

Finite Mixture Models

Partha Deb

Hunter College and the Graduate Center, CUNY
NBER

July 2008

Introduction

- The finite mixture model provides a natural representation of heterogeneity in a finite number of latent classes
- It concerns modeling a statistical distribution by a mixture (or weighted sum) of other distributions

Introduction

- The finite mixture model provides a natural representation of heterogeneity in a finite number of latent classes
- It concerns modeling a statistical distribution by a mixture (or weighted sum) of other distributions
- Finite mixture models are also known as
 - latent class models
 - unsupervised learning models
- Finite mixture models are closely related to
 - intrinsic classification models
 - clustering
 - numerical taxonomy

- Heterogeneity of effects for different “classes” of observations
 - wine from different vineyards
 - healthy and sick individuals
 - normal and complicated pregnancies

Introduction

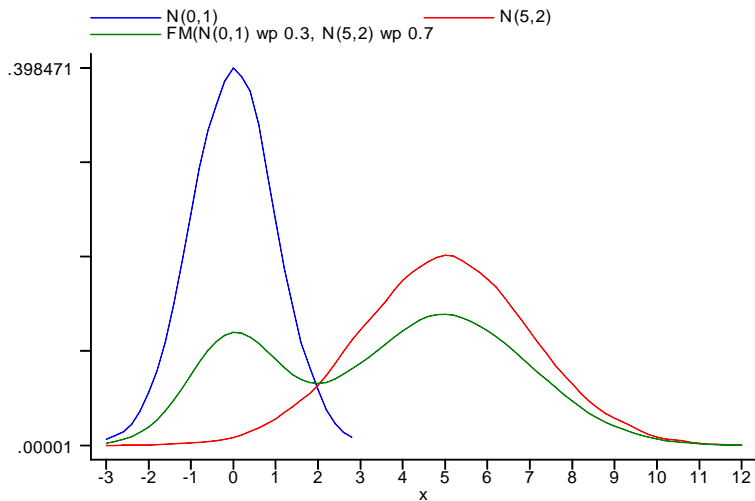
Canonical Example

- Estimating parameters of the distribution of lengths of halibut
- It is known that female halibut is longer, on average, than male fish and that the distribution of lengths is normal
- Gender cannot be determined at measurement
- Then distribution is a 2-component finite mixture of normals

- Estimating parameters of the distribution of lengths of halibut
- It is known that female halibut is longer, on average, than male fish and that the distribution of lengths is normal
- Gender cannot be determined at measurement
- Then distribution is a 2-component finite mixture of normals
- A finite mixture model allows one to estimate:
 - mean lengths of male and female halibut
 - mixing probability

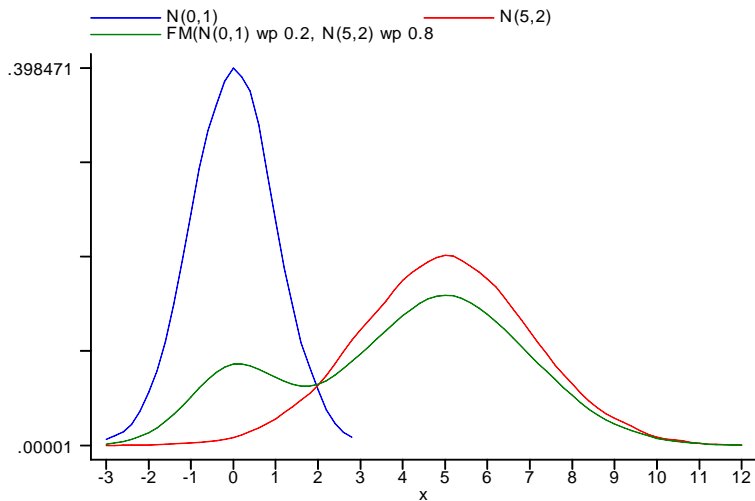
Introduction

A graphical view



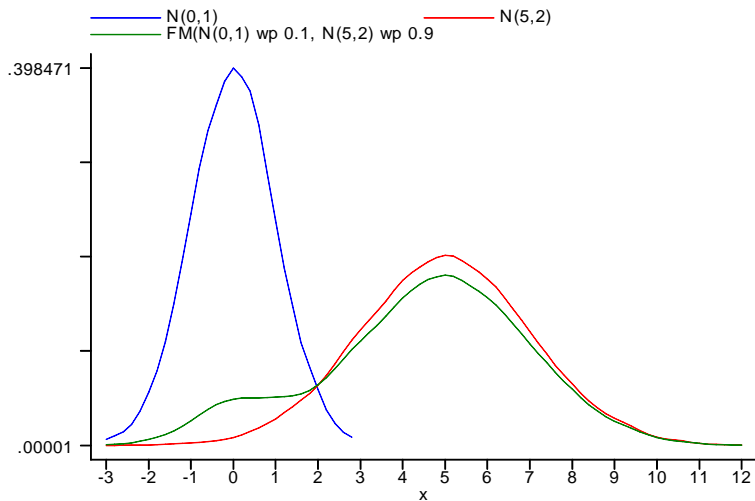
Introduction

A graphical view



Introduction

A graphical view



Introduction

Examples

- Characteristics of wine by cultivar
- Infant Birthweight - two types of pregnancies “normal” and “complicated”
- Medical Services - two types of consumers “healthy” and “sick”
- Public goods experiments - selfish, reciprocal, and altruist
- Stock Returns in “typical” and “crisis” times
- Using somatic cell counts to classify records from healthy or infected goats
- Models of internet traffic

Introduction

More generally from a statistical perspective

- FMM is a semiparametric / nonparametric estimator of the density (Lindsay)
- Experience suggests that usually only few latent classes are needed to approximate density well (Heckman)
- In practice FMM are flexible extensions to basic parametric models
 - can generate skewed distributions from symmetric components
 - can generate leptokurtic distributions from mesokurtic ones

- Introduction
- Model
 - Formulation
 - Estimation
 - Popular densities
 - Properties
- Examples
 - Color of wine
 - Birthweight and prenatal care
 - Medical care

- The density function for a C -component finite mixture is

$$f(y|\mathbf{x}; \theta_1, \theta_2, \dots, \theta_C; \pi_1, \pi_2, \dots, \pi_C) = \sum_{j=1}^C \pi_j f_j(y|\mathbf{x}; \theta_j)$$

where $0 < \pi_j < 1$, and $\sum_{j=1}^C \pi_j = 1$

- The density function for a C -component finite mixture is

$$f(y|\mathbf{x}; \theta_1, \theta_2, \dots, \theta_C; \pi_1, \pi_2, \dots, \pi_C) = \sum_{j=1}^C \pi_j f_j(y|\mathbf{x}; \theta_j)$$

where $0 < \pi_j < 1$, and $\sum_{j=1}^C \pi_j = 1$

- More generally

$$f(y|\mathbf{x}; \mathbf{z}; \theta_1, \theta_2, \dots, \theta_C; \pi_1, \pi_2, \dots, \pi_C) = \sum_{j=1}^C \pi_j(\mathbf{z}) f_j(y|\mathbf{x}; \theta_j)$$

- Maximum likelihood

$$\max_{\pi, \theta} \ln L = \sum_{i=1}^N \left(\log \left(\sum_{j=1}^C \pi_j f_j(y | \theta_j) \right) \right)$$

- Maximum likelihood

$$\max_{\pi, \theta} \ln L = \sum_{i=1}^N \left(\log \left(\sum_{j=1}^C \pi_j f_j(y | \theta_j) \right) \right)$$

- Trick to ensure $0 < \pi_j < 1$, and $\sum_{j=1}^C \pi_j = 1$

$$\pi_j = \frac{\exp(\gamma_j)}{\exp(\gamma_1) + \exp(\gamma_2) + \dots + \exp(\gamma_{C-1}) + 1}$$

- Maximum likelihood

$$\max_{\pi, \theta} \ln L = \sum_{i=1}^N \left(\log \left(\sum_{j=1}^C \pi_j f_j(y | \theta_j) \right) \right)$$

- Trick to ensure $0 < \pi_j < 1$, and $\sum_{j=1}^C \pi_j = 1$

$$\pi_j = \frac{\exp(\gamma_j)}{\exp(\gamma_1) + \exp(\gamma_2) + \dots + \exp(\gamma_{C-1}) + 1}$$

- EM
- Bayesian MCMC

Model

Popular mixture component densities

- Normal (Gaussian)
- Poisson
- Gamma
- Negative Binomial
- Student-t
- Weibull

Model

Some basic properties

- Conditional mean:

$$E(y_i | \mathbf{x}_i) = \sum_{j=1}^C \pi_j \lambda_j \text{ where } \lambda_j = E_j(y_i | \mathbf{x}_i)$$

Model

Some basic properties

- Conditional mean:

$$E(y_i | \mathbf{x}_i) = \sum_{j=1}^C \pi_j \lambda_j \text{ where } \lambda_j = E_j(y_i | \mathbf{x}_i)$$

- Marginal effects:

$$\frac{\partial E_j(y_i | \mathbf{x}_i)}{\partial \mathbf{x}_i} = \frac{\partial \lambda_j}{\partial \mathbf{x}_i} \longrightarrow \text{within component}$$

$$\frac{\partial E(y_i | \mathbf{x}_i)}{\partial \mathbf{x}_i} = \sum_{j=1}^C \pi_j \frac{\partial \lambda_j}{\partial \mathbf{x}_i} \longrightarrow \text{overall}$$

Model

Some basic properties

- Prior probability that observation y_i belongs to component c :

$$\Pr[y_i \in \text{population } c | \mathbf{x}_i, \boldsymbol{\theta}] = \pi_c$$

$$c = 1, 2, \dots, C$$

Model

Some basic properties

- Prior probability that observation y_i belongs to component c :

$$\Pr[y_i \in \text{population } c | \mathbf{x}_i, \boldsymbol{\theta}] = \pi_c$$

$$c = 1, 2, \dots, C$$

- Posterior probability that observation y_i belongs to component c :

$$\Pr[y_i \in \text{population } c | \mathbf{x}_i, y_i, \boldsymbol{\theta}] = \frac{\pi_c f_c(y_i | \mathbf{x}_i, \boldsymbol{\theta}_c)}{\sum_{j=1}^C \pi_j f_j(y_i | \mathbf{x}_i, \boldsymbol{\theta}_j)}$$

$$c = 1, 2, \dots, C$$

Model

Estimation challenges

- The number of components has to be specified - we usually have little theoretical guidance
- Even if prior theory suggests a particular number of components we may not be able to reliably distinguish between some of the components
- In some cases additional components may simply reflect the presence of outliers in the data
- Likelihood function may have multiple local maxima

- Parameterize $\gamma_j = Z\alpha_j$ in

$$\pi_j = \frac{\exp(\gamma_j)}{\exp(\gamma_1) + \exp(\gamma_2) + \dots + \exp(\gamma_{C-1}) + 1}$$

- Parameterizing mixing probabilities
 - may lead to finite sample identification issues
 - may lead to computational difficulties

Model

Selecting number of components

- Estimate models with 2 and then more components
- At each step calculate

$$AIC = -2 \log(L) + 2K$$

$$BIC = -2 \log(L) + K \log(N)$$

- Pick the model with the smallest AIC , BIC

- Stata package `fmm`

```
fmm depvar [indepvars] [if] [in] [weight],  
components(#) mixtureof(density)
```

- where density is one of

```
gamma  
negbin1  
negbin2  
normal  
poisson  
studentt
```

- `predict` and `mfx` give predictions and marginal effects of means, component means, prior and posterior probabilities

Example 1

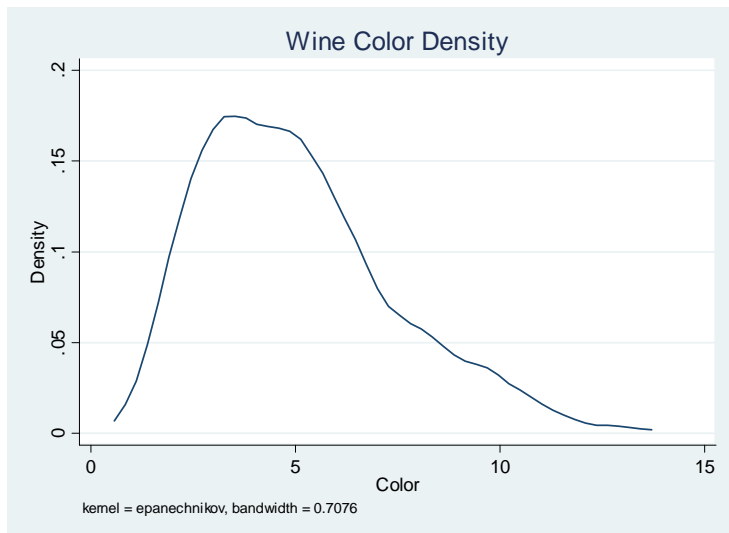
Color of Wine

Results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars

Cultivar	Freq.	% of total	Color intensity (mean)
1	59	33.15	5.528
2	71	39.89	3.086
3	48	26.97	7.396
Total	178	100	5.058

Example 1

Color of Wine



Example 1

Color of Wine

- Finite mixture of Normals with 3 components

$$f(y_i | \theta_1, \theta_2, \dots, \theta_C; \pi_1, \pi_2, \dots, \pi_C) \\ = \sum_{j=1}^C \pi_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{1}{2\sigma_j^2}(y_i - x_i\beta_j)^2\right)$$

Example 1

Color of Wine

Parameter	component 1	component 2	component 3
Constant	4.929 (0.334)	7.548 (0.936)	2.803 (0.244)
π	0.365 (0.176)	0.312 (0.117)	0.323 (0.107)

Example 1

Color of Wine

Parameter	component 1	component 2	component 3
Constant	4.929 (0.334)	7.548 (0.936)	2.803 (0.244)
π	0.365 (0.176)	0.312 (0.117)	0.323 (0.107)

Cultivar	Freq.	% of total	Color (mean)
1	59	33.15	5.528
2	71	39.89	3.086
3	48	26.97	7.396
Total	178	100	5.058

Example 1

Color of Wine

Cultivar	Posterior probability (median)		
	component 1	component 2	component 3
1	0.737	0.195	9.00e-5
2	0.048	0.023	0.923
3	0.030	0.970	7.54e-14

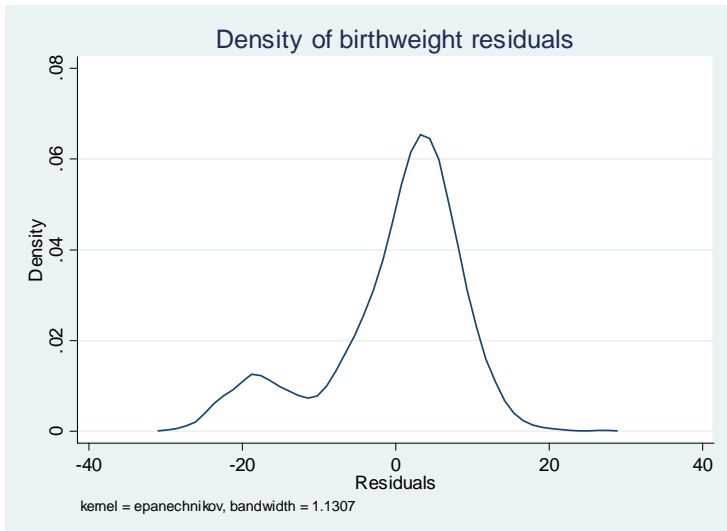
Example 2

Infant Birthweight and Prenatal Care

- Data from the National Maternal and Infant Health Survey
- Number of observations: 5219
- Number of covariates: 12

Example 2

Infant Birthweight and Prenatal Care



Example 2

Infant Birthweight and Prenatal Care

Variable	OLS	FMM	
		component 1	component 2
black	-1.213** (0.312)	-1.231** (0.215)	-0.775* (0.393)
edu	0.353** (0.074)	0.292** (0.050)	0.040 (0.102)
numdead	-1.181** (0.163)	-0.170 (0.117)	-0.585** (0.171)
onsethat	-0.501** (0.183)	-0.294* (0.127)	0.006 (0.234)
π		0.864	0.136
se(π)		(0.005)	(0.005)

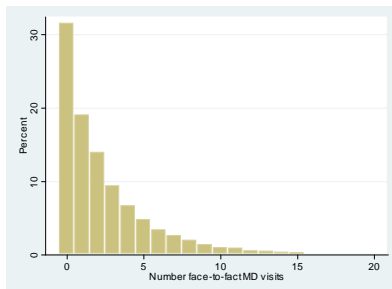
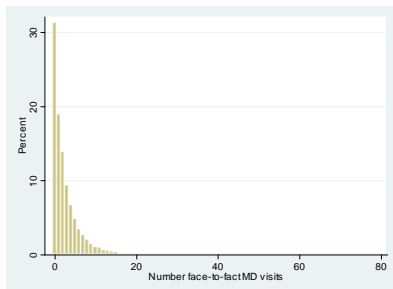
Example 3

Medical Care Use

- Data from the RAND Health Insurance Experiment
- Number of observations: 20186
- Number of covariates: 17

Example 3

Medical Care Use



Example 3

Medical Care Use

- The density of the C -component finite mixture is specified as

$$f(y_i | \theta_1, \theta_2, \dots, \theta_C; \pi_1, \pi_2, \dots, \pi_C) \\ = \sum_{j=1}^C \pi_j \frac{\Gamma(y_i + \psi_{j,i})}{\Gamma(\psi_{j,i})\Gamma(y_i + 1)} \left(\frac{\psi_{j,i}}{\lambda_{j,i} + \psi_{j,i}} \right)^{\psi_{j,i}} \left(\frac{\lambda_{j,i}}{\lambda_{j,i} + \psi_{j,i}} \right)^{y_i}$$

where $\lambda = \exp(x\beta)$ and $\psi = (1/\alpha)\lambda^k$

- $k = 1$ NB-2
- $k = 0$ NB-1
- $k = 0$ fits best

Example 3

Medical Care Use

	Parameter Estimates		
	nb1	fmm nb1	
		component 1	component 2
logc	-0.149*	-0.203*	-0.024
	(0.012)	(0.020)	(0.031)
educdec	0.023*	0.027*	0.015
	(0.003)	(0.005)	(0.010)
disea	0.021*	0.019*	0.033*
	(0.001)	(0.002)	(0.004)
π		0.802	0.198
		(0.037)	(0.037)
log L	-42405	-42037	
BIC	84999	84461	

Example 3

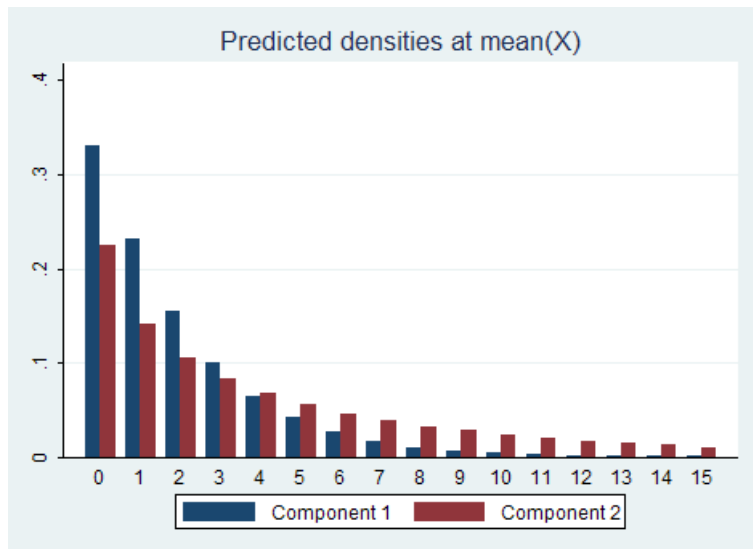
Medical Care Use

Marginal Effects

	nb1		fmm nb1	
	overall	overall	component 1	component 2
$E(y \bar{x})$	2.561	2.511	1.887	5.038
logc	-0.382* (0.030)	-0.331* (0.032)	-0.382* (0.032)	-0.121 (0.158)
educdec	0.058* (0.007)	0.056* (0.009)	0.052* (0.008)	0.073 (0.053)
disea	0.054* (0.003)	0.062* (0.004)	0.036* (0.004)	0.167* (0.024)

Example 3

Medical Care Use



Example 3

Medical Care Use

Prior and posterior probabilities

