Exploring Heterogeneity of Treatment Effects using Finite Mixture Models

Partha Deb

Hunter College and the Graduate Center, CUNY
NBER

August 2013

Introduction: Heterogeneity of treatment effects

- There are many intuitive reasons for expecting that treatment effects are not constant
- In most pragmatic trials in which experimental conditions are not tightly controlled
  - actual treatment is heterogeneous across sites or geographies
  - intensity of treatment (dose) varies
  - compliance to treatment varies by individual or group characteristics
  - effects of treatment varies by biological characteristics

Introduction: Heterogeneity of treatment effects

- Heterogeneities in each of these dimensions lead to heterogeneity of treatment effects (HTE).

- When treatment effects are heterogeneous, the typically small benefits found in studies can be misleading because small average effects may reflect a mixture of substantial benefits for some, little benefit for many, and even possibly harmful for a few

- Heterogeneous treatment effects can be categorized into three groups (Gadbury and Iyer, 2000; Gadbury, Iyer, and Albert, 2004; Shen, et al. 2013)

Introduction: Heterogeneity of treatment effects

- Let Y_1 and Y_0 be the potential outcomes under the intervention and the control
- Benefit is defined as $Y_1 - Y_0 > 0$
- Harm occurs when $Y_1 - Y_0 < 0$

Introduction: Heterogeneity of treatment effects

- Let Y_1 and Y_0 be the potential outcomes under the intervention and the control
- Benefit is defined as $Y_1 - Y_0 > 0$
- Harm occurs when $Y_1 - Y_0 < 0$
- Treatment benefit rate $TBR = Pr(Y_1 - Y_0 > 0)$
- Treatment harm rate $THR = Pr(Y_1 - Y_0 < 0)$
- For the fraction $1 - TBR - THR$ of the population, $Y_1 - Y_0 = 0$

Introduction: Heterogeneity of treatment effects

- TBR and THR may be characterized by a small set of observed covariates in some instances

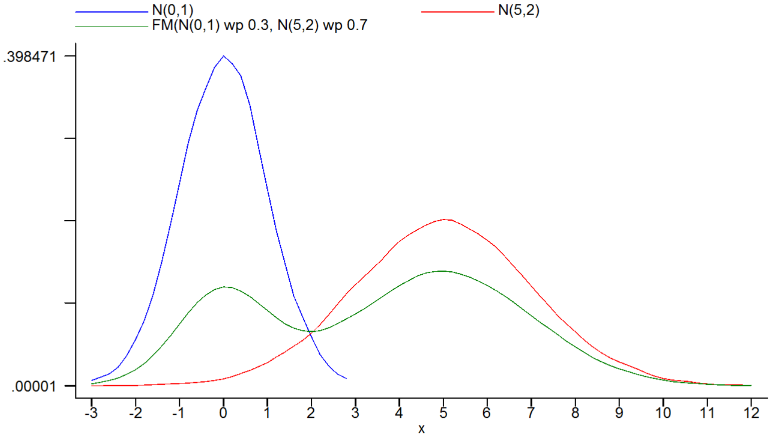
- TBR and THR are more likely to be determined by complex configurations of observed and unobserved covariates

Introduction: Finite mixture models

- When data has been drawn from a finite number of distinct populations but in which the sub-populations are not identified

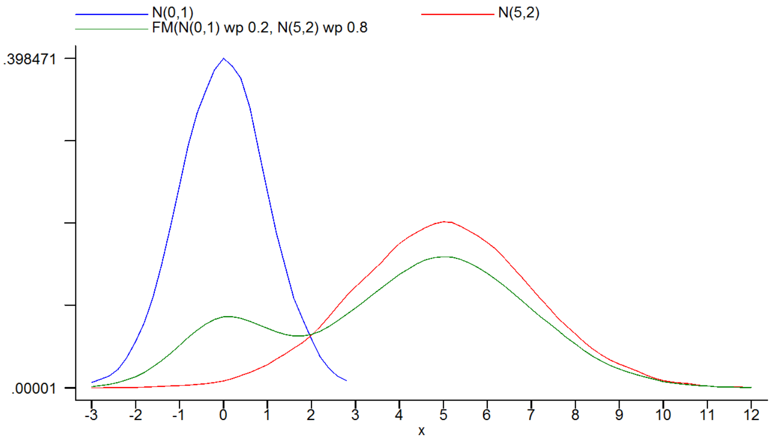
- A finite mixture model allows one to identify and estimate the parameters of interest for each sub-population in the data, not just of the overall mixed population

Introduction: A graphical view



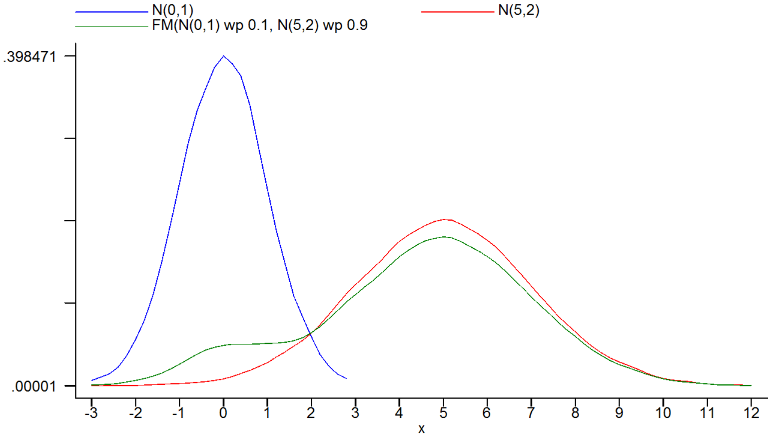
Mixture of normal densities

Introduction: A graphical view



Mixture of normal densities

Introduction: A graphical view



Mixture of normal densities

Introduction: Published applications

- Effects of job loss on BMI and alcohol consumption – Deb, Gallo, Ayyagari, Fletcher and Sindelar, *Journal of Health Economics* 2011
- Effect of prenatal care on birth weight – Conway and Deb, *Journal of Health Economics* 2005
- Price elasticities of medical care use – Deb and Trivedi, *Journal of Applied Econometrics* 1997; Deb and Trivedi, *Journal of Health Economics* 2002

Outline of talk

- Introduction

Outline of talk

- Introduction
- Model
  - Mixture density
  - Properties
  - Complications

Outline of talk

- Introduction
- Model
  - Mixture density
  - Properties
  - Complications
- Example: color of wine
- Application: healthcare expenditures in the RAND Health Insurance Experiment

Model: Common mixture component densities

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

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

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

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

Model: Basic properties

- The conditional mean of the outcome in a finite mixture model is a linear combination of component means:

$$\mathrm{E}(y_i|\mathbf{x}_i) = \sum_{j=1}^{C} \pi_j \lambda_{ij} \text{ where } \lambda_{ij} = \mathrm{E}_j(y_i|\mathbf{x}_i)$$

Model: Basic properties

- The conditional mean of the outcome in a finite mixture model is a linear combination of component means:

$$\mathrm{E}(y_i|\mathbf{x}_i) = \sum_{j=1}^{C} \pi_j \lambda_{ij} \text{ where } \lambda_{ij} = \mathrm{E}_j(y_i|\mathbf{x}_i)$$

- The marginal effect of a covariate in a finite mixture model is a linear combination of component marginal effects:

$$\frac{\partial \mathrm{E}_j(y_i|\mathbf{x}_i)}{\partial \mathbf{x}_i} = \frac{\partial \lambda_{ij}}{\partial \mathbf{x}_i} \longrightarrow \text{within component}$$

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

Model: Basic properties

- Prior probability that observation y_i belongs to component c is often specified as a constant

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

$$c = 1, 2, ..C$$

- It cannot be used to classify *individual* observations into types

Model: Basic properties

- Prior probability that observation y_i belongs to component c is often specified as a constant

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

$$c = 1, 2, ..C$$

- It cannot be used to classify *individual* observations into types
- Posterior probability that observation y_i belongs to component c:

$$\Pr[y_i \in \text{population } c|\mathbf{x}_i, y_{i;}\Theta] = \frac{\pi_c f_c(y_i|\mathbf{x}_i, \Theta_c)}{\sum_{j=1}^{C} \pi_j f_j(y_i|\mathbf{x}_i, \Theta_j)}$$

$$c = 1, 2, ..C$$

- It can be used to classify *individual* observations into types

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

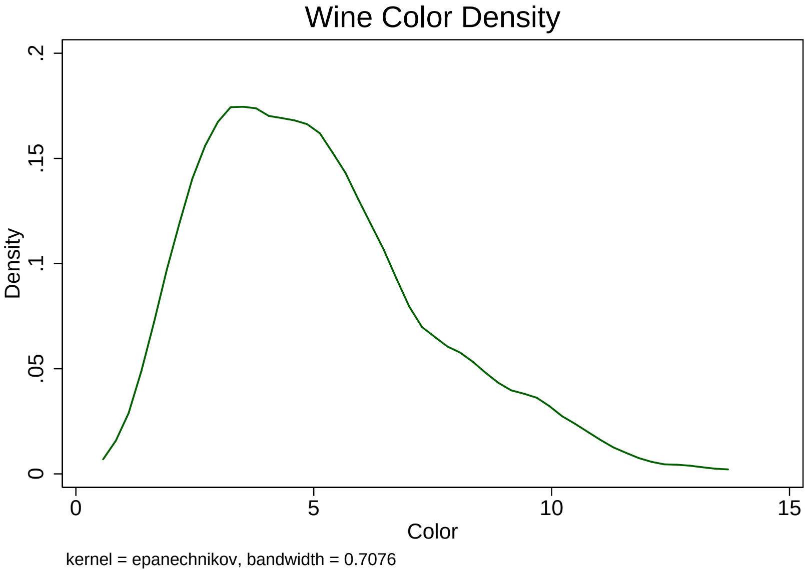
Example: Color of wine

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

Data characteristics

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: Color of wine



Example: Color of wine

- Suppress information on class (cultivar)
- Estimate a Finite mixture of Normals with 3 components
- Use estimates of posterior probabilities to assign observations into one of 3 classes

Example: Color of wine

Estimates from finite mixture of normals with 3 components

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

Data characteristics

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: Color of wine

Cultivar		Predicted Cultivar 1	2	3	Total
1	No.	42	3	14	59
	%	71.2	5.1	23.7	100.0
2	No.	15	56	0	71
	%	21.1	78.9	0.0	100.0
3	No.	19	0	29	48
	%	39.6	0.0	60.4	100.0

Application: Medical Care Use

- Heterogeneity of insurance effects on healthcare expenditure

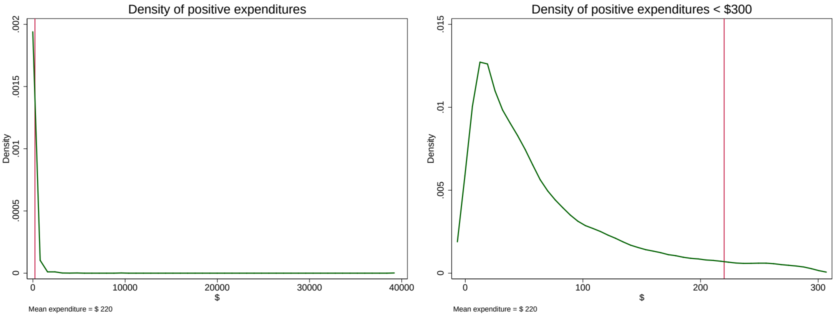
Application: Medical Care Use

- Heterogeneity of insurance effects on healthcare expenditure
- Data from the Rand Health Insurance Experiment (RHIE)
- Conducted by the RAND Corporation from 1974 to 1982
- Individuals were randomized into insurance plans
- Widely regarded as the basis of the most reliable estimates of price elasticities

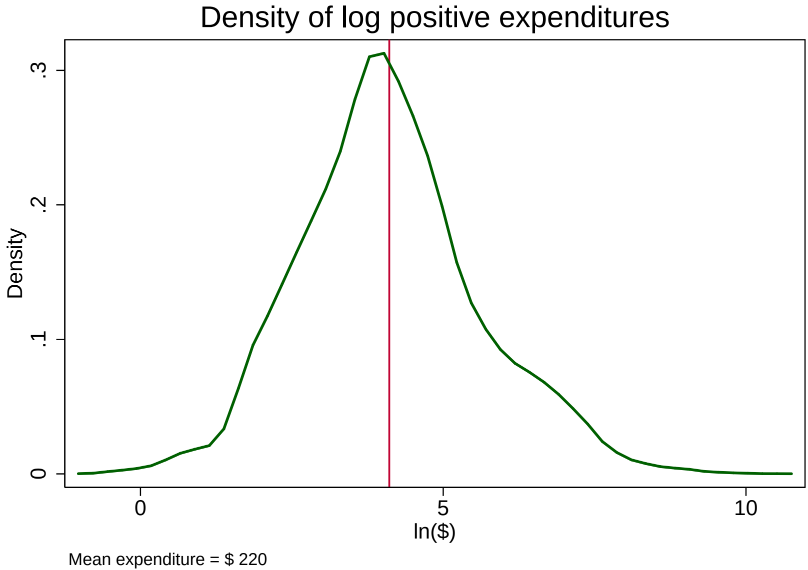
Application: Medical Care Use

- Data collected from about 8,000 enrollees in 2,823 families from six sites across the country
- Each family was enrolled in one of fourteen different insurance plans for either three or five years
- The FFS plans ranged from free care to 95% coinsurance
- Data from all 5 years of the experiment
- Number of observations: 20,186

Application: Medical Care Use



Application: Medical Care Use



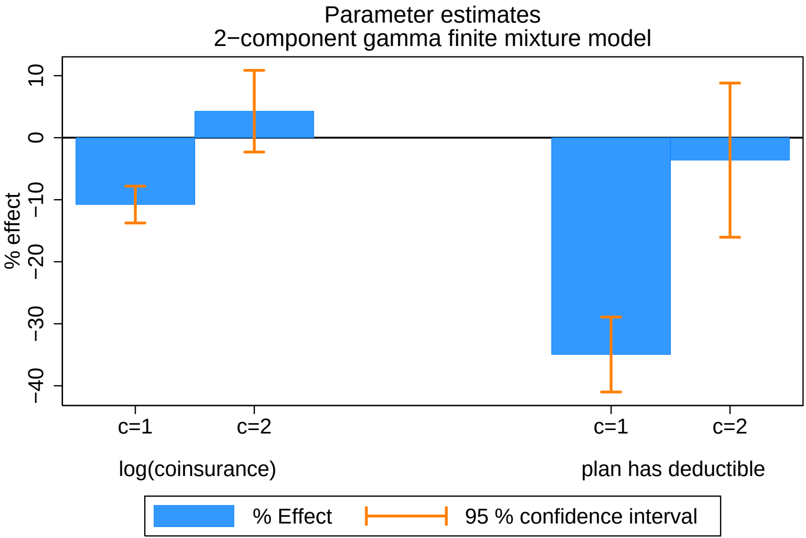
Application: Medical Care Use

Explanatory variables	
IDP	1 if individual deductible plan, 0 otherwise
LC	ln(coinsurance+1), 0≤coinsurance≤100
LPI	f(annual participation incentive payment)
FMDE	f(maximum dollar expenditure)
LINC	ln(family income)
LFAM	ln(family size)
EDUCDEC	education of the household head in years
PHYSLIM	1 if the person has a physical limitation
NDISEASE	number of chronic diseases
PHINDEX	index of health (larger is worse)
AGE, FEMALE, CHILD, BLACK	

Application: Medical Care Use

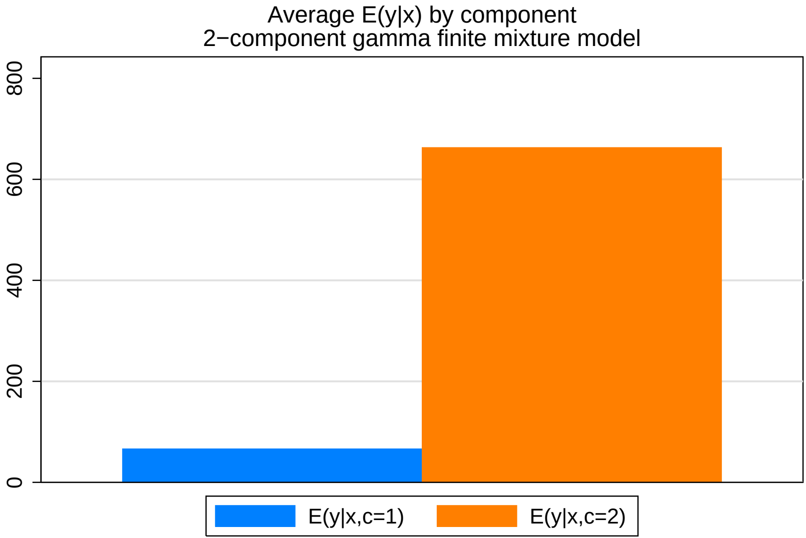
- Estimate a 2-component finite mixture of gamma densities
  - Highlight differences in treatment effects by component (class)
  - Highlight differences in distributions of expenditures by class
  - Explore sources / correlates of class differences
- Estimate a 3-component finite mixture of gamma densities

Application: Medical Care Use



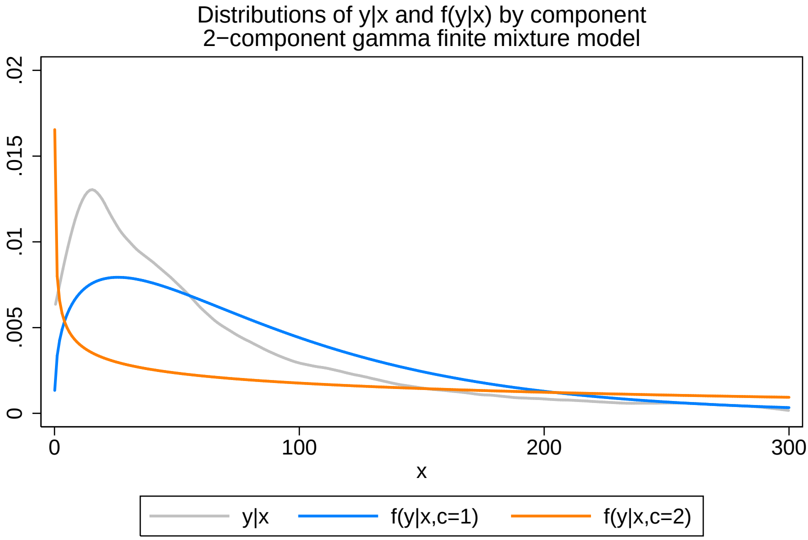
$\pi_1 = 0.75$, $\pi_2 = 0.25$

Application: Medical Care Use



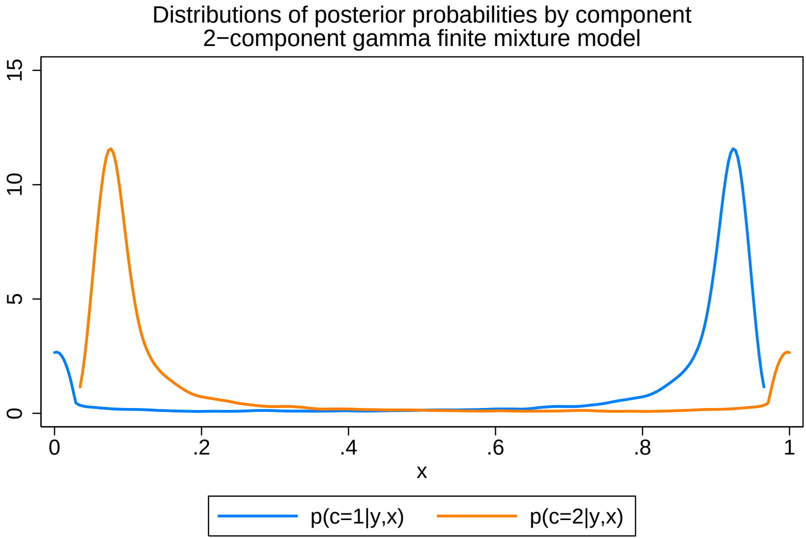
$\pi_1 = 0.75$, $\pi_2 = 0.25$

Application: Medical Care Use



$\pi_1 = 0.75$, $\pi_2 = 0.25$

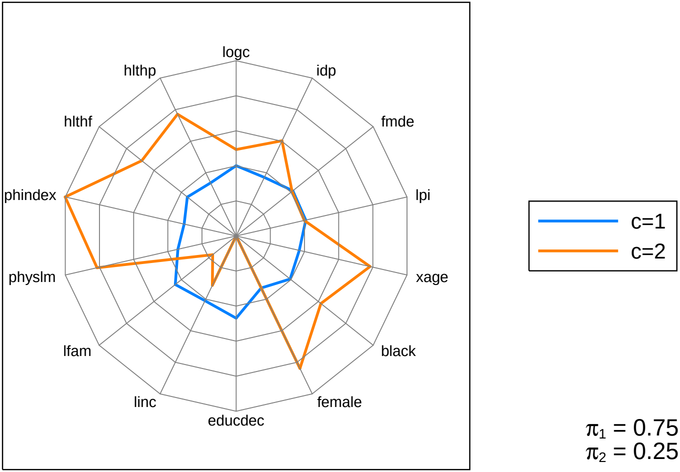
Application: Medical Care Use



$\pi_1 = 0.75$, $\pi_2 = 0.25$

Application: Medical Care Use

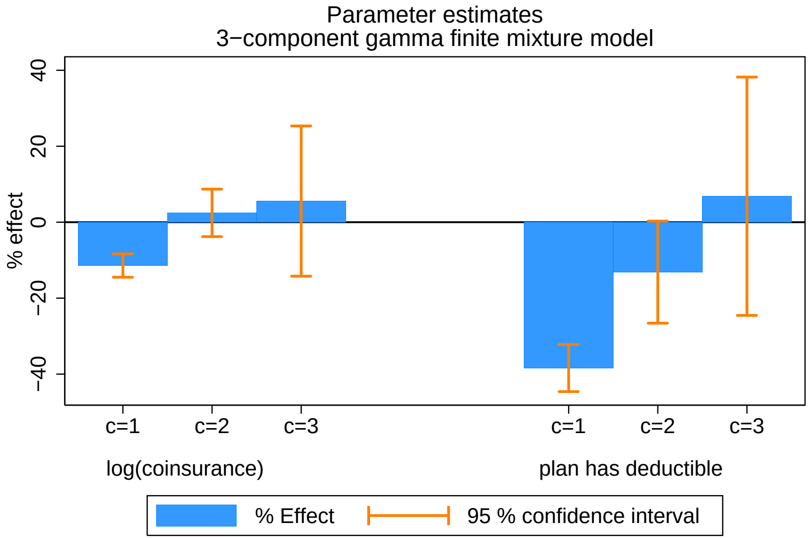
Average covariate values by component assignment
2−component gamma finite mixture model



$\pi_1 = 0.75$
$\pi_2 = 0.25$

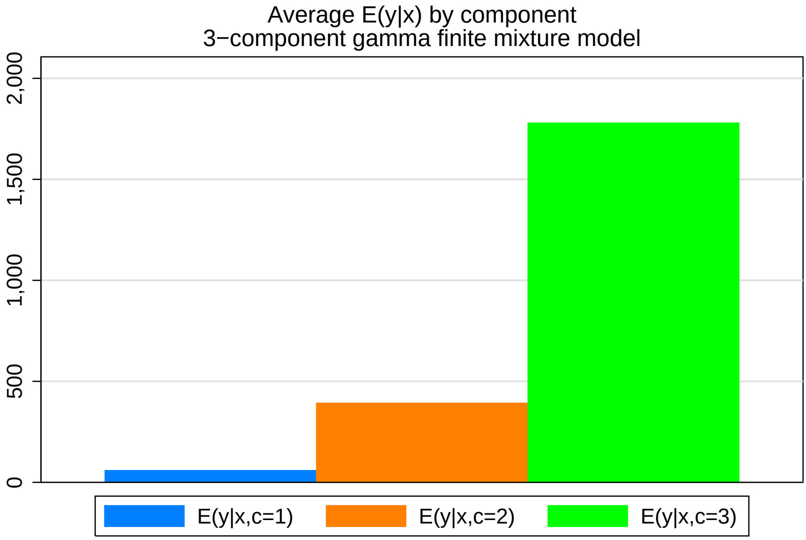
Values of covariates are standardized
An observation is classified into class c if p(c|y,x)>0.7)

Application: Medical Care Use



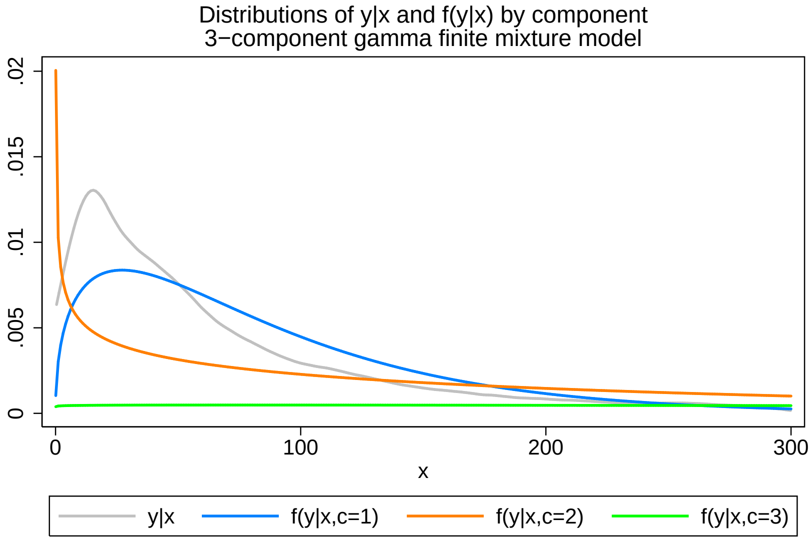
$\pi_1 = 0.71$, $\pi_2 = 0.25$, $\pi_3 = 0.04$

Application: Medical Care Use



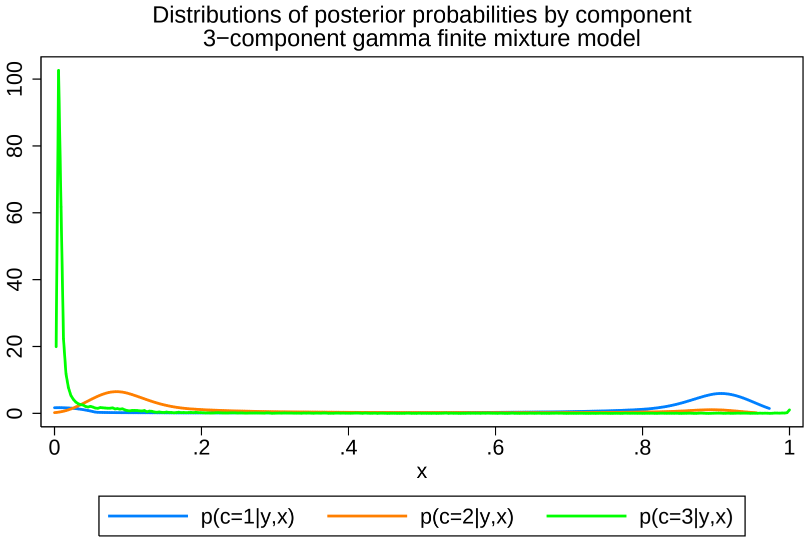
$\pi_1 = 0.71$, $\pi_2 = 0.25$, $\pi_3 = 0.04$

Application: Medical Care Use



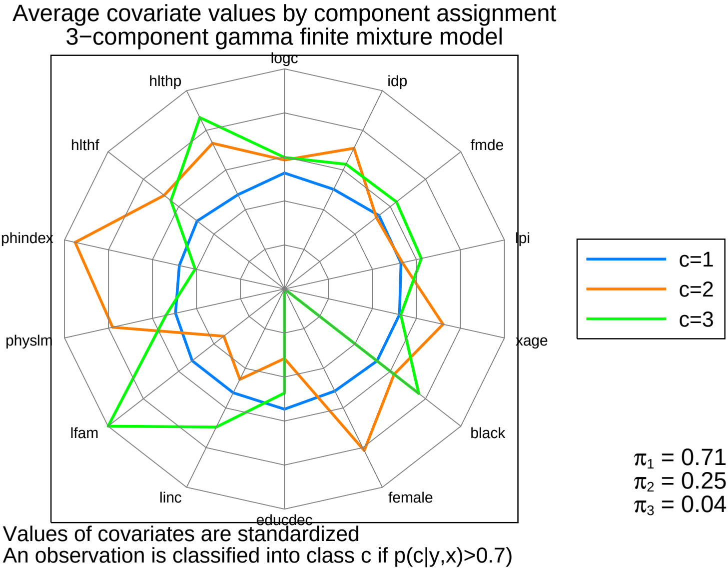
$\pi_1 = 0.71$, $\pi_2 = 0.25$, $\pi_3 = 0.04$

Application: Medical Care Use



$\pi_1 = 0.71$, $\pi_2 = 0.25$, $\pi_3 = 0.04$

Application: Medical Care Use



$\pi_1 = 0.71$
$\pi_2 = 0.25$
$\pi_3 = 0.04$

Conclusions

- Heterogeneity in treatment effects can be important to identify and explore
- When treatment effects are heterogeneous, the typically small benefits found in studies can be misleading
- Finite mixture models are a useful way to model heterogeneous treatment effects
- FMM can uncover otherwise hidden heterogeneity

Conclusions

- Heterogeneity in treatment effects can be important to identify and explore
- When treatment effects are heterogeneous, the typically small benefits found in studies can be misleading
- Finite mixture models are a useful way to model heterogeneous treatment effects
- FMM can uncover otherwise hidden heterogeneity
- As described, FMM can be applied when outcomes are continuous or discrete
- But not for binary or "severely" limited outcomes
- Extension to collection of binary outcomes is referred to as the Grade Of Membership model