Local independence

Jeroen K. Vermunt & Jay Magidson

Local independence is the basic assumption underlying LATENT VARIABLE models, such as FACTOR ANALYSIS, LATENT TRAIT ANALYSIS, LATENT CLASS ANALYSIS, and LATENT PROFILE ANALYSIS. The axiom of local independence states that the observed items are independent of each other given an individual's score on the latent variable(s). This definition is a mathematical way of stating that the latent variable explains why the observed items are related to one another.

We will explain the principle of local independence using an example taken from Lazarsfeld and Henry (1968). Suppose that a group of 1,000 persons are asked whether they have read the last issue of magazines A and B. Their responses are:

	Read A	Did not Read A	Total
Read B	260	240	500
Did not read B	140	360	500
Total	400	600	1000

It can easily be verified that the two variables are quite strongly related to one another. Readers of A tend to read B more often (65%) than do non-readers (40%). The value of phi coefficient, the product-moment correlation coefficient in a 2-by-2 table, equals .245.

Suppose that we have information on the respondents' educational levels, dichotomized as high and low. When the 1,000 people are divided into these two groups, readership is observed to be:

	High education			Low education		
	Read A	Did not read A	Total	Read A	Did not read A	Total
Read B	240	60	300	20	80	100
Did not read B	160	40	200	80	320	400
Total	400	100	500	100	400	500

Thus, in each of the 2-by-2 tables in which education is held constant, there is no association between the two magazines: the phi coefficient equals 0 in both tables. That is, reading A and B are independent within educational levels.

The reading behavior is, however, very different in the two subgroups: the high-education group has much higher probabilities of reading for both magazines (.60 and .80) than the low-education group (.20 and .20). The association between A and B can thus fully be explained from the dependence of A and B on the third factor education. Note that the cell entries in the marginal AB table, $N(AB)$, can be obtained by:

$$N(AB) = N(H) \cdot P(A|H) \cdot P(B|H) + N(L) \cdot P(A|L) \cdot P(B|L),$$

where H and L denotes high and low education, respectively.

In latent variable models, the observed variables are assumed to be locally independent given the latent variable(s). This means that the latent variables have the same role as education in the example. It should be noted that this assumption not only makes sense from a substantive point of view, but it is also necessary to identify the unobserved factors.

The fact that local independence implies that the latent variables should fully account for the associations between the observed items suggests a simple test of this assumption. The estimated two-way tables according to the model, computed as shown above for $N(AB)$, should be similar to the observed two-way tables. In the case of continuous indicators, we can compare estimated with observed correlations.

Variants of the various types of latent variable models have been developed that make it possible to relax the local independence assumption for certain pairs of variables. Depending on the scale type of the indicators, this is accomplished by inclusion of direct effects or correlated errors in the model.

References

Lazarsfeld, P.F., and Henry, N.W. (1968). *Latent Structure Analysis.* Boston: Houghton Mill.

Bartholomew, D.J., and Knott, M. (1999). *Latent Variable Models and Factor Analysis.* London: Arnold.