

Variable-Specific Entropy Contribution

Tihomir Asparouhov and Bengt Muthén

October 7, 2014

In latent class analysis it is useful to evaluate a measurement instrument in terms of how well it identifies the latent classes. This is typically done by computing the entropy

$$E = 1 + \frac{1}{N \log(k)} \left(\sum_{i=1}^N \sum_{k=1}^K P(C = k|U_i) \log(P(C = k|U_i)) \right)$$

where C is the latent variable, K is the number of classes, N is the sample size and U_i is the vector of all latent class indicator variables and the probabilities $P(C = k|U_i)$ are computed from the estimated model. The larger the entropy is the more clear the latent class identification is. The entropy value is between 0 and 1. Entropy with values approaching 1 indicate clear separation of the classes.

The entropy evaluates the quality of the measurement instrument as a whole. Often there is a need to evaluate the quality of individual items. Here we define the univariate entropy which can be computed for each latent class indicator U_j

$$E_j = 1 + \frac{1}{N \log(k)} \left(\sum_{i=1}^N \sum_{k=1}^K P(C = k|U_{ij}) \log(P(C = k|U_{ij})) \right)$$

where $P(C = k|U_{ij})$ is also computed from the estimated model. The univariate entropies E_j are directly comparable among each other. The higher E_j is the more informative the indicator U_j is in identifying the latent classes. Latent class indicators with univariate entropies near 0 can probably be removed from the model as they do not provide much information about the latent class variable. On the other hand latent class indicators with large univariate entropies can be considered good indicators. Thus the univariate entropy can be used in evaluating specific items within a latent class measurement instrument. Note here that the univariate entropy is not additive, i.e., the total entropy for an LCA model is not the sum of the univariate entropies for all of the items.

We illustrate the univariate entropy using example 7.3 from the user's guide but we treat all of the variables in the data file as class indicators. In order to obtain the univariate entropy the option **ENTROPY** has to be added to the **OUTPUT** command.

```

Univariate Entropy
  U1          U2          U3          U4          X1
  -----
  0.553      0.545      0.577      0.567      0.909

Univariate Entropy
  X2          X3          X4          X5          X6
  -----
  0.935      0.535      0.526      0.167      0.164

Univariate Entropy
  X7          X8          X9          X10
  -----
  0.501      0.561      0.913      0.934

```

From this output one can conclude that the best class indicators are X1, X2, X9 and X10 while the worst class indicators are X5 and X6. This result is accurate. The data generation is described in mcex7.3.inp file in the Mplus installation directory. The variables X5 and X6 are independent of the latent class variable and carry no information about it, while the variables X1, X2, X9 and X10 have a mean of -2 in one of the classes and 2 in the other and thus are highly correlated with the latent class variable.

In simple 2 or 3 class latent variable models it could be easy to evaluate an indicator simply by considering the estimated parameters, however, in more complex models this might be hard. The univariate entropy value gives a new and simple method for evaluating the quality of a latent class indicator. In Mplus Version 7.3 this feature is available for continuous and categorical indicators.