Latent Class Analysis

Lecture 11

March 16, 2006

Clustering and Classification

Today's Lecture

- Latent Class Analysis (LCA).
- LCA as a specific case of a Finite Mixture Model.
- How to do LCA.

LCA Introduction

- From here out, we will consider clusters to be synonymous with classes.
- Latent class models are commonly attributed to Lazarsfeld and Henry (1968).
- Like K-means and hierarchical clustering techniques, the final number of latent classes is not usually predetermined prior to analysis with latent class models.
  - The number of classes is determined through comparison of posterior fit statistics.
  - The characteristics of each class is also determined following the analysis.

Variable Types Used in LCA

- As it was originally conceived, LCA is an analysis that uses:
  - A set of binary-outcome variables - values coded as zero or one. Examples include:
    - Test items - scored correct or incorrect.
    - True/false questions.
    - Gender.
    - Anything else that has two possible outcomes.
- The number of classes (an integer ranging from two through...) must be specified prior to analysis.

LCA Process

- For a specified number of classes, LCA attempts to:
  - For each class, estimate the probability that each variable is equal to one.
  - Estimate the probability that each observation falls into each class.
    - For each observation, the sum of these probabilities across classes equals one.
    - This is different from K-means where an observation is a member of a class with certainty.
  - Across all observations, estimate the probability that *any* observation falls into a class.

LCA Estimation

- Estimation in LCA is more complicated than in previous methods discussed in this course.
  - In agglomerative hierarchical clustering, a search process was used with new distance matrices being created for each step.
  - K-means used more of a brute-force approach - trying multiple starting points.
  - Both methods relied on distance metrics to find clustering solutions.
- LCA estimation uses distributional assumptions to find classes.
- The distributional assumptions provide the measure of "distance" in LCA.

LCA Distributional Assumptions

- Because (for today) we have discussed LCA with binary-outcome variables, the distributional assumptions of LCA must use a binary-outcome distribution.
- Within each latent class, the variables are assumed to:
  - Be independent.
  - (Marginally) be distributed as Bernoulli:
    - The Bernoulli distribution states:

$$f(x_i) = (\pi_i)^{x_i} (1 - \pi_i)^{(1-x_i)}$$

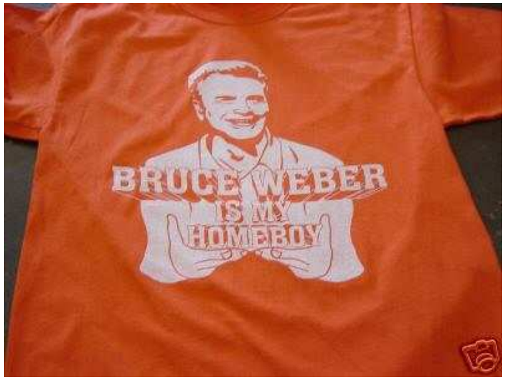
- The Bernoulli distribution is a simple distribution for a single event - like flipping a coin.

Bernoulli Distribution Illustration

- To illustrate the Bernoulli distribution (and statistical likelihoods in general, consider the following example.

- As you may or may not know, today is a big day in collegiate athletics.
  - The NCAA Mens Basketball Tournament has now started.
  - It is killing me to be here at this very moment.
  - The match up I am most concerned with is tonight at 6:30:

Illinois v. Air Force



Bernoulli Distribution Illustration

- For fun, I like to try to forecast the results of sporting events using statistical data on past performances of the teams.
  - You can see my predictions for today's games on my website, `http://www.people.ku.edu/~jtemplin/sports`.
- To illustrate the Bernoulli distribution, consider the Illinois/Air Force game as a binary-response item.
  - Lets say $X =$ if Illinois wins, and $X = 0$ otherwise.
  - My prediction is that Illinois has about an 0.87 percent chance of winning the game.
  - So, $\pi = 0.87$.
- Likewise, $P(X = 1) = 0.87$ and $P(X = 0) = 0.13$.

Bernoulli Distribution Illustration

- The likelihood function for X looks similar:
- If $X = 1$, the likelihood is:

$$f(x_i = 1) = (0.87)^1 (1 - 0.87)^{(1-1)} = 0.87$$

- If $X = 0$, the likelihood is:

$$f(x_i = 0) = (0.87)^1 (1 - 0.87)^{(1-0)} = 0.13$$

- This example shows you how the likelihood function of a statistical distribution gives you the likelihood of an event occurring.
- In the case of discrete-outcome variables, the likelihood of an event is synonymous with the probability of the event occurring.

Independent Bernoulli Variables

- Although it makes little sense for this example, consider if we were to play this game over and over again, with each trial being independent.

- The probability of observing two Illinois wins would be the product of the probability of having a single Illinois win.

$$P(X_1 = 1, X_2 = 1) = \pi_1 \pi_2 = 0.87 \times 0.87 = 0.76$$

- We can think about this as we would flipping a coin.
  - The probability of observing two heads in two flips of a "fair" coin is $0.5 \times 0.5 = 0.25$.

Finite Mixture Models

- Recall from last time that we stated that a finite mixture model expresses the distribution of **X** as a function of the sum of weighted distribution likelihoods:

$$f(\mathbf{X}) = \sum_{g=1}^{G} \eta_g f(\mathbf{X}|g)$$

- We are now ready to construct the LCA model likelihood.
- Here, we say that the conditional distribution of **X** given g is a sequence of independent Bernoulli variables.

Latent Class Analysis as a FMM

Using some notation of Bartholomew and Knott, a latent class model for the response vector of p variables ($i = 1, \ldots, p$) with K classes ($j = 1, \ldots, K$):

$$f(\mathbf{x}_i) = \sum_{j=1}^{K} \eta_j \prod_{i=1}^{p} \pi_{ij}^{x_i} \left(1 - \pi_{ij}\right)^{1-x_i}$$

- η_j is the probability that any individual is a member of class j (must sum to one).
- x_i is the observed response to item i.
- π_{ij} is the probability of a positive response to item i from an individual from class j.

LCA Example

- To illustrate the process of LCA, consider the example presented in Bartholomew and Knott (p. 142).

- The data are from a four-item test analyzed with an LCA by Macready and Dayton (1977).

Table 6.3: Macready and Dayton's (1977) data with posterior probabilities of belonging to the mastery state

Response pattern	Frequency	Expected frequency	$P\{\text{Master}\mid\mathbf{x}\}$
1111	15	15.0	1.00
1110	7	6.2	1.00
1101	23	19.7	1.00
1100	7	8.9	0.91
1011	1	4.2	1.00
1010	3	1.9	0.90
1001	6	6.1	0.90
1000	13	12.9	0.18
0111	4	4.9	1.00
0110	2	2.1	0.97
0101	5	6.6	0.98
0100	6	5.6	0.47
0011	4	1.4	0.97
0010	1	1.3	0.44
0001	4	4.0	0.45
0000	41	41.0	0.02
	142	142	

LCA Example

- Recall, we have three pieces of information we can gain from an LCA:
  - Sample information - proportion of people in each class.
  - Item information - probability of correct response for each item from examinees from each class.
  - Examinee information - posterior probability of class membership for each examinee in each class.
- Here, we will look at output from a program I wrote to do LCA (you will get a copy next time).

```
Class Probabilities
Class   Probability
1   0.586523874
2   0.413476126

Item Parameters
class:         1
item    prob    ase(prob)
1   0.75345 0.05125
2   0.78029 0.05108
3   0.43161 0.05625
4   0.70757 0.05445

class:         2
item    prob    ase(prob)
1   0.20861 0.06047
2   0.06834 0.04848
3   0.01793 0.02949
4   0.05228 0.04376
```

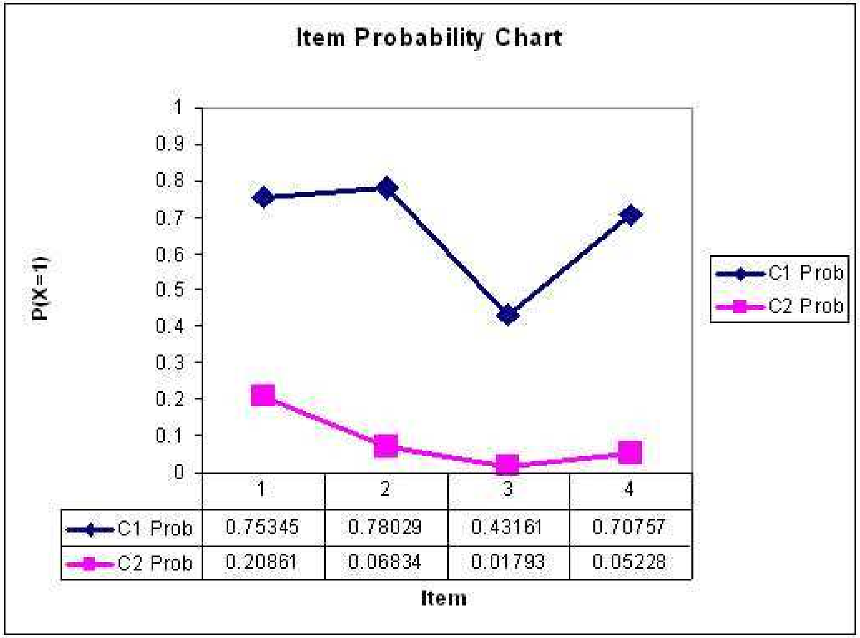
Examinee Information

Posterior Probabilities

examinee	1	2	ML Class	Max Pro
73	0.17747	0.82253	2	0.82253
74	0.17747	0.82253	2	0.82253
75	0.17747	0.82253	2	0.82253
76	0.99939	0.00061	1	0.99939
77	0.99939	0.00061	1	0.99939
78	0.99939	0.00061	1	0.99939
79	0.99939	0.00061	1	0.99939
80	0.974	0.026	1	0.974
81	0.974	0.026	1	0.974
82	0.97532	0.02468	1	0.97532
83	0.97532	0.02468	1	0.97532
84	0.97532	0.02468	1	0.97532
85	0.97532	0.02468	1	0.97532
86	0.97532	0.02468	1	0.97532
87	0.47399	0.52601	2	0.52601
88	0.47399	0.52601	2	0.52601

Interpreting Classes

- After the analysis is finished, we need to examine the item probabilities to gain information about the characteristics of the classes.



- An easy way to do this is to look at a chart of the item response probabilities by class.
- Here, we would say that Class 1 represents students who have mastered the material on the test.
- We would say that Class 2 represents students who have not mastered the material on the test.

How Many Classes?

- For our second example, we will consider data from a large, standardized test.
  - A total of 38 items.
  - A sample of 2,952 examinees.
- Test originally developed to measure reading ability on a single latent continuum.
- Imagine we wanted to determine groups of examinees based on their performance on this test.

How Many Classes?

- In a latent class analysis, the first question asked is “how many classes are needed to describe my data?”

- The answer to this comes from fitting models with increasing numbers of classes and examining the relative fit of the models.

- An index of fit sometimes used is the BIC (lowest is best - we will revisit this next class):

Classes	BIC
1	130,990.893
2	117,824.327
3	115,401.287
4	115,072.590
5	115,152.977

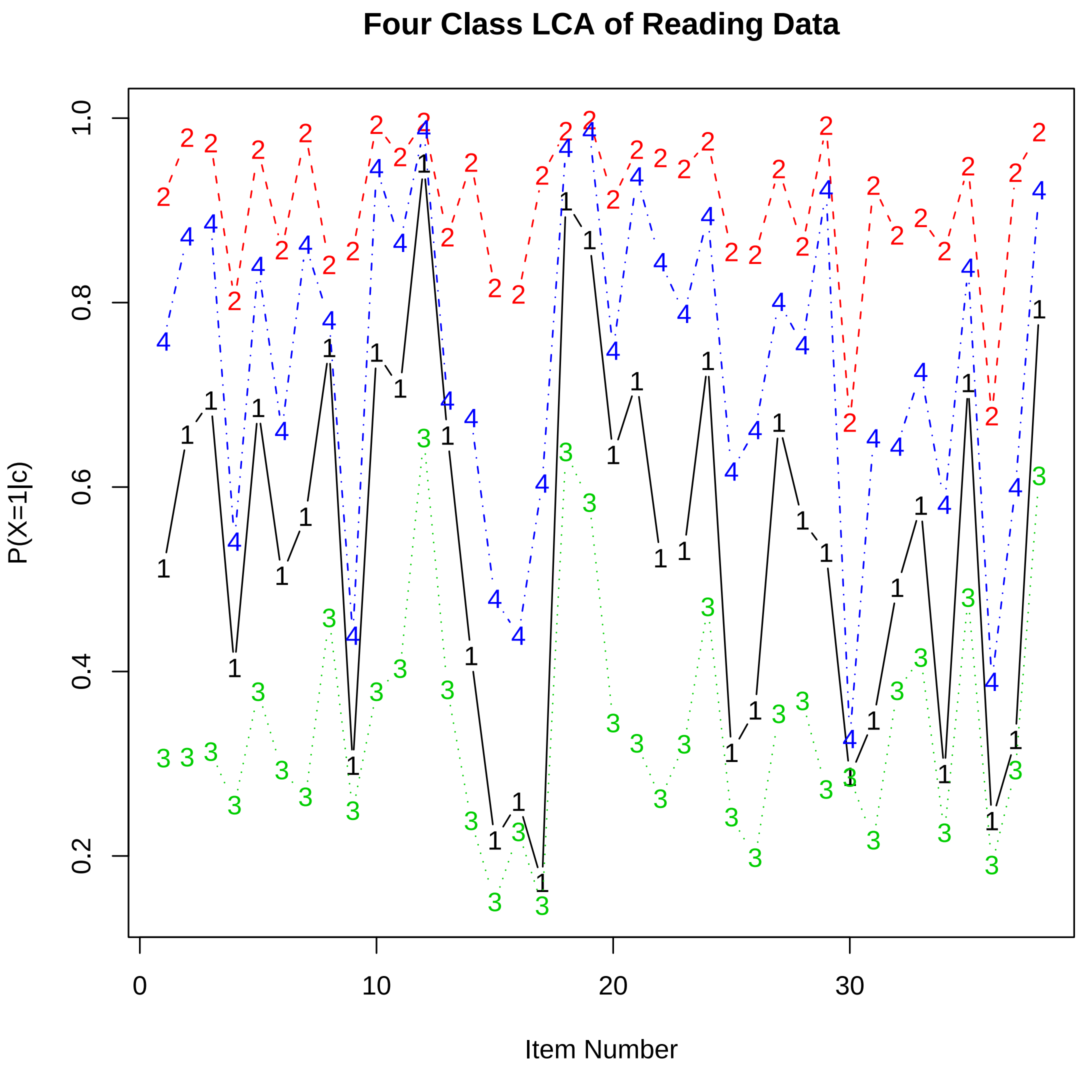
- For this application, the four-class model is considered the best fitting by the BIC.

Four Class Solution

Estimated Class Membership Probabilities:

c	$P(c)$
1	0.263
2	0.255
3	0.134
4	0.348

Estimated Item Response Probabilities:



LCA Local Independence

- LCA has the property of local independence - that given class, item responses are independent.

- To give an example, consider Items 1 and 2:

$$P(X_j = 1|c)$$

Item j	$c = 1$	$c = 2$	$c = 3$	$c = 4$
1	0.512	0.915	0.306	0.758
2	0.657	0.979	0.307	0.872

	Item 1		
Item 2	0	1	Marginal
0	0.164	0.139	0.343
1	0.321	0.336	0.657
Marginal	0.475	0.512	1.000

LCA Local Independence

- LCA has the property of local independence - that given class, item responses are independent.
- To give an example, consider Items 1 and 2:

	$P(X_j = 1\|c)$			
Item j	$c = 1$	$c = 2$	$c = 3$	$c = 4$
1	0.512	0.915	0.306	0.758
2	0.657	0.979	0.307	0.872

	Item 1		
Item 2	0	1	Marginal
0	0.002	0.019	0.021
1	0.083	0.896	0.979
Marginal	0.085	0.915	1.000

LCA Local Independence

- LCA has the property of local independence - that given class, item responses are independent.

- To give an example, consider Items 1 and 2:

	$P(X_j = 1 \mid c)$			
Item j	$c = 1$	$c = 2$	$c = 3$	$c = 4$
1	0.512	0.915	0.306	0.758
2	0.657	0.979	0.307	0.872

	Item 1		
Item 2	0	1	Marginal
0	0.481	0.212	0.693
1	0.213	0.094	0.307
Marginal	0.694	0.306	1.000

LCA Local Independence

- LCA has the property of local independence - that given class, item responses are independent.

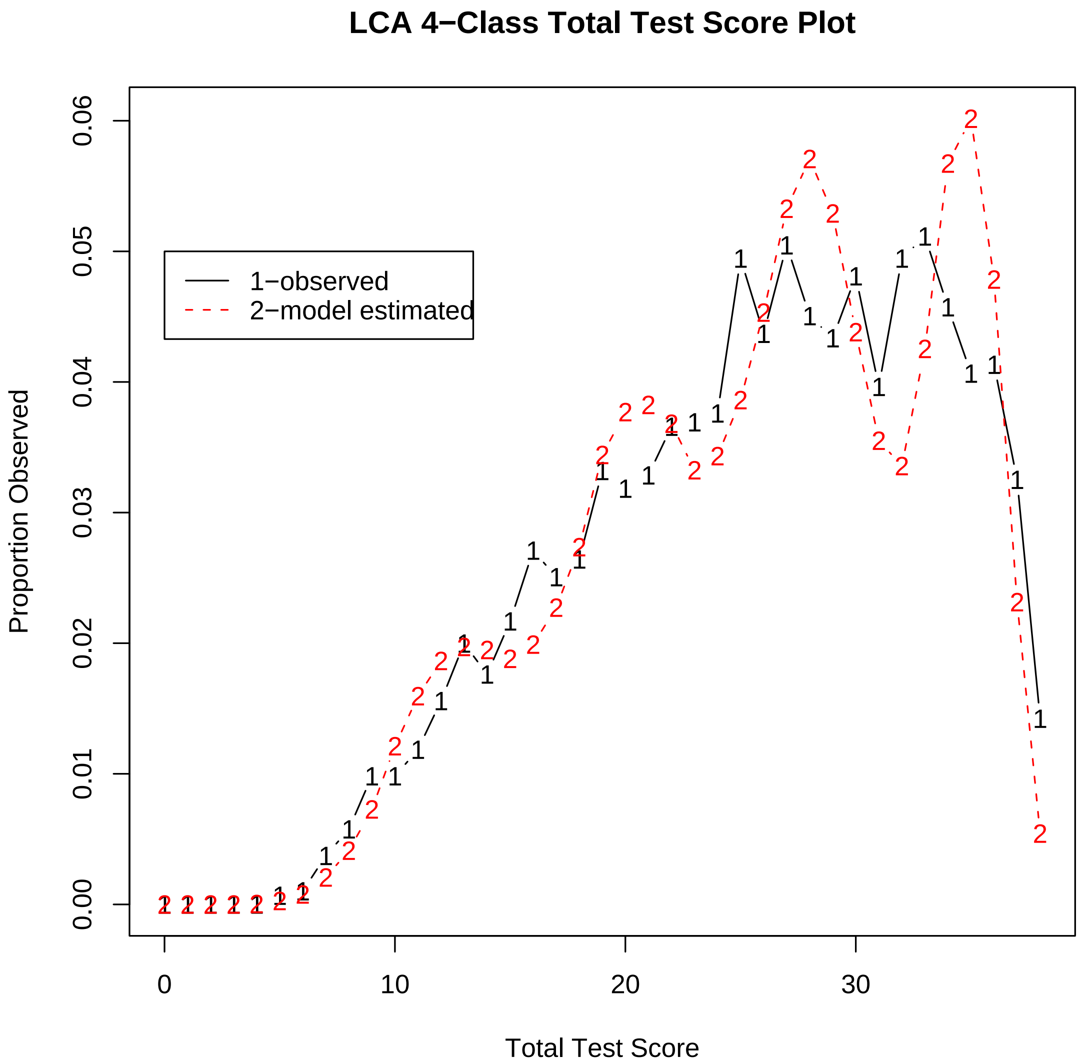
- To give an example, consider Items 1 and 2:

	$P(X_j = 1 \mid c)$			
Item j	$c = 1$	$c = 2$	$c = 3$	$c = 4$
1	0.512	0.915	0.306	0.758
2	0.657	0.979	0.307	0.872

	Item 1		
Item 2	0	1	Marginal
0	0.031	0.097	0.128
1	0.211	0.661	0.872
Marginal	0.242	0.758	1.000

Validation - Total Test Score

As a measure of validation, consider examining the total test score as observed and predicted by our model estimates (we will revisit this next time, too):

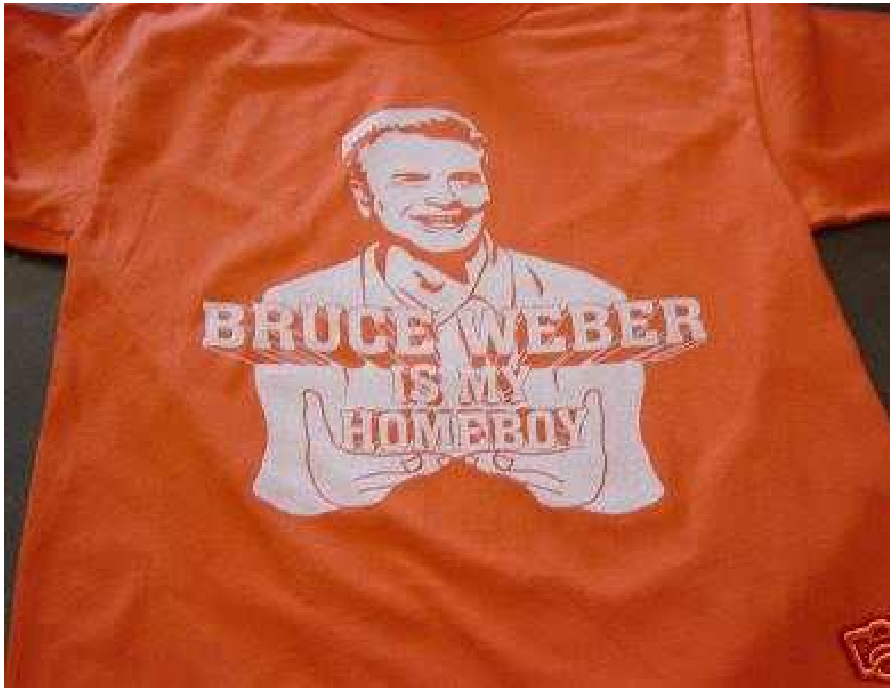


LCA Limitations

- LCA has limitations which make its general application to educational measurement difficult:
  - Classes not known prior to analysis.
  - Class characteristics not know until after analysis.
- Both of these problems are related to LCA being an exploratory procedure for understanding data.
- Cognitive diagnosis models can be thought of as confirmatory versions of LCA.
  - By placing constraints on the class item probabilities and specifying what our classes mean prior to analysis.

Final Thought

- LCA is a wonderful technique to use to find classes with very specific types of data.
- We have only scratched the surface of LCA techniques.



- We will discuss estimation and other models in the weeks to come.

Next Time

- No class next week (Spring Break!).
- Our next class:
  - Empirical research article on LCA:
  - More LCA examples/facets.