## Bayesian models for mixtures (Part II)

- Sensitivity
- Simulation study

  statistical performance

  inference on $k$
- Semi-parametric framework

  measurement error models

  performance with mixture priors
- Discussion



## Sensitivity to prior specification

Focus on sensitivity of posterior distribution of $k$ and parameters to

- prior information on the variances
- prior model for the means



### Prior models for the variances

| Fixed $\beta$ | | Random $\beta$ |
|---|---|---|
| $\sigma_j^{-2} \sim \Gamma(\alpha, \beta)$ | versus | $\sigma_j^{-2} \sim \Gamma(\alpha, \beta)$ |
| | | with $\beta \sim \Gamma(g, h)$ |

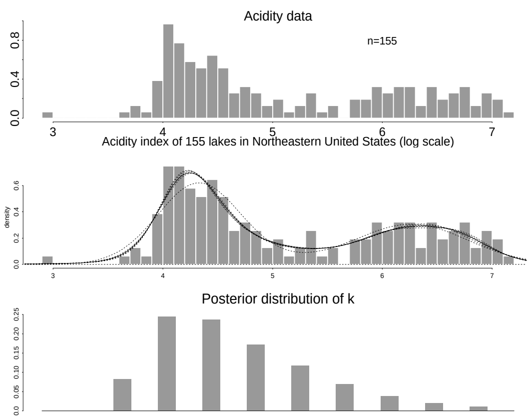
- Difficult to be weakly informative with fixed $\beta$:

  increasing number of components as $\beta/\alpha$ (typical value for $\sigma_j^2$) is reduced
- Hierarchical model with random $\beta$ successfully expresses weak prior information:

  posterior of $k$ insensitive to ratio $g/h$

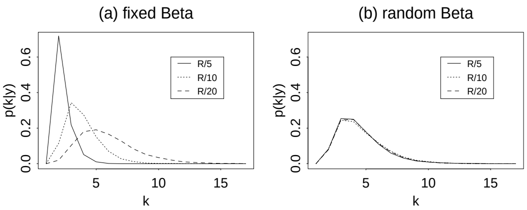
Similar sensitivity pattern between fixed and random $\beta$ cases for the posterior densities of parameters





## Comparison between fixed and random $\beta$ model

Posterior distribution of $k$:




(a) $\alpha = 2$ and $\sqrt(\beta/\alpha)$ varying between $R/5$, $R/10$ and $R/20$.
(b) $\alpha = 2, g = 0.2$ and $\sqrt(E(\beta)/\alpha) = \sqrt(g/h\alpha)$, varying between $R/5$, $R/10$ and $R/20$.



## Prior distribution of the means

$$\mu_j \sim N(\xi, \kappa^{-1}) \quad \text{drawn independently}$$

with $\quad \xi =$ midpoint

**Sensitivity to value of $\kappa$:**

- $\kappa^{-1/2} = R$ (default) does not encourage the fitting of mixtures with very close $\mu_j$
- *at first*, reducing $\kappa^{-1/2}$ is more *permissive*

  $\longrightarrow$ increase in the number of components ($\kappa^{-1/2}$ between $R$ and $R/5$)
- further reduction of $\kappa^{-1/2}$ actively *prohibits* components with extreme means

  $\longrightarrow$ *shrinkage effect* and decrease of $k$ ($\kappa^{-1/2}$ smaller than $R/5$)

Separate means explicitly by using *dependent* priors for the means?



## Joint prior model for the means

For $s \geq 1$, define

$\{\mu_j; j = 1, 2, \ldots, k\}$ to be distributed as the $\{s, 2s; \ldots, ks\}$ order statistics from a sample of size $ks + s - 1$ of $N(\xi, \kappa^{-1})$

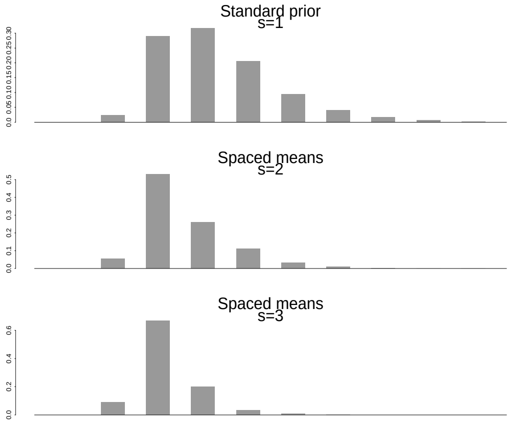
$\longrightarrow$ $s > 1$ encourages separation of the means
$\longrightarrow$ tighter posterior distribution for $k$



## Influence of spacing on $p(k|y)$

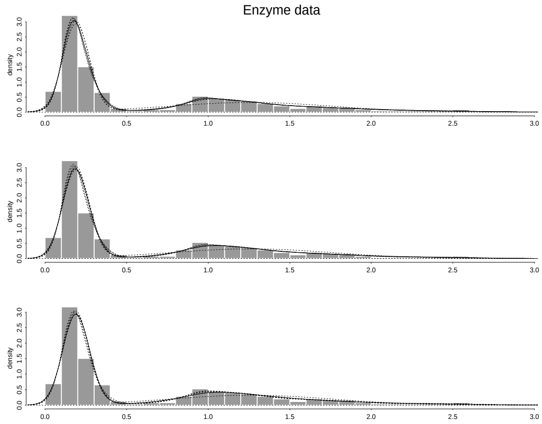
Enzyme data

| | $s=1$ | $s=2$ | $s=3$ |
|---|---|---|---|
| $p(2\|y)$ | 0.02 | 0.05 | 0.09 |
| $p(3\|y)$ | 0.29 | 0.53 | 0.67 |
| $p(4\|y)$ | 0.32 | 0.26 | 0.20 |
| $p(5\|y)$ | 0.21 | 0.11 | 0.03 |
| $p(6\|y)$ | 0.09 | 0.03 | 0.01 |
| $p(7\|y)$ | 0.04 | 0.01 | - |






**Enzyme data.**
**Predictive densities corresponding to increased spacing of the means**






**Simulation study**

**4 types of mixture were simulated**

- separated bimodal $(k = 2)$
- skewed bimodal $(k = 2)$
- kurtotic unimodal $(k = 2)$
- trimodal (not well separated) $(k = 4)$

In each case,
data set $n = 200$ and $50$ replications.

**Focus on**

- statistical performance of methodology
- inference on $k$



**Statistical performance**

- posterior distributions of $k$
- predictive densities

**Inference on $k$**

- mode of $p(k|y)$
- mean square differences (evaluated on a fixed grid) between conditional predictive densities $E(f(\cdot|k, w, \theta)|k, y)$ for consecutive $k's$
- overlap of distributions of posterior deviances $-2 \log p(y|k, w, \theta)$ for consecutive $k's$
  $\Rightarrow$ quantified this roughly by evaluating :
  area of deviance distribution $|k$ above
  5% percentile of deviance distribution $|k - 1$.

  area of deviance distribution $|k - 1$ below
  95% percentile of deviance distribution $|k$.

**Percentage of replications with $k$ such that $p(k|y)$ is highest**

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| separated bimodal (true $k=2$) | 0 | 98 | 2 | 0 | 0 | 0 |
| skewed bimodal (true $k=2$) | 0 | 90 | 10 | 0 | 0 | 0 |
| kurtotic bimodal (true $k=2$) | 0 | 88 | 12 | 0 | 0 | 0 |
| trimodal (true $k=4$) | 0 | 0 | 38 | 38 | 18 | 6 |



**Overlap of distributions of posterior deviances**

| % overlap | $1:2$ | $2:3$ | $3:4$ | $4:5$ | $5:6$ |
|---|---|---|---|---|---|
| separated bimodal | - | 93 | 95 | 95 | 95 |
| (true $k=2$) | - | 95 | 95 | 95 | 96 |
| skewed bimodal | 48 | 95 | 95 | 95 | 95 |
| (true $k=2$) | 65 | 95 | 95 | 95 | 95 |
| kurtotic bimodal | - | 97 | 96 | 95 | 96 |
| (true $k=2$) | - | 94 | 95 | 95 | 94 |
| trimodal | - | 11 | 92 | 95 | 96 |
| (true $k=4$) | - | 22 | 95 | 96 | 96 |



**Semi-parametric framework**

Bayesian modelling of complex data involves often the introduction of latent structures

It is of interest to have flexible models for the distribution of latent structures.

**Measurement error problems involve**

- $Y$ known outcome
- $X$ true (latent) covariate
- $U$ observed surrogate for $X$

**Define local submodels between components**

- $[Y|X, \beta]$ regression model
- $[U|X, \lambda]$ measurement model

Distribution for $X$ ?



**Typically in epidemiological studies**

- $Y$ disease status
- $X$ true risk factor (exposure)
- $U$ surrogate measure of risk factors
- $C$ known covariates

**Global structure**

3 submodels are linked with conditional independence assumptions ensuring non differentiable measurement error.

**Where does quantitative information on measurement model come from ?**

- a priori external information
- study designs

$\longrightarrow$ repeated measures for unbiased measuring instrument

$\longrightarrow$ reference method which can be used to apprehend $X$ on a subgroup where both $X$ and $U$ are recorded :

designs with validation group



**Designs with a validation group**

$\longrightarrow$ Information on $\lambda$ will be transferred from the validation group to the main study.

$\longrightarrow$ Provide information on relations between $X$ and $U$ in main study.

$\longrightarrow$ Strengthen inference about regression parameters $\beta$.



- in many cases it is difficult to specify a parametric model for $[X]$

  'shapeless' distribution of $U$

  large measurement errors

  small or non existent validation group

- alternatives

  pseudo-likelihood *(Carroll, 1993)* based on plugging an empirical estimate of $[X]$ based on the validation subgroup

  non parametric modelling of $[X]$ *(Roeder, Carroll, Lindsay, JASA 1996)*

  semi-parametric model for $[X]$ via a mixture of gaussian distributions with an unknown number of components



**Combining measurement error and mixtures**
**Joint work with L. Leblond**

*Mixture model for prior distribution of $X$*

$$X \sim \sum_{j=1}^{k} w_j f(\cdot|\theta_j),\ k \text{ unknown, } f(\cdot|\theta_j) \sim N(\mu_j, \sigma_j^2)$$

*Measurement error model : normal or lognormal*
$U_i \sim N(\alpha_0+\alpha_1 X_i, \gamma^{-1})$or$U_i \sim logN(\alpha_0+\alpha_1 X_i, \gamma^{-1})$

*Regression model for binary outcome (disease status): logistic model*

$$Y \sim B(p) \quad \text{with } logit\ p = \beta^T(X, C)$$

**Performance of mixture priors in measurement error models**

Influence of misspecification evaluated by comparing analyses using mixture priors for $X$ versus analyses using a gaussian prior for $X$.

*Simulation set up* 50 replications
main study $n = 500$

validation subgroup $n = 50$

prior of $X$ : $0.5N(-2, 1) + 0.5N(2, (0.75)^{-1})$

measurement model : $U \sim N(X, \theta^{-1})$, $\theta = 0.06$

logistic disease model : $logitP(D|X) = \beta_0 + \beta_1 X$, $\beta_0 = -0.8, \beta_1 = 0.9$

| Analysis | mixture prior | gaussian prior |
|---|---|---|
| $\theta$ | 0.65 (0.10) | 0.65 (0.11) |
| $\beta_0$ | - 0.76 (0.18) | -0.79 (0.21) |
| $\beta_1$ | 0.93 (0.12) | 1.07 (0.18) |
| MSE | 0.02 | 0.04 |



**Performance comparison between mixture and non-parametric approaches**

*Simulation set-up as in Roeder et al (1996)*

main study : 90 cases 90 controls

validation subgroup: 30 cases 30 controls

prior distribution of $X$ : $logN(-0.58, 1.16)$

measurement model : $logU \sim N(logX, \theta^{-1})$

logistic disease model with $\beta_0 = -3.09, \beta_1 = 0.5$

| Analysis | mixture | P.L. | NP |
|---|---|---|---|
| For true value $\theta = 4$ | | | |
| $\overline{\theta}$ | 3.57 (0.64) | - | - |
| $\overline{\beta_1}$ | 0.53 (0.14) | 0.47 | 0.47 |
| MSE (50 replications) | 0.021 | 0.034 | 0.027 |
| For true value $\theta = 1.77$ | | | |
| $\overline{\theta}$ | 1.64 (0.27) | - | - |
| $\overline{\beta_1}$ | 0.56 (0.16) | 0.49 | 0.49 |
| MSE (50 replications) | 0.028 | 0.067 | 0.048 |



**Discussion**

**Hierarchical model for mixtures**

- interaction between model set-up and number of components

  **–** weakly informative priors based on observed or expected range
  **–** motivated by analysis of heterogeneity

- hierarchical model for variances reduces sensitivity

- influence of prior structures for the means

  from independent to dependent priors

- good performance on simulated data

- sample based approach provide multiple criteria which can be used for inference on $k$



**Extensions**

**Other distributions**

- discrete distributions
- 2 parameter density suitably reparametrised
- multivariate normals

**Other mixture models**

- modelling the weights or allocations (genetic, spatial applications...)
- thorough statistical comparison with mixture models based on a Dirichlet process prior.

**Robust prior modelling in Bayesian analyses**

use of mixtures with $k$ unknown as a prior model for random effects, frailty models...