# Latent class models

Key concepts
& study plan

Experimental
design

Data collection
& processing

**Model specification
& estimation**

Interpretation
& application

# The basics

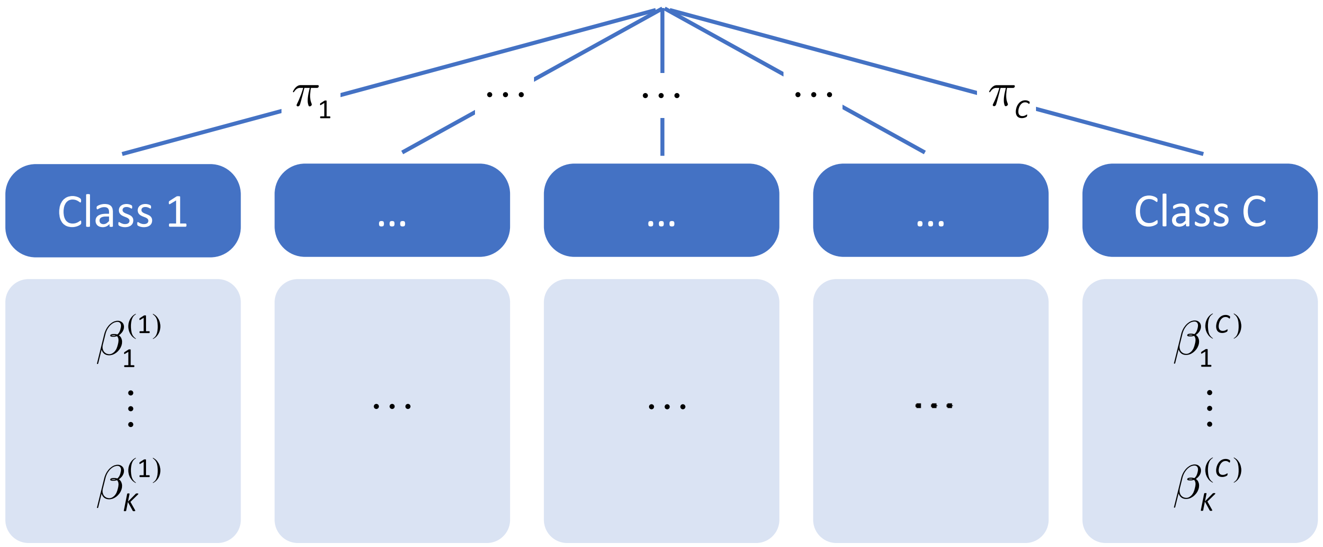
## Latent classes

- Discrete groups of decision-makers with specific preferences, called classes
- Classes are unobserved (latent)
  - Decision-makers belong to one of the classes
  - Data allows one to obtain an estimated probability of class membership
- Parameters for all classes are estimated simultaneously

# The basics

## Notation

- Assume $C$ classes, indexed as $c = 1, \ldots, C$
- Each class has class-specific parameters $\beta_1^{(c)}, \ldots, \beta_K^{(c)}$
- Each class has membership probability $\pi_c$ such that $\sum_{c=1}^{C} \pi_c = 1$



# The basics

## Example

- Consider a choice model with the following utility functions
  - $V_A = \beta_1 + \beta_2 \cdot X_A$
  - $V_B = \beta_3 \cdot X_B$
- Assume 3 classes

| | Class 1 | Class 2 | Class 3 | | MNL model |
|---|---|---|---|---|---|
| $\pi$ | 0.25 (0.04) | 0.44 (0.08) | 0.31 (0.03) | *31% of population belongs to class 3* | |
| $\beta_1$ | −0.18 (0.06) | −0.29 (0.08) | −0.09 (0.08) | compare | −0.23 (0.05) |
| $\beta_2$ | 0.45 (0.14) | 0.27 (0.14) | 0.65 (0.15) | | 0.51 (0.11) |
| $\beta_3$ | −0.23 (0.04) | −0.30 (0.09) | 0.07 (0.01) | | −0.19 (0.02) |

*standard errors between parentheses*

# Estimating latent class models

Key concepts & study plan

Experimental design

Data collection & processing

**Model specification & estimation**

Interpretation & application

# Estimating latent class models

## Recall maximum likelihood estimation procedure

- Exogenous data **x**
- Choice observations **y**

- Specify utility functions with parameters $\boldsymbol{\beta}$

- Find parameter estimates $\hat{\boldsymbol{\beta}}$ that
maximise the (log)likelihood of observing **y** based on data **x**

$$\hat{\boldsymbol{\beta}} = \max_{\boldsymbol{\beta}} \sum_{n=1}^{N} \log P_n(\mathbf{x}, \mathbf{y}, \boldsymbol{\beta})$$

where $P_n(\mathbf{x}, \mathbf{y}, \boldsymbol{\beta}) = \prod_{t=1}^{T} \prod_{j=1}^{J} \left(P_{ntj}(\mathbf{x}, \boldsymbol{\beta})\right)^{y_{ntj}}$

# Estimating latent class models

## Maximum likelihood estimation for latent class models

- Exogenous data **x**
- Choice observations **y**

- Specify utility functions with parameters $\boldsymbol{\beta}$
- Specify number of classes $C$
- Find class-specific parameter estimates $\hat{\boldsymbol{\beta}}^{(c)}$ and class membership probabilities $\pi^{(c)}$ that maximise the (log)likelihood of observing **y** based on data **x** using class-weighted probabilities

$$(\hat{\boldsymbol{\beta}}^{(1)}, \ldots, \hat{\boldsymbol{\beta}}^{(C)}, \hat{\pi}^{(1)}, \ldots, \hat{\pi}^{(C)}) = \max_{\boldsymbol{\beta},\boldsymbol{\pi}} \sum_{n=1}^{N} \log \sum_{c=1}^{C} \pi^{(c)} \cdot P_n(\mathbf{x}, \mathbf{y}, \boldsymbol{\beta}^{(c)}) \quad \text{where} \quad \sum_{c=1}^{C} \pi^{(c)} = 1$$

# Estimating latent class models

## Selection of model type for each class

- Choice probabilities $P_{ntj}\left(\mathbf{x},\boldsymbol{\beta}^{(c)}\right)$ of each latent class can be derived from
  - Multinomial logit (MNL)
  - Nested logit
  - Heteroskedastic logit
  - Etc.

- Combinations can provide flexibility

- Most common: MNL for each class

$$P_{ntj}\left(\mathbf{x},\boldsymbol{\beta}^{(c)}\right)=\frac{\exp\left(\boldsymbol{\beta}^{(c)}\mathbf{x}_{ntj}\right)}{\sum_i \exp\left(\boldsymbol{\beta}^{(c)}\mathbf{x}_{nti}\right)}$$

# Estimating latent class models

## Class membership probabilities

- Instead of estimating $\pi^{(c)}$ directly, they are typically estimated indirectly
- Use logit model with only constants $\delta^{(c)}$

$$\pi^{(c)} = \frac{\exp(\delta^{(c)})}{\sum_{c'} \exp(\delta^{(c')})}$$

- Class membership constant of one class must be normalised to zero
  - For example, for the last class, *C*, we set $\delta^{(C)} = 0$
  - It does not matter which class you choose for normalisation

# Estimating latent class models

## Example

- Estimate constants in logit model for class membership
- Normalise class membership constant of Class 3 to zero
- Most people belong to Class 2, least people belong to Class 1

| | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| $\delta$ | $-0.21$ (0.02) | 0.34 (0.05) | 0 (--) |
| $\beta_1$ | $-0.18$ (0.06) | $-0.29$ (0.08) | $-0.09$ (0.08) |
| $\beta_2$ | 0.45 (0.14) | 0.27 (0.14) | 0.65 (0.15) |
| $\beta_3$ | $-0.23$ (0.04) | $-0.30$ (0.09) | 0.07 (0.01) |

*31% of population belongs to class 3*

$$\pi^{(3)} = \frac{e^{0}}{e^{-0.21} + e^{0.34} + e^{0}} = 0.31$$

# Estimating latent class models

## General class membership functions

- Class membership probabilities $\pi^{(c)}$ may depend on
  - socio-demographic variables **z**
  - scenario variables **w**
  - other variables
- Use logit model with class membership functions $M^{(c)}$

$$\pi^{(c)} = \frac{\exp(M^{(c)})}{\sum_{c'} \exp(M^{(c')})}, \qquad M^{(c)} = \delta^{(c)} + \boldsymbol{\gamma}^{(c)}\mathbf{z} + \boldsymbol{\omega}^{(c)}\mathbf{w}$$

- One needs to normalise parameters of one class to zero
  - For example, for the last class, *C*, we set $\delta^{(C)} = 0$, $\boldsymbol{\gamma}^{(C)} = \mathbf{0}$, and $\boldsymbol{\omega}^{(C)} = \mathbf{0}$
  - It does not matter which class you choose for normalisation

# Estimating latent class models

## Example

- Assume the following class membership functions
  - $M^{(1)} = \delta^{(1)} + \gamma_{\text{Age}}^{(1)} \cdot \text{Age}$
  - $M^{(2)} = \delta^{(2)} + \gamma_{\text{Age}}^{(2)} \cdot \text{Age}$
  - $M^{(3)} = 0$

| | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| $\delta$ | $-0.20$ (0.02) | 0.31 (0.06) | 0 (--) |
| $\gamma_{\text{Age}}$ | 0.02 (0.01) | $-0.03$ (0.01) | 0 (--) |
| $\beta_1$ | $-0.17$ (0.07) | $-0.28$ (0.07) | $-0.09$ (0.06) |
| $\beta_2$ | 0.46 (0.13) | 0.25 (0.13) | 0.62 (0.12) |
| $\beta_3$ | $-0.20$ (0.04) | $-0.31$ (0.09) | 0.05 (0.02) |

$$\pi^{(3)} = \frac{e^{0}}{e^{-0.20+0.02\cdot\text{Age}} + e^{0.31-0.03\cdot\text{Age}} + e^{0}}$$

$\text{age} = 20: \quad \pi^{(3)} = 0.34$

$\text{age} = 70: \quad \pi^{(3)} = 0.22$

# Estimating latent class models

## Starting values for parameters when estimating latent class models

- Use parameter estimates from MNL model
- For each class, make small deviations to avoid getting 'stuck' during model estimation
- Try multiple starting values

| | Class 1 | Class 2 | Class 3 | MNL model |
|---|---|---|---|---|
| $\delta$ | $-0.01$ | 0.01 | 0 | |
| $\gamma_{\text{Age}}$ | 0.01 | $-0.01$ | 0 | |
| $\beta_1$ | $-0.22$ | $-0.24$ | $-0.23$ | $-0.23$ |
| $\beta_2$ | 0.50 | 0.51 | 0.52 | 0.51 |
| $\beta_3$ | $-0.18$ | $-0.20$ | $-0.19$ | $-0.19$ |

# Estimating latent class models

## Choosing number of classes, *C*

- Start with two classes
- Gradually increase the number of classes and re-estimate the model
- Compare models with different number of classes based on model fit and interpretability
  - More classes always improves the *LL* value but requires estimating (many) more parameters
  - Models with low AIC or BIC are preferred
  - Models with meaningful/explainable insights are preferred

# Application to Swiss value of time study

Key concepts & study plan

Experimental design

Data collection & processing

**Model specification & estimation**

Interpretation & application

# Application to Swiss value of time study

## Model

- Utility functions:

$$V(\text{RouteA}) = \beta_{\text{ASC}} + \beta_{\text{TT}} \cdot \text{TravelTime} + \beta_{\text{TC}} \cdot \text{TravelCost} + \beta_{\text{HW}} \cdot \text{Headway} + \beta_{\text{CH}} \cdot \text{Interchanges}$$
$$V(\text{RouteB}) = \beta_{\text{TT}} \cdot \text{TravelTime} + \beta_{\text{TC}} \cdot \text{TravelCost} + \beta_{\text{HW}} \cdot \text{Headway} + \beta_{\text{CH}} \cdot \text{Interchanges}$$

- Assume two classes, with only constants in membership functions:

$$M^{(1)} = \delta^{(1)}$$
$$M^{(2)} = 0$$

# Application to Swiss value of time study

## Estimation results

- Class 1 members are much more cost sensitive
- Class 1 members are much more averse to interchanges
- Class membership probability is not statistically different across classes (so about 50-50%)

| | Class 1 | Class 2 |
|---|---|---|
| $\delta$ | $-0.039$ (0.268) | 0 (--) |
| $\beta_{\text{ASC}}$ | $-0.045$ (0.048) | |
| $\beta_{\text{TT}}$ | $-0.098$ (0.014) | $-0.074$ (0.009) |
| $\beta_{\text{TC}}$ | $-0.534$ (0.094) | $-0.096$ (0.016) |
| $\beta_{\text{HW}}$ | $-0.047$ (0.006) | $-0.040$ (0.004) |
| $\beta_{\text{CH}}$ | $-2.168$ (0.185) | $-0.764$ (0.105) |

*51% of population belongs to class 2*

$$\pi^{(2)} = \frac{e^{0}}{e^{-0.039} + e^{0}} = 0.51$$

*ASC assumed generic across both classes*

# Application to Swiss value of time study

## Model comparison

- Which model is preferred?

- MNL model
  - 1 class
  - 5 parameters

```
LL(final)                                : -1665.62
Rho-squared vs equal shares               :  0.3119
Adj.Rho-squared vs equal shares           :  0.3098
Rho-squared vs observed shares            :  0.3118
Adj.Rho-squared vs observed shares        :  0.3102
AIC                                      :  3341.24
BIC                                      :  3372.03
```

- Latent class model
  - 2 classes
  - 10 parameters

```
LL(final, whole model)                   : -1562.08
Rho-squared vs equal shares               :  0.3546
Adj.Rho-squared vs equal shares           :  0.3505
Rho-squared vs observed shares            :  0.3546
Adj.Rho-squared vs observed shares        :  0.3513
AIC                                      :  3144.16
BIC                                      :  3205.74
```

# Application to Swiss value of time study

## Model comparison

- Which model is preferred?

- MNL model
  - 1 class
  - 5 parameters

```
LL(final)                                  : -1665.62
Rho-squared vs equal shares                 :  0.3119
Adj.Rho-squared vs equal shares             :  0.3098
Rho-squared vs observed shares              :  0.3118
Adj.Rho-squared vs observed shares          :  0.3102
AIC                                        :  3341.24
BIC                                        :  3372.03
```

- Latent class model
  - 2 classes
  - 10 parameters

```
LL(final, whole model)                     : -1562.08
Rho-squared vs equal shares                 :  0.3546
Adj.Rho-squared vs equal shares             :  0.3505
Rho-squared vs observed shares              :  0.3546
Adj.Rho-squared vs observed shares          :  0.3513
AIC                                        :  3144.16   ✓
BIC                                        :  3205.74   ✓
```