

Chapter 6

Association Models

Abstract The association models, appropriate for the analysis of ordinal contingency tables, are presented for two-way and multi-way contingency tables. Their features, properties, and the associated graphs are discussed. The models of uniform association (U), row effect (R), column effect (C), multiplicative row–column effect (RC), and the more general $RC(M)$ model are illustrated with examples in terms of fit, presentation, and interpretation. They are all worked out in \mathbb{R} , through functions provided for their fit and the construction of their scores' plots.

Keywords Association models: U, R, C, $RC(M)$ • Graphs for the $RC(2)$ model • Association models for multi-way tables

6.1 Basic Association Models for Two-way Tables

We realized so far that in the context of classical log-linear models there are just two options for modeling two-way contingency tables: the parsimonious but restrictive model of independence (4.1) and the saturated. Association models fill the gap between these two extreme cases by imposing a special structure on the association and reducing the number of interaction parameters, providing thus intermediate models of dependence. For ease in understanding but also for interpretation purposes, it is convenient to think in terms of local associations in the table and first define the models on local odds ratios rather than cell frequencies. Recall that for models applied on an $I \times J$ contingency table there always exists an equivalent expression defining them on the $(I - 1) \times (J - 1)$ table of the corresponding set of local odds ratios.

In most of the cases association models apply to ordinal classification variables and are thus usually introduced as *models for ordinal data*. However, some of them do not require ordinality, as we shall see later on in this chapter.

6.1.1 Linear-by-Linear Association Model

We have seen in Sect. 2.2.5 that for an $I \times J$ contingency table and under the model of independence, all the local odds ratios are equal to 1, i.e., $\theta_{ij}^L = 1$, $i = 1 \dots, I-1$; $j = 1 \dots, J-1$. Whenever the model of independence is of poor fit, the only alternative in the framework of classical log-linear models is the saturated model (4.5), which assumes all θ_{ij}^L 's to be free parameters and is noneffective in summarizing the underlying significant association. A natural way to proceed is to assume a pattern for this underlying association. This way, the number of parameters to be estimated is reduced and, most important, we can provide a meaningful interpretation. The easiest pattern to think of, which is meanwhile of clear and strong interpretational power, is that of constant θ_{ij}^L 's, as under independence, but different than 1. That is, to introduce the model

$$\theta_{ij}^L = c, \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad (6.1)$$

for some $c > 0$, to be estimated. This model allows for interaction while remains parsimonious, since it has just one parameter more than the independence model, the parameter c . Under the independence model, all possible odds ratios $\theta_{ij}^{k\ell}$ of the table are equal to 1. Under (6.1), local association is uniform, since all the local odds ratios are equal to c . This property characterizes model (6.1), which is therefore called *uniform* association model, denoted as U. When it comes to the odds ratio $\theta_{ij}^{k\ell}$ of any 2×2 subtable of our initial table, through (2.46), (6.1) takes the form

$$\theta_{ij}^{k\ell} = c^{(k-i)(\ell-j)}, \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad i < k \leq I, \quad j < \ell \leq J,$$

and in log-scale

$$\log \theta_{ij}^{k\ell} = (k-i)(\ell-j) \log c, \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad (6.2)$$

for $i < k \leq I$, $j < \ell \leq J$. Under the U model, the general $\theta_{ij}^{k\ell}$ odds ratio is influenced by the categories of each classification variable but only through their distances. Hence, odds ratios formed by cells further apart will exhibit stronger association. Measuring thus how far apart are two categories of a classification variable is crucial. Distances between categories are meaningful only when the corresponding classification variable is ordinal. Hence the U model makes sense to be considered only for tables with both classification variables ordinal or with one ordinal and the other binary.

The U model assumes that all successive categories of a classification variable are equidistant. However, there can arise ordinal variables of non-equidistant successive categories. A typical example of this type is a categorized income variable, which is actually interval scaled with categories corresponding to intervals of unequal length. A flexible way to handle such situations is to assign scores to the categories of the classification variables and express their distances by the corresponding scores'

differences. Thus, let $\{\alpha_1, \alpha_2, \dots, \alpha_J\}$ and $\{v_1, v_2, \dots, v_J\}$ be the scores assigned to the row and column categories, respectively. The simplest and most natural choice for the scores is $\alpha_i = i$ ($i = 1, \dots, I$) and $v_j = j$ ($j = 1, \dots, J$), which corresponds to model (6.2). Allowing the scores to take other values as well and setting $\varphi = \log c$, we are led to model

$$\log \theta_{ij}^{k\ell} = \varphi(\alpha_k - \alpha_i)(v_\ell - v_j), \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad (6.3)$$

with $i < k \leq I$, $j < \ell \leq J$, for which scores of successive row or column categories are not necessarily equidistant. Their distance is meant in terms of their similarity as they interact with the other classification variable. Thus different scores may be assigned to the same levels of a classification variable X when interacting with different variables Y or Z . This will be illustrated in a three-way contingency table example in Sect. 6.7.1. Regarding the scores' assignment, refer also to the related discussion in Sect. 2.3.1.

For non-equidistant scores for successive categories, the local odds ratios under (6.3) are no more all equal but proportional (in log-scale) to the distance between the enrolled categories of each classification variable. Due to this linear dependence on each of the classification variables, model (6.3) is called the *linear-by-linear* association model (LL).

Though the interpretation of these models is clear and natural when formulated in terms of the local odds ratios, the development of inferential aspects and model fitting is more straightforward for their equivalent formulation in terms of expected cell frequencies. Recalling that the saturated log-linear model in terms of θ_{ij} is provided by (4.7) and equating (4.7) to (6.3) for $k = i + 1$ and $\ell = j + 1$, we conclude that the (i, j) th interaction term under the LL model has the form $\lambda_{ij}^{XY} = \varphi \alpha_i v_j$. Hence, the equivalent expression of LL model (6.3) in terms of expected cell frequencies is

$$\log m_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \varphi \alpha_i v_j, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad (6.4)$$

where the overall mean and the main effects parameters are those of the classical log-linear model.

Model (6.4) reduces to the U model and is thus equivalent to (6.2), not just for $\alpha_i = i$, ($i = 1, \dots, I$) and $v_j = j$, ($j = 1, \dots, J$) but for any choice of row and column scores $\{\alpha_1, \alpha_2, \dots, \alpha_J\}$ and $\{v_1, v_2, \dots, v_J\}$, as long as they are both equidistant for successive categories. This is due to model's LL property of being invariant in linear transformation of the scores, as we shall see in Sect. 6.4. Thus, for identifiability purposes, usually the scores are set to satisfy the sum-to-zero and the sum of squares-to-one constraints

$$\sum_{i=1}^I \alpha_i = 0 \quad \text{and} \quad \sum_{i=1}^I \alpha_i^2 = 1, \quad (6.5)$$

$$\sum_{j=1}^J v_j = 0 \quad \text{and} \quad \sum_{j=1}^J v_j^2 = 1. \quad (6.6)$$

For scores satisfying the (6.5) and (6.6) constraints, multiplying (6.4) by $\varphi_i v_j$ and adding over i and j leads to

$$\varphi = \sum_{i,j} \varphi_i v_j \log m_{ij}, \quad (6.7)$$

i.e., φ measures the correlation between row and column scores, fact that justifies its characterization as *intrinsic association* parameter.

6.1.2 Example 6.1

We shall demonstrate the utility and interpretation power of this parsimonious association model with just one degree of freedom less than complete independence by an example. We shall first focus on explaining the nature and use of such a model and we will provide inferential details and application in software at a later stage. The data used are from a survey on the use of cannabis among students, conducted at the University of Ioannina (Greece) in 1995 and published in Marselos et al. (1997). The students' frequency of alcohol consumption is measured on a four-level scale ranging from at most once per month up to more frequent than twice per week while their trial of cannabis through a three-level variable (never tried–tried once or twice–more often). These two ordinal variables are cross-classified leading to a 4×3 table provided in Table 6.1.

These data provide strong evidence against the independence model (4.1), since the corresponding LR test statistic is $G^2(I) = 152.793$, which is highly significant with an asymptotic p -value < 0.00005 ($df = 6$). In the context of classical log-linear models the only alternative is to add the interaction term λ_{ij}^{XY} in the model and end up thus to the saturated model.

Taking advantage of the ordinal nature of the classification variables, we apply the U model to the data of Table 6.1, by fitting model (6.4) with $\varphi_i = i$ ($i = 1, \dots, 4$) and $v_j = j$ ($j = 1, 2, 3$). Thus, we introduce just one additional parameter to the independence model, the φ . The LR test statistic for model (6.4) equals $G^2(U) = 1.469$, leading to a reduction of 151.324 from $G^2(I)$ by sacrificing just 1 df . This model is of impressive fit with p -value = 0.92. The cell estimates under U are provided in parentheses in Table 6.1.

As already mentioned, under the U model, the local odds ratios θ_{ij}^I are constant all over the table. The corresponding sampling values for the local odds ratios are provided in Table 6.2. In this case, the association parameter φ is estimated as $\hat{\varphi} = 0.803$ and furthermore $\hat{\theta}_{ij}^I = \hat{\theta} = \exp(\hat{\varphi}) = \exp(0.803) = 2.23$, for all $i = 1, 2, 3$ and $j = 1, 2$. This means that the odds of having tried cannabis once or twice vs. never tried is 2.23 times higher for students who drink twice a month than those who drink at most once a month. The same comparison holds for any odds ratio comparing successive row and successive column categories.

Table 6.1 Students’ survey about cannabis use at the University of Ioannina, Greece (1995)

| Alcohol consumption | I tried cannabis. . . | | | Total |
|---------------------|-----------------------|---------------|------------|-------|
| | Never | Once or twice | More often | |
| At most once/month | 204 (204.4) | 6 (5.7) | 1 (0.9) | 211 |
| Twice/month | 211 (211.4) | 13 (13.1) | 5 (4.5) | 229 |
| Twice/week | 357 (352.8) | 44 (48.8) | 38 (37.4) | 439 |
| More often | 92 (95.3) | 34 (29.4) | 49 (50.3) | 175 |
| Total | 864 | 97 | 93 | 1054 |

In parentheses are given the maximum likelihood estimates under the model of uniform association (U)

Table 6.2 Sample local odds ratios for the students’ survey about cannabis use at the University of Ioannina, Greece (1995)

| Alcohol consumption | I tried cannabis. . . | | |
|---------------------|-----------------------|---------------|------------|
| | Never | Once or twice | More often |
| At most once/month | 2.10 | 2.31 | |
| Twice/month | 2.00 | 2.25 | |
| Twice/week | 3.00 | 1.67 | |
| More often | | | |

If we would like to compare any non-successive categories, the results can be adjusted accordingly. For example, for the odds ratio formed by the corner (“extreme”) cells of the table, it holds

$$\frac{\hat{\pi}_{11}\hat{\pi}_{43}}{\hat{\pi}_{13}\hat{\pi}_{41}} = \exp(\hat{\phi}(\alpha_4 - \alpha_1)(v_3 - v_1)) = \exp(\hat{\phi} \cdot 3 \cdot 2) = 123.387 ,$$

meaning that the odds of using often cannabis instead of never tried is 123 times higher for student who drink more often than twice a week than for students who drink at most once a month.

6.1.3 Row and Column Effect Models

The LL model presented above is a very parsimonious and useful model of strong interpretation power when it is applicable. However, often it is proved insufficient. It can be the case that the structure of model (6.4) is appropriate but there is no obvious way of deciding about the scores of one of the classification variables. It is then natural to broaden model (6.4) to a class of more flexible association models by relaxing the assumptions about known scores. Model (6.4) with unknown row scores $\{\alpha_1, \dots, \alpha_r\}$ and thus parameters to be estimated is the *row effect* association model, to be denoted as R. Under this model, the odds defined over the column classification variable vary from row to row, i.e., the effect of the row classification

variable on the column odds is significant but unknown. This effect is reflected in the row scores and more precisely in the unknown (and unequal) distances between successive row categories. Model R has $I - 2$ additional parameters than model LL, corresponding to the row scores. The number of parameters is reduced by two, due to the identifiability constraints (6.5) that hold. Thus, the associated df of model R equal $(I - 1)(J - 2)$. Analogously, the *column effect* association model C is defined by expression (6.4) for known row scores and unknown column scores $\{v_1, \dots, v_J\}$. It models the effect of the column classification variable on the row odds. The associated df are $df(C) = (I - 2)(J - 1)$.

We have seen in the context of the U association model that its definition in terms of odds ratios (6.1) is more natural with respect to interpretation. The R model is equivalently defined in terms of local odds ratios as

$$\theta_{ij}^L = c_{1i}, \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1, \quad (6.8)$$

and the C model as

$$\theta_{ij}^L = c_{2j}, \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1. \quad (6.9)$$

Expression (6.8) reveals the dependence of the column odds on the row category while the analogue statement is true for the C model (6.9).

In terms of local odds ratios and categories' scores, model R is expressed as

$$\log(\theta_{ij}^L) = \varphi(\alpha_{i+1} - \alpha_i)(v_{j+1} - v_j), \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1, \quad (6.10)$$

with parametric row scores $\{\alpha_i, i = 1, \dots, I\}$ and known (equidistant) column scores $\{v_j, j = 1, \dots, J\}$. Analogously, model C is (6.10) with parametric column scores and known (equidistant) row scores.

6.1.4 Row by Column Effect Model

The LL, U, R, and C models considered so far are special types of log-linear models. The LL model is applicable on two-way tables when both the classification variables are ordinal. The R and C models are less restrictive about the nature of the underlying classification variables and thus also less parsimonious. They allow the row or column classification variable, respectively, to be ordinal but with unknown distances between the scores assigned to its successive categories or even nominal. This is achieved by considering the row or the column scores as unknown parameters to be estimated. Furthermore, a more flexible model can be defined by (6.4), considering the row and the column score vectors to be both unknown parameters. Thus, we model a multiplicative *row by column* association. This model, denoted by RC, is no more linear in its parameters and their estimation is not straightforward. The estimation problem will be faced in Sect. 6.2.

Table 6.3 Association models and related *df*. The U model is a special LL model

| Association model | $\alpha = (\alpha_1, \dots, \alpha_I)$ | $\nu = (\nu_1, \dots, \nu_J)$ |
|--------------------------------|--|-------------------------------|
| Linear \times linear (LL) | Known | Known |
| Row effect (R) | Unknown | Known |
| Column effect (C) | Known | Unknown |
| Multiplicative row-column (RC) | Unknown | Unknown |

| Model | Parameters additional to independence | d.f. |
|-------|---------------------------------------|----------------------|
| LL | 1 | $(I - 1)(J - 1) - 1$ |
| R | $1 + (I - 2)$ | $(I - 1)(J - 2)$ |
| C | $1 + (J - 2)$ | $(I - 2)(J - 1)$ |
| RC | $1 + (I - 2) + (J - 2)$ | $(I - 2)(J - 2)$ |

In terms of local odds ratios, the RC model is defined by

$$\theta_{ij}^L = c_{1i}c_{2j}, \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1,$$

allowing the effect of each classification variable on the odds defined by the other one to vary from category to category. In log-scale, the RC model is given by (6.10) with parametric (unknown) row and column scores. The RC model does not require ordinality for any of the classification variables. Thus it can be applied in tables of nominal variables as well. Of course, scores' assignment is more natural for ordinal variables.

The association models considered so far are all defined in terms of expected cell frequencies by

$$\log m_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \varphi \alpha_i \nu_j, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \tag{6.11}$$

i.e., by expression (6.4). Thus, all association models considered so far are defined by the same expression (6.4) and differentiated by the assumptions made for the nature of the scores, known or unknown parameters. They are summarized in Table 6.3. The U model is a special LL model and is not listed in the table.

6.1.5 Example 6.1 (Revisited)

Revisiting the cannabis example, we fit in Table 6.1 the R, C, and RC models. The test statistic values along with their corresponding significance are provided in Table 6.4. The estimates of parametric scores as well as the values of the fixed scores for these models are provided in Table 6.5. The estimated score parameters for the rows and the columns are close to be equidistant for two successive score parameters. Thus it seems not to be worth to adopt a more complex model than U.

Table 6.4 LR goodness-of-fit tests for the independence and the association models applied in Table 6.1

| Model | G^2 | d.f. | p -value |
|-------|----------|------|------------|
| I | 152.7933 | 6 | 0.0000 |
| U | 1.4687 | 5 | 0.9167 |
| C | 1.1004 | 4 | 0.8942 |
| R | 1.2964 | 3 | 0.7230 |
| RC | 0.6044 | 2 | 0.7392 |

Table 6.5 ML estimates for parameters and fixed scores values for the U, R, C, and RC models applied in Table 6.1. Values in *italics* correspond to fixed scores

| | U | R | C | RC |
|--------------|----------------|----------------|----------------|---------|
| $\hat{\phi}$ | 2.5382 | 2.4776 | 2.4638 | 2.3191 |
| α_1 | <i>-0.6708</i> | -0.6640 | <i>-0.6708</i> | -0.6494 |
| α_2 | <i>-0.2236</i> | -0.2238 | <i>-0.2236</i> | -0.2365 |
| α_3 | <i>0.2236</i> | 0.2043 | <i>0.2236</i> | 0.1880 |
| α_4 | <i>0.6708</i> | 0.6836 | <i>0.6708</i> | 0.6979 |
| ν_1 | <i>-0.7071</i> | <i>-0.7071</i> | -0.7331 | -0.7447 |
| ν_2 | <i>0.0000</i> | <i>0.0000</i> | 0.0553 | 0.0825 |
| ν_3 | <i>0.7071</i> | <i>0.7071</i> | 0.6779 | 0.6622 |

This is verified also from the corresponding goodness-of-fit tests, where we see that moving from the simple U model to less parsimonious association models, the fit improvement is very minor. A more detailed discussion on association model selection will be carried out in Sect. 6.3.

Focusing on the C and RC models, the estimated local odds ratios under these models are provided in Table 6.6. Recall that under the U model, the common local odds ratios estimate is 2.23. We can verify that the expected local odds ratios under the C model are column dependent, i.e., the value is common in each column but differs from column to column while under the more general RC model they are row and column dependent, thus all different to each other. However, the estimated local odds ratios are not that different to justify the use of more complicated models than the simple U model, which was of impressive fit. Under the C model, the odds of having tried cannabis once or twice vs. never tried is 2.38 times higher for those who are one level higher in the alcohol consumption scale, no matter what this level is. The odds in the second column of Table 6.6 can be interpreted similarly. In this example we did not refer at all at model R since it is less parsimonious of C and of worse fit.

Table 6.6 Estimated local odds ratios under the RC model and under the C model (in parentheses) for the students' survey about cannabis use at the University of Ioannina, Greece (1995)

| Alcohol consumption | I tried cannabis... | | |
|---------------------|---------------------|---------------|------------|
| | Never | Once or twice | More often |
| At most once/month | 2.21 (2,38) | 1,74 (1,99) | |
| Twice/month | 2,26 (2,38) | 1,77 (1,99) | |
| Twice/week | 2,66 (2,38) | 1,98 (1,99) | |
| More often | | | |

6.2 Maximum Likelihood Estimation for Association Models

For any association model, the maximum likelihood estimation approach is that described in Sect. 4.2 for log-linear models. Thus, independent of the underlying sampling scheme, ML estimates of an association model's parameters and eventually of its expected cell frequencies m_{ij} are achieved by maximizing the *Poisson log-likelihood kernel* (4.13) with respect to the parameters of the model. Substituting m_{ij} by the association model expression and equating the partial derivative of (4.13) with respect to a parameter of the model to zero, one is led to the likelihood equation corresponding to this parameter.

The likelihood equations with respect to the main effect parameters of model (6.4) are the same as the corresponding of the two-way standard log-linear model, given in (4.14). For the more general RC model where both set of scores are parameters, the likelihood equations for the row scores $(\alpha_1, \dots, \alpha_I)$ are derived as

$$\sum_j v_j(\hat{m}_{ij} - n_{ij}) = 0, \quad i = 1, \dots, I \tag{6.12}$$

while for the column scores (v_1, \dots, v_J) as

$$\sum_i \alpha_i(\hat{m}_{ij} - n_{ij}) = 0, \quad j = 1, \dots, J. \tag{6.13}$$

Finally, the likelihood equation corresponding to the intrinsic association parameter φ is

$$\sum_{i,j} \alpha_i v_j(\hat{m}_{ij} - n_{ij}) = 0. \tag{6.14}$$

For the rest of the association models defined by (6.11) by considering the row or column scores or both of them as fixed, the likelihood equations are derived from the above set by eliminating the equations corresponding to known scores. Thus, the likelihood equations for the U model are (4.14) and (6.14) while for the R model (4.14), (6.12), and (6.14). Analogously, the likelihood equations for model C

are (4.14), (6.13), and (6.14). Note that the likelihood equation (6.14) is redundant given (6.12) or (6.13). This means that parameter φ is redundant whenever at least one set of scores is parametric and can thus be eliminated. At this point it is worth mentioning that the RC model was introduced by Goodman (1979b, model (4.1b)) in terms of the non-redundant parameters, as

$$m_{ij} = \alpha_i \beta_j \exp(\alpha_i v_j), \quad i = 1, \dots, I; \quad j = 1, \dots, J, \quad (6.15)$$

with unconstrained row and column scores. Introducing the intrinsic association parameter φ with the cost of imposing constraints (6.5) and (6.6) on the scores, Goodman (1979b, model (4.5b)) proposed the equivalent expression

$$m_{ij} = \alpha_i \beta_j \exp(\varphi \alpha_i v_j), \quad i = 1, \dots, I; \quad j = 1, \dots, J. \quad (6.16)$$

This way the RC model and its scores are comparable to other standard models (Chap. 7). The multiplicative form (6.16) is also equivalent to the log-form (6.4).

The ML estimates of the parameters of any association model cannot be derived in closed-form expression and the corresponding likelihood equations have to be solved iteratively. The simplest iterative procedure for association models ML estimation is based on the *Newton's unidimensional method*. The updating equations (at the t th iteration) for the RC model parameters' estimation, based on expression (6.16), are

$$\begin{aligned} \alpha_i^{(t)} &= \alpha_i^{(t-1)} \frac{n_{i+}}{\tilde{m}_{i+}}, & i &= 1, \dots, I, \\ \beta_j^{(t)} &= \beta_j^{(t-1)} \frac{n_{+j}}{\tilde{m}_{+j}}, & j &= 1, \dots, J, \\ \alpha_i^{(t)} &= \alpha_i^{(t-1)} + \frac{\sum_j v_j^{(t-1)} (n_{ij} - \tilde{m}_{ij})}{\tilde{\varphi}^{(t-1)} \sum_j \left(v_j^{(t-1)} \right)^2 \tilde{m}_{ij}}, & i &= 1, \dots, I, \\ v_j^{(t)} &= v_j^{(t-1)} + \frac{\sum_i \alpha_i^{(t-1)} (n_{ij} - \tilde{m}_{ij})}{\tilde{\varphi}^{(t-1)} \sum_i \left(\alpha_i^{(t-1)} \right)^2 \tilde{m}_{ij}}, & j &= 1, \dots, J, \\ \varphi^{(t)} &= \varphi^{(t-1)} + \frac{\sum_i \alpha_i^{(t-1)} v_j^{(t-1)} (n_{ij} - \tilde{m}_{ij})}{\tilde{\varphi}^{(t-1)} \sum_{i,j} \left(\alpha_i^{(t-1)} v_j^{(t-1)} \right)^2 \tilde{m}_{ij}}, \end{aligned}$$

where \tilde{m}_{ij} stands for the ML estimate of m_{ij} , recalculated at each step of the iterations (Goodman 1979b).

As in every iterative procedure the assignment of initial values to the parameters' estimates is crucial. In this setup, a reasonable choice for the main effects is

$$\alpha_i^{(0)} = \exp\left(\frac{\ell_{i+}}{J} - \frac{\bar{\ell}}{2}\right) \quad \text{and} \quad \beta_j^{(0)} = \exp\left(\frac{\ell_{+j}}{I} - \frac{\bar{\ell}}{2}\right),$$

where $\ell_{ij} = \log(n_{ij})$ and $\bar{\ell} = \frac{\ell_{++}}{IJ}$. A natural choice for the initial estimates of the parametric scores is to consider them equidistant for successive categories, i.e., as if the U model was applied. In this case, starting by considering the scores equal to the corresponding category index and rescaling them linearly so that constraints (6.5) and (6.6) are satisfied, we conclude to

$$\alpha_i^{(0)} = \sqrt{\frac{3}{I(I^2-1)}}(2i - I - 1) \text{ and } v_j^{(0)} = \sqrt{\frac{3}{J(J^2-1)}}(2j - J - 1).$$

A compatible then choice for $\varphi^{(0)}$ would be $\varphi^{(0)} = \sum_{i,j} \alpha_i v_j \log n_{ij}$; see (6.7). The algorithm convergence is checked through the change in the log-likelihood value (4.13), calculated after each parameters' estimates updating circle.

The standard algorithms normally applied are the *Newton–Raphson's* or the *Fisher's scoring* algorithm (see Sect. 5.3.1). In this context, the parameters of the under consideration association model has to be written in a vector form. For example, for the U model the parameter vector is $\boldsymbol{\beta} = (\lambda, \lambda_1^X, \dots, \lambda_{I-1}^X, \lambda_1^Y, \dots, \lambda_{J-1}^Y, \varphi)$. The Newton's unidimensional method is simpler, since it does not require matrix inversion but for this with the drawback that it does not estimate the s.e. of the parameters.

Information on available software and special programs for estimation of association models based on each of these algorithms will be provided in Sect. 6.6.

6.3 Association Model Selection

We have already faced in the context of the cannabis example the problem of selecting the appropriate association model when more than one of them is of adequate fit. The problem of model selection in the framework of association models is connected to the analysis of association (ANOAS) in a contingency table and is based on the interconnection between the models. In particular, it holds

$$I \subset U \text{ (or LL)} \subset R \text{ (or C)} \subset RC.$$

Indeed, the I model is the U (or LL) model with $\varphi = 0$, while the C model, for example, is the RC model for a specific choice for the row scores. This means that

$$G^2(I) > G^2(U) > G^2(C) > G^2(RC),$$

for example, with the analogous results for the LL or the R model. The crucial question at this point is whether the reduction in G^2 value as we move to less parsimonious models is worth, justifying the loss in df and simplicity. The answer is provided through the conditional testing procedure (see Sects. 4.6 and 5.3.4). As soon as we detect the simplest association model \mathcal{M}_1 of adequate fit, we abandon it in favor of a more complicated \mathcal{M}_2 ($\mathcal{M}_1 \subset \mathcal{M}_2$) only if the reduction in G^2 is statistically significant. Hence, we proceed testing the fit of \mathcal{M}_1 conditional on the fact that \mathcal{M}_2 holds by (4.34).

Thus, for example, given that the U, R, or C models hold, one could propose the conditional tests of independence $G^2(I|U)$, $G^2(I|R)$, or $G^2(I|C)$, being asymptotically distributed as χ^2 with df equal to 1, $I - 1$, or $J - 1$, respectively. These conditional tests of independence, given that model U, R, or C holds, are of greater asymptotic power, compared to the traditional unconditional test of independence (Gross 1981; Agresti 1983a). The tests $I|U$ and $I|LL$ are special mentioned since they are most powerful as 1 df tests. In this context it is important to note that the conditional test $I|RC$ is not that straightforward since $G^2(I|RC) = G^2(I) - G^2(RC)$ is *not* asymptotically χ^2 distributed with $df = df(I) - df(RC)$ as probably expected. The asymptotic null distribution of $G^2(I|RC)$ for testing independence is that of the largest eigenvalue from a Wishart distributed matrix (Haberman 1981). Gradually conditional testing from the RC to I, such as $I|U$, $U|R$, and $R|RC$, is possible and provides an analysis of association (ANOAS) table, throwing light on the underlying association structure of the table and analyzing deviance from independence in terms of source (overall, row, interaction) in a manner analogous to the ANOVA table (Goodman 1981a).

6.3.1 Model Selection for Example 6.1

We have already seen that for the cannabis data set all association models provide an acceptable fit. It seems natural to favor the C model over the R, due to parsimony and better fit. Thus, the choice lies between the U, C, and RC models. By the conditional testing procedure one has $G^2(C|RC) = G^2(C) - G^2(RC) = 0.496$, which is non-significant based on the χ^2_2 distribution (p -value=0.7804). Thus, there is no point in adopting the RC model since it does not provide a significant improvement of the fit over the C model. Further on, since $G^2(U|C) = G^2(U) - G^2(C) = 0.3683$ is again non-significant (p -value=0.5439, $df = 1$), the model that seems to be appropriate for this data set is the simple U model, with just 1 df less than the independence and a straightforward interpretation of constant local association all over the table. This sequence of conditional testing is summarized in the ANOAS table, provided for this example at the end of Sect. 6.6.1.

6.4 Features of Association Models

We have mentioned that the LL model (U as well) is invariant under linear transformations of its scores. Actually, this property holds for all association models considered so far. Let $\alpha_i^* = a_1 \alpha_i + b_1$ and $v_j^* = a_2 v_j + b_2$ be any choice of linear rescaling for the row and column scores, respectively. Then in terms of the local odds ratios and the new scores, the association model would be defined as

$$\log \theta_{ij}^{kl} = \varphi^*(\alpha_k^* - \alpha_i^*)(v_\ell^* - v_j^*), \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1,$$

for $i < k \leq I$ and $j < \ell \leq J$. This is further transformed to

$$\log \theta_{ij}^{k\ell} = a_1 a_2 \varphi^* (\alpha_k - \alpha_i)(v_\ell - v_j),$$

which for $\varphi^* = \frac{\varphi}{a_1 a_2}$ is equivalent to (6.3). Thus, without affecting the expected cell frequencies, their estimates, and consequently the fit of the model, we can replace the normalizing constraints on the scores by the *weighted normalizing constraints*:

$$\sum_i w_{1i} \alpha_i = \sum_j w_{2j} v_j = 0 \quad \text{and} \quad \sum_i w_{1i} \alpha_i^2 = \sum_j w_{2j} v_j^2 = 1. \quad (6.17)$$

Although the choice of weights does not affect the model fit, it has an impact on the scores' values and thus issues related to or depending on them. The most common choices for weights are uniform ($w_{1i} = w_{2j} = 1, i = 1, \dots, I, j = 1, \dots, J$) or *marginal* ($w_{1i} = \pi_{i+}, i = 1, \dots, I$ and $w_{2j} = \pi_{+j}, j = 1, \dots, J$). Uniform weights are preferred when the marginal distributions are not fixed and interest lies on comparing tables with unequal marginal distributions. The marginal weights are the choice when scores of association models have to be compared to correspondence analysis results (see Sect. 7.2) or when merging rows and/or columns of a table is the issue (see Sect. 7.5). For a more detailed discussion on the choice of the weighting system, please see Goodman (1985, 1991) or Becker and Clogg (1989).

Replacing the standard constraints (6.5) and (6.6) by the more general (6.17) and working analogously as for deriving (6.7), the intrinsic association parameter φ satisfies

$$\varphi = \sum_{i,j} w_{1i} w_{2j} \alpha_i v_j \log \pi_{ij},$$

i.e., it is a weighted measure of correlation between the row and columns of the table. However, as already stated, parameter φ is redundant in models R, C, and RC.

Models LL, U, R, and C are log-linear while RC is log-multiplicative (not linear in its parameters). As already mentioned, models LL and U require that both classification variables of the contingency table are ordinal and thus are sensitive in re-ordering of rows or columns. Similarly, model R (C) is invariant in re-ordering of the rows (columns) of the table and the corresponding classification variable needs not necessarily be ordinal. Ordinality is required only for columns (rows). Finally, the RC model is invariant in re-ordering of columns or rows. Hence, it can also be applied to tables with nominal classification variables. Overall, parametric scores in an association model can correspond either to nominal underlying classification variable or to ordinal with unknown distances between successive categories. Thus, the parametric scores of models R, C, and RC need not necessarily be monotone. Lack of monotonicity implies non-monotone association, in the sense that local association will be positive in some areas of the table and negative in others.

Thus, monotonicity of the row and column scores is naturally connected to positive dependence and stochastic ordering of the conditional distributions in rows or columns of the table. In particular, Goodman (1981a) showed that under the RC model, association is isotropic and tables possessing this property are TP₂, i.e.,

$\theta_{ij} \geq 1$ for all $i = 1, \dots, I-1, j = 1, \dots, J-1$ with at least one strict inequality (see also Sect. 2.5.5). As indicated by (6.10), in case the row and column scores are both ordered and of the same ordering (i.e., both increasing or both decreasing), $\varphi > 0$ is equivalent to positive dependence and consequently the conditional row or column probabilities are stochastically ordered. This means that if \mathbf{X}_i and $\mathbf{X}_{i'}$ are the conditional row distributions of rows i and i' with $i < i'$, then the positive dependence implies $\mathbf{X}_i \leq_{\text{st}} \mathbf{X}_{i'}$, e.g., \mathbf{X}_i is stochastically smaller than $\mathbf{X}_{i'}$. The distribution of \mathbf{X}_i is said to be stochastically smaller than that of $\mathbf{X}_{i'}$, if $F_{\mathbf{X}_i}(t) \geq F_{\mathbf{X}_{i'}}(t)$, for all $t = 1, \dots, J$, where $F_{\mathbf{X}}$ is the cumulative distribution function of \mathbf{X} .

In general, it is not ensured that the ML estimates of monotone parametric scores will be monotone as well. If the ML estimates are non-monotonic, then one can proceed in order-restricted estimation of the corresponding association model (see Sect. 6.8.2).

Another nice property of association models is their connection to the bivariate normal distribution. In fact, association models lead to very good approximations of the discretized bivariate normal distribution (Goodman 1981b, 1985; Wang 1987; Becker 1989a; Rom and Sarkar 1990). To see this, consider the bivariate normal density

$$f(x, y; \alpha_x, \alpha_y, \sigma_x, \sigma_y, \rho) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left(-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\alpha_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\alpha_x}{\sigma_x}\right)\left(\frac{y-\alpha_y}{\sigma_y}\right) + \left(\frac{y-\alpha_y}{\sigma_y}\right)^2 \right]\right)$$

and partition the \mathfrak{R}^2 surface in small rectangular regions $(a_{i-1} \times a_i) \times (b_{j-1} \times b_j)$, where $i = 1, \dots, I, j = 1, \dots, J, a_0 = b_0 = -\infty$, and $a_I = b_J = +\infty$. Then, the U model, or more precisely the symmetric U model (with $I = J$ and $\alpha_i = v_i$), applied in the table formed by this partition approximated well the discretization of the above density. For standardized scores, parameter φ is analogue to $\frac{\rho}{1-\rho^2}$ of the normal density.

Finally, we would like to emphasize that beyond the sophisticated insight in the structure of the underlying association, if significant, one of the major strong points of the association models is the ability for conditional testing of independence, as already discussed in Sect. 6.3.

6.5 Association Models of Higher Order: The RC(M) Model

The RC model, though the less parsimonious association model considered so far and in spite of its impressive abilities and often impressive fit, is not always adequate. RC itself imposes a restrictive structure which can sometimes be insufficient to model the underlying association. It leaves $(I-2)(J-2) df$, enough space for more in-between models for building up the interaction until the saturated model is reached.

Indeed, one could consider to add more multiplicative terms of the RC-type. For example, the next model to consider would be

$$\log m_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \varphi_1 \varpi_1 v_{j1} + \varphi_2 \varpi_2 v_{j2}, \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

In fact, this idea can be extended further, as long as I and J are large enough, since in the saturated model, there are $(I - 1)(J - 1)$ association parameters. Thus, considering M association terms for the general case, we are led to the model

$$\log m_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \sum_{m=1}^M \varphi_m \varpi_m v_{jm}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad (6.18)$$

denoted by RC(M).

How large can M be? To answer this question one must see what M represents. The concept behind the RC(M) general association model is that of dimensionality of the underlying association and its decomposition in axes. The idea is the same as in other well-known methods of reduction of dimensionality, such as factor analysis and principal component analysis. As in these methods, for identifiability purposes as well as for convenience of interpretation, the axes to which the association is decomposed are considered to be orthogonal. In our framework, the key for this decomposition is the *singular value decomposition* (SVD) of the interaction parameters matrix $\mathbf{A} = \left(\lambda_{ij}^{XY} \right)_{I \times J}$ of the saturated log-linear expression. Thus, M is the rank of matrix \mathbf{A} , the parameters φ_m ($m = 1, \dots, M$) are the associated eigenvalues while the row and column scores for a certain m are the components of the m th corresponding eigenvector. In particular, the SVD of the interaction matrix \mathbf{A} gives

$$\mathbf{A} = \mathbf{M}\boldsymbol{\varphi}\mathbf{N}'$$

where $\boldsymbol{\varphi} = \text{diag}(\varphi_1, \dots, \varphi_M)$ with $\varphi_1 \geq \dots \geq \varphi_M > 0$ are the eigenvalues while the eigenvectors $\varpi_m = (\varpi_{1m}, \dots, \varpi_{jm})$ and $\mathbf{v}_m = (v_{1m}, \dots, v_{jm})$, associated to the m th eigenvalue, form the matrices $\mathbf{M}_{J \times M} = (\varpi_{jm})$ and $\mathbf{N}_{J \times M} = (v_{jm})$, respectively. \mathbf{M} and \mathbf{N} are orthonormal, e.g., they satisfy

$$\mathbf{M}'\mathbf{M} = \mathbf{N}'\mathbf{N} = \mathbf{I}_M$$

where \mathbf{I}_M is the M th order identity matrix. The maximum possible value for the dimension of the decomposition M is $M^* = \min(I, J) - 1$. Thus, model (6.18) can be considered for $0 \leq M \leq M^*$. The associated degrees of freedom equal $df[\text{RC}(M)] = (I - M - 1)(J - M - 1)$. Model RC(0) is the independence model, RC(1) is the RC while RC(M^*) is the saturated model. The orthonormality of the eigenvectors is equivalently expressed as

$$\begin{aligned}\sum_i \alpha_{im} &= \sum_j v_{jm} = 0, \\ \sum_i \alpha_{im}^2 &= \sum_j v_{jm}^2 = 1, \quad m, \ell = 1, \dots, M, \\ \sum_i \alpha_{im} \alpha_{i\ell} &= \sum_j v_{jm} v_{j\ell} = 0, \quad m \neq \ell.\end{aligned}$$

Note that the first two restrictions are the identifiability constraints we have already imposed on the row and column scores of the RC model for uniform weights while the last one corresponds to the orthogonality of the dimensions. In order to generalize the above constraints and allow the use of weights, the *generalized singular value decomposition* (GSVD) of the interaction matrix Λ has to be applied instead of the SVD. By GSVD, \mathbf{M} and \mathbf{N} are orthonormalized with respect to the weights

$$\mathbf{W}_1 = \text{diag}(w_{11}, \dots, w_{1I}) \text{ and } \mathbf{W}_2 = \text{diag}(w_{21}, \dots, w_{2J}),$$

e.g., they satisfy

$$\mathbf{M}'\mathbf{W}_1\mathbf{M} = \mathbf{N}'\mathbf{W}_2\mathbf{N} = \mathbf{I}_M,$$

or equivalently, the row and column scores satisfy the constraints:

$$\begin{aligned}\sum_i w_{1i} \alpha_{im} &= \sum_j w_{2j} v_{jm} = 0, \quad m = 1, \dots, M, \\ \sum_i w_{1i} \alpha_{im} \alpha_{i\ell} &= \sum_j w_{2j} v_{jm} v_{j\ell} = \delta_{m\ell}, \quad m, \ell = 1, \dots, M,\end{aligned}\tag{6.19}$$

where $\delta_{m\ell}$ is Kronecker's delta.

Analogously to the RC, the RC(M) model can alternatively be expressed by the multiplicative form, used by Goodman:

$$m_{ij} = \alpha_i \beta_j \exp \left(\sum_{m=1}^M \varphi_m \alpha_{im} v_{jm} \right), \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

However, the most convenient expression for physical interpretation is in terms of the local odds ratios

$$\log \theta_{ij}^L = \sum_{m=1}^M \varphi_m (\alpha_{im} - \alpha_{i+1,m}) (v_{jm} - v_{j+1,m}), \tag{6.20}$$

for $i = 1, \dots, I-1$, $j = 1, \dots, J-1$.

6.5.1 Maximum Likelihood Estimation of the RC(M) Model

The estimation procedure for the RC(M) follows the lines of the procedure described in Sect. 6.2 for the simple RC model. The extension for the RC(M) model is straightforward. Thus it can be proved that the likelihood equations for the RC(M) model are the (4.14) for the main effects while the likelihood equations corresponding to the row and column scores and the association parameters ϕ_m 's are

$$\sum_j v_{jm}(\hat{m}_{ij} - n_{ij}) = 0, \quad i = 1, \dots, I, \quad m = 1, \dots, M, \quad (6.21)$$

$$\sum_i \infty_{im}(\hat{m}_{ij} - n_{ij}) = 0, \quad j = 1, \dots, J, \quad m = 1, \dots, M, \quad (6.22)$$

$$\sum_{i,j} \infty_{im} v_{jm}(\hat{m}_{ij} - n_{ij}) = 0, \quad m = 1, \dots, M, \quad (6.23)$$

i.e., straightforward extensions of (6.12), (6.13), and (6.14), respectively.

In practice, the updating equations of the simple Newton's unidimensional method for the interaction parameters of RC(M) are direct extensions of the corresponding updating equations for the RC model, presented in Sect. 6.2, while the updating equations for the main effects remain the same. The orthonormal constraints that must be satisfied by the scores of RC(M) need not to be enrolled in the iterative procedure. Since it is only a matter of parameters' identifiability and rescaling that does not affect the cell estimates, it is sufficient if they are fulfilled by the initial values and if the final estimated interaction parameters are rescaled by SVD at the final stage, after the convergence of the algorithm is achieved. The initial values $\phi_m^{(0)}$, $\infty_{im}^{(0)}$, and $v_{jm}^{(0)}$ ($m = 1, \dots, M$) can easily be obtained as the corresponding values of the first M terms of the SVD of the observed interaction matrix, e.g., matrix Γ with entries $\gamma_{ij} = \frac{n_{ij}}{\alpha_i^{(0)}\beta_j^{(0)}}$. The extension of the Newton-Raphson algorithm, presented in Sect. 6.2, is also straightforward.

6.5.2 Example 6.2

The data considered in Table 6.7 are from Wermuth and Cox (1998) and cross-classify people in West Germany (Central archive, 1993) according to their type of schooling completed and their age in a 5×5 table. As can be observed in Table 6.8, there exists a highly significant association between age and type of schooling which is not captured by the RC model. Hence, the consideration of an association model RC(M) with $M > 1$ is necessary. The RC(2) model is of very good fit and is the model we propose for this data set and base inference on.

In the context of the RC model, we have seen that the important information lies not on the values of the row and column scores themselves but on their distances

Table 6.7 Cross-classification of 3,673 subjects according to their age and type of school attended, West Germany 1991/92

| Type of schooling | Age group | | | | |
|-------------------|------------------|------------------|------------------|------------------|------------------|
| | 18–29 | 30–44 | 45–59 | 60–74 | >74 |
| Basic, incomplete | 12 (11.943) | 13 (13.068) | 12 (12.008) | 20 (20.994) | 7 (5.987) |
| Basic, complete | 215 (215.823) | 507 (504.677) | 493 (495.243) | 460 (458.267) | 137 (137.990) |
| Medium | 277 (273.739) | 300 (309.576) | 192 (182.462) | 126 (129.431) | 38 (37.792) |
| Upper medium | 52 (51.859) | 91 (91.307) | 47 (46.776) | 15 (16.149) | 6 (4.909) |
| Intensive | 233 (235.637) | 225 (217.372) | 102 (109.510) | 74 (70.158) | 19 (20.323) |

In parentheses are given the ML estimates under the RC(2) model

Table 6.8 G^2 statistics for the fit of independence and association models applied in Table 6.7

| Model | G^2 | d.f. | p -value |
|-------|---------|------|------------|
| I | 357.146 | 16 | 0.000 |
| RC | 24.275 | 9 | 0.039 |
| RC(2) | 2.599 | 4 | 0.627 |

for successive categories. Distances are the quantities that are interpreted in terms of closeness of the effect of the underlying categories on odds formed by categories of the other classification variable. For the RC(M) model with $M > 1$, the logic of interpretation is the same, as can easily be verified by definition (6.20). However, distances between rows (or columns) are now defined by the Euclidean distance in the M dimensional space. For $M = 2$, this is easily visualized on the two-dimensional space, with the i -th row ($i = 1, \dots, I$) and the j -th column ($j = 1, \dots, J$) being represented by the points $(\hat{\alpha}_{i1}, \hat{\alpha}_{i2})$ and $(\hat{\nu}_{j1}, \hat{\nu}_{j2})$, respectively. For our example, Fig. 6.1 presents such graphs for scores satisfying constraints (6.19) subject to the uniform (left) or marginal (right) weights. The MLEs of the scores (under uniform weights) are provided in Table 6.9.

These two graphs, though they obviously refer to different scores' values, they correspond to equivalent expressions of the RC(2) estimates just differently scaled through the choice of the weights. It is evident from the plots that the 2nd dimension captures the differentiation of row 1 from 2 (incomplete from complete basic education) and 4 from 3 and 5 (upper medium from medium and intensive education). The closeness of columns 4 and 5 (ages 60–74 or > 74) is remarkable, especially in the marginal weights plot, where they are almost indistinguishable. This observation motivates Sect. 7.5 on merging categories, where this example is revisited (Sect. 7.5.1). Though the marginal weighted scores are preferred for comparisons in rows (or in columns), the uniform weighted are more appropriate for investigating the row–column combinations of strong association. We can observe,

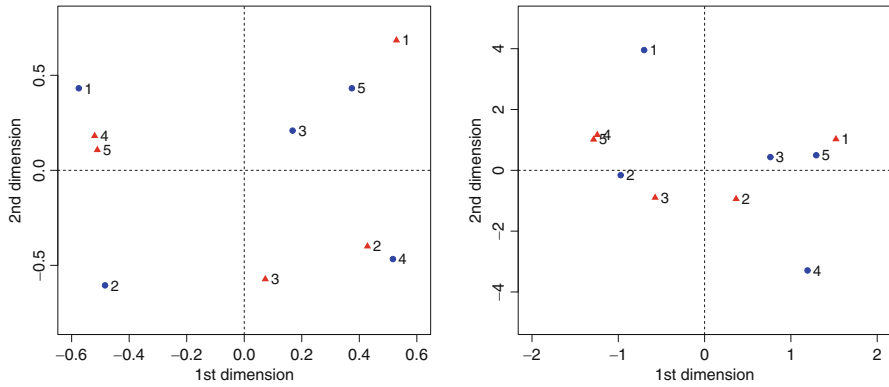


Fig. 6.1 Plots of the estimated row (*bullets*) scores ($\hat{\alpha}_{i1}, \hat{\alpha}_{i2}$), $i = 1, \dots, 5$, and column (*triangles*) scores ($\hat{\nu}_{j1}, \hat{\nu}_{j2}$), $j = 1, \dots, 5$, under the RC(2) model applied to Table 6.7, with respect to uniform (*left*) and marginal (*right*) weights

Table 6.9 Association parameters’ ML estimates for the RC(2) model applied in Table 6.7

| | $\varphi_1 = 1.7830$ | | $\varphi_2 = 0.6904$ | |
|---|----------------------|---------|----------------------|---------|
| | ∞-scores | | ν-scores | |
| | $m = 1$ | $m = 2$ | $m = 1$ | $m = 2$ |
| 1 | -0.575 | 0.431 | 0.529 | 0.684 |
| 2 | -0.484 | -0.605 | 0.428 | -0.400 |
| 3 | 0.168 | 0.209 | 0.073 | -0.573 |
| 4 | 0.517 | -0.467 | -0.520 | 0.181 |
| 5 | 0.374 | 0.432 | -0.511 | 0.107 |

The estimates are subject to the orthonormal constraints with uniform weights

for example, that upper medium education (row 4) is stronger associated with people aged 30–44 (column 2). Also, as expected basic incomplete education (row 1) is more often among elder people (columns 4 and 5). Alternatively, one could apply Correspondence Analysis (CA) and conclude to very similar results. The CA of this data set is provided in Sect. 7.2.2.

6.6 Software Applications for Association Models

Association models, though so powerful tools in modeling the association in contingency tables, did not receive the attention one would expect. The major reason for that is the fact that their fit is not provided as a standard option in statistical software. They can be fitted in statistical packages, but some extra programming is required. Additionally, a Fortran algorithm for ML estimation of the RC(M) model by the Newton’s unidimensional method has been provided by Becker (1990a) while the Newton–Raphson algorithm has been implemented in Fortran by Haberman

(1995) and a Fisher's scoring type algorithm using the weighted least squares as initial estimates by Ait-Sidi-Allal et al. (2004). As already mentioned in Sect. 6.2, the above algorithms are appropriate for the estimation and fit of models linear in their parameters, i.e., models U, R, and C. In case of $RC(M)$, $M \geq 1$, we still apply these methods by considering at each step of estimation for the row (columns) scores that the column (row) scores are fixed at the estimated value of the previous step. This procedure is continued until convergence is achieved.

Association models which are linear in their parameters, e.g., the models U (and LL), R, and C, are log-linear and can be fitted as GLM by any available software, adopting the procedure described next for R. In the web appendix (see Sect. A.4) are also available syntax codes for automatized fitting of all the association models in SPSS, including the $RC(M)$.

6.6.1 Association Models in R: Example 6.1

The simple association models U (or LL), R, and C can be fitted in R straightforward, in the generalized linear models framework by the `glm()` function of R, as described below.

First of all, the data have to be in the standard format for fitting classical log-linear models. Thus, let `freq`, `row`, and `col` be the usual variables of a data frame corresponding to the vectors of observed frequencies and row and column classification variables, respectively. Then, we have to construct the variables of row and column scores, $\mu_i = \text{row}$ and $\nu_j = \text{col}$, respectively. This way, the row and column scores are set equal to $\alpha_i = i$, $i = 1, \dots, I$ and $\nu_j = j$, $j = 1, \dots, J$. In the sequel, `row` and `col` have to be defined as factors and then the U, R, and C models are the log-linear models with terms in the model `row + col + mu:nu`, `row + col + row:nu` and `row + col + mu:col`, respectively. The LL model can be fitted as the U model with the only difference that the score variables `mu` and `nu` will now contain the values of the prefixed, not equidistant scores for the corresponding row and column categories.

To illustrate, let us consider the cannabis example. The data are saved under the data frame `cannabis.fr`. Models U, R, and C are fitted by `glm()` as follows:

```
> freq <- c(204, 6, 1, 211, 13, 5, 357, 44, 38, 92, 34, 49)
> row <- rep(1:4, each=3); col <- rep(1:3, 4)
> mu <- row; nu <- col
> row <- factor(row); col <- factor(col)
> cannabis.fr <- data.frame(freq, row, col, mu, nu)
> model.U <- glm(freq~row+col+mu:nu, poisson, data=cannabis.fr)
> model.R <- glm(freq~row+col+row:nu, poisson, data=cannabis.fr)
> model.C <- glm(freq~row+col+mu:col, poisson, data=cannabis.fr)
```

Table 6.10 Output of the U model fit in R for the cannabis data (Table 6.1)

```

Call:
glm(formula = freq~row + col + mu:nu, family = poisson, data = cannabis.fr)

Deviance Residuals:
    1     2     3     4     5     6     7
-0.03133  0.13352  0.13252 -0.02757 -0.02850  0.23335  0.22138
    8     9    10    11    12
-0.69889  0.10336 -0.34184  0.82418 -0.17906

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.51766    0.10017  45.098 < 2e-16 ***
row2         -0.76921    0.12181  -6.315 2.70e-10 ***
row3         -1.05962    0.17968  -5.897 3.70e-09 ***
row4         -3.17104    0.30478 -10.404 < 2e-16 ***
col2         -4.38621    0.25357 -17.298 < 2e-16 ***
col3         -7.06112    0.53471 -13.205 < 2e-16 ***
mu:nu        0.80265    0.07827  10.255 < 2e-16 ***
--
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1357.7001 on 11 degrees of freedom
Residual deviance: 1.4687 on 5 degrees of freedom
AIC: 79.655
Number of Fisher Scoring iterations: 4

```

for models U, R, and C, respectively. The output is then derived, for the U model, for example, through the command

```
> summary(model.U)
```

and is provided in Table 6.10.

Note that by the above procedure, the scores involved in the interaction term are not standardized, they are equal to the corresponding category index and thus they do not satisfy constraints (6.5) and (6.6). For these scores, we have $\hat{\phi} = 0.80265$, as given in Sect. 6.1.2. This, however, does not affect the ML estimates of the common value of all expected under U local odds ratios or of the expected cell frequencies, which can be obtained by

```
> MLE.U <- xtabs(model.U$fitted.values~row+col)
```

verifying the corresponding entries of Table 6.1.

The R and C models fitted above are defined by (6.15), the parameterization without the intrinsic association parameter ϕ . If we want the models to be in the form (6.4) and the scores to be standardized, then μ and ν have to be rescaled appropriately before applying the model, while the estimates of the parametric scores have to be rescaled at a final stage as well. To simplify this procedure, we conducted for each association model the corresponding R function, namely

the `fit.U()`, `fit.R()` and `fit.C()`, to be found in the web appendix (see Sect. A.3.5). They fit the corresponding model subject to the general constraints (6.17), controlling the weights used by the parameter `iflag`, with the option of uniform (=0) or marginal (=1) weights. Hence, the U model on the cannabis example with marginal weights could be fitted by this function as

```
> U <- fit.U(freq, NI=4, NJ=3, iflag=1)
```

where $NI=I$ and $NJ=J$. Under U, additional to the standard `glm` output, the standardized scores are saved under `U$mu` and `U$nu`, respectively, as well as $\hat{\phi}$ (`U$phi`), the G^2 value (`U$G2`), the degrees of freedom (`U$df`), the p -value (`U$p.value`), and the ML estimates of the expected cell frequencies (`U$fit.freq`). Functions `fit.R()` and `fit.C()` are called analogously and give output of the same format.

The RC and more generally the $RC(M)$ models, $M \geq 1$, cannot be fitted by `glm`, since they are not linear in their parameters and thus not in the GLM family. Thus, these models need special treatment. They can be fitted through functions available in special packages developed for nonlinear models, such as `gnm`, developed by Turner and Firth. An overview of version 1.0–6 is provided by Turner and Firth (2012a). An alternative choice is the `VGAM` package, which deals with Vector Generalized Additive Models (Yee and Wild 1996). For a short presentation of the package, see Yee (2008).

We will illustrate association models by the `gnm` package, based on Turner and Firth (2007). It is designed for models multiplicative in their parameters and defines the product of parameters, corresponding to factors `f1` and `f2`, respectively, through `Mult(f1, f2)`. Thus, the RC model is fitted on our cannabis example by

```
> library(gnm)
```

```
> RC.model <- gnm(freq ~ row + col + Mult(row, col), family = poisson)
```

Recall that `row` and `col` have to be defined as factors before calling the model. Output is printed on the screen by typing

```
> RC.model
```

The output is provided in Table 6.11.

The ML estimates of the expected cell frequencies under the RC model are provided by

```
> predict(RC.model, type = "response", se.fit = TRUE)
```

| | | | | | | |
|-------------------------------|------------|-----------|-----------|------------|-----------|-----------|
| <code>\$fit</code> | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | 204.249905 | 5.393465 | 1.356630 | 211.279836 | 12.320820 | 5.399344 |
| | 7 | 8 | 9 | 10 | 11 | 12 |
| | 355.826805 | 46.847419 | 36.325776 | 92.643454 | 32.438296 | 49.918250 |
| <code>\$se.fit</code> | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | 14.2819358 | 1.9407176 | 0.8911325 | 14.5093040 | 2.8007752 | 1.9625848 |
| | 7 | 8 | 9 | 10 | 11 | 12 |
| | 18.7968741 | 5.6652758 | 5.5879960 | 9.5879954 | 5.3135278 | 6.9616519 |
| <code>\$residual.scale</code> | | | | | | |
| | [1] 1 | | | | | |

The ML estimates of the parameters of the model are printed on screen by typing:

```
> coefficients(RC.model)
```

Table 6.11 Output for the RC model applied on the cannabis example (data in Table 6.1) by `gnm`

```

Call:
gnm(formula=freq~row+col+Mult(row,col), family=poisson)

      (Intercept)          row2          row3
      4.97406         0.20900         0.91031
      row4          col2          col3
     -0.21912        -2.07242        -2.35822
Mult(.,col).row1  Mult(.,col).row2  Mult(.,col).row3
     -0.67224         -0.33122         0.01931
Mult(.,col).row4  Mult(row,.)col1    Mult(row,.)col2
     0.44034         -0.51364         1.80955
Mult(row,.)col3
     3.43750

Deviance: 0.5888162
Pearson chi-squared: 0.5802588
Residual df: 2

```

Furthermore, the command `coef()` gives the ability to save the ML estimates of a parameter in a separate vector in order to be handy for further use. For example, the row main effects estimates can be saved under the vector `a`:

```

> a<-c(0,coef(model.RC)[2:4])
> a

```

| | row2 | row3 | row4 |
|-----------|-----------|-----------|------------|
| 0.0000000 | 0.2090018 | 0.9103069 | -0.2191229 |

Note that the model is fitted through (6.15) and the scores' estimates are not with respect to the constraints (6.17). They can be rescaled linearly though, in order to satisfy them. The `getContrasts()` command of the `gnm` package provides this facility. Thus, for uniform weights, the rescaling is achieved as

```

mu<-getContrasts(model.RC, pickCoef(model.RC,"[.]row"),
+                  ref="mean", scaleWeights="unit")

```

and

```

> nu<-getContrasts(model.RC, pickCoef(model.RC,"[.]col"),
+                  ref="mean", scaleWeights="unit")

```

for the row and column scores, respectively, leading to

```

> mu

```

| | Estimate | Std. Error |
|-------------------|------------|------------|
| Mult(., col).row1 | -0.6494141 | 0.07259224 |
| Mult(., col).row2 | -0.2364548 | 0.10092333 |
| Mult(., col).row3 | 0.1880143 | 0.05142579 |
| Mult(., col).row4 | 0.6978546 | 0.04442136 |

```
> nu
```

| | Estimate | Std. Error |
|-------------------|-------------|------------|
| Mult(row, .).col1 | -0.74474733 | 0.04043242 |
| Mult(row, .).col2 | 0.08252274 | 0.09813196 |
| Mult(row, .).col3 | 0.66222459 | 0.05769954 |

For marginal weights, the vectors of row and column marginal probabilities have to be computed first:

```
> rowProbs<-with(cannabis.fr, tapply(freq,row,sum)/sum(freq))
> colProbs<-with(cannabis.fr, tapply(freq,col,sum)/sum(freq))
```

The rescaling follows then analogously:

```
> mu<-getContrasts(model.RC, pickCoef(model.RC,"[.]row"),
+                   ref=rowProbs, scaleWeights=rowProbs)
> nu<-getContrasts(model.RC, pickCoef(model.RC,"[.]col"),
+                   ref=colProbs, scaleWeights=colProbs)
```

For our example, this leads to

```
>mu
```

| | Estimate | Std. Error |
|-------------------|------------|------------|
| Mult(., col).row1 | -1.5106642 | 0.17070781 |
| Mult(., col).row2 | -0.5686042 | 0.22575503 |
| Mult(., col).row3 | 0.3997124 | 0.08382567 |
| Mult(., col).row4 | 1.5627815 | 0.14715755 |

and

```
>nu
```

| | Estimate | Std. Error |
|-------------------|------------|-------------|
| Mult(row, .).col1 | -0.2849555 | 0.003252048 |
| Mult(row, .).col2 | 0.8920811 | 0.141339476 |
| Mult(row, .).col3 | 1.7168786 | 0.117206021 |

Alternatively, the RC model can be fitted by the function `fit.RC()`, provided in the web appendix (see Sect. A.3.5), with the option of selecting marginal or uniform weights for the constraints (6.17) on the scores. The function is called exactly as `fit.U()` and provides the same type of output. However, this function does not provide the standard errors of the parametric scores. For this, the `getContrasts()` function described above is needed.

The conditional testing between nested association models, when allowed, can be performed by function `anova()`. Thus, for our cannabis example, the ANOAS table based on the conditional tests $G^2(I|U)$, $G^2(U|C)$, and $G^2(C|RC)$ (see Sect. 6.3.1) is produced by

```
> I<-glm(freq~row+col, family=poisson)
> m1<- fit.U(freq,4,3,1)
> m2<- fit.C(freq,4,3,1)
> m3<- fit.RC(freq,4,3,1)
> anova(I,m1$model,m2$model,m3$model,test="Chisq")
```


| Analysis of Deviance Table | | | | | | |
|--|-----------|------------|----|----------|------------|--|
| Model 1: freq ~ row + col | | | | | | |
| Model 2: freq ~ X + Y + mu:nu | | | | | | |
| Model 3: freq ~ X + Y + Y:mu | | | | | | |
| Model 4: freq ~ X + Y + Mult(X, Y) | | | | | | |
| | Resid. Df | Resid. Dev | Df | Deviance | P(> Chi) | |
| 1 | 6 | 152.793 | | | | |
| 2 | 5 | 1.469 | 1 | 151.325 | <2e-16 *** | |
| 3 | 4 | 1.100 | 1 | 0.368 | 0.5439 | |
| 4 | 2 | 0.589 | 2 | 0.512 | 0.7743 | |
| -- | | | | | | |
| Signif. codes: | | | | | | |
| 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | | | |

6.6.2 The RC(M) Model in R: Example 6.2

Association models of order M , $M > 1$, can be fitted in the `gnm` package applying the `instances` argument for the multiplicative term of the model. Thus, for our Example 6.2 (Table 6.7), the RC(2) model can be fitted as

```
> RC2.model <- gnm(freq ~ row+col+instances(Mult(row,col),2),
+               family=poisson)
```

where `freq` is the vector of cell frequencies while `row` and `col` are the factors corresponding to the rows (type of schooling) and columns (age group) of the table, respectively. The `fit.RCm()` function in the web appendix (see Sect. A.3.5) fits the RC(M) model ($M \geq 1$) on a contingency table, read in vector form, and rescales the row and column score vectors through singular value decomposition of the appropriate table, so that constraints (6.19) hold for uniform or marginal weights.

Thus, for Example 6.2, the 5×5 data table is provided in vector form (by rows) as

```
> WCox <- c(12,13,12,20,7,215, 507,493,460,137,
+          277,300,192,126,38,52,91,47,15,6,233,225,102,74,19)
```

and the RC(2) model is fitted by

```
> m <- 2
> RC.m <- fit.RCm(freq=WCox, NI=5, NJ=5, m=2, iflag=1)
```

where the parameter `m` specifies the order of the association model. The derived score vectors are subject to constraints (6.19) with marginal weights. Changing the last argument of `fit.RCm()` from 1 to 0, the uniform weights are applied.

One can save the scores' estimates in order to proceed with the presentation of the results, for example, through appropriate graphs. For Example 6.2, the vectors of row and column scores subject to marginal weights can be saved in vectors `mu1` and `nu1` as

```
> mu1 <- RC.m$mu ; nu1 <- RC.m$nu
```

while subject to uniform weights in `mu0` and `nu0` as

```
> RC.m0 <- fit.RCm(freq=WCoX, NI=5, NJ=5, m=2, iflag=1)
> mu0 <- RC.m0$mu ; nu0 <- RC.m0$nu
respectively.
```

The plot, for example, of the row and column scores' coordinates under uniform weights (Fig. 5.1 (left)) can easily be obtained through the standard `plot()` command, applied on `mu0` and `nu0`. This plot is produced by function `plot_2dim()`, provided in the web appendix (see Sect. A.3.5). The plot in Fig. 5.1 (left) is obtained by calling this function as

```
> plot_2dim(mu0, nu0, -0.6, 0.6, -0.8, 0.8, -0.7, 1.1, 1.2)
```

The parameters of this function following `nu0` control the plot appearance. Thus, $(-0.6, 0.6)$ and $(-0.8, 0.8)$ define the range of values of the first and second axis, respectively. The value set -0.7 leaves a gap of 70% of the text width between the category label and the corresponding plotted symbol. According to the case, it can be adjusted each time for the better appearance of the graph. The size of text characters in axes and labels is set to be 1.1 times the default text size while the size of the symbols and their categories' labels are 1.2 times the default text size. Analogously, the plot in Fig. 5.1 (right) is obtained through

```
> plot_2dim(mu1, nu1, -2, 2, -5, 5, -0.7, 1.1, 1.2)
```

6.6.3 Example 2.4 (Revisited)

Recall the data set on varicella disease in Table 2.5, where 170 children are cross-classified by complication occurrence and age (in a 2×4 table). Independence was rejected (p -value=0.040) and the linear trend test suggested that the linear association is non-significant (p -value=0.104). Fitting association models on this example, we confirm the inappropriateness of linear association since the U model is rejected with $G^2 = 7.093$ (p -value=0.029, $df = 2$). Note that because $I = 2$, the R model is equivalent to the U while the C model is saturated ($G^2 = 0$, $df = 0$). However, derivation of the column scores of the C model is very informative on comparing the different age groups in terms of their association to the complication response. The C model is fitted by function `fit.c` of the web appendix (see Sect. A.3.5). From the corresponding output, the coefficients along with their standard errors and significances are provided below.

| Coefficients: (1 not defined because of singularities) | | | | |
|--|----------|------------|---------|--------------|
| | Estimate | Std. Error | z value | Pr(> z) |
| (Intercept) | 2.1503 | 0.2759 | 7.794 | 6.47e-15 *** |
| X2 | -0.2063 | 0.1944 | -1.062 | 0.2885 |
| Y2 | 0.3980 | 0.3399 | 1.171 | 0.2416 |
| Y3 | 0.2939 | 0.3395 | 0.866 | 0.3867 |
| Y4 | 1.9272 | 0.2759 | 6.986 | 2.84e-12 *** |
| Y1:mu | -0.1522 | 0.2759 | -0.552 | 0.5811 |
| Y2:mu | 0.6024 | 0.2415 | 2.495 | 0.0126 * |
| Y3:mu | 0.2470 | 0.2409 | 1.025 | 0.3053 |
| Y4:mu | NA | NA | NA | NA |
| -- | | | | |
| Signif. codes: | | | | |
| 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

In the estimation procedure above, the parametric score v_4 is redundant and is the reference category (coefficient for $Y_4:\mu$, shown as not defined). The fixed row scores used are $\alpha_1 = -1$ and $\alpha_2 = 1$. The estimated column scores are rescaled to satisfy the marginal weighted constraints (6.17). The rescaled scores and the interaction parameter are also part of the output. In particular, $\hat{\phi} = 0.231$, $\hat{v}_1 = -1.127$, $\hat{v}_2 = 2.136$, $\hat{v}_3 = 0.560$, and $\hat{v}_4 = -0.468$.

Observing that only the column score \hat{v}_2 is significantly different from the others, we conclude that only the category “1–2 years old” relates differently to complications than all other age categories. Thus, we proceed by applying the LL model with the constraint $v_1 = v_3 = v_4$. In R this is easily achieved, as shown next. Before fitting the LL model, we rescale the simple raw scores 1,2, assigned initially to the rows and column categories, through the function `rescale` of the web appendix (see Sect. A.3.5), so that the $\hat{\phi}$, derived by `glm()`, corresponds to the marginally weighted scores:

```
> NI <- 2
> NJ <- 4
> freq <- c(10, 7, 9, 59, 6, 19, 12, 48)
> row<-gl(NI,NJ,length=NI*NJ)
> col<-gl(NJ,1,length=NI*NJ)
> dtable <- data.frame(freq,row,col)
> mu0<-c(1,2)
> nu0<-c(1,2,1,1)
> mu<-rep(rescale(mu0, dtable, 1, 1)$score,each=NJ)
> nu<-rep(rescale(nu0, dtable, 1, 0)$score, NI)
> LL.model <- glm(freq~row+col+mu:nu,poisson)
```

From the summary output, obtained by `summary(LL.model)`, we see that the model is acceptable, since $G^2 = 1.572$ (p -value=0.456, $df = 2$). Commands

```
MLEs <- xtabs(LL.model$fitted.values ~ row + col)
stdres <- xtabs(rstandard(LL.model) ~ row + col)
```

express in table form the ML estimates of the expected frequencies and the corresponding standardized residuals. None of the standardized residuals exceeds 1.96; thus all cells are fitted satisfactorily by the model.

```

> MLEs
      col
row    1      2      3      4
1    8.666667  7.000000 11.375000 57.958333
2    7.333333 19.000000  9.625000 49.041667
> stdres
      col
row    1      2      3      4
1    0.6924565 0.0000000 -1.1684753  0.3975276
2   -0.7328918 0.0000000  1.0833608 -0.4001372

```

The interaction parameter is estimated as $\hat{\phi} = 0.210$ (coefficient for `mu:nu` in the output) while the row and column scores are $\alpha_1 = -1$, $\alpha_2 = 1$, $\nu_1 = \nu_3 = \nu_4 = -0.4249$ and $\nu_2 = 2.3534$ (saved in vectors `mu` and `nu`).

This model provides a clear and strong interpretation. The equality restrictions among the column scores impose on the expected local odds ratios the restrictions $\theta_{13}^L = 1$ and $\theta_{11}^L = 1/\theta_{12}^L$. The odds ratios θ_{11}^{23} and θ_{11}^{24} , opposing age categories 1 to 3 and 1 to 4, respectively, are also equal to 1 and $\theta_{12}^{24} = \theta_{12}^L$. Thus we conclude that the odds of complication occurrence for children 1–2 years old is $\hat{\theta}_{11}^L = e^{\hat{\phi}(\alpha_2 - \alpha_1)(\nu_2 - \nu_1)} = e^{1.1669} = 3.2$ times higher than for children of any other age. The $\hat{\theta}_{1j}^L$, $j = 1, 2, 3$, could also have been computed by the `local.odds.DM()` function (see Sect. A.3.2), implemented as follows:

```

> NI <- 2;  NJ <- 4;  C <- local.odds.DM(NI, NJ)
> LO <- as.vector(C%*%log(LL.model$fitted.values))
> exp(t(matrix(LO, NJ-1)))

```

| | [,1] | [,2] | [,3] |
|------|----------|-----------|------|
| [1,] | 3.207792 | 0.3117409 | 1 |

6.6.4 Association Models Fitted on the Local Odds Ratios

Association models can also be fitted directly on the local odds ratios through the generalized log-linear model GLLM (5.28) and implemented in R by Lang's `mph` package. The GLLM turns to a model on local odds ratios by eliminating matrix **M** and appropriately defining matrix **C**, so that $\mathbf{C} \log(\mathbf{m})$ becomes the vector of the expected local odds ratios under the assumed model, where **m** is the vector of expected cell frequencies. For an $I \times J$ table, **C** and **m** are of size $(I-1)(J-1) \times IJ$ and $IJ \times 1$, respectively. This matrix **C** for an $I \times J$ table is produced by function `local.odds.DM()` of the web appendix (see Sect. A.3.2). The design matrix **X** specifies the restrictions imposed on the local odds ratios by the model under consideration and is of size $(I-1)(J-1) \times s$, where s is the number of parameters in the model.

We illustrate this option fitting the U model on our cannabis example. Under the U model a common value is assumed for all local odds ratios; thus the parameter β is scalar and the design matrix \mathbf{X} is the $(I-1)(J-1) \times 1$ vector of 1's. Recall (see Sect. 5.6) that `mph` needs to be actualized in R and that the data are read as a vector saved in matrix form. The way we define the \mathbf{C} matrix requires the data to be read by rows. For the cannabis example,

```
> y <- c(204, 6, 1, 211, 13, 5, 357, 44, 38, 92, 34, 49)
> y <- matrix(y); NI <- 4; NJ <- 3; dim1 <- (NI-1) * (NJ-1)
> X <- matrix(rep(1, dim1))
```

In our context, matrix \mathbf{C} is

```
> C <- local.odds.DM(NI, NJ)
```

and the link of the GLLM model is defined by the function

```
> L.fct <- function(m) {C%*%log(m)}
```

Finally, the U model is fitted by

```
> mph.out <- mph.fit(y=y, L.fct=L.fct, X=X)
> mph.summary(mph.out, cell.stats=T, model.info=T)
```

The derived output provides goodness-of-fit statistics for the model; estimate of the β (the log of the common local odds ratio under U), and estimates of the expected cell frequencies and the associated residuals. Further informations, as for example on the algorithm's convergence, are also provided.

In case of one or more sampling zeros, when working with odds ratios and for ensuring their existence, we set

```
> z <- y+0.000001
> mph.out <- mph.fit(y=z, L.fct=L.fct, X=X)
```

6.7 Association Models for Multi-way Tables

Association models can also be applied on contingency tables of higher dimension. Consider a $I \times J \times K$ contingency table with classification variables X , Y , and Z , respectively. Association models can be derived by replacing one or more of the interaction terms of any hierarchical log-linear model by multiplicative terms based on scores, leading thus to more parsimonious models of special structure, in analogy to two-way association models.

For example, consider the model

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \varphi^{XZ} \omega_i \tau_k + \varphi^{YZ} \nu_j \tau_k, \quad (6.24)$$

$$i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K,$$

with $(\omega_1, \dots, \omega_i)$, (ν_1, \dots, ν_j) , and (τ_1, \dots, τ_k) sets of known scores assigned to the categories of the classification variables X , Y , and Z , respectively, all equidistant for successive categories. This model is a special type of conditional XY independence model, derived from the (XZ, YZ) log-linear model by replacing the λ_{ik}^{XZ} and λ_{jk}^{YZ} interaction terms by the uniform (U)-type terms $\varphi^{XZ} \omega_i \tau_k$ and $\varphi^{YZ} \nu_j \tau_k$, respectively.

For this, it will be denoted as (XZ_U, YZ_U) . It is very parsimonious, having $df = IJK - I - J - K$, just 2 less than the complete independence model (X, Y, Z) .

More options are available by considering some of the scores to be parametric. Assuming thus an R-type interaction only for the term λ_{ik}^{XZ} , the $(\infty_1, \dots, \infty_7)$ scores would be considered as parameters in (6.24) and the model would then be (XZ_X, YZ_U) . In terms of notation, an interaction term without an index is of log-linear model type, with U of uniform association type, while when it is of row or column effect type, the variable of parametric scores is set as an index. A multiplicative RC-type has also a multiplicative index. Thus (XZ_{XZ}, YZ_U) is the model defined by (6.24) with parametric ∞ - and τ -scores of the XZ interaction and fixed equidistant scores for the ν - and τ -scores of the YZ interaction term. Were the layer scores parametric in both interaction terms, they could be homogeneous or not. The model for parametric nonhomogeneous τ -scores, (XZ_Z, YZ_Z) , is

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \phi^{XZ} \infty_i \tau_k^{XZ} + \phi^{YZ} \nu_j \tau_k^{YZ},$$

while with the additional restrictions $\tau_k^{XZ} = \tau_k^{YZ}$, $k = 1, \dots, K$, the homogeneous (XZ_Z, YZ_Z) is derived.

A flexible form of association model, including three-factor interaction, is

$$\begin{aligned} \log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \phi^{XY} \infty_i^{XY} \nu_j^{XY} + \phi^{XZ} \infty_i^{XZ} \tau_k^{XZ} + \quad (6.25) \\ \phi^{YZ} \nu_j^{YZ} \tau_k^{YZ} + \phi^{XYZ} \infty_i^{XYZ} \nu_j^{XYZ} \tau_k^{XYZ}, \end{aligned}$$

which offers a variety of model options, depending on the combinations of assumptions about the scores.

The most general expressions for imposing association structures on the two-factor interaction terms of a three-way log-linear model are

$$\begin{aligned} \lambda_{ij}^{XY} &= \sum_{m=1}^{M_1} \phi_m^{XY} \infty_{im}^{XY} \nu_{jm}^{XY}, & \lambda_{jk}^{YZ} &= \sum_{m=1}^{M_2} \phi_m^{YZ} \nu_{jm}^{YZ} \tau_{km}^{YZ}, & (6.26) \\ \lambda_{ik}^{XZ} &= \sum_{m=1}^{M_3} \phi_m^{XZ} \infty_{im}^{XZ} \tau_{km}^{XZ}, \end{aligned}$$

with $1 \leq M_1 \leq \min(I, J) - 1$, $1 \leq M_2 \leq \min(J, K) - 1$, and $1 \leq M_3 \leq \min(I, K) - 1$. The three-factor interaction can be decomposed in an analogue manner

$$\lambda_{ijm}^{XYZ} = \sum_{m=1}^{M_4} \phi_m \infty_{im} \nu_{jm} \tau_{km}. \quad (6.27)$$

The consideration of (6.27) for $M_4 > 1$ as well as other options for decomposing three-way arrays, known as *trilinear decomposition*, is beyond the scopes of this book (see Sect. 6.8.1).

The scores are subject to constraints analogue to (6.19) of the two-way case. When $M_i = 1$ ($i = 1, \dots, 4$), we conclude to model (6.25). Furthermore, the scores can be considered known and lead to terms of U-, R-, C-, or L- (for the layer scores) type.

The idea extends analogously to contingency tables of higher dimension. However, the number of possible association models augments with the dimension of the table. It is difficult to control all possible combinations of assumptions regarding the interaction terms of a multi-way table, so an automated stepwise association model selection procedure is not feasible. In practice we start by selecting the appropriate hierarchical log-linear model by a stepwise procedure and then try to conclude to a more parsimonious model by imposing special structures to some of the interaction terms. In this procedure, conditional tests between nested models are helpful. Finally, we can test whether parametric scores of the same classification variable but on different interaction terms are homogeneous.

Multi-way association models and their physical interpretations will be illustrated with two examples that follow.

6.7.1 Example 6.3

In a study, 16,236 teenagers in Holland are cross-classified in a $6 \times 7 \times 2$ table by their educational level after 4 years of second-level education, their test for intellectual capacity (TIC) score, and their gender (Siciliano and Mooijaart 1997). The data are provided in Table 6.12.

In the framework of hierarchical log-linear model, we do not have another option for this data set than the saturated model, since the three-factor interaction is significant. We can verify that the model of homogeneous association (EI, EG, IG) is rejected with $G^2(GE, GI, EI) = 61.517$ (p -value=0.001, $df=30$). It is notable however that the highly significant G^2 value is also affected by the large sample size of the table. The corresponding dissimilarity index is $\hat{\Delta} = 0.02$, at the limit for a satisfying data representation by this model (see Sects. 4.2 and 4.2.1 for calculation in R). The significance of each term in the log-linear model is summarized in the analysis of deviance table of the saturated model, derived as shown below.

Provided that the data are given in vector `freq`, expanded by rows, followed by columns and layers, we program in R

```
> G<-factor(rep(1:2, each=42)); I<-factor(rep(1:7, 12))
> E<-factor(rep(1:6, 2, each=7)); educ.fr<-data.frame(freq, E, I, G)
> sat.glm <- glm(freq ~ E*I*G, family=poisson, data=educ.fr)
> anova(sat.glm, test="Chisq")
```

and get the output of Table 6.13.

Since all interaction terms are significant, the basis for selecting the appropriate association model will be the saturated. The simplest model expression of this type is the

$$\log(m_{ijk}) = \lambda + \lambda_i^E + \lambda_j^I + \lambda_k^G + \varphi^{EI} \alpha_j v_j + \varphi^{IG} v_j \tau_k + \varphi^{EG} \alpha_j \tau_k + \varphi^{EIG} \alpha_j v_j \tau_k, \quad (6.28)$$

Table 6.12 Cross-classification of 16,236 teenagers in Holland by their educational level after 4 years of second-level education, their test for intellectual capacity (TIC) score, and their gender (Siciliano and Mooijaart 1997)

| Gender | | TIC | | | | | | Total | |
|--------|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | | 7 |
| Boys | Education | | | | | | | | |
| | DO | 75 (57.24) | 77 (79.28) | 105 (127.13) | 125 (128.12) | 89 (87.64) | 38 (37.99) | 17 (8.60) | 526 |
| | LBO | 216 (212.43) | 305 (304.47) | 495 (505.33) | 522 (527.06) | 389 (373.10) | 168 (167.38) | 34 (39.23) | 2129 |
| | MAVO | 67 (71.70) | 144 (131.16) | 267 (277.81) | 368 (369.79) | 339 (334.07) | 194 (191.27) | 54 (57.20) | 1433 |
| | MBO | 51 (49.32) | 84 (98.25) | 239 (226.62) | 345 (328.50) | 301 (323.18) | 208 (201.50) | 65 (65.63) | 1293 |
| | HAVO | 26 (31.04) | 65 (71.86) | 200 (192.65) | 332 (324.57) | 383 (371.13) | 258 (268.95) | 98 (101.81) | 1362 |
| | VWO | 12 (8.81) | 27 (27.28) | 104 (97.81) | 216 (220.36) | 325 (336.94) | 321 (326.51) | 178 (165.28) | 1183 |
| | Total | 447 | 702 | 1410 | 1908 | 1826 | 1187 | 446 | 7926 |
| Girls | Education | | | | | | | | |
| | DO | 51 (44.21) | 60 (65.96) | 115 (113.96) | 123 (123.73) | 78 (91.18) | 56 (42.58) | 9 (10.39) | 492 |
| | LBO | 144 (154.72) | 223 (221.27) | 382 (366.45) | 370 (381.38) | 290 (269.39) | 107 (120.59) | 26 (28.20) | 1542 |
| | MAVO | 60 (64.61) | 134 (128.52) | 288 (296.03) | 424 (428.51) | 442 (420.98) | 266 (262.10) | 72 (85.24) | 1686 |
| | MBO | 75 (80.81) | 167 (152.43) | 320 (332.96) | 458 (457.05) | 428 (425.81) | 258 (251.41) | 72 (77.54) | 1778 |
| | HAVO | 23 (25.43) | 68 (65.95) | 211 (198.08) | 373 (373.88) | 450 (478.94) | 402 (388.83) | 169 (164.89) | 1696 |
| | VWO | 5 (4.67) | 9 (16.59) | 77 (68.17) | 183 (176.07) | 307 (308.64) | 326 (342.88) | 209 (198.98) | 1116 |
| | Total | 358 | 661 | 1393 | 1931 | 1995 | 1415 | 557 | 8310 |

In parentheses are given the fitted values under the association model (6.29). Educational-level scale: (1) DO, dropped out; (2) LBO, junior level of education for professions; (3) MAVO, medium level of general education; (4) MBO, senior level of education for professions; (5) HAVO, high level of general education; and (6) VWO, general education preparing for university

with all the involved set of scores known. Considering the scores in each set equidistant for successive categories, model (6.28), denoted by (EIG_U) , is the most parsimonious three-way association model in the class of models with up to three-factor interaction, having just 4 parameters more than the model of complete independence (E, I, G) .

Table 6.13 Decomposition of the deviance for Table 6.12

| Analysis of Deviance Table | | | | | | |
|--|----|----------|-----------|------------|---------------|--|
| Model: poisson, link: log | | | | | | |
| Response: freq | | | | | | |
| Terms added sequentially (first to last) | | | | | | |
| | Df | Deviance | Resid. Df | Resid. Dev | P(> Chi) | |
| NULL | | | 83 | 9063.2 | | |
| E | 5 | 1910.5 | 78 | 7152.7 | < 2.2e-16 *** | |
| I | 6 | 4508.9 | 72 | 2643.8 | < 2.2e-16 *** | |
| G | 1 | 9.1 | 71 | 2634.7 | 0.002580 ** | |
| E:I | 30 | 2330.2 | 41 | 304.5 | < 2.2e-16 *** | |
| E:G | 5 | 222.3 | 36 | 82.2 | < 2.2e-16 *** | |
| I:G | 6 | 20.6 | 30 | 61.6 | 0.002154 ** | |
| E:I:G | 30 | 61.6 | 0 | 0.0 | 0.000584 *** | |
| -- | | | | | | |
| Signif. codes: | | | | | | |
| 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | | | |

In order to fit association models in R, we create the known score vectors for the classification variables. For simplicity, we set each score equal to the index of the category it corresponds to. We compute

```
> mu<-rep(1:6,2,each=7); nu<-rep(1:7,12); tau<-rep(1:2,each=42)
and extend the data frame
> educ.fr<-data.frame(freq,E,I,G,mu,nu,tau)
```

Model (EIG_U) is then fitted as

```
> EIG.U <- glm(freq~E+I+G+mu:nu+mu:tau+nu:tau+mu:nu:tau,
+ poisson, data=educ.fr)
```

It is of very bad fit, with $G^2(EIG_U) = 450.179$ (p -value< 0.0005, $df=67$), but reduces the G^2 statistic drastically, compared to complete independence ($G^2(E, I, G) = 2634,719, df=71$).

This means that some of the row and/or column scores in (6.28) have to be considered parametric. Since G is binary, the corresponding scores (τ_1, τ_2) cannot be parametric and their choice does not affect the model fit.

Considering that only the TIC effect is parametric on all the interaction terms, model (6.28) extends to

$$\log(m_{ijk}) = \lambda + \lambda_i^G + \lambda_j^E + \lambda_k^I + \alpha_i v_j^{EI} + v_j^{IG} \tau_k + \phi^{EG} \alpha_i \tau_k + \alpha_i v_j^{EIG} \tau_k ,$$

denoted by (EI_I, EG_U, IG_I, EIG_I). This last model expression employs non-standardized parametric scores and therefore the redundant intrinsic association ϕ -parameters are absorbed. It is fitted in R by

```
> EIG.I <- glm(freq~E+I+G+mu:I+mu:tau+I:tau+mu:I:tau,
+ poisson, data=educ.fr)
```

Table 6.14 ML estimates of the parametric scores of model (6.29) fitted on the data in Table 6.12 for equidistant scores $v_j = j$ ($j = 1, \dots, 7$) and $\tau_k = k$ ($k = 1, 2$)

| $i =$ | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------------|--------|--------|--------|--------|--------|-------|
| $\hat{\alpha}_i^{EI}$ | -0.742 | -0.631 | -0.473 | -0.250 | -0.268 | 0.000 |
| $\hat{\alpha}_i^{EG}$ | 0.438 | 0.456 | 0.583 | 1.319 | 0.458 | 0.000 |
| $\hat{\alpha}_i^{EIG}$ | 0.000 | -0.077 | 0.009 | -0.129 | 0.039 | 0.062 |

Its bad fit ($G^2(EI_I, EG_U, IG_I, EIG_I) = 426.6, p\text{-value} < 0.0005, df=52$) provides evidence that the education scores in some or all interaction terms should be considered as unknown parameters.

Thus, we try next the model

$$\log(m_{ijk}) = \lambda + \lambda_i^G + \lambda_j^E + \lambda_k^I + \alpha_i^{EI} v_j + \varphi^{IG} v_j \tau_k + \alpha_i^{EG} \tau_k + \alpha_i^{EIG} v_j \tau_k, \quad (6.29)$$

where only the education level (E) effect is parametric for all interaction terms. This is denoted as $(EI_E, EG_E, IG_U, EIG_E)$ and fitted in R by

```
> EIG.E <- glm(freq~E+I+G+E:nu+E:tau+nu:E:tau+nu:tau+E:nu:tau,
+             poisson, data=educ.fr)
```

exhibiting an adequate fit with $G^2(EI_E, EG_E, IG_U, EIG_E) = 61.074$ ($p\text{-value}=0.267, df=55$).

The fitted cell frequencies under $(EI_E, EG_E, IG_U, EIG_E)$ are provided in Table 6.12 in parentheses. For equidistant scores $v_j = j$ ($j = 1, \dots, 7$) and $\tau_k = k$ ($k = 1, 2$), the ML estimate of the intrinsic association parameter φ^{IG} is $\hat{\varphi}^{IG} = 0.0745$ while the ML estimates of the parametric scores are given in Table 6.14.

The interpretation of parameters needs caution and has to be done locally due to the non-monotonicity of the parametric scores. For interpretation, the fitted odds ratios under the model have to be considered. Model (6.29) in terms of the conditional EI log local odds ratios and for the choice of known scores given above is expressed as

$$\begin{aligned} \log\left(\theta_{ij(k)}^{EI}\right) &= \log\left(\frac{m_{ijk} \cdot m_{i+1,j+1,k}}{m_{i+1,j,k} \cdot m_{i,j+1,k}}\right) = (\alpha_{i+1}^{EG} - \alpha_i^{EG}) + (\alpha_{i+1}^{EIG} - \alpha_i^{EIG})k \\ &= \log\left(\theta_{i(k)}^{EI}\right), \quad i = 1, \dots, I-1, j = 1, \dots, J-1, k = 1, 2, \end{aligned}$$

i.e., independent of j , as expected since successive v_j scores are equidistant. This means that under (6.29) the fitted local odds ratios are constant within rows. In our case, the $\hat{\theta}_{i(k)}^{EI}$ row values are given in Table 6.15. Thus, we see that for boys, the strongest association between educational level and the TIC score is between HAVO and VWO. The odds of a boy achieving a category of TIC score vs. the immediate previous one is 1.34 times higher for a boy having general education preparing for university (VWO) than high level of general education (HAVO). The corresponding odds ratio for girls is 1.37. The conditional (within gender) association between TIC

Table 6.15 ML estimates of the $\theta_{i(k)}^{EI}$, $i = 1, \dots, 5$, $k = 1, 2$, under model (6.29) for the example in Table 6.12

| Education | Boys | Girls |
|-----------|-------|-------|
| DO | 1.035 | 0.958 |
| LBO | 1.276 | 1.391 |
| MAVO | 1.089 | 0.948 |
| MBO | 1.162 | 1.375 |
| HAVO | 1.337 | 1.368 |
| VWO | | |

score and educational level is positive for boys although not equally strong for all educational levels while for girls it is negative (though weak) when comparing DO to LBO and MAVO to MBO.

Based on model (6.29), one could further try if more parsimonious models are preferable. More parsimonious models are obtained by imposing homogeneity constraints among the vectors of the unknown education scores or considering one of them as being equidistant. It could also be tested whether less parsimonious non-log-linear models involving interaction terms multiplicative in their parameters (i.e., of RC-type) lead to a significant improvement of the fit.

6.7.2 Homogeneous Uniform Association

Consider a $I \times J \times K$ contingency table, consisting of K independent strata. Then the simplest association structure is to consider that for each XY partial table, all local odds ratios are equal, i.e., assume that the U model holds for each stratum. This model is defined by

$$\theta_{ij(k)}^{XY} = \theta_k^{XY}, \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1, \quad k = 1, \dots, K. \tag{6.30}$$

Local odds ratios however from different strata may vary and (6.30) is the *nonhomogeneous* U model. An even simpler model is the *homogeneous* U model, assuming that all strata have a common local odds ratio

$$\theta_{ij(k)}^{XY} = \theta^{XY}, \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1, \quad k = 1, \dots, K. \tag{6.31}$$

To illustrate these models, consider the data in Table 6.16. The first stratum of this $4 \times 3 \times 2$ data table is the cannabis example of Table 6.1 while the second stratum corresponds to an analogue survey among students of another university (artificial data).

The U model, fitted on the 4×3 partial table of the second stratum (data in Table 6.16, stratum (2)), with one of the procedures described in Sect. 6.6.1 for Example 6.1, is of good fit with $G_2^2 = 8.284$ (p -value= 0.141, $df = 5$). Under this model, the MLE of the common local odds ratio in log-scale is $\log \hat{\theta}_J^I = 0.749$, close to the corresponding estimate for the data in the first stratum ($\log \hat{\theta}_1^I = 0.803$), for which we had $G_1^2 = 1.469$ (p -value= 0.917, $df = 5$).

Table 6.16 Students' survey about cannabis use at two universities

| Alcohol consumption | I tried cannabis ... | | |
|---------------------|----------------------|---------------|------------|
| | Never | Once or twice | More often |
| Stratum (1) | | | |
| ≤ Once/month | 204 | 6 | 1 |
| Twice/month | 211 | 13 | 5 |
| Twice/week | 357 | 44 | 38 |
| More often | 92 | 34 | 49 |
| Stratum (2) | | | |
| ≤ Once/month | 311 | 5 | 4 |
| Twice/month | 339 | 19 | 12 |
| Twice/week | 429 | 66 | 57 |
| More often | 134 | 51 | 74 |

Stratum (1) is the data of Table 6.1; stratum (2) is artificial

The nonhomogeneous U model (6.30) for data in Table 6.16 can be derived in mph by

```
> source("c://Program Files//R//mph.Rcode.txt")
> freq <-c(204,6,1,211,13,5,357,44,38,92,34,49)
> freq2<-c(311,5,4,339,19,12,429,66,57,134,51,74)
> y<- matrix(append(freq,freq2))
> NI<-4; NJ<-3; dim1<-(NI-1)*(NJ-1); dim2<-2*dim1
> zer<-matrix(rep(0,NI*NJ*(NI-1)*(NJ-1)),(NI-1)*(NJ-1))
> C0<-local.odds.DM(NI,NJ); C1<-cbind(C0,zer)
> C2<-cbind(zer,C0); C<-rbind(C1,C2)
> L.fct <- function(m){C%*%log(m)}
> X<-matrix(rep(1,dim1)); Z<-matrix(rep(0,dim1))
> X2<-rbind(cbind(X,Z),cbind(Z,X))
> mph.out <- mph.fit(y=y,L.fct=L.fct,X=X2)
> mph.summary(mph.out,cell.stats=T,model.info=T)
```

leading to $G^2 = 9.752$ (p -value=0.462) with corresponding residual $df=10$. In this case, since the two strata are independent, model (6.30) is equivalent to fitting the U model independently to each of the partial two-way tables. Indeed, we can verify that $G_1^2 + G_2^2 = 9.752 = G^2$ and that the ML estimates of $\log \theta_k^{XY}$, $k = 1, 2$ ($\log \hat{\theta}_1^{XY} = \hat{\beta}_1 = 0.803$, $\log \hat{\theta}_2^{XY} = \hat{\beta}_2 = 0.749$) coincide with the corresponding $\log \hat{\theta}_k^L$, $k = 1, 2$.

The homogeneous U model (6.31) can be fitted as follows. The L.fct function is defined as above but the design matrix X2 is replaced by X1, defining thus a univariate parameter β instead of the bivariate (β_1, β_2) above:

```
> X1<-rbind(X,X) # homogeneous U model for both layers
> mph.out <- mph.fit(y=y,strata=2,L.fct=L.fct,X=X1)
> mph.summary(mph.out,cell.stats=T,model.info=T)
```

Selected parts of the output are provided in Tables 6.17 and 6.18.

Table 6.17 Output of the `mph` function for the homogeneous U model, applied on the $4 \times 3 \times 2$ data of Table 6.16: the observed local log odds ratios (OBS LINK) are listed by rows, along with the ML estimate of the common under the assumed model local log odds ratio value, its s.e. and the standardized link residuals

```

MODEL GOODNESS OF FIT: Test of Ho: h(p)=0 vs. Ha: not Ho...

Likelihood Ratio Stat (df=11):  Gsq = 10.05081 (pval = 0.5258 )
Pearson's Score Stat (df=11):  Xsq = 10.28396 (pval = 0.505 )
Generalized Wald Stat (df=11):  Wsq = 9.83676 (pval = 0.5451)

Adj Resids: -1.708 -1.452 ... 1.546 1.705,
Number |Adj Resid| > 2: 0

SAMPLING PLAN INFORMATION...
Number of strata: 1
Strata identifiers: 2
Strata with fixed sample sizes: all
Observed strata sample sizes: 2555
LINEAR PREDICTOR MODEL RESULTS...
      BETA  StdErr(BETA)  Z-ratio p-value
beta1 0.7691    0.0472   16.2809  0
      OBS LINK ML LINK StdErr(L) LINK RESID
link1  0.7395  0.7691   0.0472  -0.0596
link2  0.8362  0.7691   0.0472   0.0563
link3  0.6934  0.7691   0.0472  -0.2391
link4  0.8089  0.7691   0.0472   0.0701
link5  1.0981  0.7691   0.0472   1.2801
link6  0.5121  0.7691   0.0472  -0.8144
link7  1.2488  0.7691   0.0472   1.2460
link8  -0.2364  0.7691   0.0472  -1.1251
link9  1.0098  0.7691   0.0472   0.9707
link10 0.3129  0.7691   0.0472  -1.0692
link11 0.9058  0.7691   0.0472   0.6372
link12 0.5188  0.7691   0.0472  -0.9681
    
```

The equivalent expression of model (6.30) in terms of expected cell frequencies is

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \varphi^{XY} \alpha_i v_j + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \varphi^{XYZ} \alpha_i v_j \tau_k, \quad (6.32)$$

$$i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K,$$

where the set of scores $(\alpha_1, \dots, \alpha_I)$, (v_1, \dots, v_J) , and (τ_1, \dots, τ_K) are all known and equidistant for successive categories. They can be considered subject to standardization constraints or set equal to the corresponding category index.

The conditional local odds ratios under this model are fixed within partial tables equal to

$$\theta_{(k)}^{XY} = \exp((\varphi^{XY} + \varphi^{XYZ} \tau_k) \Delta_1 \Delta_2), \quad (6.33)$$

$$i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad k = 1, \dots, K,$$

Table 6.18 Output of the `mph` function: observed and ML fitted cell frequencies under the homogeneous U model applied on the $4 \times 3 \times 2$ data of Table 6.16, along with ML estimates of the cell probabilities, standard errors, and standardized residuals

| CELL-SPECIFIC STATISTICS... | | | | | | | |
|-----------------------------|--------|-----|----------|-----------|--------|-------------|------------|
| | strata | OBS | FV | StdErr.FV | PROB | StdErr.PROB | ADJ.RESIDS |
| y1 | 2 | 204 | 203.9403 | 13.4845 | 0.0798 | 0.0053 | 0.0247 |
| y2 | 2 | 6 | 6.0413 | 0.9179 | 0.0024 | 0.0004 | -0.0181 |
| y3 | 2 | 1 | 1.0183 | 0.2575 | 0.0004 | 0.0001 | -0.0188 |
| y4 | 2 | 211 | 210.6392 | 13.4128 | 0.0824 | 0.0052 | 0.0987 |
| y5 | 2 | 13 | 13.4638 | 1.6978 | 0.0053 | 0.0007 | -0.1431 |
| y6 | 2 | 5 | 4.8970 | 0.8612 | 0.0019 | 0.0003 | 0.0506 |
| y7 | 2 | 357 | 352.2819 | 16.4381 | 0.1379 | 0.0064 | 0.8152 |
| y8 | 2 | 44 | 48.5868 | 5.1333 | 0.0190 | 0.0020 | -0.9936 |
| y9 | 2 | 38 | 38.1313 | 4.4575 | 0.0149 | 0.0017 | -0.0312 |
| y10 | 2 | 92 | 97.1386 | 8.4659 | 0.0380 | 0.0033 | -1.1012 |
| y11 | 2 | 34 | 28.9081 | 3.4680 | 0.0113 | 0.0014 | 1.2515 |
| y12 | 2 | 49 | 48.9533 | 5.7150 | 0.0192 | 0.0022 | 0.0119 |
| y13 | 2 | 311 | 308.3105 | 16.1909 | 0.1207 | 0.0063 | 0.8984 |
| y14 | 2 | 5 | 9.8678 | 1.3053 | 0.0039 | 0.0005 | -1.7076 |
| y15 | 2 | 4 | 1.8218 | 0.4348 | 0.0007 | 0.0002 | 1.7054 |
| y16 | 2 | 339 | 337.4154 | 16.4817 | 0.1321 | 0.0065 | 0.3441 |
| y17 | 2 | 19 | 23.3021 | 2.4133 | 0.0091 | 0.0009 | -1.0353 |
| y18 | 2 | 12 | 9.2825 | 1.4482 | 0.0036 | 0.0006 | 1.0162 |
| y19 | 2 | 429 | 432.2293 | 17.6141 | 0.1692 | 0.0069 | -0.4620 |
| y20 | 2 | 66 | 64.4085 | 5.6842 | 0.0252 | 0.0022 | 0.2883 |
| y21 | 2 | 57 | 55.3622 | 5.2807 | 0.0217 | 0.0021 | 0.3195 |
| y22 | 2 | 134 | 135.0448 | 10.0966 | 0.0529 | 0.0040 | -0.2050 |
| y23 | 2 | 51 | 43.4217 | 4.3209 | 0.0170 | 0.0017 | 1.5465 |
| y24 | 2 | 74 | 80.5335 | 7.5990 | 0.0315 | 0.0030 | -1.4519 |

where $\Delta_1 = \alpha_{i+1} - \alpha_i$ and $\Delta_2 = v_{j+1} - v_j$. Distances Δ_1 and Δ_2 are constant over i and j , respectively, since the corresponding scores are equidistant. In case the scores equal their categories' indexes, (6.33) is simplified to

$$\theta_{(k)}^{XY} = \exp(\varphi^{XY} + k\varphi^{XYZ}), \quad i = 1, \dots, I - 1, \quad j = 1, \dots, J - 1, \quad k = 1, \dots, K.$$

Eliminating the three-factor interaction term in (6.32), the model of homogeneous uniform association (6.31) is derived in its equivalent expression

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \varphi^{XY} \alpha_i v_j + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \quad (6.34)$$

and $\theta^{XY} = \exp(\varphi^{XY} \Delta_1 \Delta_2)$.

Replacing λ_{ik}^{XZ} and/or λ_{jk}^{YZ} in (6.32) or (6.34) by $\varphi^{XZ} \alpha_i \tau_k$ and/or $\varphi^{YZ} v_j \tau_k$, respectively, more parsimonious models of uniform or homogeneous uniform

association are derived that consider U-type structure also for one at least of the other two-factor interactions. The options for models of this type do not restrict to XZ and/or YZ interactions of U-type. They could be of any other type (R, C, RC, or $RC(M)$). Such special models of uniform and homogeneous uniform association cannot be captured via the odds ratio formulation (6.30) or (6.31).

Finally, the simplest homogeneous uniform XY association model is obtained when X is jointly independent from X and Y , i.e., the model

$$\log m_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \varphi^{XY} \alpha_i v_j, \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1,$$

with $k = 1, \dots, K$. For the example above, this model is the best option, with $G^2 = 18.165$ (p -value=0.314, $df = 16$), giving $\log \hat{\theta}^{XY} = 0.7688$.

6.8 Overview and Further Reading

Association models, in their dominant form, have been mainly developed by the fundamental and inspiring work of Goodman (1979b, 1981a, 1985, 1986, 1991, 1996) and thus it is common to refer to them also as Goodman's models. For an overview, we refer to the 1986 and 1991 discussion papers and the review Goodman (1982). Significant in the development of association models was continuation work of Haberman (1979, 1995), Clogg (1982a), Becker and Clogg (1988, 1989), and Becker (1989a, 1990a, 1992) as well as the contribution of Anderson and Philips (1981) and Anderson (1984). Association models are presented in the book by Clogg and Shihadeh (1994). An overview of association models with formulation and interpretation based on odds ratios is provided by Breen (2008) along with social sciences-orientated illustrations and references. For their connection to *latent class models*, see Sect. 10.3.1.

To be fair, we must say that the basis for the development of the association models lies back to Tukey's 1 d.f. test (Tukey 1949) and in a different form they have been considered earlier. In particular, Nelder and Wedderburn (1972) consider the U model applied on the popular Boys' Dream Disturbance data set of Maxwell as an illustration of their GLM model on contingency tables. Simon (1974) introduced the R model (his formulation A) as well as the R model for cumulative odds (his formulation B), being the forerunner of the association models for global odds ratios (see Sect. 7.1 below). A similar model for the cumulative odds was earlier considered by Williams and Williams and Grizzle (1972). Also his analysis of information is remarkable, throwing insight into the nature of departure from independence in the direction of the ANOAS, developed later by Goodman. Other multiplicative models modeling triangular or diagonal departures from independence for square tables have been proposed by Goodman (1972). We should mention that methods that analyze contingency tables with ordinal classification variables by applying scores to their categories have been proposed much earlier, even from Yates (1948) and

Armitage (1955), not to forget the linear trend test of Mantel (1963), described in Sect. 2.3. However, these early references, after assigning scores to the categories, treated the corresponding categorical variables applying methods appropriate for continuous variable analysis.

Haberman (1974b) adopted a different approach by generating a class of models through the decomposition of the vector with elements of the expected cell frequencies $\log(m_{ij})$ on an orthonormal basis, formed by orthogonal polynomials. Special members of this class of models are the linear-by-linear association model and the row effect model. Further, he proved standard asymptotic inference results for these models and expressed them in terms of the log odds ratio, noting the importance of the difference between scores. Finally, he was the first to mention that his approach could be extended straightforward to define such models for multi-way tables. Association models for two-way and three-way tables are presented in Wong (2010).

Diagnostics for the RC association model have been discussed in Andersen (1992). De Rooij and Heiser (2005) criticized the classical graphical representation of the $RC(M)$ model and proposed the distance-association model representation, for which the distances between row and column points can be interpreted directly. Marginal association models have been considered by Lapp et al. (1998), Bartolucci et al. (2001), and Bartolucci and Forcina (2002). For ordinal tables with a response variable, Agresti (1986) proposed a regression R^2 -type measure of association, based on scores assigned to the classification variables' categories, and used the R association model to estimate these scores. Baccini and Khoudraji (1992) and Baccini et al. (2000) considered least squares estimation of association models. Beh and Farver (2009) discuss on closed-form estimation of the association parameter φ of the U model.

The $RC(M)$ are not the only models with additive multiplicative interaction terms. Goodman (1985) introduced other possible models with interaction of rank M but simpler than the $RC(M)$. For example, for $M = 2$, the R+C model is defined by the same formula as $RC(2)$ but assumes that the column scores of the first term and the row scores of the second are known; thus it has less parameters than $RC(2)$. Similarly, model U+R+C has just one parameter more than the R+C model, since the third term that is added is of uniform type, having assigned fixed row and column scores. More options of parsimonious models of higher rank for the interaction are obtained through the use of orthogonal polynomials for assigning scores (Kateri et al. 1998). For example, the model $U^{(1)}+U^{(2)}$ is of $M = 2$, but all the involved scores are fixed, assigned through orthogonal polynomials of first and second order for the (1) and (2) term, respectively, and has thus just 2 parameters more than the independence model.

In the special case of a square table with commensurable classification variables, it makes sense to assume that the row and column scores are homogeneous. Thus, the RC model with the homogeneity restriction on its scores $\alpha_i = \nu_i, \forall i = 1, \dots, I$, can be applied, which is more parsimonious than the standard RC and simultaneously of special interpretational value for such tables. On this we shall return and comment more on Chap. 9, specialized on square tables.

We have already mentioned in Chap. 2 that the log-linear models are the discrete analogue of the analysis of variance. It is interesting to note briefly at this point the analogues to association model in the two-way ANOVA framework. Special analysis of variance models that impose a structure on the interaction as that of the association models have been considered as well. Indicatively we mention the early work of Williams (1952), who used the multiplicative term for the interaction, and that by Gollob (1968), who introduced the more general term of the $RC(M)$ type. Goodman and Haberman (1990) proved asymptotic normality for the scores of the $RC(M)$ ANOVA model and provided asymptotic confidence intervals for the estimated scores. Furthermore, they developed the asymptotic conditional tests of the appropriateness of a simpler association model of the type U, R, or C given that the RC holds. Finally, they extended their results for the more general $RC(M)$ ANOVA model. Viele and Srinivasan (2000) proceeded to the Bayesian analysis of the $RC(M)$ ANOVA model. Speaking about analogies to the continuous case, Jones (1998) noted that the constant local dependence for continuous bivariate random variables is the continuous analogue of the U model.

We have seen in Sect. 6.3 that for ordinal contingency tables, conditional tests of independence given that the U, R, or C model holds, i.e., testing independence against a directed alternative, are more asymptotically powerful. Alternative approaches for strengthening the power of the classical Pearson's X^2 test of independence are based on the decomposition of Pearson's X^2 into orthogonal components in terms of assigned scores to the categories of the ordinal classification variables. For example, Best and Rayner (1996) and Rayner and Best (2000) considered scores based on orthogonal polynomials while Nair (1986, 1987), proportional to the midrank scores. Beh (1998) studied the use of different types of scores in the correspondence analysis framework. Nair's procedure partitions the X^2 statistic value for testing independence into location, dispersion, and residual effects. It is related to the location-dispersion model of McCullagh (1980), as is also pointed out by McCullagh's and Agresti's comments in the discussion of Nair (1986). Agresti's comment related Nair's statistics also to the statistics of Koch et al. (1982) with fixed or rank-based scores, to the measure of Agresti (1986), and to association models and the models in Semanya et al. (1983). Koshimizu and Tsujitani (1998) consider association models with location and dispersion scores for singly ordered contingency tables. Their model is actually analogue to the $R^{(1)}+R^{(2)}$ model of Kateri et al. (1998) with the column scores of the first dimension being the Nair's scores instead of equidistant for successive categories.

6.8.1 Multi-way Association Models

Conditional and partial associations in multi-way tables are discussed in Clogg (1982b). Becker (1989b) introduced the no three-factor interaction model with all two-way interaction terms replaced by the general terms (6.26). Becker and Clogg (1989) considered three-way association models for the analysis of stratified

two-way tables, with and without homogeneity constraints on the scores across the strata. Association models for stratified tables focusing on detecting layer differences were developed by Goodman and Hout (1998).

On the decomposition of the three-factor interaction term (6.27) focused Goodman (1983, 1986), Agresti and Kezouh (1983), Choulakian (1988), Anderson (1996), and Siciliano and Mooijaart (1997). A review is provided by Wong (2001). Methods of decomposing three-way arrays are reviewed in Ten Berge (2011).

6.8.2 Order-Restricted Inference

In case of association models with parametric scores, the monotonicity of the scores is not ensured by the standard estimation procedures. Since their monotonicity is related to stochastic ordering of the corresponding classification variable (Goodman 1981a), it is usually natural to expect the scores for ordinal classification variables to be monotonic. Estimation procedures subject to order constraints for the parametric scores have been developed for the R (or C) model by Agresti et al. (1987), based on isotonic regression. The RC model with order-restricted row and column scores has been considered by Ritov and Gilula (1991). A test of independence, conditional on the order-restricted RC model, is discussed in Kuriki (2005). Alternative algorithms for fitting the order-restricted RC model have been proposed and compared by Galindo-Garre and Vermunt (2004). Order restrictions yield also for an extended RC model, introduced by Bartolucci and Forcina (2002).

Ordinary or order-restricted inferences for these models rely on large-sample asymptotic methods. As it is stated in Galindo-Garre and Vermunt (2004), these methods do not work well for sparse tables or small sample sizes, common in social and biomedical applications, where the usual asymptotic chi-squared p -values are known to be inaccurate. A promising alternative is the Bayesian approach (see Sect. 10.5).

6.8.3 Comparison of Two Ordinal Responses

The problem of comparing two ordinal responses is very old and of special interest in many fields, especially in biomedical applications. The need to compare the response to a treatment of two independent groups of patients, defined, for example, by the presence of a prognostic factor, is obvious. Another common situation is to compare two different treatments applied on two independent samples with the corresponding responses measured on a common scale. The ordinality of the response scale has to be taken into consideration in handling the problem and answering to the question “Which group of patients benefits more from the treatment?” or “Which treatment is superior?.” The underlying sampling scheme can be multinomial or product multinomial. The first is the case whenever a

sample of n subjects is cross-classified with respect to an ordinal response Y and a binary variable X indicating the two groups, while the second when two independent multinomial samples of the same ordinal response and of sizes n_1 and n_2 are available. If the ordinal response has J categories, then the above described data form a $2 \times J$ contingency table. For the multinomial sampling scheme the corresponding joint distribution is $\pi = (\pi_{ij}) = P(X = i, Y = j)$, $i = 1, 2$, $j = 1, \dots, J$. In case of two independent multinomials, the row marginals are also fixed, $n_{i+} = \sum_{j=1}^J n_{ij} = n_i$ ($i = 1, 2$).

The problem of comparing two response profiles is equivalent to the stochastic comparison of the two row distributions of the abovementioned $2 \times J$ contingency table and as such has been faced by a variety of methods. The related bibliography is very rich and an extended critical review of the available methods can be found in Agresti and Coull (2002).

The hypothesis that two multinomial distributions are identical against an ordered alternative is mainly tested through LR, Wald, and score tests or through linear rank tests. It is well known that restricting the alternative hypothesis leads to more powerful tests than the standard chi-squared test of independence. These approaches are all asymptotic, while the linear rank tests depend on the choice of the scores assigned to the ordered categories. Characteristic references of LR tests are Grove (1980, 1984) and Robertson and Wright (1981), while the approaches of Emerson and Moses (1985), Graubard and Korn (1987), and Gautam (1997) are based on linear rank tests. To deal with the sensitivity of the linear rank tests on the scores, Kimeldorf et al. (1992) proposed the min-max scoring and Gautam et al. (2001) the iso-chi-square approach. Nonlinear rank tests have also been proposed. For example, Hilton et al. (1994) and Nikiforov (1994) applied the Smirnov test while Berger (1998) proposed the convex hull methodology that leads to admissible tests. Properties and power of the convex hull test applied on $2 \times J$ tables are further studied in Berger et al. (1998) (see also Cohen and Sackrowitz 1998; Cohen et al. 2000). An interesting approach is provided by Permutt and Berger (2000), who reviewed various rank tests, classified them as Smirnov-like or Wilcoxon-like, and compared them. However, the nonlinear tests are not easy to compute for $J > 3$. It is important to note that “with few exceptions there is no optimal test for this problem,” as stated by Berger and Ivanova (2002). Tests based on log-linear models were developed by Agresti and Coull (1998).

The connection of association models to the stochastic ordering of the conditional row (or column) distributions of the contingency table has been discussed in Sect. 6.4. In case of the $2 \times J$ tables, the RC model coincides with the C model, which is saturated. For $2 \times J$ contingency tables with $\alpha_1 < \alpha_2$ and $\phi > 0$, positive dependence is equivalent to $v_j \leq v_{j+1}$ ($j = 1, \dots, J-1$) with $v_1 < v_J$. Thus, *monotonicity* of the column scores $\{v_j : j = 1, \dots, J\}$ implies *stochastic ordering* of the probabilities

$$\pi_i = \left\{ \frac{\pi_{ij}}{\pi_{i+}}, j = 1, \dots, J \right\}, i = 1, 2.$$

Thus the distribution of the response Y for the second group (row) is stochastically larger than the one of the first group (row).

The comparison of the two row distributions can be further enriched with the option of umbrella ordering as an alternative when stochastic ordering is rejected (Kateri 2011). Umbrella ordering means that the distribution in the first row is stochastically smaller than the one in the second up to a level of the ordinal scale that defines the column categories and stochastically larger after this level (or the opposite). In terms of physical interpretation, when comparing two alternative treatments, umbrella ordering of their response distributions corresponds to cases where one treatment is better over the other up to a certain level of the response scale while the situation changes after this point. In a retrospective study context cross-classifying the “cured”– “not-cured” groups with the J levels of a prognostic factor, this could mean higher risk for the very low and very high levels of the prognostic factor. Umbrella ordering essentially reveals a dispersion effect for the group comparison. Dispersion effects for ordinal responses have been handled by the generalized cumulative link model, introduced by McCullagh (1980). Umbrella ordering can be captured by the C model, with adequately constrained column scores.

6.8.4 Cell Frequencies vs. Local Odds Ratios Modeling

We have seen so far that models applied on contingency tables can be expressed in terms of expected cell frequencies or equivalently in terms of local odds ratios. The choice depends on issues of interpretation and on convenience of model formulation. For example, the association models are easier interpreted through the local odds ratios. On the other hand, the quasi-independence model can be expressed in terms of local odds ratios but is too complicated to compete with (5.24).

A clarifying and inspiring insight into the possible different views of log-linear models is provided by Goodman (1981d), who considers three alternative views, depending on the purpose of the analysis. The model imposed on the cell frequencies is preferred whenever the purpose is the examination of the joint distribution of the contingency table. Local odds ratio formulation of the model is employed when interest lies on the association between the two variables that are cross-classified. In both cases, the classification variables of the table are treated symmetrically. If there exists a response variable, then modeling the possible dependence of the response variable on the explanatory one is more adequate than the symmetric approaches and leads to more direct interpretations. This constitutes the third view and corresponds to modeling the odds for the response variable, given that the explanatory variable is at a fixed, prespecified level. Such models are presented in Chap. 8. Goodman (1981d) discussed the connections between these different approaches of log-linear modeling and illustrated them on characteristic examples. These comments apply also to the special models for square tables in Chap. 9.