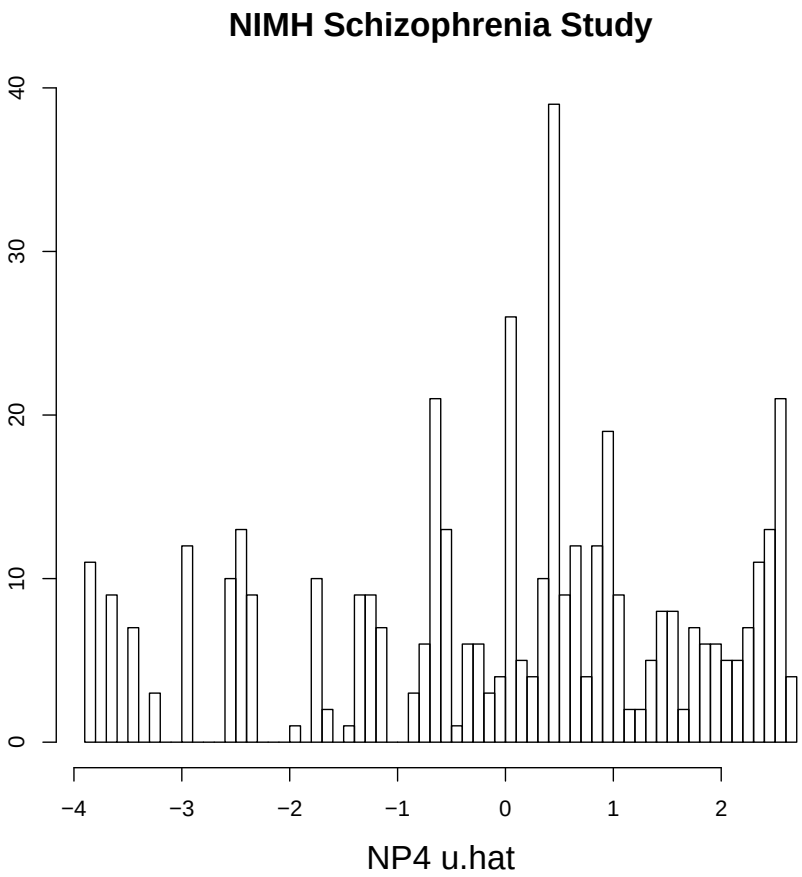
NIMH Schizophrenia Study: Results

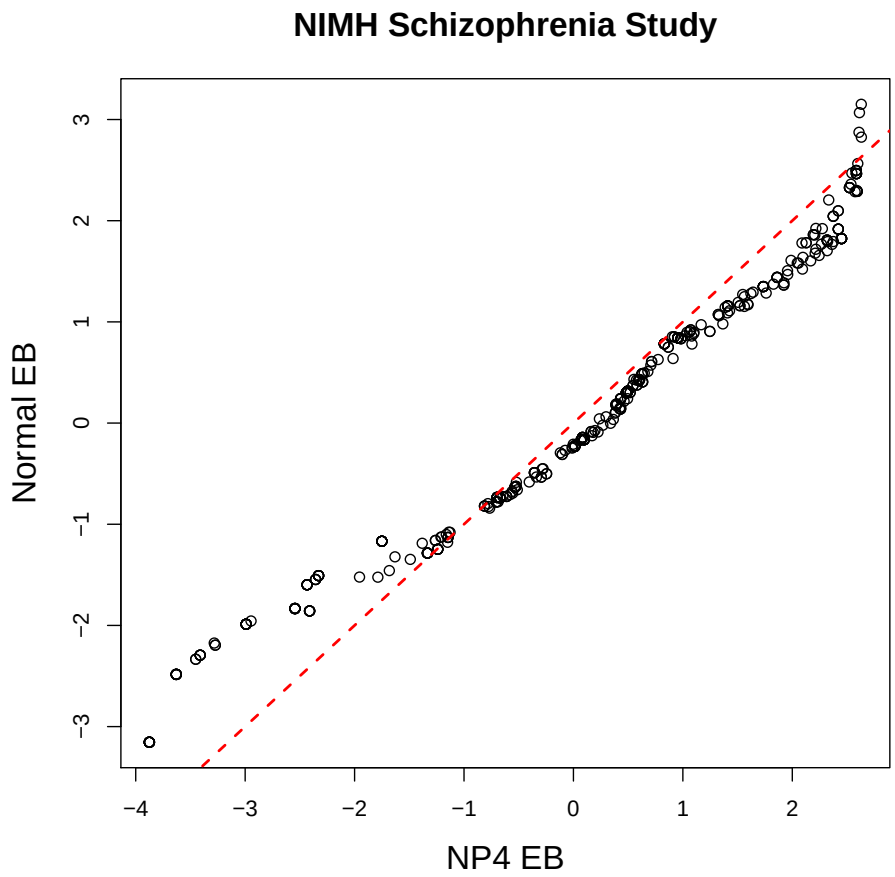
	Normal	SNP 2	SMNP 2	NP 4
θ_1	$-5.17(0.28)$	$-5.63(0.325)$	-5.81(0.339)	$-5.94(0.34)$
θ_2	$-2.45(0.25)$	$-2.73(0.272$	-2.82(0.287)	$-2.94(0.29)$
θ_3	$-0.51(0.25)$	$-0.63(0.246)$	-0.68(0.262)	$-0.76(0.26)$
β_{trt}	$-0.004(0.28)$	$-0.18(0.264)$	-0.07(0.293)	0.08(0.29)
β_{wk}	0.67(0.12)	0.77(0.122)	0.78(0.120)	0.78(0.12)
β_{trt*wk}	**1.07(0.14)**	**1.18(0.13)**	**1.21(0.13)**	**1.21(0.13)**
$Var(u_i)$	2.75(0.26)	4.07(0.570)	4.13(0.572)	4.08(0.50)
AIC	3417	3416	3406	3395

NIMH Study: Estimated Random Effects Densities



NIMH Study: Estimated Random Effects





Drosophila Melongaster Data

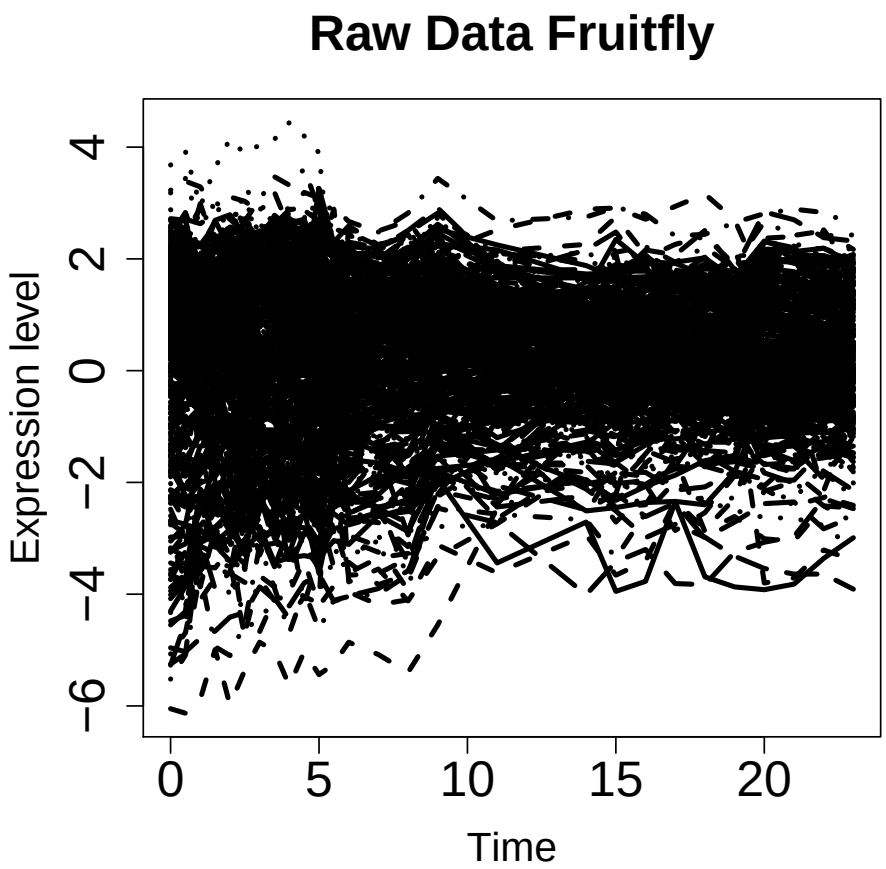


Figure: Gene expression profiles of genes in the embryo phase of *Drosophila Melongaster* (fruitfly) data.

Time-course gene expression data

- Gene expression over time can be thought of as arising from a smooth underlying curve/function.
- Expression values usually measured with error/noise.
- For each gene use the model

$$y_j = g(t_j) + \varepsilon_j, \quad j = 1, \ldots, n \tag{1}$$

  j denotes time t_j, $g(t)$ = smooth expression profile, ε_j is measurement error.
- Need to remove noise and estimate smooth expression profiles $g(t)$.
- Do this using basis function expansions.

P-splines as mixed model

- Can represent P-spline smoothing as a linear mixed effects model.
- Mixed effects model has the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}.$$

- For simplicity assume $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$.
- For smoothing must also assume $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I})$.
- Estimates of $\boldsymbol{\beta}$, σ_ε^2, σ_u^2 and $\mathbf{u}$ determined using (RE)ML and BP.

P-splines as mixed model

- Divide our basis into 2 parts - unpenalised part (fixed effects) and penalised part (random effects).
- Let

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \text{and} \quad \mathbf{u} = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1L} \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} 1 & t_1 & \cdots & t_1^p \\ 1 & t_2 & \cdots & t_2^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & \cdots & t_n^p \end{pmatrix} \qquad \mathbf{Z} = \begin{pmatrix} (t_1 - \kappa_1)_+^p & \cdots & (t_1 - \kappa_L)_+^p \\ (t_2 - \kappa_1)_+^p & \cdots & (t_2 - \kappa_L)_+^p \\ \vdots & \ddots & \vdots \\ (t_n - \kappa_1)_+^p & \cdots & (t_n - \kappa_L)_+^p \end{pmatrix}$$

P-splines as mixed model

- Let $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2_\varepsilon \mathbf{I})$ and $\mathbf{u} \sim N(\mathbf{0}, \sigma^2_u \mathbf{I})$.
- Since $\mathbf{u} \sim N(\mathbf{0}, \sigma^2_u \mathbf{I}) \Rightarrow \lambda^{2p} = \sigma^2_\varepsilon / \sigma^2_u$, (i.e. λ chosen automatically via (RE)ML).
- Full mixed model specification:

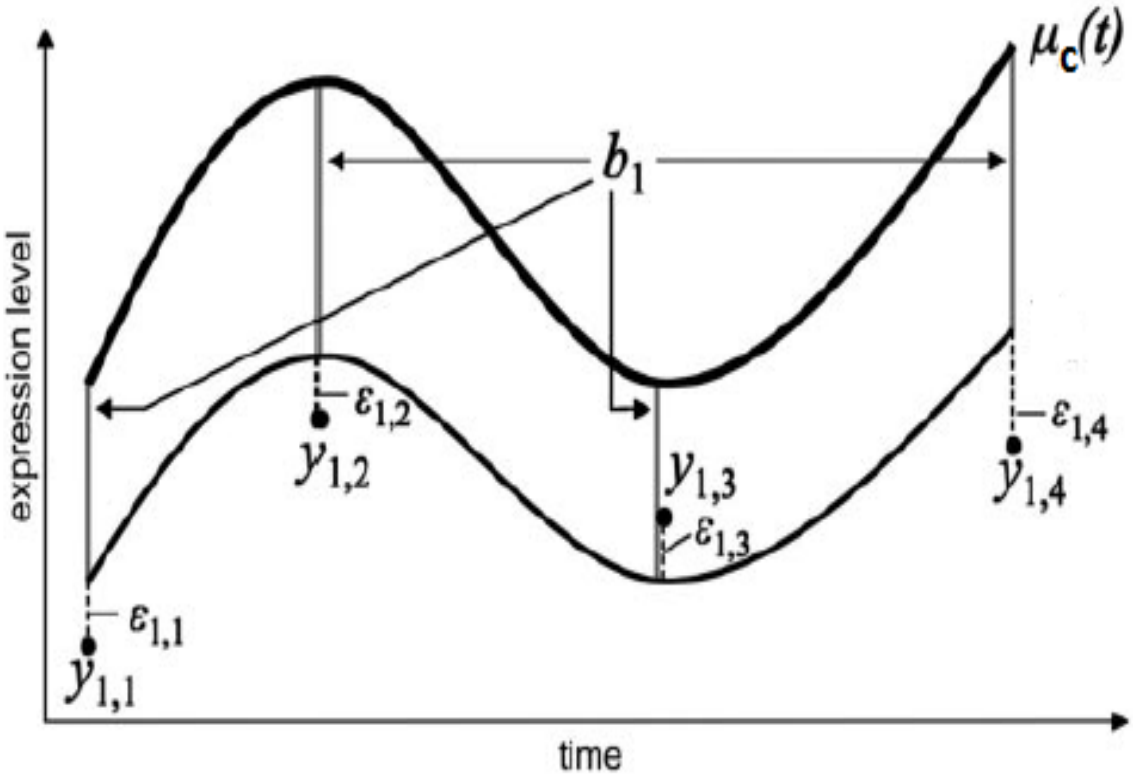
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \text{Var}\begin{pmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \sigma^2_u \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma^2_\varepsilon \mathbf{I} \end{pmatrix}$$

Advantages of mixed model representation

- Have a model for gene expression over time that can be fitted using readily available software, e.g. SAS, R, S-Plus, etc..
- Incorporates smoothing $\Rightarrow$ removes measurement error.
- Accounts for correlation between observations made on the same gene over time.
- Can handle missing values.
- Smoothing parameter λ chosen via (RE)ML.
- When clustering, have a model for the smooth cluster mean but can also incorporate additional random effects (e.g. random intercepts) to estimate individual gene profiles.
- Easy to implement mixtures of mixed effects models.

Modelling gene expression clusters

- Use P-splines to model mean curve $\mu_c(t)$ of cluster c.
- Assumed the expression profile of gene i in cluster c follows the shape of the mean curve of that cluster, but with a gene-specific shift (b_i) from that mean.



Modelling gene expression clusters

- Write the expression level for gene i in cluster c at time j as

$$y_{ij} = \mu_c(t_{ij}) + b_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i,$$

  where $b_i \sim N(0, \sigma^2_{bc})$.
- b_i allows for gene-specific shift from the mean.
- Equivalent to saying the ith gene in cluster c is distributed as

$$\mathbf{y}_i \sim N(\boldsymbol{\mu}_c, \mathbf{V}_c),$$

  where $\mathbf{V}_c = \sigma^2_{bc} E_{n_i \times n_i} + \sigma^2_{\varepsilon c} \mathbf{I}_{n_i \times n_i}$.
- $E_{n_i \times n_i}$ is a matrix where all the entries are 1.

Modelling gene expression clusters

- Stack all of the data from all N_c genes in cluster c.

$$\mathbf{Y}_c = \underbrace{\mathbf{X}_{c,s}\beta_{c,s} + \mathbf{Z}_{c,s}\mathbf{u}_{c,s}}_{\mu_c(t)} + \varepsilon_c$$

$$\mathbf{Y}_c = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{N_c} \end{pmatrix} \quad \mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} \quad \mathbf{X}_{c,s} = \begin{pmatrix} \mathbf{X}_{1,s} \\ \mathbf{X}_{2,s} \\ \vdots \\ \mathbf{X}_{N_c,s} \end{pmatrix} \quad \mathbf{X}_{i,s} = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix}$$

$$\mathbf{Z}_{c,s} = \begin{pmatrix} \mathbf{Z}_{1,s} \\ \mathbf{Z}_{2,s} \\ \vdots \\ \mathbf{Z}_{N_c,s} \end{pmatrix} \quad \mathbf{Z}_{i,s} = \begin{pmatrix} (t_{i1} - \kappa_1)_+ & (t_{i1} - \kappa_2)_+ & \cdots & (t_{i1} - \kappa_L)_+ \\ (t_{i2} - \kappa_1)_+ & (t_{i2} - \kappa_2)_+ & \cdots & (t_{i2} - \kappa_L)_+ \\ \vdots & \vdots & \ddots & \vdots \\ (t_{in_i} - \kappa_1)_+ & (t_{in_i} - \kappa_2)_+ & \cdots & (t_{in_i} - \kappa_L)_+ \end{pmatrix}$$

Modelling gene expression clusters

- Need to incorporate random intercept into model.
- For each gene, define a second design matrix of dimension $n_i \times 1$ for the random intercepts,

$$\mathbf{Z}_{i,b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

- For fitting, stack all $\mathbf{Z}_{i,b}$ matrices.

Modelling gene expression profiles

- Full mixed model for estimating mean curve and individual expression profiles in cluster c

$$\mathbf{Y}_c = \underbrace{\mathbf{X}_{c,s}\boldsymbol{\beta}_{c,s} + \mathbf{Z}_{c,s}\mathbf{u}_{c,s}}_{\mu_c(t)} + \mathbf{Z}_{c,b}\mathbf{b}_c + \boldsymbol{\varepsilon}_c,$$

- $\mathbf{Y}_c$, $\mathbf{X}_{c,s}$, $\mathbf{Z}_{c,s}$ as defined previously.

$$\mathbf{Z}_{c,b} = \begin{pmatrix} \mathbf{Z}_{1,b} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{2,b} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_{N_c,b} \end{pmatrix}$$

$$\mathbf{u}_{c,s} \sim N(\mathbf{0}, \sigma^2_{uc}\mathbf{I}), \quad \mathbf{b}_c \sim N(\mathbf{0}, \sigma^2_{bc}\mathbf{I}) \quad \boldsymbol{\varepsilon}_c \sim N(\mathbf{0}, \sigma^2_{\varepsilon c}\mathbf{I})$$

Mixtures of mixed effects models

- In practice do not know cluster membership.
- Assume $\mathbf{y}_i$ comes from a mixture of C clusters/groups

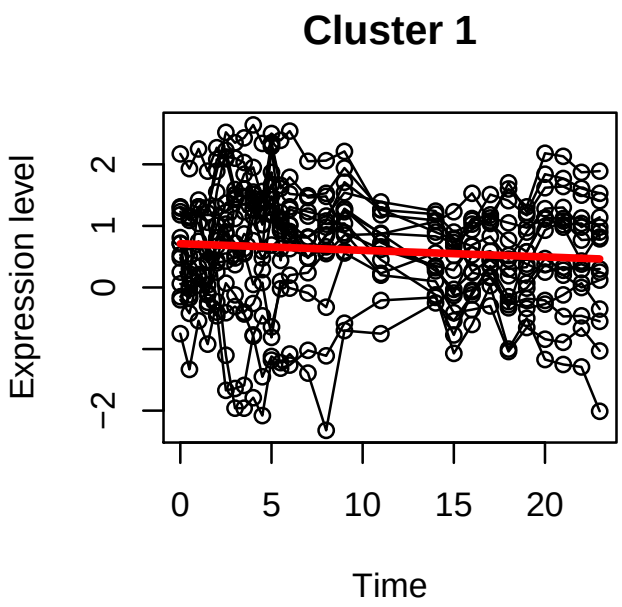
$$\mathbf{y}_i \sim \pi_1 N(\boldsymbol{\mu}_1, \mathbf{V}_1) + \pi_2 N(\boldsymbol{\mu}_2, \mathbf{V}_2) + \ldots + \pi_C N(\boldsymbol{\mu}_C, \mathbf{V}_C).$$

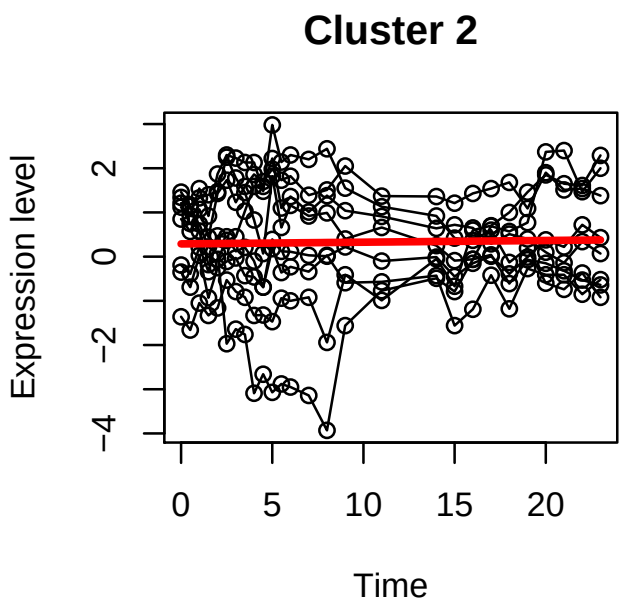
- $\pi_1, \pi_2, \ldots, \pi_C$ are mixing proportions such that $\sum_{c=1}^{C} \pi_c = 1$.
- Give the (posterior) probability that gene i comes from cluster c.
- Need to be estimated for each gene.
- Also need to estimate $(\boldsymbol{\mu}_1, \mathbf{V}_1), \ldots, (\boldsymbol{\mu}_C, \mathbf{V}_C)$.
- Use the EM algorithm.

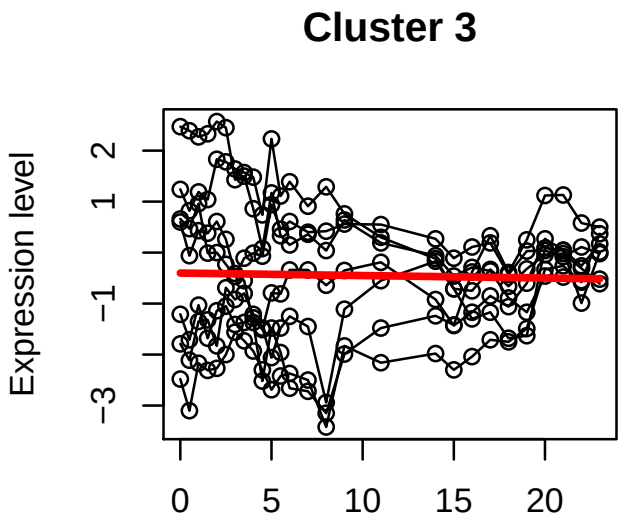
EM algorithm

- **Step 1:** Randomly assign genes to C clusters. Fit a mixed effects model to each cluster to get initial estimates of $(\boldsymbol{\mu}_1, \mathbf{V}_1), \ldots, (\boldsymbol{\mu}_C, \mathbf{V}_C)$.
- **E-step:** Based on current estimates, calculate posterior probability that gene i belongs to cluster c.
- **M-step:** Re-fit mixed effects model in each cluster using posterior probabilities as weights to update estimates of $(\boldsymbol{\mu}_1, \mathbf{V}_1), \ldots, (\boldsymbol{\mu}_C, \mathbf{V}_C)$.
- Iterate between E-step and M-step until convergence.
- Never know true number of clusters.
- Repeat process using 2, 3, 4,... clusters and compare with BIC.

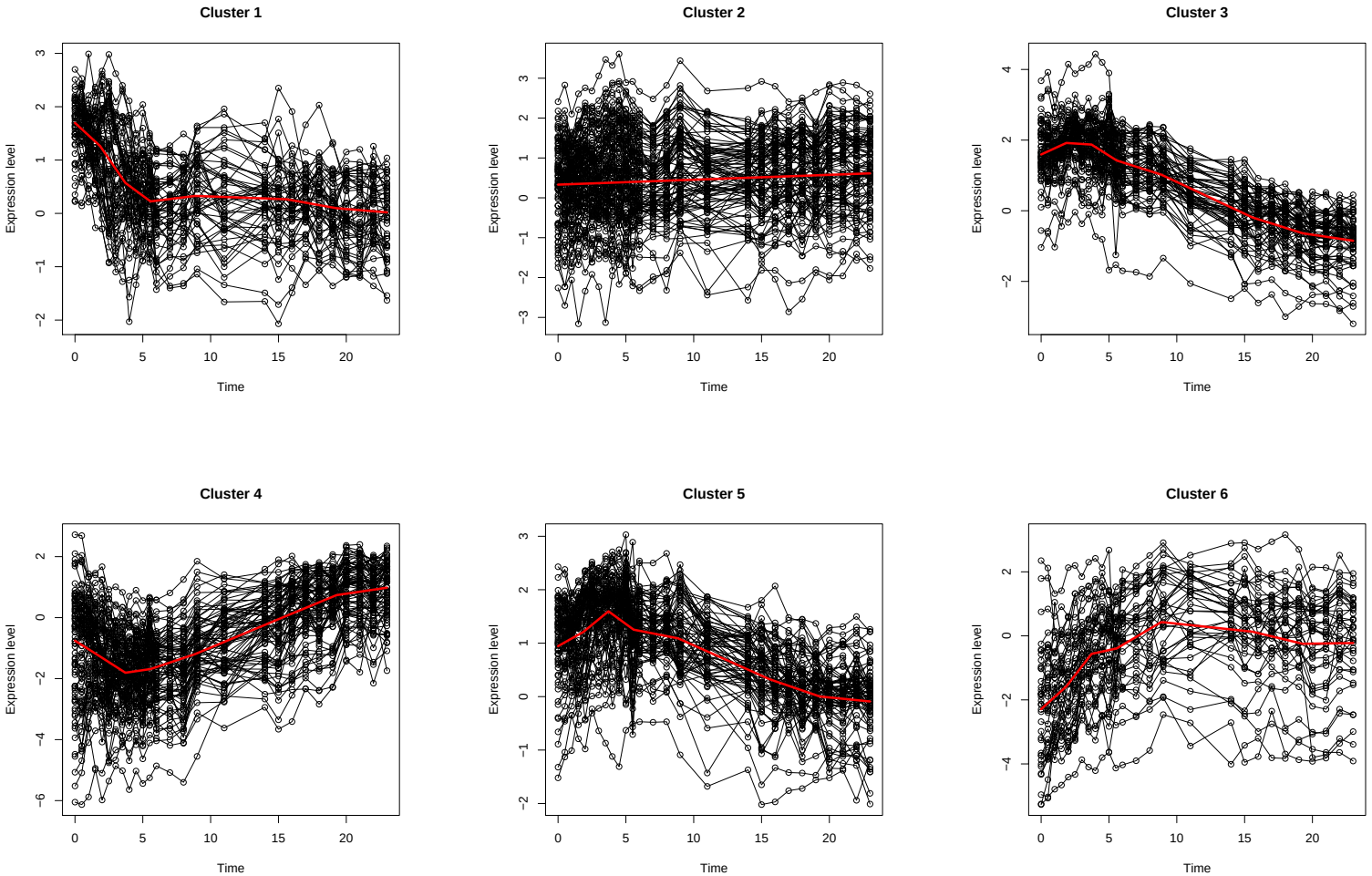
Example


Cluster 1
Expression level
−2
0
1
2
0
5
10
15
20
Time


Cluster 2
Expression level
−4
−2
0
2
0
5
10
15
20
Time


Cluster 3
Expression level
−3
−1
1
2
0
5
10
15
20

Results *Drosophila Melongaster* Data



References

Aitkin, M, Francis, B, Hinde, J, Darnell, R (2009)
Statistical Modelling in R, Oxford.

Also ...

Aitkin, M, Francis, B, Hinde, J (2005)
Statistical Modelling in Glim4, 2nd Edition, Oxford.

Forthcoming ...

Hinde, J, Demétrio, C (2012?)
Overdispersion: Models and Estimation, Taylor & Francis.

Software

R-code and examples available from:

NUI, Galway

`http://www.nuigalway.ie/maths/jh/npml.html`

CRAN

glms: `npmlreg`

multicategory data: `mixcat`