Wald Test of Mean Equality for Potential Latent Class Predictors in Mixture Modeling

November 16, 2007

For mixture models, frequently there can be many potential class predictors. In this note we describe a procedure that can be used to select the most important of these variables. Including a large number of class predictors in the model during the estimation could potentially lead to very slow computation as well as unclear results due to collinearity in the predictors. We can instead estimate the mixture model without the covariates as a first step. Consequently for each potential class predictor variable X we can evaluate the conditional class specific means for that variable $E(X|C)$ using the estimated model with a categorical latent class variable C. Typically, the stronger the predictor X, the bigger the differences between these conditional means across the different classes. We can evaluate the statistical significance of the class specific mean differences by conducting a Wald test for the following hypothesis:

$$E(X|C=1) = E(X|C=2) = ... = E(X|C=k),$$

where k is the number of classes in the estimated model. Mplus will compute this Wald test for each variable included in the *auxiliary =* command and marked with *(e)*. The stronger the evidence that this hypothesis is rejected, i.e., the larger the test value and the smaller the p-value, the more promising and useful the covariate is. Denote by $\mu_j = E(X|C=j)$. The Wald test statistic is computed as follows

$$W = \hat{D}V^{-1}\hat{D}^T$$

where

$$\hat{D} = (\hat{\mu}_1 - \hat{\mu}_2, \hat{\mu}_2 - \hat{\mu}_3, ..., \hat{\mu}_{k-1} - \hat{\mu}_k)$$

and V is the asymptotic variance covariance matrix of the difference vector $\hat{D}$. The parameters μ_j are not model parameters because the covariate X is not a part of the model. Thus we need to estimate the parameters μ_j and their asymptotic variance/covariance using the pseudo-class draw technique, see Wang et al. (2005).

Using the estimated model we first compute the posterior class distribution. Denote by C_i the latent class variable for individual i and let

$$p_{ij} = P(C_i = j|*).$$

Here we condition on the estimated model and all observed data. Using this posterior distribution we generate L pseudo draws for the class variable C_i. We denote these by C_{il} for $l = 1, ..., L$. The variables are random samples taken from the discrete distribution with probabilities p_{i1},...,p_{ik}. Given the pseudo draws we can then compute the corresponding class specific sample mean values of X for each of the L pseudo draw. For $l = 1, ..., L$ we get

$$\hat{\mu}_{jl} = \frac{1}{n_{jl}} \sum_{i|C_{il}=j} X_i$$

where n_{jl} is the number of individuals in class j for the l-th pseudo draw. Similarly the sample variance is computed for each pseudo draw

$$\hat{\sigma}_{jl} = \frac{1}{n_{jl}} \sum_{i|C_{il}=j} (X_i - \hat{\mu}_{jl})^2.$$

Consistent estimates for the parameter μ_j are then obtained by averaging the estimates across the pseudo draws

$$\hat{\mu}_j = \frac{1}{L} \sum_l \hat{\mu}_{jl}$$

as well as the difference vector

$$\hat{D} = \frac{1}{L} \sum_l \hat{D}_l$$

where

$$\hat{D}_l = (\hat{\mu}_{1l} - \hat{\mu}_{2l}, \hat{\mu}_{2l} - \hat{\mu}_{3l}, ..., \hat{\mu}_{k-1l} - \hat{\mu}_{kl}).$$

The asymptotic variance for $\hat{D}$ is obtained similar to multiple imputation variance estimation

$$\hat{V} = \hat{V}_1 + \left(1 + \frac{1}{L}\right)\hat{V}_2.$$

where $\hat{V}_1$ and $\hat{V}_2$ are the within and the between variance components. These are computed as follows. First we compute the variance of D_l by the delta method

$$W_l = MAM^T$$

where M is a $k-1$ by k matrix

$$M = \begin{pmatrix} 1 & -1 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & -1 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & 1 & -1 \end{pmatrix}$$

and $A = diag\{\hat{\sigma}_{jl}\}$ is the diagonal k dimensional matrix with $\hat{\sigma}_{jl}$ on the diagonal, for $j = 1, ..., k$, and represents the variance of the vector $(\hat{\mu}_{1l}, ..., \hat{\mu}_{kl})$. Now $\hat{V}_1$ and $\hat{V}_2$ are computed by

$$\hat{V}_1 = \frac{1}{L}\sum_l W_l$$

$$\hat{V}_2 = \frac{1}{L-1}\sum_l (\hat{D}_l - \hat{D})(\hat{D}_l - \hat{D})^T$$

Finally the Wald test statistic is computed as

$$\hat{W} = \hat{D}\hat{V}^{-1}\hat{D}^T.$$

Mplus uses $L = 20$ pseudo draws, which is recommenced in general, see Wang et al. (2005).

References

Wang C-P, Brown CH, Bandeen-Roche K (2005). Residual Diagnostics for Growth Mixture Models: Examining the Impact of a Preventive Intervention on Multiple Trajectories of Aggressive Behavior. Journal of the American Statistical Association 100(3):1054-1076.