Latent class analysis

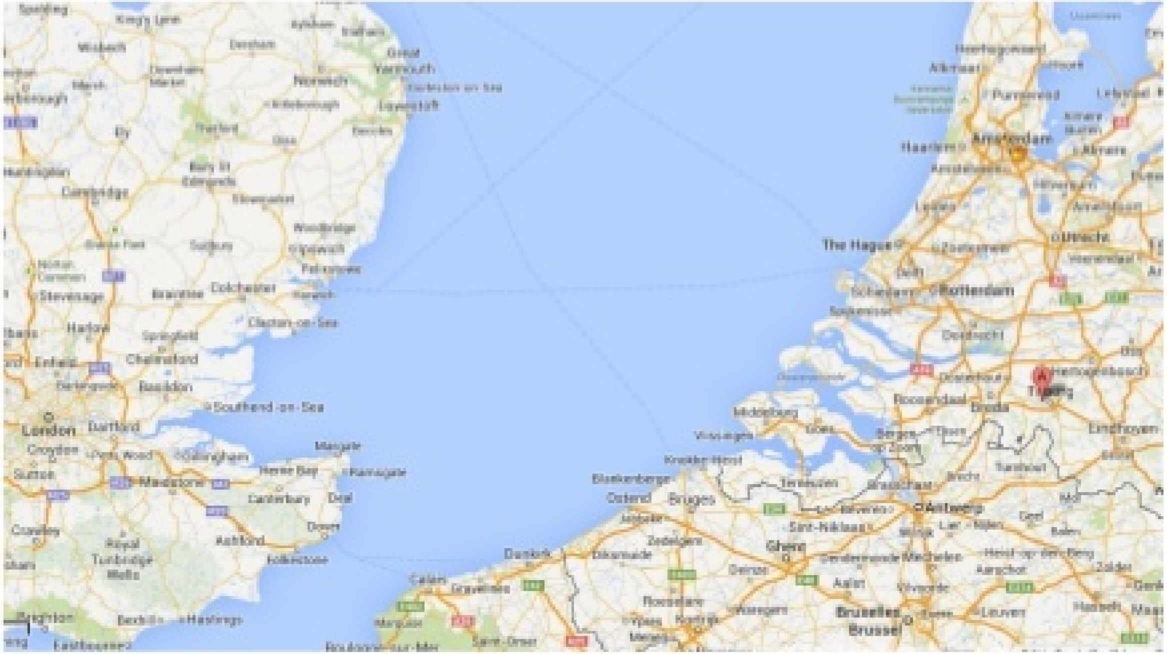
Daniel Oberski
Dept of Methodology & Statistics
Tilburg University, The Netherlands

(with material from Margot Sijssens-Bennink & Jeroen Vermunt)

TILBURG UNIVERSITY

ESRA

About Tilburg University Methodology & Statistics



About Tilburg University Methodology & Statistics

“Home of the latent variable”

Major contributions to latent class analysis:

Jacques Hagenaars *(emeritus)*	Jeroen Vermunt	Marcel Croon *(emeritus)*

More latent class modeling in Tilburg

Guy
Moors
(extreme
respnse)

Klaas
Sijtsma
(Mokken;
IRT)

Wicher
Bergsma
(marginal
models)
(@LSE)

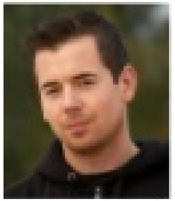
Daniel
Oberski
(local fit of
LCM)

Recent PhD's

Zsuzsa
Bakk
(3step LCM)

Dereje
Gudicha
(power
analysis in
LCM)

Margot
Sijssens-
Bennink
*(micro-
macro LCM)*



Daniel van
der Palm
(divisive
LCM)

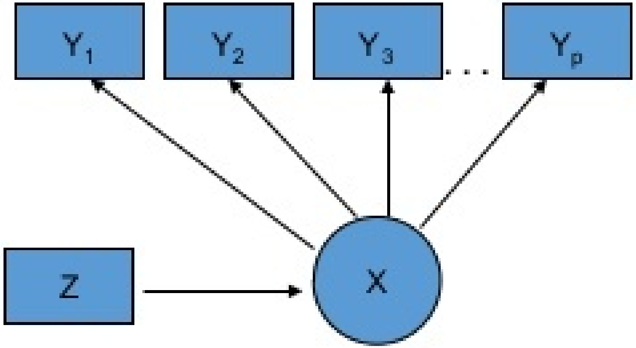
What is a latent class model?

Statistical model in which parameters of interest differ across unobserved subgroups (“latent classes”; “mixtures”)

Four main application types:

- Clustering (model based / probabilistic)
- Scaling (discretized IRT/factor analysis)
- Random-effects modelling (mixture regression / NP multilevel)
- Density estimation

The Latent Class Model



- Observed Continuous or Categorical Items
- Categorical Latent Class Variable (X)
- Continuous or Categorical Covariates (Z)

Adapted from: Nylund (2003) Latent class anlalysis in Mplus. URL: http://www.ats.ucla.edu/stat/mplus/seminars/lca/default.htm

Four main applications of LCM

- Clustering (model based / probabilistic)
- Scaling (discretized IRT/factor analysis)
- Random-effects modelling (mixture regression / nonparametric multilevel)
- Density estimation

Why would survey researchers need latent class models?

For substantive analysis:

- Creating typologies of respondents, e.g.:
    - McCutcheon 1989: tolerance,
    - Rudnev 2015: human values
    - Savage et al. 2013: “A new model of Social Class”
    - …
- Nonparametric multilevel model (Vermunt 2013)
- Longitudinal data analysis
    - Growth mixture models
    - Latent transition (“Hidden Markov”) models

Why would survey researchers need latent class models?

For survey methodology:

- As a method to evaluate questionnaires, e.g.
  - Biemer 2011: Latent Class Analysis of Survey Error
  - Oberski 2015: latent class MTMM
- Modeling extreme response style (and other styles), e.g.
  - Morren, Gelissen & Vermunt 2012: extreme response
- Measurement equivalence for comparing groups/countries
  - Kankaraš & Moors 2014: Equivalence of Solidarity Attitudes
- Identifying groups of respondents to target differently
  - Lugtig 2014: groups of people who drop out panel survey
- Flexible imputation method for multivariate categorical data
  - Van der Palm, Van der Ark & Vermunt

Latent class analysis at ESRA!

latent class — Search for session or paper

Search for paper author or session convenor

Paper(s)

- Apathy is the Enemy. A study of UK environmental concern and its complicated relationship with pro-environmental behaviour. (Rebecca Rhead)
- Aspects of Validity: Scenario-Technique, Self-Report & Social Desirability (Lena Verneuer)
- Developing a diagnostic tool for detecting response styles, and a demonstration of its use in comparative research of single item measurements (Eva Van vlimmeren)
- Elimination and Selection by aspects decision rules in discrete choice experiments (Seda Erdem)
- Measurement equivalence in cross-cultural surveys: multigroup latent class analysis and MIMIC-models in prejudice research (Ekaterina Lytkina)
- Policy-Culture Gaps and the Role of Gender Norms (Daniela Grunow)
- Testing the Invariance of the Value Typology of Europeans Across Time Points (Maksim Rudnev)
- Testing the Theory of Social Integration (Ashley Amaya)
- Validating Schwartz value theory with confirmatory latent class analysis (Marko Sömer)

Software

Commercial

- Latent GOLD
- Mplus
- gllamm in Stata
- PROC LCA in SAS

Free (as in beer)

- ℓem

Open source

- R package poLCA
- R package flexmix
- (with some programming) OpenMx, stan
- Specialized models: HiddenMarkov, depmixS4,

A small example

(showing the basic ideas and interpretation)

Small example: data from GSS 1987

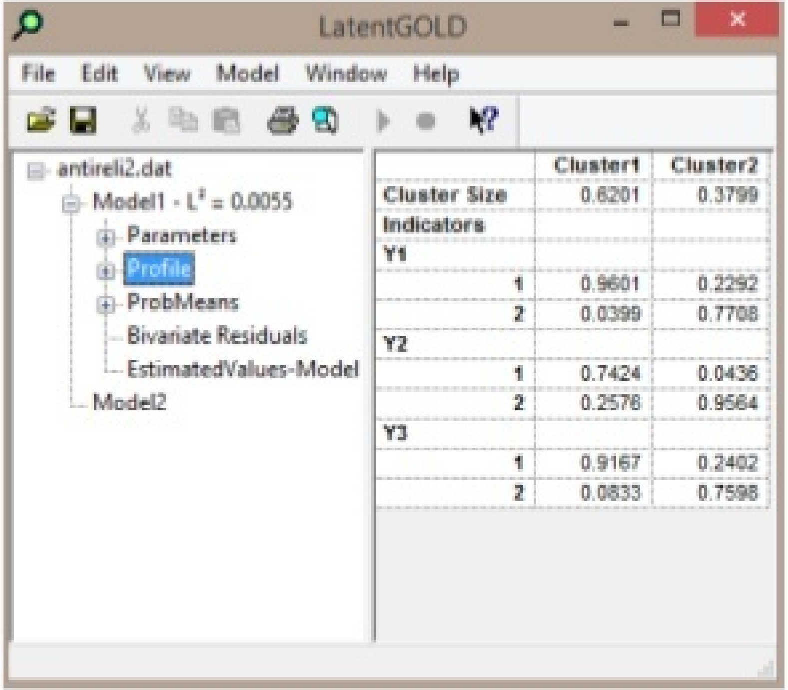
Y1: "allow anti-religionists to speak" (1 = allowed, 2 = not allowed),
Y2: "allow anti-religionists to teach" (1 = allowed, 2 = not allowed),
Y3: "remove anti-religious books from the library" (1 = do not remove, 2 = remove).

	Y1	Y2	Y3	Observed frequency (n)	Observed proportion (n/N)
	1	1	1	696	0.406
	1	1	2	68	0.040
	1	2	1	275	0.161
	1	2	2	130	0.076
	2	1	1	34	0.020
	2	1	2	19	0.011
	2	2	1	125	0.073
	2	2	2	366	0.214

N = 1713

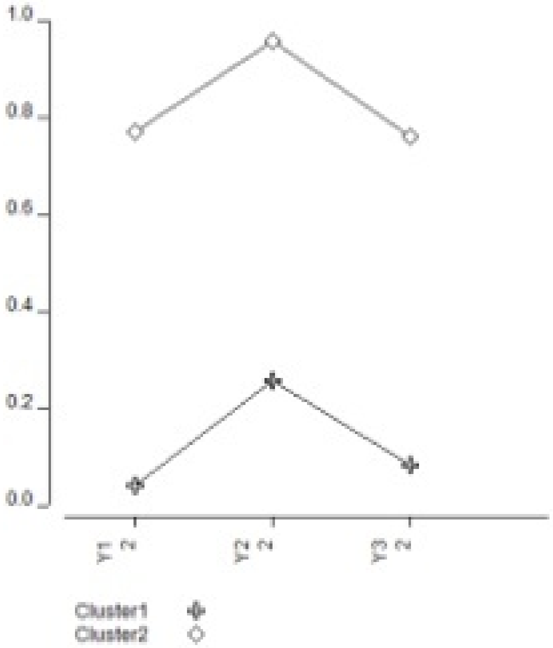
2-class model in Latent GOLD

Profile for 2-class model



	Cluster1	Cluster2
Cluster Size	0.6201	0.3799
Indicators		
Y1		
1	0.9601	0.2292
2	0.0399	0.7708
Y2		
1	0.7424	0.0436
2	0.2576	0.9564
Y3		
1	0.9167	0.2402
2	0.0833	0.7598

Profile plot for 2-class model



Estimating the 2-class model in R

```
antireli  <- read.csv("antireli_data.csv")

library(poLCA)

M2 <- poLCA(cbind(Y1, Y2, Y3)~1, data=antireli, nclass=2)
```

Profile for 2-class model

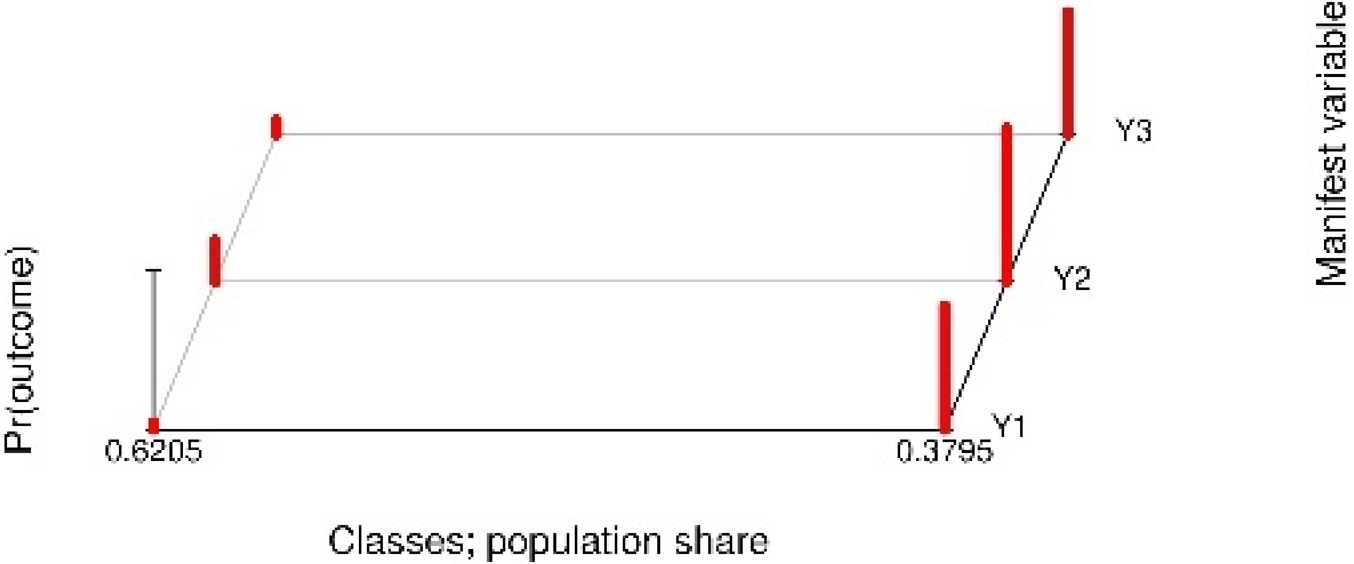
```
$Y1
           Pr(1)  Pr(2)
class 1:  0.9601 0.0399
class 2:  0.2284 0.7716

$Y2
           Pr(1)  Pr(2)
class 1:  0.7424 0.2576
class 2:  0.0429 0.9571

$Y3
           Pr(1)  Pr(2)
class 1:  0.9166 0.0834
class 2:  0.2395 0.7605

Estimated class population shares
 0.6205 0.3795
```

```
> plot(M2)
```



Y3
Y2
Y1
Manifest variables
Pr(outcome)
0.6205
0.3795
Classes; population share

Model equation
for 2-class LC model for 3 indicators

Model for

$$P(y_1, y_2, y_3)$$

the probability of a particular response pattern.

For example, how likely is someone to hold the opinion
“allow speak, allow teach, but remove books from library:

P(Y1=1, Y2=1, Y3=2) = ?

Two key model assumptions

(X is the latent class variable)

1. (MIXTURE ASSUMPTION)
Joint distribution mixture of 2 class-specific distributions:

$$P(y_1, y_2, y_3) = P(X=1)P(y_1, y_2, y_3 \mid X=1) + P(X=2)P(y_1, y_2, y_3 \mid X=2)$$

2. (LOCAL INDEPENDENCE ASSUMPTION)
Within class X=x, responses are independent:

$$P(y_1, y_2, y_3 \mid X=1) = P(y_1 \mid X=1)P(y_2 \mid X=1)P(y_3 \mid X=1)$$

$$P(y_1, y_2, y_3 \mid X=2) = P(y_1 \mid X=2)P(y_2 \mid X=2)P(y_3 \mid X=2)$$

Example: model-implied proprtion

	X=1	X=2
P(X)	0.620	0.380
P(Y1=1|X)	0.960	0.229
P(Y2=1|X)	0.742	0.044
P(Y3=1|X)	0.917	0.240

P(Y1=1, Y2=1, Y3=2) =

(Mixture assumption)

P(Y1=1, Y2=1, Y3=2 | X=1) P(X=1) +

P(Y1=1, Y2=1, Y3=2 | X=2) P(X=2)

Example: model-implied proprtion

	X=1	X=2
P(X)	0.620	0.380
P(Y1=1|X)	0.960	0.229
P(Y2=1|X)	0.742	0.044
P(Y3=1|X)	0.917	0.240

P(Y1=1, Y2=1, Y3=2) =

(Mixture assumption)

P(Y1=1, Y2=1, Y3=2 | X=1) 0.620 +

P(Y1=1, Y2=1, Y3=2 | X=2) 0.380 =

(Local independence assumption)

P(Y1=1|X=1) P(Y2=1|X=1) P(Y2=2|X=1) 0.620 +

P(Y1=1|X=2) P(Y2=1|X=2) P(Y2=2|X=2) 0.380

Example: model-implied proprtion

	X=1	X=2
P(X)	0.620	0.380
P(Y1=1|X)	0.960	0.229
P(Y2=1|X)	0.742	0.044
P(Y3=1|X)	0.917	0.240

P(Y1=1, Y2=1, Y3=2) =

(Mixture assumption)

$$P(Y1=1, Y2=1, Y3=2 \mid X=1)\ 0.620\ +$$
$$P(Y1=1, Y2=1, Y3=2 \mid X=2)\ 0.380\ =$$

(Local independence assumption)

$$(0.960)\ (0.742)\ (1-0.917)\ (0.620)\ +$$
$$(0.229)\ (0.044)\ (1-0.240)\ (0.380) \approx$$

$$\approx \mathbf{0.0396}$$

Small example: data from GSS 1987

Y1: "allow anti-religionists to speak" (1 = allowed, 2 = not allowed),
Y2: "allow anti-religionists to teach" (1 = allowed, 2 = not allowed),
Y3: "remove anti-religious books from the library" (1 = do not remove, 2 = remove).

	Y1	Y2	Y3	Observed frequency (n)	Observed proportion (n/N)
	1	1	1	696	0.406
	1	1	2	68	0.040
	1	2	1	275	0.161
	1	2	2	130	0.076
	2	1	1	34	0.020
	2	1	2	19	0.011
	2	2	1	125	0.073
	2	2	2	366	0.214

N = 1713

Implied is 0.0396, observed is 0.040.

More general model equation

Mixture of C classes

$$P(\mathbf{y}) = \sum_{x=1}^{C} P(X = x) P(\mathbf{y} \mid X = x)$$

Local independence of K variables

$$P(\mathbf{y} \mid X = x) = \prod_{k=1}^{K} P(y_k \mid X = x)$$

Both together gives the likelihood of the observed data:

$$P(\mathbf{y}) = \sum_{x=1}^{C} P(X = x) \prod_{k=1}^{K} P(y_k \mid X = x)$$

“Categorical data” notation

- In some literature an alternative notation is used

- Instead of Y1, Y2, Y3, variables are named A, B, C
- We define a model for the joint probability

$$P(A = i, B = j, C = k) := \pi_{ijk}^{ABC}$$

$$\pi_{ijk}^{ABC} = \sum_{t=1}^{T} \pi_t^X \pi_{ijkt}^{ABC|X} \quad \text{with} \quad \pi_{ijkt}^{ABC|X} = \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X}$$

Loglinear parameterization

$$\pi_{ijk\,t}^{ABC|X} = \pi_{i\,t}^{A|X}\pi_{j\,t}^{B|X}\pi_{k\,t}^{C|X}$$

$$\ln(\pi_{ijk\,t}^{ABC|X}) = \ln(\pi_{i\,t}^{A|X}) + \ln(\pi_{j\,t}^{B|X}) + \ln(\pi_{k\,t}^{C|X})$$
$$:= \lambda_{i\,t}^{A|X} + \lambda_{j\,t}^{B|X} + \lambda_{k\,t}^{C|X}$$

The parameterization actually used in most LCM software

$$P(y_k \mid X = x) = \frac{\exp(\beta_{0\,y_k}^{k} + \beta_{1\,y_k x}^{k})}{\sum_{m=1}^{M_k} \exp(\beta_{0m}^{k} + \beta_{1mx}^{k})}$$

$\beta_{0\,y_k}^{k}$ Is a logistic intercept parameter

$\beta_{1\,y_k x}^{k}$ Is a logistic slope parameter (loading)

So just a series of **logistic regressions**, with X as independent and Y dep't!
Similar to CFA/EFA (but logistic instead of linear regression)

A more realistic example

(showing how to evaluate the model fit)

One form of political activism

61.31%

38.69%

Another form of political activism

There are different ways of trying to improve things in [country] or help prevent[9] things from going wrong. During the last 12 months, have you done any of the following?
Have you…**READ OUT…**

		Yes	No	(Don't know)
B13	…contacted a politician, government or local government official?	1	2	8
B14	…worked in a political party or action group?	1	2	8
B15	…worked in another organisation or association?	1	2	8
B16	…worn or displayed a campaign badge/sticker?	1	2	8
B17	…signed a petition?	1	2	8
B18	…taken part in a lawful public demonstration?	1	2	8
B19	…boycotted certain products?	1	2	8

Data from the European Social Survey round 4 Greece

contplt	wrkprty	wrkorg	badge	sgnptit	pbldmn	bctprd	clsprty
2	2	2	2	2	2	1	2
2	2	2	2	2	2	1	1
2	2	2	2	2	1	1	1
2	2	2	2	2	2	2	1
2	2	2	2	2	2	2	1
2	2	2	2	2	2	1	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	1
2	2	2	2	2	2	2	2

```
library(foreign)
ess4gr <- read.spss("ESS4-GR.sav", to.data.frame = TRUE,
      use.value.labels = FALSE)


K <- 4      # Change to 1,2,3,4,..
MK <- poLCA(cbind(contplt, wrkprty, wrkorg,
                  badge, sgnptit, pbldmn, bctprd)~1,
                  ess4gr, nclass=K)
```

Evaluating model fit

In the previous small example you calculated the model-implied (expected) probability for response patterns and compared it with the observed probability of the response pattern:

observed - expected

The small example had $2^3 – 1= 7$ unique patterns and 7 unique parameters, so df = 0 and the model fit perfectly.

observed – expected = 0 <=> df = 0

Evaluating model fit

Current model (with 1 class, 2 classes, …)

Has $2^7 - 1 = 128 - 1 = 127$ unique response patterns

But much fewer parameters

So the model can be **tested**.

Different models can be compared with each other.

Evaluating model fit

- Global fit
- Local fit
- Substantive criteria

Global fit

Goodness-of-fit chi-squared statistics

- H0: model with C classes; H1: saturated model
- $L^2 = \sum 2\, n \ln (n / (P(y)*N))$
- $X^2 = \sum (n - P(y)*N)^2 / (P(y)*N)$
- df = number of patterns -1 - Npar
- Sparseness: bootstrap *p*-values

Information criteria

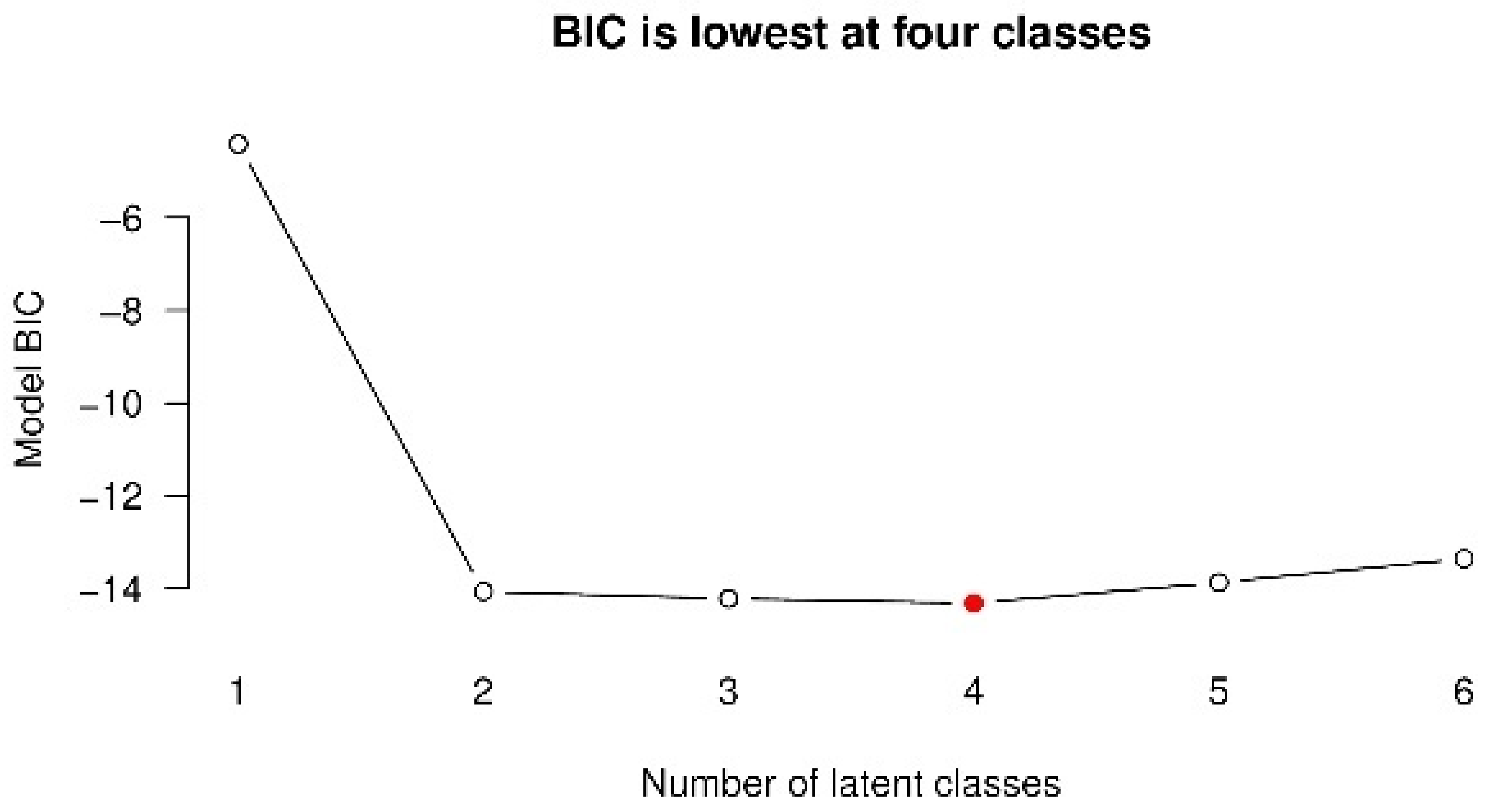
- for model comparison
- parsimony versus fit

Common criteria

- BIC(LL) = -2LL + ln(N) * Npar
- AIC(LL) = -2LL + 2 * Npar
- AIC3(LL) = -2LL + 3 * Npar
- BIC(L2) = L2 - ln(N) * df
- AIC(L2) = L2 - 2 * df
- AIC3(L2) = L2 - 3 * df

Model fit comparisons

	L^2	BIC(L^2)	AIC(L^2)	df	p-value
1-Cluster	1323.0	-441.0	861.0	120	0.000
2-Cluster	295.8	-1407.1	-150.2	112	0.001
3-Cluster	219.5	-1422.3	-210.5	104	0.400
4-Cluster	148.6	-1432.2	-265.4	96	1.000
5-Cluster	132.0	-1387.6	-266.0	88	1.000
6-Cluster	122.4	-1336.1	-259.6	80	1.000


BIC is lowest at four classes
Model BIC
−6
−8
−10
−12
−14
1
2
3
4
5
6
Number of latent classes

Local fit

Local fit: bivariate residuals (BVR)

Pearson “chi-squared” comparing observed and estimated frequencies in 2-way tables.

Expected frequency in two-way table:

$$N \cdot P(y_k, y_{k'}) = N \cdot \sum_{x=1}^{C} P(X = x) P(y_k \mid X = x) P(y_{k'} \mid X = x)$$

Observed:

Just make the bivariate cross-table from the data!

Example calculating a BVR

Observed

	No	Yes
No	3250	280
Yes	123	216

Expected

	No	Yes
No	3217	313
Yes	156	183

Bivariate residuals

	No	Yes
No	32.6	-32.6
Yes	-32.6	32.6

$$\mathrm{BVR}_{1,3} = {r_{11}}^2 \sum_{k,l} \hat{\mu}_{kl}^{-1} = (32.6)^2 \sum_{k,l} \hat{\mu}_{kl}^{-1} \approx 1063(0.0154) \approx 16.3$$

1-class model BVR’s

	contplt	wrkprty	wrkorg	badge	sgnptit	pbldmn	bctprd
contplt	.						
wrkprty	342.806	.					
wrkorg	133.128	312.592	.				
badge	203.135	539.458	396.951	.			
sgnptit	82.030	152.415	372.817	166.761	.		
pbldmn	77.461	260.367	155.346	219.380	272.216	.	
bctprd	37.227	56.281	78.268	65.936	224.035	120.367	.

2-class model BVR's

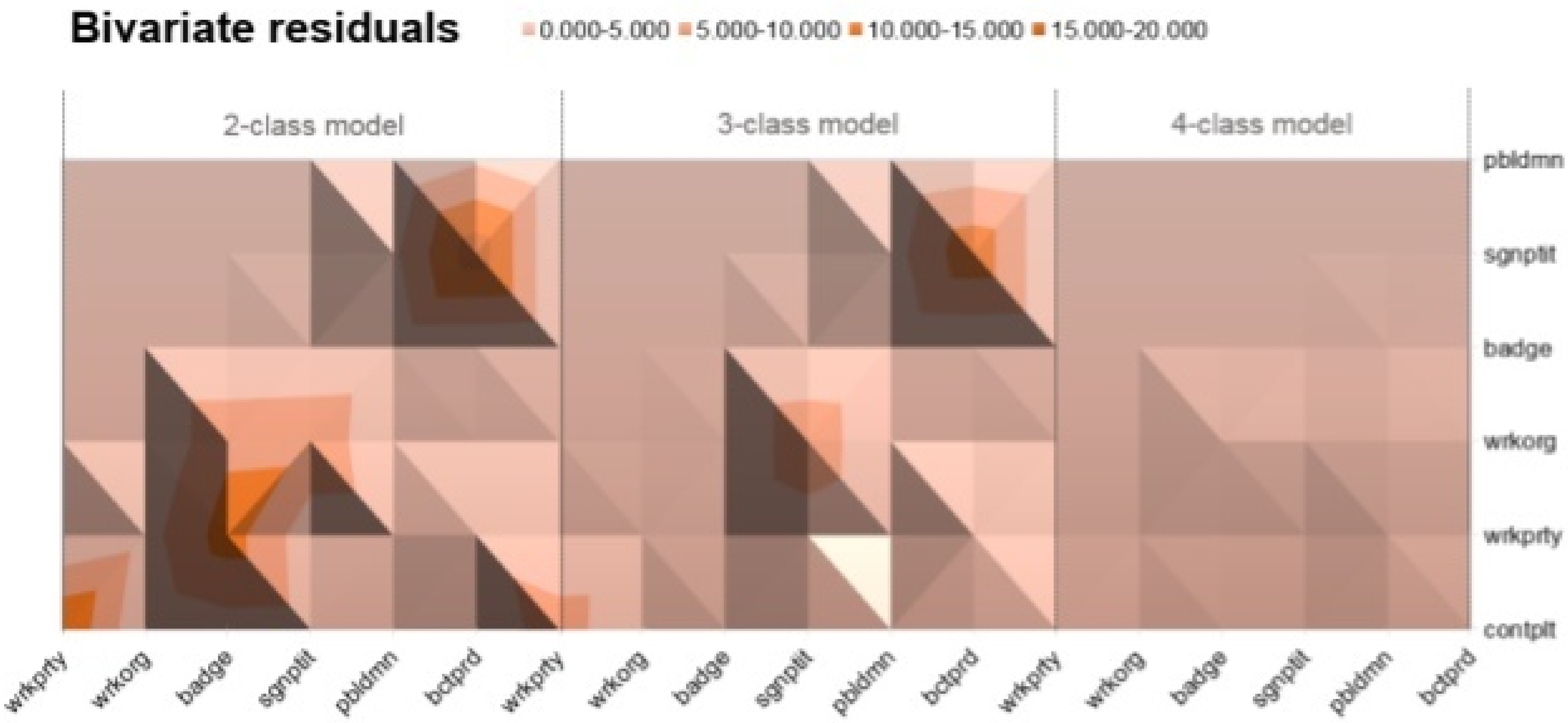
	contplt	wrkprty	wrkorg	badge	sgnptit	pbldmn	bctprd
contplt	.						
wrkprty	15.147	.					
wrkorg	0.329	2.891	.				
badge	2.788	12.386	8.852	.			
sgnptit	2.402	1.889	9.110	0.461	.		
pbldmn	1.064	1.608	0.108	0.945	3.957	.	
bctprd	1.122	2.847	0.059	0.717	18.025	4.117	.

3-class model BVR's

	contplt	wrkprty	wrkorg	badge	sgnptit	pbldmn	bctprd
contplt	.						
wrkprty	7.685	.					
wrkorg	0.048	0.370	.				
badge	0.282	0.054	0.273	.			
sgnptit	2.389	2.495	8.326	0.711	.		
pbldmn	2.691	0.002	0.404	0.086	2.842	.	
bctprd	2.157	2.955	0.022	0.417	13.531	1.588	.

4-class model BVR's

	contplt	wrkprty	wrkorg	badge	sgnptit	pbldmn	bctprd
contplt	.						
wrkprty	0.659	.					
wrkorg	0.083	0.015	.				
badge	0.375	0.001	1.028	.			
sgnptit	0.328	0.107	0.753	0.019	.		
pbldmn	0.674	0.939	0.955	0.195	0.004	.	
bctprd	0.077	0.011	0.830	0.043	0.040	0.068	.


Bivariate residuals
0.000-5.000
5.000-10.000
10.000-15.000
15.000-20.000
2-class model
3-class model
4-class model
pbldmn
sgnptit
badge
wrkorg
wrkprty
contplt
wrkprty
wrkorg
badge
sgnptit
pbldmn
bctprd
wrkprty
wrkorg
badge
sgnptit
pbldmn
bctprd
wrkprty
wrkorg
badge
sgnptit
pbldmn
bctprd

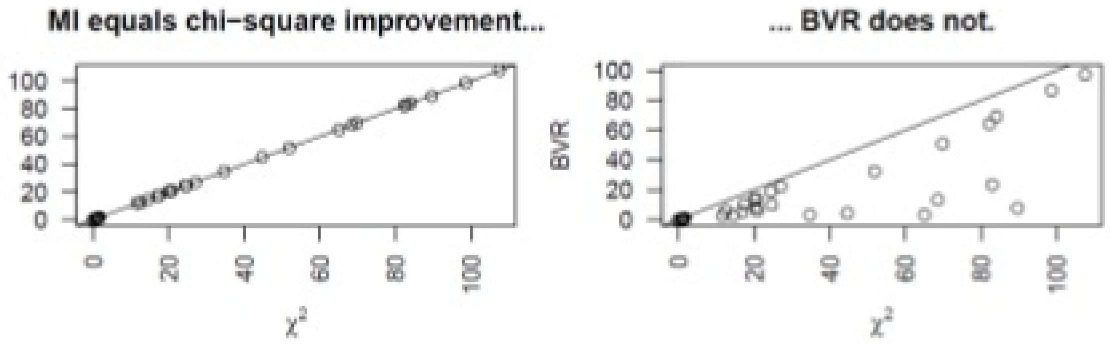
Local fit: beyond BVR

The bivariate residual (BVR) is not actually chi-square distributed!

(Oberski, Van Kollenburg & Vermunt 2013)

Solutions:

- Bootstrap p-values of BVR (LG5)
- “Modification indices” (score test) (LG5)

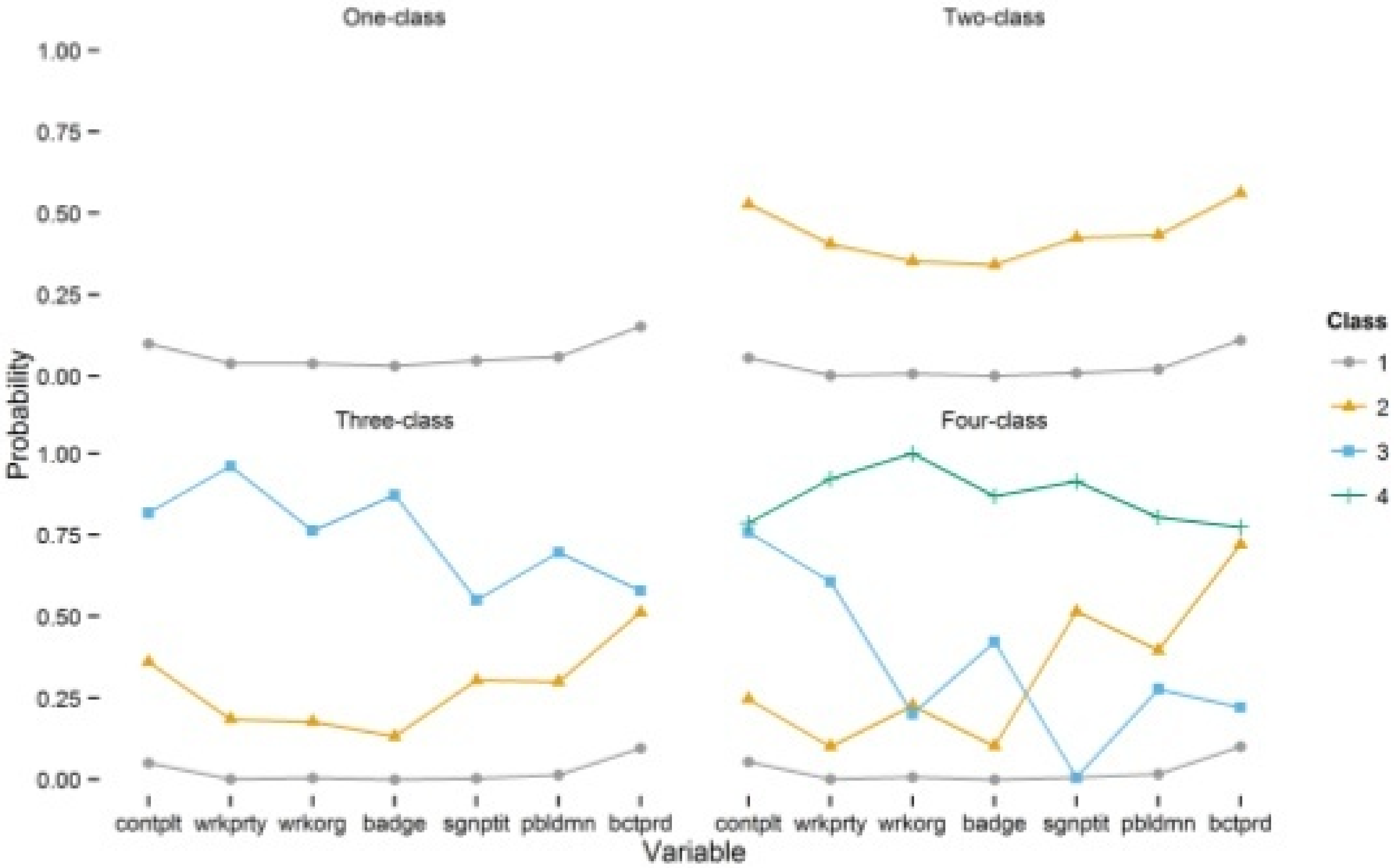


Example of modification index (score test) for 2-class model

Covariances / Associations							
term			coef	EPC(self)	Score	df	BVR
contplt	<->	wrkprty	0	1.7329	28.5055	1	15.147
wrkorg	<->	wrkprty	0	0.6927	**4.3534**	1	**2.891**
badge	<->	wrkprty	0	1.3727	16.7904	1	12.386
sgnptit	<->	bctprd	0	1.8613	37.0492	1	18.025

wrkorg <-> wrkparty is “not significant” according to BVR but is when looking at score test!
(but not after adjusting for multiple testing)

Interpreting the results and using substantive criteria


One-class
Two-class
Three-class
Four-class
1.00
0.75
0.50
0.25
0.00
Probability
contplt
wrkprty
wrkorg
badge
sgnptit
pbldmn
bctprd
Variable
Class
1
2
3
4

EPC-interest for looking at change in substantive parameters

After fitting two-class model, how much would loglinear “loadings” of the items change if local dependence is accounted for?

term			Y1	Y2	Y3	Y4	Y5	Y6	Y7
contplt	<->	wrkprty	-0.44	-0.66	0.05	1.94	0.05	0.02	0.00
wrkorg	<->	wrkprty	0.00	-0.19	-0.19	0.63	0.02	0.01	0.00
badge	<->	wrkprty	0.00	-0.37	0.03	-1.34	0.03	0.01	0.00
sgnptit	<->	bctprd	0.01	0.18	0.05	1.85	-0.58	0.02	-0.48

See Oberski (2013); Oberski & Vermunt (2013); Oberski, Moors & Vermunt (2015)

Model fit evaluation: summary

Different types of criteria to evaluate fit of a latent class model:

- **Global**
  BIC, AIC, L2, X2

- **Local**
  Bivariate residuals, modification indices (score tests), and expected parameter changes (EPC)

- **Substantive**
  Change in the solution when adding another class or parameters

Model fit evaluation: summary

- Compare models with different number of classes using BIC, AIC, bootstrapped L2
- Evaluate overall fit using bootstrapped L2 and bivariate residuals
- Can be useful to look at the profile of the different solutions: if nothing much changes, or very small classes result, fit may not be as useful

Classification

(Putting people into boxes, while admitting uncertainty)

Classification

- After estimating a LC model, we may wish to classify individuals into latent classes

- The latent classification or **posterior** class membership probabilities $P(X = x \mid \mathbf{y})$ can be obtained from the LC model parameters using Bayes' rule:

$$P(X = x \mid \mathbf{y}) = \frac{P(X = x)P(\mathbf{y} \mid X = x)}{P(\mathbf{y})} = \frac{P(X = x)\prod_{k=1}^{K} P(y_k \mid X = x)}{\sum_{c=1}^{C} P(X = c)\prod_{k=1}^{K} P(y_k \mid X = c)}$$

Small example: posterior classification

Y1	Y2	Y3	P(X=1 | Y)	P(X=2 | Y)	Most likely (but not sure!)
1	1	1	0.002	0.998	2
1	1	2	0.071	0.929	2
1	2	1	0.124	0.876	2
1	2	2	0.832	0.169	1
2	1	1	0.152	0.848	2
2	1	2	0.862	0.138	1
2	2	1	0.920	0.080	1
2	2	2	0.998	0.003	1

Classification quality

Classification Statistics

- classification table: true vs. assigned class
- overall proportion of classification errors

Other reduction of "prediction" errors measures

- How much more do we know about latent class membership after seeing the responses?
- Comparison of $P(X=x)$ with $P(X=x \mid \mathbf{Y}=\mathbf{y})$
- R-squared-like reduction of prediction (of X) error

```
posteriors <- data.frame(M4$posterior, predclass=M4$predclass)

classification_table <-
    ddply(posteriors, .(predclass), function(x) colSums(x[,1:4]))

> round(classification_table, 1)
  predclass post.1 post.2 post.3 post.4
1         1 1824.0   34.9    0.0   11.1
2         2    7.5   87.4    1.1    3.0
3         3    0.0    1.0   19.8    0.2
4         4    4.0    8.6    1.4   60.1
```

Classification table for 4-class

	post.1	post.2	post.3	post.4
1	**0.99**	0.26	0.00	0.15
2	0.00	**0.66**	0.05	0.04
3	0.00	0.01	**0.89**	0.00
4	0.00	0.07	0.06	**0.81**
	1	1	1	**1**

Total classification errors:

```
> 1 - sum(diag(classification_table)) / sum(classification_table)
[1] 0.0352
```

Entropy R^2

```
entropy <- function(p) sum(-p * log(p))
error_prior <- entropy(M4$P) # Class proportions
error_post <- mean(apply(M4$posterior, 1, entropy))

R2_entropy  <- (error_prior - error_post) / error_prior
```

```
> R2_entropy
[1] 0.741
```

This means that we know a lot more about people's political participation class after they answer the questionnaire.

Compared with if we only knew the overall proportions of people in each class

Classify-analyze does not work!

- You might think that after classification it is easy to model people's latent class membership
- "Just take assigned class and run a multinomial logistic regression"
- Unfortunately, this **does not work** (biased estimates and wrong se's) *(Bolck, Croon & Hagenaars 2002)*
- (Many authors have fallen into this trap!)
- Solution is to **model class membership and LCM simulaneously**
- (Alternative is 3-step analysis, not discussed here)

Predicting latent class membership

(using covariates; concomitant variables)

Fitting a LCM in poLCA with gender as a covariate

```
M4 <- poLCA(
     cbind(contplt, wrkprty, wrkorg,
          badge, sgnptit, pbldmn, bctprd) ~ gndr,
          data=gr, nclass = 4, nrep=20)
```

This gives a **multinomial logistic regression** with X as dependent and gender as independent (“concomitant”; “covariate”)

Predicting latent class membership from a covariate

$$P(X = x \mid Z = z) = \frac{\exp(\gamma_{0x} + \gamma_{zx})}{\sum_{c=1}^{C} \exp(\gamma_{0c} + \gamma_{zc})}$$

γ_{0x} Is the logistic intercept for category x of the latent class variable X

γ_{zx} Is the logistic slope predicting membership of class x for value z of the covariate Z

```
=========================================================
Fit for 4 latent classes:
=========================================================
2 / 1
             Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)     -0.35987     0.37146   -0.969     0.335
gndrFemale      -0.34060     0.39823   -0.855     0.395
=========================================================
3 / 1
             Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)      2.53665     0.21894   11.586     0.000
gndrFemale       0.21731     0.24789    0.877     0.383
=========================================================
4 / 1
             Coefficient  Std. error  t value  Pr(>|t|)
(Intercept)     -1.57293     0.39237   -4.009     0.000
gndrFemale      -0.42065     0.57341   -0.734     0.465
=========================================================
```

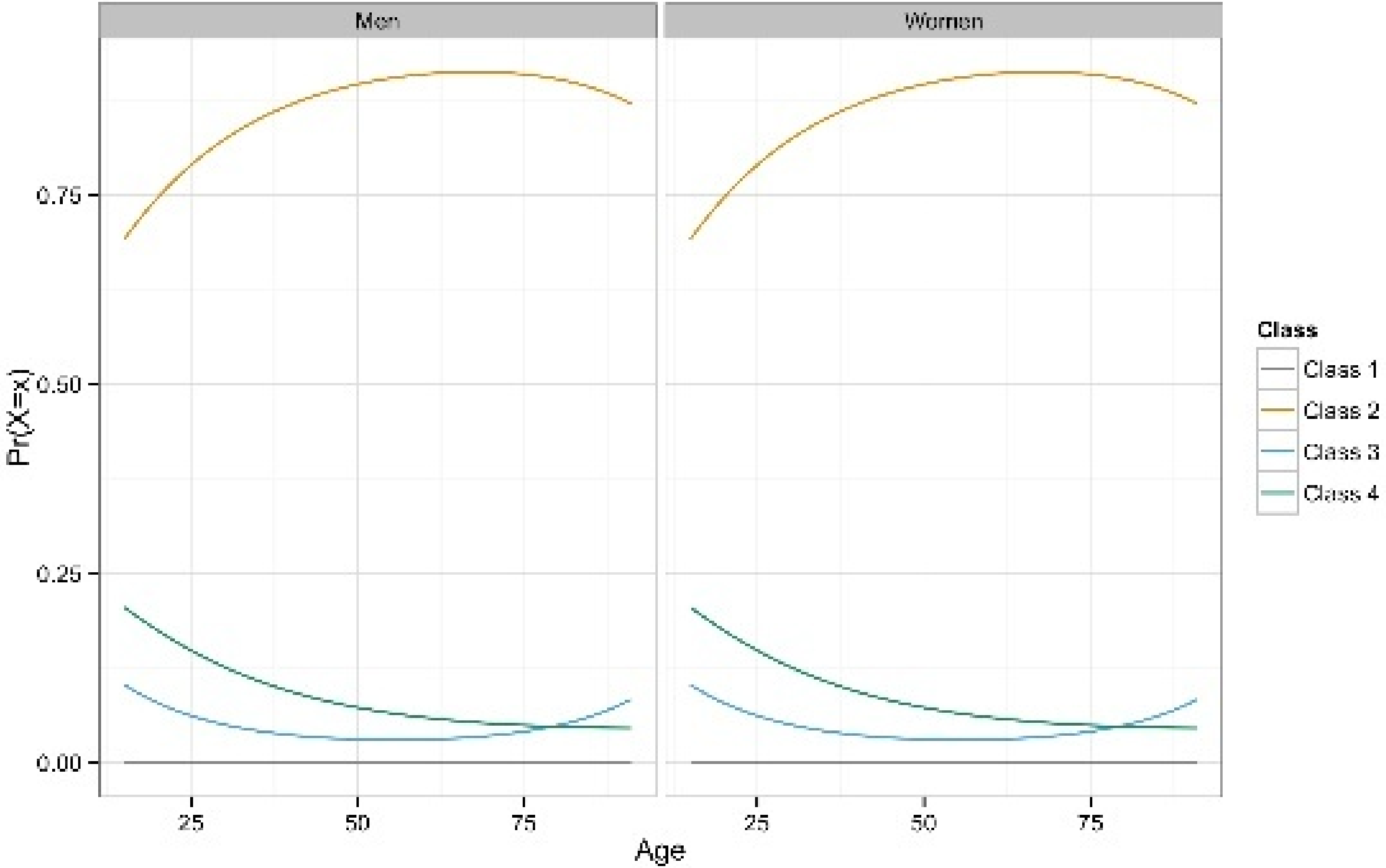
Class 1 Modern political participation
Class 2 Traditional political participation
Class 3 No political participation
Class 4 Every kind of political participation

Women more likely than men to be in classes 1 and 3
Less likely to be in classes 2 and 4

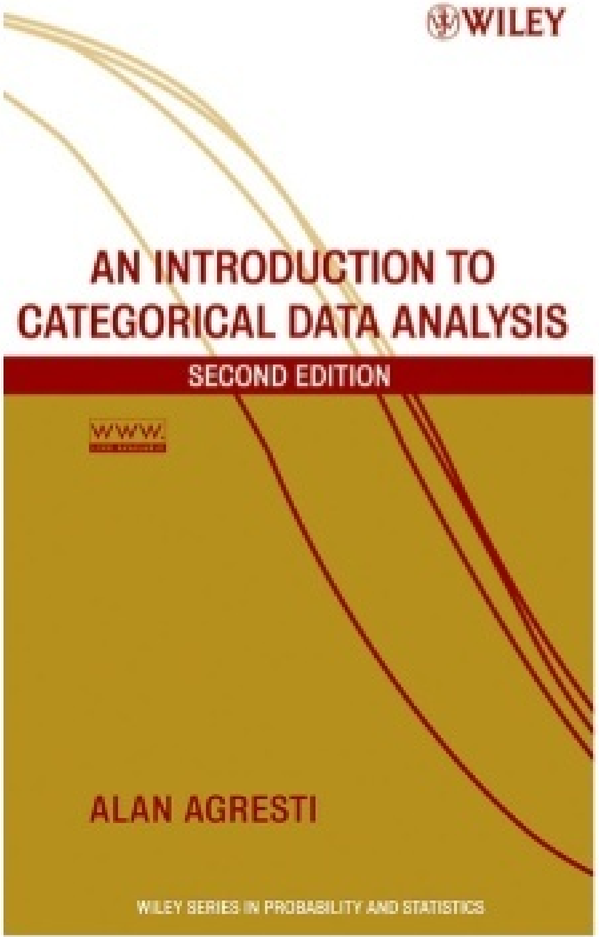
Multinomial logistic regression refresher

For example:

- Logistic multinomial regression coefficient equals -0.3406
- Then log odds ratio of being in class 2 (compared with reference class 1) is -0.3406 smaller for women than for men
- So odds ratio is smaller by a factor exp(-0.3406) = 0.71
- So odds are 30% smaller for women


Men
Women
0.75
0.50
0.25
0.00
Pr(X=x)
25
50
75
25
50
75
Age
Class
Class 1
Class 2
Class 3
Class 4

Even more (re)freshing:



Problems you will encounter when doing latent class analysis

(and some solutions)

Some problems

- Local maxima
- Boundary solutions
- Non-identification

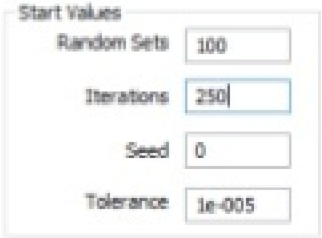
Problem: Local maxima

Problem: there may be different sets of “ML” parameter estimates with different L-squared values we want the solution with lowest L-squared (highest log-likelihood)

Solution: multiple sets of starting values

```
poLCA(cbind(Y1, Y2, Y3)~1, antireli, nclass=2, nrep=100)


Model 1: llik = -3199.02 ... best llik = -3199.02
Model 2: llik = -3359.311 ... best llik = -3199.02
Model 3: llik = -2847.671 ... best llik = -2847.671
Model 4: llik = -2775.077 ... best llik = -2775.077
Model 5: llik = -2810.694 ... best llik = -2775.077
....
```



Problem: boundary solutions

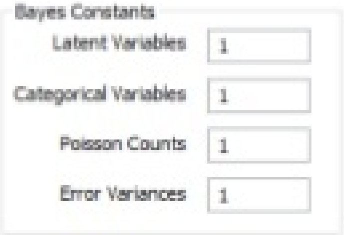
Problem: estimated probability becomes zero/one, or logit parameters extremely large negative/positive

Example:

```
$badge
            Pr(1)  Pr(2)
class 1:   0.8640 0.1360
class 2:   0.1021 0.8979
class 3:   0.4204 0.5796
class 4:   0.0000 1.0000
```

Solutions:

1. Not really a problem, just ignore it;
2. Use priors to smooth the estimates
3. Fix the offending probabilities to zero (classical)



Problem: non-identification

- Different sets of parameter estimates yield the same value of L-squared and LL value: estimates are not unique
- Necessary condition DF>=0, but not sufficient
- Detection: running the model with different sets of starting values or, formally, checking whether rank of the Jacobian matrix equals the number of free parameters
- “Well-known” example: 3-cluster model for 4 dichotomous indicators

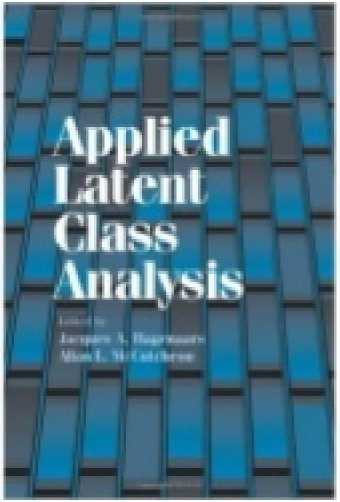
What we did not cover

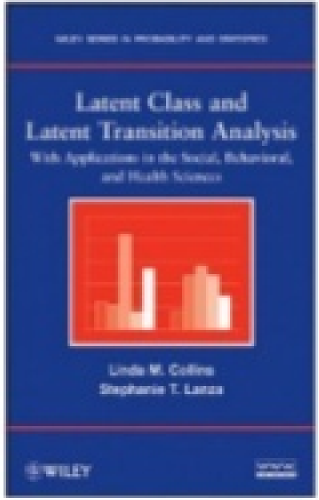
- 1 step versus 3 step modeling
- Ordinal, continuous, mixed type indicators
- Hidden Markov (“latent transition”) models
- Mixture regression

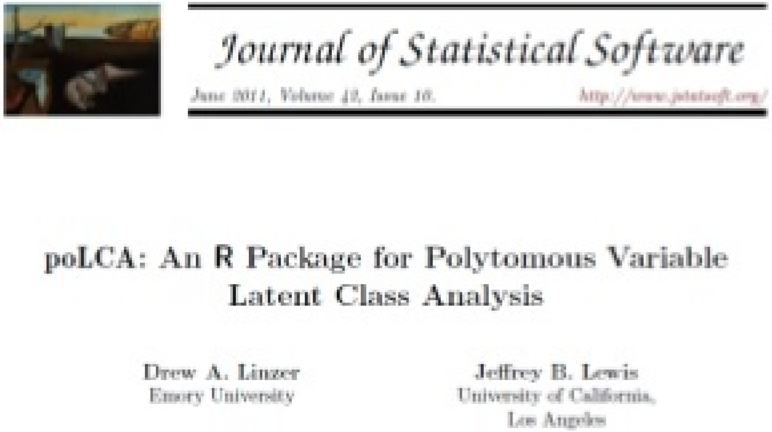
What we did cover

- Latent class “cluster” analysis
- Model formulation, different parameterizations
- Model interpretation, profile
- Model fit evaluation: global, local, and substantive
- Classification
- Common problems with LCM and their solutions

Further study






Journal of Statistical Software

June 2011, Volume 42, Issue 10. http://www.jstatsoft.org/

poLCA: An R Package for Polytomous Variable Latent Class Analysis

Drew A. Linzer
Emory University

Jeffrey B. Lewis
University of California, Los Angeles

Thank you for your attention!

@DanielOberski

http://daob.nl

doberski@uvt.nl