Latent Class Analysis: Evaluating Model Fit

Lecture 13

March 30, 2006

Clustering and Classification

Today's Lecture

- Evaluating the fit of an LCA.
- Model-based measures of "fit:"
  - The Log-likelihood of a model.
  - Chi-square tests.
- Model comparison measures:
  - Akaike Information Criterion (AIC).
  - Bayesian Information Criterion (BIC, A.K.A. Schwarz's Criterion).
- Distributional comparisons:
  - Predicted item means.
  - Predicted item covariance/correlations.
- Entropy (a measure of classification certainty).

Upcoming Schedule

Date	Topic
3/30	Other Topics in Latent Class Analysis
4/4	Latent Profile Analysis
4/6,4/11	No Class
4/13	Model Estimation: EM Algorithm
4/18	Model Estimation: MCMC algorithm
4/20, 4/25	General Finite Mixture Models
4/27	Growth Mixture Models
5/2, 5/4	Models for Cognitive Diagnosis
5/9, 5/11	More Cognitive Diagnosis (or spill over time)

Note: Projects due 5/16.

Model Based Measures

- Recall the standard latent class model: 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 ($j = 1, \ldots, K$):

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

- Model based measures of fit revolve around the model function listed above.

- With just the function above, we can compute the probability of *any* given response pattern.

Model Chi-squared Test

- The χ^2 test compares the sets of response patterns that were observed with the set of response patterns expected under the model.
- To form the χ^2 test, one must first compute the probability of each response pattern using the latent class model equation displayed on the last slide.
- The hypothesis tested is that the observed frequency is equal to the expected frequency.
- If the test has a low p-value, the model is said to not fit.
- To demonstrate the model χ^2 test, lets consider the results of the latent class model fit to the data from our running example (from Macready and Dayton, 1977).

Chi-squared Test Example

```
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
```

Chi-squared Test Example

- To begin, compute the probability of observing the pattern $[1111]$...
- Then, to find the expected frequency, multiply that probability by the number of observations in the sample.
- Repeat that process for all cells...
- The compute $\chi^2_p = \sum_r \frac{(O_r - E_r)^2}{E_r}$, where r represents each response pattern.
- The degrees of freedom are equal to the number of response patterns minus model parameters minus one.
- Then find the p-value, and decide if the model fits.

Likelihood Ratio Chi-squared

- The likelihood ratio Chi-square is a variant of the Pearson Chi-squared test, but still uses the observed and expected frequencies for each cell.

- The formula for this test is:

$$G = 2 \sum_r O_r \ln \left(\frac{O_r}{E_r} \right)$$

- The degrees of freedom are still the same as the Pearson Chi-squared test, however.

Chi-square Problems

- The Chi-square test is reasonable for situations where the sample size is large, and the number of variables is small.
  - If there are too many cells where the observed frequency is small (or zero), the test is not valid.
- Note that the total number of response patterns in an LCA is 2^I, where I is the total number of variables.
- For our example, we had four variables, so there were 16 possible response patterns.
- If we had 20 variables, there would be a total of 1,048,576.
  - Think about the number of observations you would have to have if you were to observe at least one person with each response pattern.
  - Now think about if the items were highly associated (you would need even more people).

Model Comparison

- So, if model-based Chi-squared tests are valid only for a limited set of analyses, what else can be done?

- One thing is to look at comparative measures of model fit.

- Such measures will allow the user to compare the fit of one solution (say two classes) to the fit of another (say three classes).

- Note that such measures are only valid as a means of relative model fit - what do these measures become if the model fits perfectly?

Log Likelihood

- Prior to discussing anything, let's look at the log-likelihood function, taken across all the observations in our data set.
- The log likelihood serves as the basis for the AIC and BIC, and is what is maximized by the estimation algorithm.
- The likelihood function is the model formulation across the joint distribution of the data (all observations):

$$L(\mathbf{x}_i) = \prod_{k=1}^{N} \left[\sum_{j=1}^{K} \eta_j \prod_{i=1}^{p} \pi_{ij}^{x_{ki}} \left(1 - \pi_{ij}\right)^{1-x_{ki}} \right]$$

Log Likelihood

- The log likelihood function is the log of the model formulation across the joint distribution of the data (all observations):

$$LogL(\mathbf{x}_i) = \log\left(\prod_{k=1}^{N}\left[\sum_{j=1}^{K}\eta_j\prod_{i=1}^{p}\pi_{ij}^{x_{ki}}\left(1-\pi_{ij}\right)^{1-x_{ki}}\right]\right)$$

$$LogL(\mathbf{x}_i) = \sum_{k=1}^{N}\log\left(\sum_{j=1}^{K}\eta_j\prod_{i=1}^{p}\pi_{ij}^{x_{ki}}\left(1-\pi_{ij}\right)^{1-x_{ki}}\right)$$

- Here, the log function taken is typically base e - the natural log.
- The log likelihood is a function of the observed responses for each person and the model parameters.

Information Criteria

- The Akaike Information Criterion (AIC) is a measure of the goodness of fit of a model that considers the number of model parameters (q).

$$AIC = 2q - 2\log L$$

- Schwarz's Information Criterion (also called the Bayesian Information Criterion or the Schwarz-Bayesian Information Criterion) is a measure of the goodness of fit of a model that considers the number of parameters (q) and the number of observations (N):

$$BIC = q\log(N) - 2\log L$$

- Other information criteria exist - most are based on the concept of entropy.
- For another example, look up Shannon Entropy.

Information Criteria

- When considering which model “fits” the data best, the model with the lowest AIC or BIC should be considered.
- As you have seen (from our empirical article), not all solutions are preferred.
- Again, although AIC and BIC are based on good statistical theory, neither is a gold standard for assessing which model should be chosen.
- Furthermore, neither will tell you, overall, if your model estimates bear any decent resemblance to your data.
- You could be choosing between two (equally) poor models - other measures are needed.

Distributional Measures of Model Fit

- The model-based Chi-squared provided a measure of model fit, while narrow in the times it could be applied, that tried to map what the model said the data looked like to what the data actually looked like.

- The same concept lies behind the ideas of distributional measures of model fit - use the parameters of the model to "predict" what the data should look like.

- In this case, measures that are easy to attain are measures that look at:
  - Each variable marginally - the mean (or proportion).
  - The bivariate distribution of each pair of variables - contingency tables (for categorical variables), correlation matrices, or covariance matrices.

Marginal Measures

- For each item, the model-predicted mean of the item (proportion of people responding with a value of one) is given by:

$$\hat{\bar{X}}_i = \hat{E}(X_j) = \sum_{x_j=0}^{M} \hat{P}(X_i = x_i)x_i = \sum_{j=1}^{J} \hat{\eta}_j \times \hat{\pi}_{ij}$$

- Across all items, you can then form an aggregate measure of model fit by comparing the observed mean of the item to that found under the model, such as the root mean squared error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{I} (\hat{\bar{X}}_i - \bar{X}_i)^2}{I}}$$

- Often, there is not much difference between observed and predicted mean (depending on the model, the fit will always be perfect).

Bivariate Measures

- For each pair of items (say a and b, the model-predicted probability of both being one is given in the same way:

$$\hat{P}(X_a = 1, X_b = 1) = \sum_{j=1}^{J} \hat{\eta}_j \times \hat{\pi}_{aj} \times \hat{\pi}_{bj}$$

- Given the marginal means, you can now form a 2 x 2 table of the probability of finding a given pair of responses to variable a and b:

b	a: 0	a: 1	
0			$1 - \hat{P}(X_b = 1)$
1		$\hat{P}(X_a = 1, X_b = 1)$	$\hat{P}(X_b = 1)$
	$1 - \hat{P}(X_a = 1)$	$\hat{P}(X_a = 1)$	1

Bivariate Measures

- Given the model-predicted contingency table (on the last slide) for every pair of items, you can then form a measure of association for the items.
- Depending on your preference, you could use:
  - Pearson correlation.
  - Tetrachoric correlation.
  - Cohen’s kappa.
- After that, you could then summarize the discrepancy between what your model predicts and what you have observed in the data.
  - Such as the RMSE, MAD, or BIAS.

Entropy

- The entropy of a model is defined to be a measure of classification uncertainty.
- In my mind, I am not certain of it's value as a model selection criterion.
- To define the entropy of a model, we must first look at the posterior probability of class membership, let's call this $\hat{\alpha}_{ic}$ (notation borrowed from Dias and Vermunt, date unknown - online document).
- Here, $\hat{\alpha}_{ic}$ is the estimated probability that observation i is a member of class c
- The entropy of a model is defined as:

$$EN(\boldsymbol{\alpha}) = -\sum_{i=1}^{N}\sum_{j=1}^{J}\alpha_{ij}\log\alpha_{ij}$$

Relative Entropy

- The entropy equation on the last slide is bounded from $[0, \infty)$, with higher values indicated a larger amount of uncertainty in classification.

- Mplus reports the *relative* entropy of a model, which is a rescaled version of entropy:

$$E = 1 - \frac{EN(\boldsymbol{\alpha})}{N \log J}$$

- The relative entropy is defined on $[0, 1]$, with values near one indicating high certainty in classification and values near zero indicating low certainty.

Final Thought

- Evaluating the fit of a mixture model (or specifically, an LCA) can be very complicated.

- The way of doing it, preferably, would consider each of the above measures.

- Relative fit of a model is ok - but what if you are choosing between two really bad models?

Next Time

- Latent Profile Analysis.
- Who has the responsibility for the empirical article?