

Linear Logistic Scoring Equations for Latent Class and Latent Profile Models: A Simple Method for Classifying New Cases

Jeroen K. Vermunt

Tilburg University, Tilburg, The Netherlands

ORCID: <https://orcid.org/0000-0001-9053-9330>

Jay Magidson

Statistical Innovations Inc., Arlington, MA, USA

Version of March 5 2024

Corresponding author's contact address: Jeroen K. Vermunt, Tilburg School of Social and Behavioral Sciences, Department of Methodology and Statistics, PO box 90153, 5000 LE Tilburg, The Netherlands. E-mail: j.k.vermunt@uvt.nl.

Linear Logistic Scoring Equations for Latent Class and Latent Profile Models: A Simple Method for Classifying New Cases

Abstract

Researchers are often interested in using latent class or latent profile parameter estimates to obtain posterior class membership probabilities for observations other than those of the original sample. In this paper, we demonstrate that these probabilities typically take on the form of linear logistic equations with coefficients which are functions of the original model parameters. In other words, the posterior class membership probabilities can be computed with a prediction formula similar to that of a multinomial logistic regression model. We derive the scoring equations for nominal, ordinal, count, and continuous indicators, as well as investigate models with missing values on class indicators, local dependencies, covariates, or multiple latent variables. In addition to the mathematical derivations of the scoring equations, we describe how either exact or approximate scoring equations can be obtained by estimating a multinomial regression model using a weighted data set.

Keywords: Posterior class membership probabilities, Classification, Out-of-sample prediction, Multinomial logistic regression.

In applications of factor analysis, after selecting and estimating the factor model of interest, one will typically obtain (linear) factor-score equations which can be used to estimate the subjects' factor scores as a function of the original items included in the model (Bartholomew, Knott, and Moustaki, 2011). An important feature of factor-score equations is that these can be used not only for the subjects in the estimation sample, but also for new subjects, that is, for out-of-sample prediction.

When performing a latent class (LC) analysis, after selecting the final model, one may assign the individuals in the estimation sample to LCs using their posterior class membership probabilities. However, it is not well known that these posterior probabilities can be expressed exactly by a set of linear logistic equations, with “regression” weights which are functions of the original LC model parameters. More specifically, a closed-form expression for the posteriors exists, as a function of the LC model parameters, if the responses are modeled using distributions from the exponential family and with canonical link functions. Availability of a set of scoring equations makes it straightforward to compute the class membership probabilities for subjects which do not belong to the original sample used to estimate the LC model. In this way, one can realize an important goal of many LC analysis applications, namely obtaining out-of-sample class membership predictions.

As far as we know, LatentGOLD (Vermunt and Magidson, 2016, 2021) is currently the only software for LC analysis that allows one to obtain these logistic scoring equations, both in tabular form and in the form of SPSS or R syntax. The aim of this paper is to show how these equations are derived. As will be shown, the slopes of the linear logistic scoring equations are obtained easily, but the expression for the intercept terms (constants) may be somewhat more complex. In many situations, the equations for the posteriors will contain only main effects of the

response variables. However, as in quadratic discriminant analysis, when a LC model for continuous responses assumes variances to be class specific, quadratic terms also need to be included, and when the LC model contains covariances/associations which are class specific, interactions are also required. The approach can be extended easily to LC models with covariates and multiple latent variables. More complicated are situations where responses contain missing values (in which case the constants need to be adapted to the missing data pattern), where the model contains direct effects of covariates on the responses (in which case the exact logistic form may no longer hold), and where non-canonical link functions are used (in which case there is no longer any direct relation between the LC model parameters and the scoring equation).

Rather than computing the scoring equations from the LC model parameters, one can also obtain these equations by estimating a multinomial logistic regression model using the posteriors as weights, as done in the LatentGOLD Step3-Scoring option (Vermunt and Magidson, 2016, 2021). This approach has the advantage of increased flexibility in that it is also possible to obtain *approximate* equations when exact closed-form solutions are not available, or when one prefers a simpler approximate set of scoring equations over more complex exact equations.

Below, we present the scoring equations for models for categorical responses (nominal, ordinal, and counts), models for continuous responses, models with local dependencies, models with covariates, models with missing values on responses, and models with multiple latent variables. We also discuss how the scoring equations can be obtained using a weighted multinomial logistic regression analysis.

Latent Class Models for Categorical Responses

Let Y_j denote one of J response variables (or indicators), with $1 \leq j \leq J$. A particular response for and the number of categories of the j th response variable are referred to as y_j and R_j ,

respectively, with $1 \leq y_j \leq R_j$. The probability of having a particular set of responses \mathbf{y} is denoted by $P(\mathbf{Y} = \mathbf{y})$. The discrete latent variable is denoted by X , a particular latent class by k , and the number of classes by K .

Nominal Responses

The standard LC model for nominal responses has the following form (Collins and Lanza, 2010; Goodman, 1974a/b; Hagenaars, 1990; McCutcheon, 1987):

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{k=1}^K P(X = k) \prod_{j=1}^J P(Y_j = y_j | X = k),$$

where $P(X = k)$ is the probability of belonging to class k and $P(Y_j = y_j | X = k)$ the conditional probability of giving response y_j on variable Y_j conditional on belonging to class k .

These probabilities are often parameterized using logistic equations (Formann, 1992; Heinen, 1996; Magidson and Vermunt, 2004); that is,

$$P(X = k) = \frac{\exp(\gamma_k)}{D}$$

$$P(Y_j = y_j | X = k) = \frac{\exp(\alpha_{y_j} + \beta_{y_j k})}{E_{jk}}$$

With

$$D = \sum_{k=1}^K \exp(\gamma_k)$$

$$E_{jk} = \sum_{y_j=1}^{R_j} \exp(\alpha_{y_j} + \beta_{y_j k})$$

Here, γ_k are intercept or constant terms in the regression model for $P(X = k)$, and α_{y_j} and $\beta_{y_j k}$ are intercept and slope parameters in the regression model for $P(Y_j = y_j | X = k)$. As

always, identifying constraints need to be imposed on the logistic parameters. Typically, they are either restricted to sum to 0 over classes and response categories (referred to as effect coding), or set to 0 for one class and one response category (called dummy coding). The terms D and E_{jk} are normalizing constants.

The posterior probability of belonging to class k conditional on response vector \mathbf{y} , denoted by $P(X = k | \mathbf{Y} = \mathbf{y})$, can be obtained as follows (Dias and Vermunt, 2008; Goodman, 1974a/b):

$$P(X = k | \mathbf{Y} = \mathbf{y}) = \frac{P(X = k) \prod_{j=1}^J P(Y_j = y_j | X = k)}{\sum_{k=1}^K P(X = k) \prod_{j=1}^J P(Y_j = y_j | X = k)}$$

Replacing the model probabilities by their logit equations yields:

$$P(X = k | \mathbf{Y} = \mathbf{y}) = \frac{\frac{\exp(\gamma_k)}{D} \prod_{j=1}^J \frac{\exp(\alpha_{y_j} + \beta_{y_j k})}{E_{jk}}}{\sum_{k=1}^K \frac{\exp(\gamma_k)}{D} \prod_{j=1}^J \frac{\exp(\alpha_{y_j} + \beta_{y_j k})}{E_{jk}}}$$

This equation can be simplified by removing D and α_{y_j} , which are redundant because they do not depend on k . Moreover, the product over the J responses can be replaced by a sum over the logs of the terms concerned. This yields

$$P(X = k | \mathbf{Y} = \mathbf{y}) = \frac{\exp(\gamma_k - \sum_{j=1}^J \log(E_{jk}) + \sum_{j=1}^J \beta_{y_j k})}{\sum_{k=1}^K \exp(\gamma_k - \sum_{j=1}^J \log(E_{jk}) + \sum_{j=1}^J \beta_{y_j k})} = \frac{\exp(\gamma_k^* + \sum_{j=1}^J \beta_{y_j k})}{\sum_{k=1}^K \exp(\gamma_k^* + \sum_{j=1}^J \beta_{y_j k})}$$

where $\gamma_k^* = \gamma_k - \sum_{j=1}^J \log(E_{jk})$. Though not really necessary, the γ_k^* parameters may be subjected to the same identifying – effect or dummy coding – constraints as the other parameters.

The above derivation shows that the posterior class membership probabilities can be written as a logistic equation with slopes equal to the LC model logistic regression slopes and with constants equal to the logistic class constants minus the sum of the logs of the normalizing

constants. The more difficult part in the computation of these scoring equations is the computation of the constants γ_k^* . But, once we have the scoring equations, we can easily compute the class membership probabilities for any response pattern, including response patterns which were not available in the original data used to estimate the LC model of interest.

It should be noted that the γ_k^* terms are identical to the one-variable parameters for the latent classes in the log-linear formulation of the LC model proposed by Haberman (1979). In this formulation, the joint distribution of the X and \mathbf{Y} , $P(X = k, \mathbf{Y} = \mathbf{y})$, is modelled as follows:

$$P(X = k, \mathbf{Y} = \mathbf{y}) = \frac{\exp(\gamma_k^* + \sum_{j=1}^J \alpha_{y_j} + \sum_{j=1}^J \beta_{y_j k})}{F}.$$

The posterior class membership probability is obtained as $P(X = k | \mathbf{Y} = \mathbf{y}) = P(X = k, \mathbf{Y} = \mathbf{y}) / \sum_{k=1}^K P(X = k, \mathbf{Y} = \mathbf{y})$. As can be seen, since $\sum_{j=1}^J \alpha_{y_j}$ and F cancel, $P(X = k | \mathbf{Y} = \mathbf{y})$ has exactly the form we derived above. Thus, when using Haberman's log-linear formulation, the constants of the scoring equations are also model parameters. However, an important disadvantage of this formulation is that it is computationally less efficient since parameter estimation involves processing the cell entries in the joint cross-tabulation of X and all Y_j variables. Therefore, this log-linear approach can be used only when the number of response variables is small.

Ordinal Responses and Counts

In the LatentGOLD program for LC analysis (Vermunt and Magidson, 2016, 2021), ordinal response variables can be modeled using an adjacent-category logit model, that is, using a canonical link function (Agresti, 2002). More specifically, these are multinomial logit models in which the class-indicator association parameters are restricted as follows: $\beta_{y_j k} = \beta_{jk} \gamma_j$; that is, to be nominal-by-linear (Goodman, 1979; Heinen, 1996). This implies that for ordinal

variables,

$$P(Y_j = y_j | X = k) = \frac{\exp(\alpha_{y_j} + \beta_{jk}y_j)}{E_{jk}}.$$

The same restrictions imposed on the β_{y_jk} parameters also apply to the scoring equations; that is, for ordinal variables, we replace the β_{y_jk} terms in the scoring equations by $\beta_{jk}y_j$. This shows that in the ordinal case, the class-membership logits are linear functions of the item responses. It should be noted that while we assumed that the category scores ranged from 1 to R_j , the adjacent-category logit model allows for any type of scoring. In its more general form, $\beta_{y_jk} = \beta_{jk}v_{y_j}$, where v_{y_j} is the score for category y_j .

When modeling ordinal variables using other (non-canonical) link functions, such as cumulative logit or cumulative probit link functions, exact expressions for the scoring equations no longer exist. As will be shown below, a possible way out is to estimate the scoring equations treating the response variables as either numeric or nominal predictors of class membership.

Also for a Poisson and binomial count variable, the scoring equations contain the term $\beta_{jk}y_j$. This can be seen from the fact that the class-specific density of a count variable Y_j takes on the following form:

$$P(Y_j = y_j | X = k) \propto \frac{\exp(\alpha_j y_j + \beta_{jk}y_j)}{E_{jk}},$$

where it should be noted that $\alpha_j y_j$ cancels from the scoring equation. The expression for $\log E_{jk}$ changes compared to the nominal and ordinal case. For Poisson counts, $\log E_{jk} = \exp(\alpha_j + \beta_{jk})e_j$, where e_j is the exposure; and for Binomial counts, $\log E_{jk} = e_j \log(1 + \exp(\alpha_j + \beta_{jk}))$, where e_j represents the number of trials. This shows that scoring equations should also include terms for the exposure (number of trials) when this number varies across individuals. When the

e_j are fixed, as for nominal and ordinal variables, the $\log E_{jk}$ terms can be included in the constants γ_k^* .

Local Dependencies

Thus far, we assumed that responses are independent within classes. Now we will look at the scoring equations for LC models with local dependencies (Hagenaars, 1988; Magidson and Vermunt, 2004; Oberski, Kollenburg, and Vermunt, 2013). In the most general case, including a local dependency between (nominal) response variables Y_j and Y_m implies that

$$P(Y_j = y_j, Y_m = y_m | X = k) = \frac{\exp(\alpha_{y_j} + \alpha_{y_m} + \beta_{y_j k} + \beta_{y_m k} + \delta_{y_j y_m} + \lambda_{y_j y_m k})}{E_{jmk}},$$

where

$$E_{jmk} = \sum_{y_j=1}^{R_j} \sum_{y_m=1}^{R_m} \exp(\alpha_{y_j} + \alpha_{y_m} + \beta_{y_j k} + \beta_{y_m k} + \delta_{y_j y_m} + \lambda_{y_j y_m k}).$$

This is a model with an association between Y_j and Y_m , $\delta_{y_j y_m}$, and an interaction with the latent classes, $\lambda_{y_j y_m k}$. In other words, it represents a model in which the strength of the local dependency is allowed to vary across classes.

As can be seen, one difference with the local independence model is that the normalizing constants entering in the γ_k^* coefficients of the scoring equations should be computed per set of locally dependent variables. The $\delta_{y_j y_m}$ term cancels from the scoring equations because it does not depend on the classes. In contrast, the term $\lambda_{y_j y_m k}$ becomes part of the scoring equations, which in the case of class-specific local dependencies not only contains main effects but also interaction terms. Note that when local dependencies are not class-specific, that is, when $\lambda_{y_j y_m k} = 0$, the only remaining difference between local independence and local dependence models concerns the computation of the constants γ_k^* .

The scoring equations in local-dependence LC models for ordinal variables are very similar to those for nominal variables. When the ordinal variables are modelled using an adjacent-category logit specification, $\delta_{y_j y_m} = \delta_{jm} y_j y_m$ and $\lambda_{y_j y_m k} = \lambda_{jmk} y_j y_m$. The scoring equations will contain the term $\lambda_{jmk} y_j y_m$ when the interaction parameters λ_{jmk} are not fixed to 0.

Missing Values

When some indicators have missing values, the LC model for the observed values \mathbf{Y}_{obs} can be defined as follows:

$$P(\mathbf{Y}_{obs} = \mathbf{y}_{obs}) = \sum_{k=1}^K P(X = k) \prod_{j=1}^J P(Y_j = y_j | X = k)^{r_j},$$

where $r_j = 1$ if the response variable concerned is observed and 0 when it has a missing value (Vermunt et al., 2008; Vermunt and Magidson, 2016). Note that this formulation implies that the product is taken over the observed responses only. Therefore, similar to subjects with complete data, the computation of the posteriors for subjects with missing values involves using only their observed responses. This means that the sum $\sum_{j=1}^J \beta_{y_j k}$ should be taken over the observed variables only or, equivalently, that $\beta_{y_j k}$ should be set to 0 for the missing value category.

However, the sum $\sum_{j=1}^J \log(E_{jk})$ which is subtracted from the constants should also be taken over the observed variables only, implying that each pattern of missing data has its own set of constants γ_k^* . A way to account for this is by using the same γ_k^* for all observations but adding a term $\log(E_{jk})$ to the scoring equation when variable j has a missing value. In other words, in order to deal with missing data, the scoring equation should be expanded to include the term $\sum_{j=1}^J \log(E_{jk}) (1 - r_j)$.

A special type of missing data occurs when the LC model is estimated using J variables, but only the first J_1 of these are to be used for classification purposes (where $J = J_1 + J_2$); for example, this situation may occur if one wishes to ignore the last J_2 variables when calculating the classification probabilities because this information will not be available when performing out-of-sample predictions. In this case, the posteriors are obtained as follows:

$$P(X = k | \mathbf{Y}_1 = \mathbf{y}_1) = \frac{P(X=k) \prod_{j=1}^{J_1} P(Y_j=y_j|X=k)}{\sum_{k=1}^K P(X=k) \prod_{j=1}^{J_1} P(Y_j=y_j|X=k)}.$$

As can be seen, only the slope parameters and the normalizing constants of the first J_1 response variables will enter into the scoring equations.

An Example with Five Dichotomous Indicators

Table 1 provides an example illustrating the computation of the scoring equations for an application with five dichotomous response variables. It concerns the model with three latent classes estimated for the LatentGOLD “political.sav” demo data set. The upper part of Table 1 gives the estimates of the model parameters γ_k , α_{y_j} , and β_{y_jk} using dummy coding with the parameters for the first class and the first item category fixed to 0. The lower part gives the values of γ_k^* and $\log(E_{jk})$, where for consistency we also use dummy coding for the $\log(E_{jk})$ terms. To obtain the $\log(E_{1k})$ values for the first item, we first compute $\exp(\alpha_{y_1} + \beta_{y_1k})$, which for all three classes equals 1.0000 for $y_1 = 1$, and 2.6533, 0.4452, and 1.4312 for $y_1 = 2$. Next, we sum the obtained values across the two item categories and take the log, yielding 1.29562, 0.3682, and 0.8884 for the three classes. Because of the dummy coding, we subtract the value of the first class, yielding the reported $\log(E_{1k})$ values 0.0000, -0.9275, and -0.4073. The $\log(E_{jk})$ values of all items are subtracted from the γ_k values to get the intercepts γ_k^* of the scoring equations, and the slopes β_{y_jk} can be used without any modification in the

scoring equations. In the case of a missing value, the slope parameters for the item concerned equal $\log(E_{jk})$. Appendix A shows R code generated for this application by LatentGOLD, which can be used for classifying new observations.

Other Types of Latent Class and Mixture Models

Continuous Responses

Now, let us turn to LC or mixture models for continuous response variables (McLachlan and Peel, 2000), also referred as latent profile models. In a local independence model with normal within-class distributions with possibly unequal variances, the response distributions have the following form:

$$P(Y_j = y_j | X = k) \propto \exp\left(-\frac{1}{2} \log \sigma_{jk}^2 - \frac{1}{2} \frac{\mu_{jk} \mu_{jk}}{\sigma_{jk}^2} + \frac{\mu_{jk} y_j}{\sigma_{jk}^2} - \frac{1}{2} \frac{y_j y_j}{\sigma_{jk}^2}\right),$$

where μ_{jk} and σ_{jk}^2 denote the mean and variance of Y_j in latent class k . It can be seen that in the construction of the logistic scoring equations, the terms $-\frac{1}{2} \log \sigma_{jk}^2$ and $-\frac{1}{2} \frac{\mu_{jk} \mu_{jk}}{\sigma_{jk}^2}$, which do not contain the response, become part of the constants. Moreover, the equations will contain the linear and quadratic terms $\frac{\mu_{jk}}{\sigma_{jk}^2} y_j$ and $-\frac{1}{2 \sigma_{jk}^2} y_j y_j$.

When variances are assumed to be equal across classes, the first and last term of the above univariate normal distribution become $-\frac{1}{2} \log \sigma_j^2$ and $-\frac{1}{2} \frac{y_j y_j}{\sigma_j^2}$, respectively, implying that these cancel from the scoring equations because they do not depend on the class. This yields a set of scoring equations similar to those obtained in linear discriminant analysis (Hastie, Tibshirani, and Friedman, 2008).

In the more general case of multivariate normal responses with unrestricted covariances Σ_k , the LC model becomes

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{k=1}^K P(X = k)P(\mathbf{Y} = \mathbf{y}|X = k),$$

with

$$P(\mathbf{Y} = \mathbf{y}|X = k) \propto \exp\left(-\frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}\boldsymbol{\mu}'_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k + \boldsymbol{\mu}'_k\boldsymbol{\Sigma}_k^{-1}\mathbf{y} - \frac{1}{2}\mathbf{y}'\boldsymbol{\Sigma}_k^{-1}\mathbf{y}\right).$$

As can be seen, the scoring equations now not only contain linear and quadratic terms, but also interaction terms are needed. More specifically, denoting an entry of $\boldsymbol{\Sigma}_k^{-1}$ by a_{mjk} , the weights for y_j , y_j^2 , and $y_j y_m$ are $\sum_{m=1}^J \mu_{mk} a_{jmk}$, $-\frac{1}{2} a_{jjk}$, and $-a_{jmk}$, respectively. The first two terms of the multivariate normal density, which do not depend on the responses, become part of the constants. When variances and covariances are equal across classes, we again have equations with main effects only, as in linear discriminant analysis.

Various kinds of restricted mixtures of multivariate normal distributions have been proposed in which constraints are imposed on the class-specific means and/or covariances. Examples include mixture factor models (McLachlan and Peel, 2000; Yung, 1997), mixture structural equation models (Dolan and Van der Maas, 1997), mixture models with constrained eigenvalue decompositions of $\boldsymbol{\Sigma}_k$ (Banfield and Raftery, 1993), and mixture growth models (Muthén, 2004). For these models, the same scoring equations can be used as when means and covariances are unrestricted.

An Example with Three Continuous Indicators

Table 2 reports the model parameters and the scoring equations for a LC model with three continuous indicators (Glucose, Insulin, and SSPG) from the LatentGOLD “diabetes.dat” demo data set. It is a three-class model with a free residual covariance between the first two class indicators and with class-specific residual (co)variances. The scoring equation for this model contains not only linear terms, but also quadratic terms as well as the interaction terms between

y_1 and y_2 .

Covariates

When covariates are included in the model, the latent class probabilities are typically modeled as a logistic function of the covariates (Bandeem-Roche et al, 1997; Dayton and McReady, 1988; Yamaguchi, 2000). That is,

$$P(X = k|\mathbf{z}) = \frac{\exp(\gamma_{0k} + \sum_{p=1}^P \gamma_{pk}z_p)}{\sum_{k=1}^K \exp(\gamma_{0k} + \sum_{p=1}^P \gamma_{pk}z_p)}$$

Here, \mathbf{z} denotes the vector of covariates, and γ_{0k} and γ_{pk} represent the constants and the regression parameters for covariate z_p .

Since the denominator does not depend on the class, it cancels from the formula for the posterior class membership probability and thus also from the scoring equations. Assuming the response variables are nominal, the posterior probability of class membership given responses and covariates becomes:

$$P(X = k|\mathbf{z}) = \frac{\exp(\gamma_{0k}^* + \sum_{j=1}^J \beta_{y_jk} + \sum_{p=1}^P \gamma_{pk}z_p)}{\sum_{k=1}^K \exp(\gamma_{0k}^* + \sum_{j=1}^J \beta_{y_jk} + \sum_{p=1}^P \gamma_{pk}z_p)}$$

That is, the covariate terms can simply be added to the scoring equations.

Covariates may also have direct effects on the indicators. Let us assume we have a single covariate z which has a direct effect on the categorical response variable Y_j ; that is,

$$P(Y_j = y_j|X = k, z) = \frac{\exp(\alpha_{y_j} + \beta_{y_jk} + \delta_{y_j}z)}{E_{jk|z}}$$

where

$$E_{jk|z} = \sum_{y_j=1}^{R_j} \exp(\alpha_{y_j} + \beta_{y_jk} + \delta_{y_j}z).$$

As can be seen, in this model, the normalizing constants depend on the covariate value, meaning

that we no longer have a single $\log E_{jk}$ which can be subtracted from γ_k . Because the $\log E_{jk|z}$ are neither linear functions of the covariate values, they cannot be added to the linear term for the covariate concerned. In other words, the exact linear logistic representation of the posterior probabilities collapses in this situation, though, as discussed below and in Appendix B, it may still be used as an approximation. An exception is the situation in which the covariate is a nominal or dichotomous variable, in which case exact scoring equations can still be obtained by subtracting $\log E_{jk|z}$ from the γ_{pk} terms of the covariate concerned.

Note that when the direct effect of a covariate is allowed to be class specific, or equivalently, when an interaction term is included, the indicator-covariate interaction should also be added to the scoring equation. Again, the scoring equations will be exact only when the covariate concerned is nominal or dichotomous. For example, this is the specification used in multiple-group LC models in which response probabilities may be allowed to differ across subgroups for one or more indicators (Clogg and Goodman, 1984; Eid, Langeheine, and Diener, 2004; Kankaras, Moors, and Vermunt, 2010).

Multiple Latent Variables

Suppose the LC model contains two latent variables X_1 and X_2 instead of one, so that we have a LC Factor or Discrete Factor model. Such a model, has the following form (Goodman, 1974b; Hagenaars, 1990; Magidson and Vermunt, 2001; Vermunt and Magidson, 2005):

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} P(X_1 = k_1, X_2 = k_2) \prod_{j=1}^J P(Y_j = y_j | X_1 = k_1, X_2 = k_2).$$

As in the single latent variable case, the $P(X_1 = k_1, X_2 = k_2)$ and $P(Y_j = y_j | X_1 = k_1, X_2 = k_2)$ can be modelled using logistic regression models (Magidson and Vermunt, 2001). For example,

$$P(X_1 = k_1, X_2 = k_2) = \frac{\exp(\gamma_{k_1}^1 + \gamma_{k_2}^2 + \gamma_{k_1 k_2}^{12})}{D}$$

$$P(Y_j = y_j | X_1 = k_1, X_2 = k_2) = \frac{\exp(\alpha_{y_j} + \beta_{y_j k_1}^1 + \beta_{y_j k_2}^2)}{E_{j k_1 k_2}}.$$

Also in this case, the posterior probabilities can be written as functions of the LC model parameters; i.e.,

$$P(X_1 = k_1, X_2 = k_2 | \mathbf{Y} = \mathbf{y}) = \frac{\exp(\gamma_{k_1 k_2}^* + \sum_{j=1}^J \beta_{y_j k_1}^1 + \sum_{j=1}^J \beta_{y_j k_2}^2)}{\sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \exp(\gamma_{k_1 k_2}^* + \sum_{j=1}^J \beta_{y_j k_1}^1 + \sum_{j=1}^J \beta_{y_j k_2}^2)},$$

where the $\gamma_{k_1 k_2}^*$ contain the γ terms and the normalizing constants $E_{j k_1 k_2}$.

Note that the above equation for two latent variables can easily be generalized to an arbitrary number of S latent variables. The logistic scoring equation then becomes:

$$P(X_1 = k_1, \dots, X_S = k_S | \mathbf{Y} = \mathbf{y}) = \frac{\exp(\gamma_{k_1 \dots k_S}^* + \sum_{s=1}^S \sum_{j=1}^J \beta_{y_j k_s}^s)}{\sum_{k_1=1}^{K_1} \dots \sum_{k_S=1}^{K_S} \exp(\gamma_{k_1 \dots k_S}^* + \sum_{s=1}^S \sum_{j=1}^J \beta_{y_j k_s}^s)}.$$

It should be noted that the marginal posterior probabilities $P(X_s = k_s | \mathbf{Y} = \mathbf{y})$, which are obtained by collapsing over the other latent variables, can not be written as logistic functions. However, logistic approximations of the marginal posteriors may be precise enough in most applications. Below, we discuss how such approximations can be obtained.

Estimating the Scoring Equations Using Logistic Regression Analysis

Rather than computing the scoring equations from the parameters of the LC model, it is also possible to obtain these equations posthoc using a standard routine for multinomial logistic regression analysis. This involves the following three steps:

1. After selection of the final LC model, save the posterior class membership probabilities to an output file. This is a feature available in all software packages for LC analysis.
2. Create an expanded data set with K records per subject, which contains a column with the class number taking on values from 1 to K , a column with the posterior probability

for the person and class concerned, and columns for the response variables and covariates used in the LC model, the latter columns containing the same values repeated in each of the K records for each subject.

3. Estimate a logistic regression model in which the posteriors are used as weights. The class number is the dependent variable, and the responses and covariates are the predictors.

Depending on the situation, in the third step the responses and covariates are modeled as either nominal or numeric predictors, quadratic and/or interaction effects are added, and missing value dummies are included. For count variables, one should include the exposures (or total number of trials) as additional numeric predictors when these differ across individuals. Steps 2 and 3 are automated in the LatentGOLD program (Vermunt and Magidson, 2016, 2021), and are called Step3-Scoring.

This approach can be used not only for the posthoc computation of the exact scoring equations, but also for obtaining approximate scoring equations. This is useful when an exact form does not exist, such as when direct effects of numeric covariates on indicators were included in the LC model or when non-canonical link functions were used for ordinal variables, as well as the situation where one prefers a set of simplified equations, say without quadratic or interaction terms, that are almost as good as the exact ones. An example of the latter can be seen in Table 3, which reports the approximate scoring equations for the diabetes.dat example presented above, but leaving out the quadratic terms of y_1 and y_2 and their interaction terms. The approximate equations predict the class memberships almost as well as the exact equations; that is, the entropy R^2 equals 0.817, while its original value equals 0.833.

When ordinal variables are modeled using non-canonical link functions, such as

cumulative logit or probit models, we have two options. Option 1 is to compute the exact scoring equations by treating the response variables as nominal predictors in the posthoc logistic regression analysis; that is, by making use of the fact that the estimated $P(Y_j = y_j|X = k)$ based on an ordinal model can be reproduced perfectly by an unrestricted multinomial model. Option 2 is to estimate the scoring equations using the response variables as numeric predictors, which in fact implies that the estimated $P(Y_j = y_j|X = k)$ from the original LC model are approximated by an adjacent-category logit model.

As shown above, in LC models with multiple latent variables, an exact set of logistic scoring equations exists for the joint class membership probabilities, but not for the marginal class membership probabilities. The posthoc estimation method can also be used to obtain approximate scoring equations based on the marginal posteriors. Applying these equations will be simpler than first computing the joint and subsequently collapsing over the other latent variables, especially with models containing more than two latent variables. The quality of the resulting approximation can be assessed by a goodness-of-fit measure.

Discussion

As in continuous latent variables models, in LC models it is important to have a simple scoring rule for predicting a person's value on the latent variable. In this paper, we showed that for LC models this scoring rule has the form of a linear logistic equation, with weights which are simple functions of the original LC model parameters. We derived the exact scoring equations for nominal, ordinal, count, and continuous response variables, for local independence and local dependence models, for models with covariates, for models with multiple latent variables, and for models with missing values on some of the indicators. Moreover, we discussed several situations in which exact scoring equations may not exist, such as LC models with direct effects

of covariates on the indicators and LC models in which the conditional response distributions are restricted using regression models based on non-canonical link functions.

We also explained how to compute exact or approximate scoring equations with the saved posterior probabilities from any LC analysis program. This can be achieved with standard routines for logistic regression analysis. In practice, this may be much easier than computing the scoring equations from the LC model parameters, where the constants from the scoring equations may be somewhat more tedious to obtain.

While not discussed explicitly, the computation of the scoring equations proceeds in exactly the same manner in LC models for mixed responses; that is, in LC models for combinations of nominal, ordinal, count, and continuous indicators (Hennig and Liao, 2013; Hunt and Jorgensen, 1999; Vermunt and Magidson, 2002). The only thing that needs to be done in the computation of the scoring equations is to collect the terms for the different indicators, irrespective of their scale types. When using the posthoc method based on a logistic regression analysis, things are even easier. Nominal indicators are used as nominal predictors, and ordinal, count and continuous indicators as numeric predictors. Depending on the situation, quadratic and/or interaction terms may also need to be included.

The scoring equations discussed in this article can be used to obtain *point* estimates for the posterior probabilities, not only for subjects in the original sample, but also for new subjects. However, an issue not dealt with in this paper is the uncertainty about those estimates. Since the “regression” weights of the scoring equation are sample estimates, when deriving a prediction, it would be better to take into account this sampling variability. Our future research will focus on this important topic.

References

- Agresti, A. (2002). *Categorical data analysis*. Second Edition, New York: Wiley.
- Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92, 1375-1386.
- Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49(3), 803–821.
- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis*. London: Arnold.
- Clogg, C. C., & Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79, 762-771.
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley.
- Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83, 173-178.
- Dias, J. G., & Vermunt, J. K. (2008). A bootstrap-based aggregate classifier for model-based clustering. *Computational Statistics*, 23, 643-659.
- Dolan, C. V., & Van der Maas, H. L. J. (1997). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, 63, 227-253.
- Eid M., Langeheine R., & Diener E. (2003). Comparing typological structures across cultures by multigroup latent class analysis. A primer. *Journal of Cross-Cultural Psychology*, 34, 195–210.

- Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87, 476–486.
- Goodman, L. A. (1974a). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Goodman, L. A. (1974b). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I - A modified latent structure approach. *American Journal of Sociology*, 79, 1179–1259.
- Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications saving ordered categories. *Journal of the American Statistical Association*, 74, 537–52.
- Haberman, S. J. (1979). *Analysis of Qualitative Data, Vol 2, New Developments*. New York: Academic Press.
- Hastie T., Tibshirani R., & Friedman J., (2008). *The Elements of Statistical Learning*. Springer.
- Hagenaars, J. A. P. (1988). Latent structure models with direct effects between indicators local dependence models. *Sociological Methods & Research*, 16(3), 379–405.
- Hagenaars, J. A. (1990). *Categorical longitudinal data - loglinear analysis of panel, trend and cohort data..* Newbury Park: Sage.
- Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks, CA: Sage.
- Hennig, C., & Liao, T. F. (2013). How to Find an Appropriate Clustering for Mixed Type Variables with Application to Socioeconomic Stratification (with discussion). *Journal of Royal Statistical Society Series C*, 62, 309-369.
- Hunt, L., & Jorgensen, M. (1999). Mixture model clustering using the MULTIMIX program. *Australian and New Zealand Journal of Statistics*, 41, 153-172.

- Kankaras, M., Moors, G., & Vermunt, J. K. (2010). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt, and J. Billiet (Eds.), *Cross-Cultural Analysis: Methods and Applications* (pp 359-384). New York: Routledge.
- Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays, *Sociological Methodology*, 31, 223-264.
- Magidson, J., & Vermunt, J. K., (2004). Latent class models. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp 175-198). Thousand Oaks: Sage Publications.
- McCutcheon, A. L. (1987). *Latent class analysis*. Newbury Park, CA: Sage.
- McLachlan, G. J. & Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons, New York.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345-368). Thousand Oakes, CA: Sage.
- Oberski, D. L, van Kollenburg, G. H., & Vermunt, J. K. (2013). A Monte Carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Classification and Data Analysis*, 7, 267-279.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis*, 89-106. Cambridge, UK: Cambridge University Press.
- Vermunt, J. K., & Magidson, J. (2005). Factor analysis with categorical indicators: A comparison between traditional and latent class approaches. In A. Van der Ark, M.A. Croon & K. Sijtsma (Eds.), *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences* (pp. 41-62). Mahwah: Erlbaum.

Vermunt, J.K. and Magidson, J. (2016). *Technical Guide for LatentGOLD 5.1: Basic, Advanced, and Syntax*. Belmont MA: Statistical Innovations Inc.

Vermunt, J.K. and Magidson, J. (2021). *Upgrade Manual for LatentGOLD Basic, Advanced/Syntax, and Choice Version 6.0*. Arlington, MA: Statistical Innovations Inc.

Vermunt, J. K., Van Ginkel, J. R., Van der Ark, L. A., & Sijtsma, K. (2008). Multiple imputation of categorical data using latent class analysis. *Sociological Methodology*, 33, 369-297.

Yamaguchi, K. (2000). Multinomial logit latent-class regression models: An analysis of the predictors of gender-role attitudes among Japanese women. *American Journal of Sociology*, 105, 1702-40.

Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297-330.

Appendix A: R Code Generated by LatentGOLD for the First Example

Application

With the LatentGOLD output options “ScoringEquations” and “WriteRsyntax=<filename>”, one can request an R syntax file that can be used to classify new observations. The variable names in the “political.sav” data file are `sys_resp`, `ideo_lev`, `rep_pot`, `prot_app`, and `conv_par`.

The `lg_scoring` function, first creates dummies for the response categories and for missing values. Then, it computes the class-specific linear terms using the dummies and the scoring equations’ parameters. Subsequently, the linear terms are exponentiated and transformed to posterior probabilities. The function’s output are the posterior probabilities and the modal class. Below, you find the R code, which ends with example code calling the `lg_scoring` function to add classification information to a data set.

```
## Scoring function to be called per record
lg_scoring<-function(dat) {
  # Create auxiliary variables
  if(is.na(dat$sys_resp)) {
    sys_resp_lg_1<-0;sys_resp_lg_2<-0;sys_resp_lg_m<-1
  }
  else {
    if(dat$sys_resp==1) {
      sys_resp_lg_1<-1;sys_resp_lg_2<-0;sys_resp_lg_m<-0
    }
    else if(dat$sys_resp==2) {
      sys_resp_lg_1<-0;sys_resp_lg_2<-1;sys_resp_lg_m<-0
    }
    else {
      sys_resp_lg_1<-0;sys_resp_lg_2<-0;sys_resp_lg_m<-1
    }
  }
}

# The same is done for the other 4 indicators
...

# Compute classification logits
Cluster_lg_1<-(0)+
```


Running head: Obtaining Scoring Equations for Latent Class Models

```
(0)*sys_resp_lg_1+(0)*sys_resp_lg_2+
(0)*ideo_lev_lg_1+(0)*ideo_lev_lg_2+
(0)*rep_pot_lg_1+(0)*rep_pot_lg_2+
(0)*prot_app_lg_1+(0)*prot_app_lg_2+
(0)*conv_par_lg_1+(0)*conv_par_lg_2+
(0)*sys_resp_lg_m+(0)*ideo_lev_lg_m+
(0)*rep_pot_lg_m+(0)*prot_app_lg_m+
(0)*conv_par_lg_m
Cluster_lg_2<-(3.4185551)+
(0)*sys_resp_lg_1+(-1.7853117)*sys_resp_lg_2+
(0)*ideo_lev_lg_1+(-3.0502076)*ideo_lev_lg_2+
(0)*rep_pot_lg_1+(0.56595846)*rep_pot_lg_2+
(0)*prot_app_lg_1+ (-0.74630356)*prot_app_lg_2+
(0)*conv_par_lg_1+(-3.039846)*conv_par_lg_2+
(-0.9274766)*sys_resp_lg_m+(-0.49927555)*ideo_lev_lg_m+
(0.11060958)*rep_pot_lg_m+ (-0.34180273)*prot_app_lg_m+
(-1.8328984)*conv_par_lg_m
Cluster_lg_3<-(3.6424675)+
(0)*sys_resp_lg_1+ (-0.61732246)*sys_resp_lg_2+
(0)*ideo_lev_lg_1+ (-0.23276128)*ideo_lev_lg_2+
(0)*rep_pot_lg_1+ (3.6818864)*rep_pot_lg_2+
(0)*prot_app_lg_1+ (3.0608866)*prot_app_lg_2+
(0)*conv_par_lg_1+ (-1.0034281)*conv_par_lg_2+
(-0.40726727)*sys_resp_lg_m+(-0.089565904)*ideo_lev_lg_m+
(1.9387485)*rep_pot_lg_m+(2.5015355)*prot_app_lg_m+
(-0.81826931)*conv_par_lg_m

# Compute odds from logits
max_lg<-Cluster_lg_1
if(Cluster_lg_2>max_lg) {
  max_lg<-Cluster_lg_2
}
if(Cluster_lg_3>max_lg) {
  max_lg<-Cluster_lg_3
}
Cluster_lg_1<-exp(Cluster_lg_1-max_lg)
Cluster_lg_2<-exp(Cluster_lg_2-max_lg)
Cluster_lg_3<-exp(Cluster_lg_3-max_lg)

# Compute modal class and probabilities from odds
max_lg<-Cluster_lg_1
Cluster_lg_modal<-1
if(Cluster_lg_2>max_lg) {
  max_lg<-Cluster_lg_2; Cluster_lg_modal<-2
}
if(Cluster_lg_3>max_lg) {
```

```

    max_lg<-Cluster_lg_3; Cluster_lg_modal<-3
  }
  sum_lg<-Cluster_lg_1+Cluster_lg_2+Cluster_lg_3
  Cluster_lg_1<-Cluster_lg_1/sum_lg
  Cluster_lg_2<-Cluster_lg_2/sum_lg
  Cluster_lg_3<-Cluster_lg_3/sum_lg

  return(list(
    "Cluster_modal"=Cluster_lg_modal,
    "Cluster_1"=Cluster_lg_1,
    "Cluster_2"=Cluster_lg_2,
    "Cluster_3"=Cluster_lg_3
  ))
}

## Example of call of scoring function in a loop over records
outdata<-inpdata
for(i in 1:nrow(outdata)){
  scoring<-lg_scoring(outdata[i,])
  outdata[i,"Cluster_modal"]<-scoring$Cluster_modal
  outdata[i,"Cluster_1"]<-scoring$Cluster_1
  outdata[i,"Cluster_2"]<-scoring$Cluster_2
  outdata[i,"Cluster_3"]<-scoring$Cluster_3
}

```

Appendix B: Taylor Approximation of the Normalizing Constants with a Covariate Having a Direct Effect on a Categorical Indicator

As shown in the main text, when a numeric covariate z has a direct effect on a categorical indicator, the normalizing constant of the indicator concerned becomes:

$$E_{jk|z} = \sum_{y_j=1}^{R_j} \exp(\alpha_{y_j} + \beta_{y_j k} + \delta_{y_j} z),$$

and will thus depend on the value of z . As a result, the scoring equations will no longer be linear logistic. However, a possible way out is to approximate this term using a Taylor expansion.

For simplicity, assume covariate z is centered, and thus has a mean of 0. The second-order Taylor approximation of $\log E_{jk|z}$ at $z = 0$ equals:

$$\log E_{jk|z} = \log E_{jk|z=0} + \frac{d \log E_{jk|z=0}}{dz} z + \frac{d^2 \log E_{jk|z=0}}{d^2 z} / 2 z^2,$$

with

$$\frac{d \log E_{jk|z=0}}{dz} = \sum_{y_j=1}^{R_j} P(Y_j = y_j | X = k, z = 0) \delta_{y_j}$$

and

$$\frac{d^2 \log E_{jk|z=0}}{d^2 z} = \sum_{y_j=1}^{R_j} P(Y_j = y_j | X = k, z = 0) \left[\delta_{y_j} - \frac{d \log E_{jk|z=0}}{dz} \right] \delta_{y_j}.$$

The term $\log E_{jk|z=0}$ is subtracted from the intercept of class k , the $\frac{d \log E_{jk|z=0}}{dz}$ are subtracted from the linear terms for z , and the $-\frac{d^2 \log E_{jk|z=0}}{d^2 z} / 2$ appear as quadratic terms in the scoring equations.

Table 1: Latent class model parameters and scoring equation parameters for the political.sav data example

LC model parameters

| | Intercept X=1 | Intercept X=2 | Intercept X=3 |
|---|---------------|---------------|---------------|
| X | 0.0000 | -0.0723 | -0.5173 |

| | Intercept | Slopes X=1 | Slopes X=2 | Slopes X=3 |
|------|-----------|------------|------------|------------|
| Y1=1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Y1=2 | 0.9758 | 0.0000 | -1.7853 | -0.6173 |
| Y2=1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Y2=2 | -0.3534 | 0.0000 | -3.0502 | -0.2328 |
| Y3=1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Y3=2 | -1.7062 | 0.0000 | 0.5660 | 3.6819 |
| Y4=1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Y4=2 | 0.2028 | 0.0000 | -0.7463 | 3.0609 |
| Y5=1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Y5=2 | 2.0140 | 0.0000 | -3.0398 | -1.0034 |

Scoring equation parameters

| | Intercept X=1 | Intercept X=2 | Intercept X=3 |
|---|---------------|---------------|---------------|
| X | 0.0000 | 3.4186 | -3.6425 |

| | E values for X=1 | E values for X=2 | E values for X=3 |
|----|------------------|------------------|------------------|
| Y1 | 0.0000 | -0.9275 | -0.4073 |
| Y2 | 0.0000 | -0.4993 | -0.0896 |
| Y3 | 0.0000 | 0.1106 | 1.9387 |
| Y4 | 0.0000 | -0.3418 | 2.5015 |
| Y5 | 0.0000 | -1.8329 | -0.8183 |

Table 2: Latent class model parameters and scoring equation parameters for the diabetes.dat data example

| <i>LC model parameters</i> | | | | |
|------------------------------------|----------------|----------------|----------------|--|
| | Intercept X=1 | Intercept X=2 | Intercept X=3 | |
| X | 0 | -0.6927 | -1.036 | |
| | Parameters X=1 | Parameters X=2 | Parameters X=3 | |
| Mean Y1 | 91.2315 | 104.0049 | 234.7598 | |
| Mean Y2 | 359.2211 | 495.0568 | 1121.0893 | |
| Mean Y3 | 163.1271 | 309.4323 | 76.9772 | |
| Variance Y1 | 76.4770 | 230.0891 | 5005.9106 | |
| Variance Y2 | 2669.7454 | 14844.5520 | 73551.0945 | |
| Variance Y3 | 2421.4506 | 22966.5152 | 2224.5020 | |
| Covariance Y1-Y2 | 96.4624 | 1279.9240 | 17910.7089 | |
| <i>Scoring equation parameters</i> | | | | |
| | Intercept X=1 | Intercept X=2 | Intercept X=3 | |
| X | 0.0000 | 42.6430 | 56.8566 | |
| | Slopes X=1 | Slopes X=2 | Slopes X=3 | |
| Y1 | 0.0000 | -0.5599 | -1.1314 | |
| Y2 | 0.0000 | -0.1066 | -0.0661 | |
| Y3 | 0.0000 | -0.0539 | -0.0328 | |
| Y1*Y1 | 0.0000 | 0.0027 | 0.0061 | |
| Y2*Y2 | 0.0000 | 0.0002 | -0.0001 | |
| Y3*Y3 | 0.0000 | 0.0001 | 0.0001 | |
| Y1*Y2 | 0.0000 | 0.0002 | 0.0000 | |

Table 3: Approximate scoring equation parameters for the diabetes.dat data example

| | Intercept X=1 | Intercept X=2 | Intercept X=3 |
|-------|------------------|------------------|------------------|
| X | 0.0000 | -9.8317 | -25.7983 |
| | Slopes X=1 | Slopes X=2 | Slopes X=3 |
| Y1 | 0.0000 | 0.0435 | 0.0626 |
| Y2 | 0.0000 | 0.0236 | 0.0416 |
| Y3 | 0.0000 | -0.0816 | -0.0173 |
| Y3*Y3 | 0.0000 | 0.0003 | -0.0001 |