Segmentation V

Advanced latent class analysis

Morten Berg Jensen

Department of Economics and Business Economics

April 22, 2024

Outline

Outcome

- This lecture will help you to understand
  - Other finite mixture models
  - Applications of mixture models

LCA with concomitant variables

- The possibility of introducing explanatory variables/covariates for the finite mixture of distributions/LCA is typically done via the class membership probabilities
- The model becomes

$$\mathbf{P}(\mathbf{y}_i|x_i) = \sum_{c=1}^{C} \pi_{c|x_i} \prod_{k=1}^{K} \mathbf{P}(y_{ik}|c)$$

which should be contrasted with the "standard" formulation

$$\mathbf{P}(\mathbf{y}_i) = \sum_{c=1}^{C} \pi_c \prod_{k=1}^{K} \mathbf{P}(y_{ik}|c)$$

- Only the class membership probabilities have been changed but we also implicitly assume that the effect of x_i on y_i is fully mediated by the latent classes

LCA with concomitant variables (cont'd)

- The latter assumption can be tested using the local fit measures and relaxed by specifying the conditional probability as $\mathbf{P}(y_{ik}|c, x_i)$ whenever necessary
- Given the requirements of probabilities the specification of the membership probabilities is usually given as

$$\pi_{c|x_i} = \frac{\exp\left(\gamma_{0c} + \sum_{p=1}^{P} \gamma_{pc} x_{ip}\right)}{\sum_{c'=1}^{C} \exp\left(\gamma_{0c'} + \sum_{p=1}^{P} \gamma_{pc'} x_{ip}\right)}$$

- For identification the membership probabilities should be equal to 0 for one of the classes or add up to 0 across the classes

LCA with concomitant variables (cont'd)

- At the outset, this would seem to be unnecessarily complex – why not just
  1. Do LCA without covariates
  2. Allocate each object to one of the classes based on the posterior probabilities
  3. Analyze the relationship between class membership and covariates
- In this way, the segmentation structure is entirely determined by the indicator variables and not by the covariates, which in many exploratory analyses would be the natural way
- The problem is that this leads to downward biased estimates of the covariate effect – but there are ways around this problem

Finite mixture of regressions

- The “classical” segmentation approach (hierarchical and non-hierarchical) resembles in many ways the latent class analysis/finite mixtures of distributions and result often in comparable solutions
- The same kind of input data can be used and the overall segmentation process is very similar
- A related set of methods, finite mixtures of regression models, takes a rather different approach
- As the name suggests, we have a dependent variable and a set of independent variables, but the functional relationship between them depends on which of a finite set of segments the object belongs to

Finite mixture of regressions (cont'd)

- This is how the model looks like

$$y_i = \beta_0^k + \beta_1^k x_{i1} + \ldots + \beta_p^k x_{ip} + \varepsilon_i^k$$

where $\varepsilon_i^k \sim N(0, \sigma_k^2)$

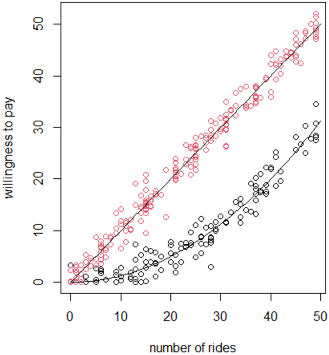
- The choice regarding the number of clusters can be based on AIC, BIC and ICL like for a mixture of distributions
- The issues with label switching are still present
- A small toy example will illustrate

Toy example

- We are interested in the relationship between the entrance fee consumers are willing to pay for a theme park and the number of rides available in the park
- The data generating process for two segments is as follows

$$\text{segment 1: } y = x + \varepsilon$$
$$\text{segment 2: } y = 0.0125x^2 + \varepsilon$$

- And this is how it looks like (colors are estimated classes)



Toy example (cont'd)

- In the simulation, ε is normally distributed with $\sigma = 2$ for both classes
- Furthermore, y was kept non-negative
- Using the `flexmix` package, here are the estimated parameters

```
> set.seed(1234)
> park.f1 <- stepFlexmix(pay ~ rides + I(rides^2),data = themepark, k = 2,
+                        nrep = 10, verbose = FALSE)
> park.f1

Call:
stepFlexmix(pay ~ rides + I(rides^2), data = themepark, k = 2, nrep = 10, verbose = FALSE)

Cluster sizes:
  1   2
119 201

convergence after 20 iterations
> # Present parameters
> parameters(park.f1)
                      Comp.1       Comp.2
coef.(Intercept)  1.60922212 0.3172187330
coef.rides       -0.11509873 0.9905142322
coef.I(rides^2)   0.01439448 0.0001851452
sigma             2.06269059 1.9898849543
```

Bijmolt et al. (2004)

- **Purpose:** Segment customers based on ownership of eight financial products taking into account that segmentation must address the customer as well as the country level differences; age, income, marital status, and type of community together with country level segment information was used as predictors for customer level segments
- **Data:** Survey data from the Eurobarometer 56.0 from 17 countries and regions; approximately 1000 consumers per country
- **Model:** Multilevel LCA with concomitant variables
- **Results:** Based on CAIC a 14 consumer level and seven country level solution is optimal and highly interpretable; demographic variables predict customers segments

De Keyser et al. (2015)

- **Purpose:** Segment customers based on use of either the Internet, a brick-and-mortar store, or a call center in each of the three stages of the customer journey – Information search, purchase, and after-sales service using five latent and four manifest covariates to predict segment membership
- **Data:** Survey data from 314 customers of a Dutch telecom retailer; mean scores used for the latent covariates based on between two and four items (seven-point Likert)
- **Model:** LCA with concomitant variables
- **Results:** Based on AIC3 a four segment solution is optimal for a model involving the first two stages whereas a six segment solution is optimal for the full three stage model; both models are easily interpretable; for the latter model, age, loyalty, and avg. revenue are significant predictors

Background and deciding (not) to segment

- McDonald's would like to know whether consumer segments exist with distinctly different images of McDonald's
- Such an understanding would inform McDonald's which segment(s) to focus on if any and what kind of communication to use
    - McDonald's can choose to cater to the entire market and hence ignore systematic differences across segments
    - The can also choose to on focus market segments with a positive perception and strengthen this perception
    - Or, focus on the segment with a negative perception and try to modify the drivers of the negative perception

Collecting data

- We have information from 1453 adult Australian consumers regarding their perception of McDonald's
- Specifically, they have indicated whether they feel McDonald's do or do not possess the following 11 attributes: YUMMY, CONVENIENT, SPICY, FATTENING, GREASY, FAST, CHEAP, TASTY, EXPENSIVE, HEALTHY, and DISGUSTING
- Given the data limitations we will use "liking McDonald's" and "frequently eating at McDonald's" as attractiveness criteria
- Finally, in addition to the two attractiveness criteria we have information about gender and age
- Had the data been collected for segmentation, additional information should have been collected about for instance dining out behaviour and use of information channels

Extracting segments

- Instead of finding market segments with similar perceptions we could consider finding segments with members whose love/hate for McDonald's is driven by similar perceptions
- In this situation, McDonald's could try to modify critical perceptions for certain segments with the purpose of improving love
- This idea can be operationalized using finite mixtures of linear regression models
- The dependent variable is the degree to which consumers love McDonald's and the independent variables are the 11 perceptions
- The segmentation variables are unobserved and consist of the regression coefficients

Extracting segments (cont'd)

- The ordinal nature of the dependent variable causes problems – the larger number of segments we investigate the more likely we end up in a situation where a segment is made up of consumers with identical rating on the dependent variable
- Such a group can be perfectly predicted and would lead to an infinite log-likelihood – a degenerate solution
- For this reason we will have to settle with a 2 segment solution

Extracting segments (cont'd)

- Group 1 like McDonald's if they perceive it as YUMMY, not FATTENING, FAST, CHEAP, TASTY, and not DISGUSTING
- Group 2 like McDonald's if they perceive it as YUMMY, CONVENIENT, not GREASY, HEALTHY, and not DISGUSTING
- So if segment 2 is targeted it is important to stress that McDonald's serves some healthy food – this is not necessary in order to target segment 1
- To target segment 1, McDonald's should stress how tasty, fast and cheap the food is

The remaining steps

- See previous slide set for considerations regarding
  - Profiling segments
  - Describing segments
  - Selecting (the) target segment(s)
  - Customising the marketing mix
  - Evaluation and monitoring