6
The Strength of Marginal Models

6.1 Introduction

In the past century, a vast part of the literature devoted to multivariate categorical data focused on describing the association structure between two or more variables. Eminent early references are Yule and Kendall (1950) and Goodman (1969, 1979, 1981a, 1981b, 1985).

Recently, the focus in multivariate categorical data has somewhat shifted to regression models, intended mainly for the analysis of longitudinal data. It is fair to say that the gap between classical contingency table and categorical data analysis on the one hand and categorical longitudinal data on the other hand is less wide than the corresponding gap for Gaussian data, where multivariate and longitudinal methods have their own focus and flavor. As a consequence, not only classically used models such as log-linear models (Cox 1972, Agresti 2002) ought to be considered, but also marginal models can be of great use. Perhaps it is not sufficiently recognized that these models provide a versatile basis, not only for regressing multiple outcomes on predictor variables, as will be done in Chapter 7, but also to study the association between two (or more) categorical variables. In other words, they can be used for the *analysis of association*. To this end, it is necessary to construct more complex association structures than are often needed for longitudinal applications. For this purpose, one can borrow flexible association structures as used in more conventional models, such as the ones described in Goodman (1981a).

In Section 6.2, we first sketch the so-called multivariate logistic models (McCullagh and Nelder 1989, Glonek and McCullagh 1995). Then we review the classical row-column (RC) association models (Goodman 1981a) and the marginal association model (Dale 1986, Molenberghs and Lesaffre 1994). It is indicated that both families can be seen as specific multivariate logistic models. This naturally leads to the observation that, within the multivariate logistic models family, very general association models can be constructed. Sections 6.3 and 6.4 present two simple but illustrative examples: the British occupational study and the Caithness data.

The fluvoxamine trial, introduced in Section 2.4, is analyzed in Section 6.5. These data are rich in the sense that two important outcomes, therapeutic effect and side effects, measured on 4-point ordinal scales, are measured repeatedly over time, and both continuous and discrete covariates are measured. Here, we first restrict attention to two-way contingency tables, and then, in Section 6.6, two extensions are presented, the first one to contingency tables in the presence of a categorical covariate, the second one to three-way tables. Both of these extensions will be put within a general framework in Chapter 7. In Section 6.7, we sketch how the association models can be embedded in families of models, arising as discretizations of continuous distributions.

6.2 Marginal Models in Contingency Tables

We first introduce the notation, needed for this chapter. A general notational framework is given in Section 7.1. The notation here is somewhat different from the notation used in the purely longitudinal chapters but allows us to efficiently deal with the contingency table nature of the data in this chapter. Suppose a contingency table arises from cross-classifying N subjects with respect to two categorical variables Y_1 and Y_2, having I and J levels respectively. It is convenient to introduce both ordinary multinomial cell counts

$$Z^*_{ijr} = \begin{cases} 1 & \text{if } Y_{1r} = i \text{ and } Y_{2r} = j, \\ 0 & \text{otherwise.} \end{cases}$$

as well as their cumulative counterparts

$$Z_{ijr} = \begin{cases} 1 & \text{if } Y_{1r} \leq i \text{ and } Y_{2r} \leq j, \\ 0 & \text{otherwise,} \end{cases} \tag{6.1}$$

with a subscript r denoting the rth subject. The corresponding probabilities are defined by $\mu^*_{ij} = \text{pr}(Z^*_{ijr} = 1)$ and $\mu_{ij} = \text{pr}(Z_{ijr} = 1)$. This notation will be used to describe the association models. Should the probabilities depend on the subject (for example, through the introduction of covariate information), then a subscript r will be added (μ^*_{ijr} and μ_{ijr}). We will

first introduce a general framework, largely due to McCullagh and Nelder (1989) and Glonek and McCullagh (1995). Then, the RC family of models (Goodman 1981a) and the Dale (1986) model are shown to fit within this framework, conditional on a slight generalization in the RC case. Finally, it is indicated how the modeling framework can be used to combine useful aspects of both subclasses to yield a very wide and versatile class, which, in addition, allows extension to covariates as well as to higher order tables.

6.2.1 Multivariate Logistic Models

McCullagh and Nelder (1989) defined a useful class of generalized linear models, by writing the vector link function in terms of the joint probabilities in the following way:

$$\boldsymbol{\eta} = C' \ln(L\boldsymbol{\mu}^*), \tag{6.2}$$

where $\boldsymbol{\mu}^*$ is the vector of joint probabilities, formed by stacking the μ^*_{ij}. The matrix L consists solely of zeros and ones, such that $L\boldsymbol{\mu}^*$ contains the probabilities necessary to construct the required link functions. Then, contrasts of log-probabilities are equated to a vector of linear predictors $\boldsymbol{\eta}$ using the contrast matrix C. Contrasts of log-probabilities encompass many commonly used links for both marginal probabilities and associations. Within this model formulation, the marginal means can be modeled via, e.g., baseline category logits, adjacent category logits, continuation-ratio logits, or cumulative logits. The association can be described in terms of, e.g., local or global cross-ratios. This means that this formulation applies to binary, ordinal, and nominal data. When cumulative logits and/or global cross-ratios are used, the model can be expressed directly in terms of the cumulative probabilities μ_{ij}, such that (6.2) becomes $\boldsymbol{\eta} = C' \ln(L\boldsymbol{\mu})$. In this case, L may contain other elements than merely zeros and ones. Alternatively, the connection between μ_{ij} and μ^*_{ij} ($\boldsymbol{\mu} = B\boldsymbol{\mu}^*$, for some constant matrix B) can be absorbed into the matrix L as well. As counterexamples, modeling the marginal distribution via, e.g., the probit or the complementary log-log link is excluded from (6.2). One usually requires that $\boldsymbol{\mu}^*$ and $\boldsymbol{\eta}$ are in 1-to-1 relationship. Model (6.2) is called the *multivariate logistic transform* by Glonek and McCullagh (1995). They illustrate its use for both marginal and conditional regression models, as well as for mixed marginal-conditional parameterizations. A general and flexible class of marginal logistic models of the form (6.2) was studied by Lang and Agresti (1994), who allow a many-to-one relationship between $\boldsymbol{\mu}^*$ and $\boldsymbol{\eta}$ because they do not require that the (higher order) associations are modeled explicitly. Examples will be given in the next two sections.

In the spirit of generalized *linear* modeling (Chapter 3), McCullagh and Nelder (1989) completed (6.2) by

$$\boldsymbol{\eta} = X\boldsymbol{\xi}, \tag{6.3}$$

i.e., by adopting a vector of linear predictors. Here, X is a known design and/or covariate matrix and $\boldsymbol{\xi}$ is a vector of parameters of direct interest. Glonek and McCullagh (1995) call the resulting family *multivariate logistic regression models*.

When not only regression aspects are of scientific interest, but focus is placed on the association structure as well, it is useful to generalize the vector of linear predictors (6.3) to the potentially non-linear class

$$\boldsymbol{\eta} = C' \ln(L\boldsymbol{\mu}^*) = \boldsymbol{g}(\boldsymbol{\xi}), \tag{6.4}$$

where $\boldsymbol{g}(\boldsymbol{\xi})$ is a known vector-valued function.

6.2.2 Goodman's Local Association Models

Goodman (1981a) defines association models in terms of log local cross-ratios for $I \times J$ tables. These log cross-ratios are given by

$$\begin{aligned} \ln \theta^*_{ij} &= \ln \left(\frac{\mathrm{pr}(Y_1 = i, Y_2 = j)\mathrm{pr}(Y_1 = i+1, Y_2 = j+1)}{\mathrm{pr}(Y_1 = i, Y_2 = j+1)\mathrm{pr}(Y_1 = i+1, Y_2 = j)} \right) \\ &= \ln \frac{\mu^*_{ij}\mu^*_{i+1,j+1}}{\mu^*_{i,j+1}\mu^*_{i+1,j}}, \end{aligned}$$

with $i = 1, \ldots, I-1$ and $j = 1, \ldots, J-1$. They naturally follow from the following closed form model for the joint cell probabilities:

$$\mu^*_{ij} = \alpha_i \beta_j e^{\phi \lambda_i \nu_j}, \tag{6.5}$$

$(i = 1, \ldots, I; j = 1, \ldots, J)$. Here, α_i and β_j are main effect parameters while λ_i, ν_j and ϕ describe the association structure. Indeed, the local cross-ratios are $\ln \theta^*_{ij} = \phi(\lambda_i - \lambda_{i+1})(\nu_j - \nu_{j+1})$. Identifiability constraints have to be imposed on the parameters in (6.5). This model is also called the row-column model (RC model).

Note that this model is not fully marginal in nature since the marginal probabilities or transformations thereof do not easily follow from the model parameters. In fact, the model has a close connection to log-linear models, which are conditional in nature. In this sense, it bridges the gap between the models treated here and those in Part III.

Model (6.5) can be seen as a member of (6.4) by setting L and C equal to the identity matrix: $\boldsymbol{\eta} = \ln \boldsymbol{\mu}^* = g(\boldsymbol{\xi})$, with $\boldsymbol{g}(\boldsymbol{\xi})$ defined by

$$g_{ij}(\boldsymbol{\xi}) = \ln \alpha_i + \ln \beta_j + \phi \lambda_i \nu_j. \tag{6.6}$$

Due to its third term, the predictor function (6.6) is non-linear. Note that (6.6) is a mixture of main effect and association parameters. By setting C equal to the identity matrix, the concept of *contrasts* of log-probabilities is not maintained and thus (6.4) is slightly extended.

An alternative association parameterization is additive in the log cross-ratios: $\ln \theta^*_{ij} = \delta_{1i} + \delta_{2j}$. This model is induced by the following expression for the cell probabilities:

$$\mu^*_{ij} = \alpha_i \beta_j \gamma^j_{1i} \gamma^i_{2j}. \tag{6.7}$$

For this parameterization (6.6) changes to

$$g_{ij}(\boldsymbol{\xi}) = \ln \alpha_i + \ln \beta_j + j \ln \gamma_{1i} + i \ln \gamma_{2j}. \tag{6.8}$$

Note that this predictor is of the linear type inln α_i, etc. Fitting algorithms for (6.5) and (6.7) can be found in Goodman (1981a).

Goodman (1981a) generalizes model (6.5) to:

$$\mu^*_{ij} = \alpha_i \beta_j \exp\left(\sum_{k=1}^{4} \phi_k \lambda_{ki} \nu_{kj}\right), \tag{6.9}$$

where λ_{1i} and λ_{3i} are linear functions of the index i and ν_{1j} and ν_{2j} are linear in j. The others are allowed to be non-linear. He shows that the log cross-ratios can be written as

$$\ln \theta^*_{ij} = \eta + \eta^I_i + \eta^J_j + \zeta^I_i \zeta^J_j. \tag{6.10}$$

This model allows the inclusion of additive effects on the association. Goodman calls it the R+C+RC model.

Although the above models provide an elegant description of the association in contingency tables, a disadvantage of the RC family is the cumbersome form they induce for the marginal distributions. The model presented next is built marginally.

6.2.3 Dale's Marginal Models

Dale (1986) and Molenberghs and Lesaffre (1994, 1999) define a marginal model for ordinal data in terms of marginal cumulative logits and global cross-ratios. We will describe it here for the purpose of our contingency table type data setting, and defer a fully general, longitudinal introduction to Chapter 7. The cumulative logits

$$\begin{aligned} \eta_{1i} &= \text{logit}[\text{pr}(Y_1 \leq i)] = \ln(\mu_{iJ}) - \ln(1 - \mu_{iJ}), & (6.11)\\ \eta_{2j} &= \text{logit}[\text{pr}(Y_2 \leq j)] = \ln(\mu_{Ij}) - \ln(1 - \mu_{Ij}), & (6.12) \end{aligned}$$

$(i = 1, \ldots, I-1; j = 1, \ldots, J-1)$, and the global cross-ratios

$$\begin{aligned} \ln \psi_{ij} &= \ln\left(\frac{\text{pr}(Y_1 \leq i, Y_2 \leq j)\text{pr}(Y_1 > i, Y_2 > j)}{\text{pr}(Y_1 \leq i, Y_2 > j)\text{pr}(Y_1 > i, Y_2 \leq j)}\right)\\ &= \ln \frac{\mu_{ij}(1 - \mu_{Ij} - \mu_{iJ} + \mu_{ij})}{(\mu_{iJ} - \mu_{ij})(\mu_{Ij} - \mu_{ij})} \end{aligned} \tag{6.13}$$

define the joint probabilities.

It is clear from (6.11), (6.12), and (6.13) that this model is a member of (6.2). For the special case of binary data ($I = J = 2$), (6.2) becomes

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \times$$

$$\times \ln \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu^*_{11} \\ \mu^*_{12} \\ \mu^*_{21} \\ \mu^*_{22} \end{pmatrix},$$

where the model is written in terms of the cell probabilities μ^*_{jk}. Because

$$\begin{pmatrix} \mu^*_{11} \\ \mu^*_{12} \\ \mu^*_{21} \\ \mu^*_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{pmatrix},$$

an expression in terms of the cumulative probabilities μ_{jk} is immediate.

Should it be thought reasonable, then local cross-ratios:

$$\ln \psi^*_{ij} = \ln \frac{\mu^*_{ij}(1 - \mu^*_{i+1,j} - \mu^*_{i,j+1} + \mu^*_{ij})}{(\mu^*_{i,j+1} - \mu^*_{ij})(\mu^*_{i+1,j} - \mu^*_{ij})} \tag{6.14}$$

can be used instead. For the particular case of binary variables, both types of cross-ratios coincide.

For the association (6.13), we will pay particular attention to

$$\ln \psi_{ij} = \phi + \rho_{1i} + \rho_{2j} + \sigma_{1i}\sigma_{2j}, \tag{6.15}$$

including a constant association parameter, row and column effects, and interactions between rows and columns, respectively. This model is identified, e.g., by imposing $\rho_{1I} = \rho_{2J} = \sigma_{1I} = \sigma_{2J} = 0$ and $\sigma_{11} = 1$. Due to the fourth term of (6.15) this parameterization is a member of the nonlinear family (6.4). It is very similar in structure to the local cross-ratios of the R+C+RC model (6.10). Of course, model (6.15) is only one of many possibilities, since there is a whole spectrum of possible models between

independence and constant association on the one hand and a saturated association model on the other hand. When the number of categories increases, it becomes more crucial to look for parsimonious association models in order to reduce the number of parameters in the model. To this end, the more flexible class (6.4) might be preferable over (6.3).

Model fitting proceeds, e.g., via Newton-Raphson or Fisher scoring techniques. Details, for the general case, can be found in Section 7.7.6. To do so, the cumulative cell probabilities need to be computed. First, note that $\mu_{IJ} = 1$. Then, μ_{iJ} and μ_{Ij} follow from η_{1i} and η_{2j}, i.e., (6.11) and (6.12) are solved for μ_{iJ} and μ_{Ij}. The other counts follow from

$$\mu_{ij} = \begin{cases} \frac{1+[\mu_{iJ}+\mu_{Ij}](\psi_{ij}-1)-S_{ij}}{2(\psi_{ij}-1)} & \text{if } \psi_{ij} \neq 1, \\ \mu_{iJ}\mu_{Ij} & \text{otherwise,} \end{cases} \tag{6.16}$$

where

$$S_{ij} = \sqrt{[1+(\psi_{ij}-1)(\mu_{iJ}+\mu_{Ij})]^2 + 4\psi_{ij}(1-\psi_{ij})\mu_{iJ}\mu_{Ij}}.$$

Molenberghs and Lesaffre (1994, 1999) show how to extend this class of models to more than two variables. They also indicate how to adopt other association measures, such as marginal correlations, which corresponds to the Bahadur (1961) model. Molenberghs (1994) and Lesaffre, Verbeke, and Molenberghs (1994) provide details on maximum likelihood estimation for the two-way and higher order versions of the model. See also Section 7.7.

6.2.4 A General Class of Models

The models described in Sections 6.2.2 and 6.2.3 differ in two respects:

1. The association in the RC model is in terms of local cross-ratios, while the Dale model is based on global cross-ratios. This difference is not essential, as we argued that local cross-ratios can be incorporated into the marginal model without problem.

2. The marginal probabilities of the RC model are complex functions of the model parameters, whereas the Dale model is expressed directly in terms of the marginal logits.

However, upon generalizing (6.4) slightly, both models are seen as subclasses of this flexible family. For both models, linear and non-linear predictors are possible. Indeed, for the RC family, (6.8) is linear whereas (6.6) is non-linear. For the Dale model, (6.15) is non-linear, but if the fourth term is dropped, it becomes a linear predictor.

The advantage of this result is that completely general models can be constructed, combining and extending interesting aspects of both the RC

TABLE 6.1. *British Occupational Study. Cross-classification of male sample according to each subject's occupational status category and his father's occupational category, using seven status categories*

Father's	Subject's status						
status	1	2	3	4	5	6	7
1	50	19	26	8	18	6	2
2	16	40	34	18	31	8	3
3	12	35	65	66	123	23	21
4	11	20	58	110	223	64	32
5	14	36	114	185	714	258	189
6	0	6	19	40	179	143	71
7	0	3	14	32	141	91	106

and the Dale model. For example, a genuine marginal model can be constructed, with an association function of the RC type. Depending on the data problem, one can opt for local or for global cross-ratios. Arguably, local cross-ratios are suitable for nominal variables, whereas global cross-ratios are a natural choice for cross-classified ordinal variables.

6.3 British Occupational Status Study

We re-analyze the data presented in Goodman (1979). Subjects are cross-classified, according to their occupational status and their father's occupational status, using seven ordered categories. The data are presented in Table 6.1

Standard RC and Dale models, fitted to Table 6.1, are presented in Table 6.2. The Dale model with row effects, column effects, and interactions, provides a good fit, based on a deviance χ^2 approach. This means that no model of the form (6.3) fits the data and that the full non-linear version (6.15) is necessary to achieve an acceptable fit. No RC model, not even the R+C+RC model, fits the data well.

6.4 The Caithness Data

Goodman (1981a) studied association models for two-way contingency tables with ordered categories. The cross-classification of eye color and hair color of 5387 children is reproduced in Table 6.3.

Goodman treated these responses as ordinal that, although sensible, might be open to discussion. We combine marginal probabilities, one set

TABLE 6.2. *British Occupational Study. Deviance χ^2 goodness-of-fit statistics for Dale and RC models, fitted to the data in Table 6.1. The models with an acceptable fit ($p > 0.05$) are indicated by an asterisk.*

	Dale		RC	
Description	df	χ^2	df	χ^2
Independence	36	897.52	36	897.52
Constant association	35	207.23	35	98.19
Row effects only	30	105.23	30	87.14
Column effects only	30	100.69	30	80.74
Row and column effects	25	42.94	25	75.59
Row, column, interactions	16	*20.11	16	38.09
Saturated model	0	0.00	0	0.00

TABLE 6.3. *Caithness Data. Eye color and hair color of 5387 children in Caithness (Goodman 1981a).*

Eye	Hair color				
color	Fair	Red	Medium	Dark	Black
Blue	326	38	241	110	3
Light	688	116	584	188	4
Medium	343	84	909	412	26
Dark	98	48	403	681	85

for each variable, with local odds ratios to describe the association. We consider two models. The first one (8 parameters) assumes a constant local odds ratio. The simpler model which assumes independence between both responses has been shown by Goodman to provide a poor fit and will not be considered here. The second, saturated, model allows an unstructured 3×4 table of local odds ratios. The marginal probabilities for *both* models are $(0.13, 0.29, 0.33, 0.25)$ for eye color and $(0.27, 0.05, 0.40, 0.26, 0.02)$ for hair color. The common local odds ratio for the first model equals 1.50. The deviance is 131.10 on 11 degrees of freedom, rejecting the constant (or uniform) association model. Note that Goodman's *conditional* model for uniform association exhibited a much poorer deviance of 265.03 on 11 degrees of freedom. The 12 local odds ratios, organized as an association

TABLE 6.4. *Fluvoxamine Trial. Cross-classification of initial severity and side effect at the second occasion. In parentheses, the fitted values for the independence model are shown.*

	Side 2			
Severity	1	2	3	4
1	1	0	1	0
	(0.86)	(0.86)	(0.18)	(0.10)
2	21	28	5	5
	(25.29)	(25.29)	(5.42)	(3.01)
3	62	62	15	7
	(62.57)	(62.57)	(13.41)	(7.45)
4	41	31	6	2
	(34.29)	(34.29)	(7.35)	(4.08)
5	1	5	0	1
	(3.00)	(3.00)	(0.64)	(0.36)

table, are:

	1/2	2/3	3/4	4/5
1/2	1.45	0.79	0.71	0.78
2/3	1.45	2.15	1.41	2.97
3/4	2.00	0.78	3.73	1.98

.

Although there is some fluctuation in the association structure, it is very hard to pinpoint a clear trend. This is typically much harder for multivariate data than for genuinely longitudinal data where, for example, exchangeable (constant) or exponentially decaying structures are commonly encountered.

6.5 Analysis of the Fluvoxamine Trial

Let us consider the fluvoxamine trial, presented in Section 2.4. Further analyses will be given in various sections of Chapter 7, as well as in the missing data Chapters 29, 30, and 31.

Because the focus here is on marginal models for contingency tables coming from repeated categorical data, we select four two-way classifications from the fluvoxamine study. We will first consider a cross-classification of side effects and initial severity (Table 6.4). Then, we cross-classify the

TABLE 6.5. *Fluvoxamine Trial. Cross-Classification of therapeutic effect at the second and third Occasion. In parentheses, the fitted values: the first entry corresponds to the constant association Dale model, while the second entry stands for the row and column local association model.*

	Ther. 3			
Ther. 2	1	2	3	4
1	13	2	0	0
	(11.64)	(2.87)	(0.49)	(0.15)
	(13.06)	(1.87)	(0.06)	(0.01)
2	37	40	8	4
	(40.46)	(39.77)	(5.50)	(1.39)
	(34.98)	(44.28)	(7.55)	(2.20)
3	13	58	18	4
	(10.09)	(53.94)	(23.38)	(4.77)
	(15.65)	(52.42)	(18.49)	(6.45)
4	1	13	36	21
	(2.68)	(16.71)	(32.52)	(21.64)
	(0.32)	(14.44)	(35.91)	(20.34)

measurements on therapeutic effect at visits 2 and 3 in Table 6.5. A similar table is constructed for side effects (Table 6.6). Finally, we consider a cross-classification of side effects and therapeutic effect, recorded at visit 2 (Table 6.7). Note that the total of Table 6.7 (299) is higher than the total of Table 6.4 (294), as there are 5 subjects with information on therapeutic and side effects, but without initial severity measurement. These tables cover different settings: a cross-classification of an outcome and a baseline variable, the same outcome at subsequent measurement times and a "cross-sectional" picture, composed of two variables measured simultaneously. Table 6.8 shows the data from Table 6.7, split by sex category.

Let us now analyze these data. Table 6.9 summarizes the deviance χ^2 goodness-of-fit statistics for the models fitted to Tables 6.4–6.8.

Table 6.4 shows a complete lack of association. As a consequence, the independence model is accepted for both the Dale and the RC model. Of course, the deviance for the independence model in both families is exactly the same. Initial severity measures symptoms present at baseline, whereas side effects measures symptoms induced by the therapy. Thus the independence model implies that incidence and intensity of side effects do not depend on the initial conditions. Note that for Tables 6.4–6.8 the R+C+RC

TABLE 6.6. *Fluvoxamine Trial. Cross-classification of side effects at the second and third occasion. In parentheses, the fitted values: the first entry corresponds to the row and column effects Dale model, while the second entry corresponds to model (6.7).*

	Side 3			
Side 2	1	2	3	4
1	105	14	0	0
	(104.98)	(13.84)	(0.16)	(0.00)
	(105.01)	(13.98)	(0.01)	(0.00)
2	34	80	7	1
	(33.88)	(80.46)	(7.27)	(0.27)
	(33.63)	(79.96)	(8.20)	(0.22)
3	2	7	10	2
	(2.09)	(7.02)	(8.76)	(2.91)
	(2.71)	(7.14)	(7.58)	(3.57)
4	3	1	0	2
	(3.14)	(1.01)	(0.00)	(2.21)
	(2.65)	(0.92)	(1.21)	(1.22)

model is overparameterized and thus coincides with the saturated model, whence it is not included in Table 6.9.

For Table 6.5, we find a strong association main effect with the Dale model. The constant global cross-ratio is very high: $\widehat{\psi} = \widehat{\psi}_{ij} = \exp(2.52) = 12.43$. Note that this model corresponds to an underlying Plackett (1965) distribution, as such a distribution is characterized by a constant global cross-ratio. The fit improves by 7.68 on 2 degrees of freedom if we add a row effect. This model deserves our preference. For the RC family, there is certainly a strong constant association effect, but the fit is not acceptable at that point. A fully satisfactory fit is provided by the row and column association model.

There is also a clear global association main effect in Table 6.6. Including this parameter improves the fit of the model dramatically, although adding both row and column effects provides a better fit. Associations are shown in Table 6.10. Some of the observed cross-ratios are infinite, due to zero cells in the contingency table. All but one associations are high to extremely high. High associations in the upper right corner are explained by the fact that side effects over time are of course highly correlated, but also tend to go down, and only rarely go up, showing that the drug has a beneficial effect. It is remarkable that no RC model fits the data well, as can be learned

TABLE 6.7. *Fluvoxamine Trial. Cross-classification of side effects and therapeutic effect at the second occasion. In parentheses, the fitted values are shown. The first model is the global association column effects model. The second global cross-ratio model includes row and column effects, as well as interactions. The third set of fitted values corresponds to the RC model (row and column effects).*

	Therapeutic 2			
Side 2	1	2	3	4
1	8	40	40	40
	(7.46)	(38.32)	(44.19)	(38.11)
	(8.10)	(39.58)	(40.41)	(39.50)
	(8.91)	(38.18)	(41.85)	(39.05)
2	7	45	51	25
	(9.73)	(45.60)	(43.33)	(29.09)
	(6.61)	(46.37)	(49.46)	(25.49)
	(6.33)	(46.92)	(48.99)	(25.75)
3	2	9	8	9
	(1.37)	(7.82)	(10.39)	(8.52)
	(2.22)	(7.49)	(9.66)	(8.64)
	(2.02)	(8.12)	(8.95)	(8.91)
4	2	1	3	9
	(0.32)	(2.15)	(4.45)	(8.15)
	(2.24)	(1.30)	(2.52)	(8.95)
	(1.74)	(1.77)	(2.20)	(9.29)

from Table 6.9. In conclusion, a marginal model such as the Dale model fits the data better than a model from the RC family. Should one choose to remain within the RC family, then a model of a more elaborate nature, such as the ones discussed in Section 6.7, might be needed. Note, again, that the R+C+RC model is no alternative, as it is overparameterized. Fitting related model (6.7) to Table 6.7 yields an acceptable fit: $\chi^2 = 6.33$ on 4 degrees of freedom ($p = 0.1760$).

Both Tables 6.5 and 6.6 are cross-classifications of an ordinal variable, recorded at two subsequent measurement times. In both cases, a parsimonious global association model explains the data well. It seems to be much harder to fit these data with local association models.

For Table 6.7, the row effects model is the most parsimonious one that provides an acceptable fit. One might argue that it is careful to retain the model adding column effects and interactions as well. Therefore, fitted frequencies for both models are shown in Table 6.7. Table 6.11 shows the global cross-ratios for the data of Table 6.7, together with the pre-

TABLE 6.8. *Fluvoxamine Trial. Cross-classification of side effects and therapeutic effect at the second occasion, split by sex.*

	Therapeutic 2			
Side 2	1	2	3	4
Male subjects				
1	4	18	12	16
2	0	9	19	9
3	0	4	3	4
4	0	1	1	5
Female subjects				
1	4	22	28	24
2	7	36	32	16
3	2	5	5	5
4	2	0	2	4

dicted values under both models. We observe two patterns in Table 6.11. First, the association increases along the main diagonal. This means that the association between the variables $I(\text{SIDE2} \leq 1)$ and $I(\text{THER2} \leq 1)$ is smaller than the association between the variables $I(\text{SIDE2} \leq 3)$ and $I(\text{THER2} \leq 3)$. Here, $I(\cdot)$ is an indicator function. Also, the association becomes "negative" (i.e., smaller than 1 on the cross-ratio scale) for pairs such as $I(\text{SIDE2} \leq 3)$ and $I(\text{THER2} \leq 1)$. The RC models, fitted to this table, suggest the selection of the row and column effects model. The fitted model is also presented in Table 6.7. All RC models are based on model (6.5).

In conclusion, the Dale model yields a non-linear association model for Tables 6.1 and 6.7, through the interaction terms in (6.15), which is a very natural association model as it is a Dale model analogue of Goodman's R+C+RC model, of which the cross-ratios are given by (6.10). For Tables 6.4–6.6, simpler association models, including at most row and/or column effects, but no interactions, are found to be acceptable. The models of RC type fitted to these data tend to be of a more complex nature, arguably because they model the association through local cross-ratios even though the data are ordered categorical.

6.6 Extensions

As mentioned earlier, the fluvoxamine study recorded more than two outcomes and further there is covariate information available. We consider

TABLE 6.9. *Fluvoxamine Trial. Deviance χ^2 goodness-of-fit statistics for Dale and RC models, fitted to Tables 6.4–6.8 (models with an acceptable fit are indicated by an asterisk).*

	Table 6.4		Table 6.5		Table 6.6		Table 6.7	
Description	df	χ^2	df	χ^2	df	χ^2	df	χ^2
Dale models								
Independence	12	*14.20	9	141.95	9	158.15	9	17.12
Constant association	11	*11.71	8	*11.48	8	18.27	8	17.12
Row effects only	8	*8.34	6	*3.80	6	14.49	6	*9.78
Column effects only	9	*11.37	6	*10.26	6	*12.29	6	16.74
Row and column effects	6	*8.03	4	*1.29	4	*2.05	4	*9.31
Row, column, interactions	2	*0.22	1	*0.31	1	*0.35	1	*0.94
Saturated model	0	0.00	0	0.00	0	0.00	0	0.00
RC models								
Independence	12	*14.20	9	141.95	9	158.15	9	17.12
Constant association	11	*12.04	8	19.46	8	48.66	8	16.71
Row effects only	8	*8.21	6	12.90	6	18.84	6	*11.69
Column effects only	9	*11.88	6	14.35	6	45.12	6	15.14
Row and column effects	6	*2.22	4	*5.16	4	10.48	4	*1.44
Saturated model	0	0.00	0	0.00	0	0.00	0	0.00

in turn two ways of extending the models described so far, while still remaining within the contingency table framework. First, in Section 6.6.1, we discuss the inclusion of a dichotomous covariates in marginal association models, followed by a generalization to three-way tables (Section 6.6.2). These extensions are members of the class (6.4). Completely general covariates, as well as multi-way tables and fully longitudinal models are the subject of Chapter 7.

6.6.1 Covariates

The marginal Dale model presented here is flexible in incorporating covariate effects. Their influence on the marginal means and on the association can be described in separate ways. For example, age could be found to influence the marginal response functions, while the association could be seen to change with sex. We will exemplify the possibilities that are brought about by this feature using two covariates. First, the data presented in Table 6.7 are split into two sex groups (Table 6.8). Second, we will add the effect

TABLE 6.10. *Fluvoxamine Trial. Global cross ratios for the classification of side effects at time 2 versus time 3 (Data in Table 6.6).*

	Side 3		
Side 2	1	2	3
	Observed		
1	21.15	$+\infty$	$+\infty$
2	6.00	31.37	41.74
3	1.17	6.05	43.17
	Row and column effects		
1	21.07	116.88	760.06
2	5.70	31.65	205.37
3	1.20	6.67	43.26

of the continuous covariate age on the responses and on the association between responses.

Let us consider sex first. Selected models, fitted to these data, are presented in Table 6.12. Obviously the marginal regressions are independent of sex, but we do find a sex effect in the association. If we add row effects (but no column effects), the fit is satisfactory ($p = 0.12$). The association structure of this model is:

$$\ln \psi_{ijr} = 1.64 - 0.88\text{sex}_r - 1.24I(i = 1) - 0.56I(i = 2),$$

where ψ_{ijr} is the global cross-ratio, depending on subject r through their sex, and $I(.)$ is the indicator function. The association is stronger for males than for females ($p = 0.0402$).

Even though Table 6.8 contains four sampling zeros, no convergence problems are encountered and all parameters lie in the interior of their space. The Dale likelihood attains its maximum in the interior of the parameter space under very mild conditions, a feature shared with univariate ordinal logistic regression, which it generalizes. First, there must not be a complete separation in the covariate space between response groups. A similar condition was derived for the multigroup logistic model by Albert and Lesaffre (1986). Second, even with zero cell counts, models can be constructed for which the MLE lies in the interior of the space. For example, in a 3×3 table with cells $(1,1)$, $(1,3)$, $(2,2)$, $(3,1)$, and $(3,3)$ equal to zero (with the other cells non-zero), a model with global cross-ratio dependent on row and column classification, yields finite estimates. We can easily include such continuous covariates as age. For 296 subjects out of 299 recorded in Table 6.7, age (in years) is recorded. Age ranges from 16

TABLE 6.11. *Fluvoxamine Trial. Global cross ratios for the classification of side effects versus therapeutic effect (both at the second occasion).*

Side 2	Therapeutic 2: 1	2	3
	Observed		
1	0.97	0.95	0.74
2	0.61	1.33	2.12
3	0.41	2.57	4.26
	Column effects only		
1	0.86	0.86	0.86
2	1.77	1.77	1.77
3	3.24	3.24	3.24
	Row, column, interaction		
1	0.92	0.86	0.80
2	0.55	1.55	1.92
3	0.37	2.17	4.00

to 75 years, with a mean of 42.2 years (median is 40.5 years). There are 97 distinct age by sex combinations, which yields an average of about 3 subjects per distinct 4×4 table! Thus, we have a generalization of a purely contingency table analysis to multivariate ordinal regression. Obviously, a saturated model is not meaningful here, as the number of covariate levels (and hence the number of cells) increases with the sample size. Derivation of formal goodness-of-fit tools, such as appropriate residuals, requires further research. The most complex model we will consider, contains sex and age effects in both the marginal mean and in the association and lets the association further depend on row and column classification. Clearly, this model could be extended (for example, by means of higher order effects of age and interactions between sex and age). Table 6.13 reports on a backward selection performed to simplify the model. In the final model, the marginal logits are simplified such that only SIDE2 depends on age. The association is independent of the column classification. Although sex and age could be omitted from the association when comparing Models 4 and 5 with 3 (numbers referring to Table 6.13), or 6 with 4 and 5, a direct comparison of Model 6 (no covariate influence on association) with Model 3 (both age and sex influence the association) is significant at the 5% level. Therefore, we prefer Model 3. The cumulative logits (6.11) and (6.12) for

TABLE 6.12. *Fluvoxamine Trial. Deviance χ^2 goodness-of-fit statistics for Dale models, fitted to Table 6.8 (distinguishing between sex groups). Models with an acceptable fit ($p > 0.05$) are indicated by an asterisk.*

Marginal model	Association model	df	χ^2
No sex effect	Constant	23	39.32
No sex effect	Sex, row	20	*27.76
No sex effect	Sex, row, column	18	*27.30
Sex effect	Constant	21	36.80
Sex effect	Sex effect	20	31.74
Sex effect	Sex, row, column	16	*25.53
Saturated	Saturated	0	0.00

subject r are

$$\begin{aligned} \eta_{1ir} &= 0.54I(i=1) + 2.63I(i=2) + 3.75I(i=3) - 0.019\text{age}_r, \\ \eta_{2jr} &= -2.69I(j=1) - 0.47I(j=2) + 0.95I(j=3), \end{aligned}$$

and the association structure is

$$\ln \psi_{ijr} = 2.94 - 0.80\text{sex}_r - 0.028\text{age}_r - 1.44I(i=1) - 0.63I(i=2).$$

The logit for side effects decreases with age, implying, e.g., that the probability of category 1 (no side effects) decreases and the probability of category 4 (highest level of side effects) increases with age. The association is stronger for males than for females (consistent with Table 6.12) and decreases with age.

6.6.2 Three-way Contingency Tables

Molenberghs and Lesaffre (1994) extended the Dale model, originally constructed for two response variables, to arbitrary dimensions. This implies that the model is suitable to analyze multi-way contingency tables. Computational details can be found in Molenberghs and Lesaffre (1994). We apply the general method technique on the fluvoxamine data set, more specifically to a cross-classification of therapeutic effect at visits 2, 3, and 4. The data are presented as a $4 \times 4 \times 4$ contingency table (Tables 6.14–6.17). There are 242 patients with measurements on all three outcomes.

Let the variables Y_1, Y_2, and Y_3 have I, J, and K levels, respectively, and define cumulative three-way probabilities μ_{ijk} ($i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K$), similar to the definition in Section 6.2.

The model extends as follows. Apart from three sets of marginal parameters, one for each measurement time:

$$\eta_{1i} = \text{logit}[\text{pr}(Y_1 \leq i)] = \ln(\mu_{iJK}) - \ln(1 - \mu_{iJK}), \qquad (6.17)$$

TABLE 6.13. *Fluvoxamine Trial. Backward selection for Dale models, fitted to Table 6.8 (including sex and age). The number of model parameters (Par), the deviance (Dev) of the model are reported. For each model comparison, the reference model (Vs), and the corresponding χ^2 statistic and p-value are reported. ('R' stands for row effects and 'C' stands for column effects.)*

Nr	Side 2	Ther. 2	Association	Par	Dev	Vs	df	χ^2	p
1	Sex, age	Sex, age	Sex, age, R, C	17	1372.5				
2	Age	—	Sex, age, R, C	14	1375.0	1	3	2.52	0.472
3	Age	—	Sex, age, R	12	1375.5	2	2	0.50	0.779
4	Age	—	Age, R	11	1379.0	3	1	3.47	0.063
5	Age	—	Sex, R	11	1379.0	3	1	3.49	0.062
6	Age	—	R	10	1382.7	4	1	3.66	0.056
6	Age	—	R	10	1382.7	3	2	7.13	0.028

$$\eta_{2j} = \text{logit}[\text{pr}(Y_2 \leq j)] = \ln(\mu_{IjK}) - \ln(1 - \mu_{IjK}), \tag{6.18}$$

$$\eta_{3k} = \text{logit}[\text{pr}(Y_3 \leq k)] = \ln(\mu_{IJk}) - \ln(1 - \mu_{IJk}), \tag{6.19}$$

$(i = 1, \ldots, I-1; j = 1, \ldots, J-1; k = 1, \ldots, K-1)$, there are also three sets of pairwise association parameters:

$$\ln \psi_{12,ij} = \ln \frac{\mu_{ijK}(1 - \mu_{IjK} - \mu_{iJK} + \mu_{ijK})}{(\mu_{iJK} - \mu_{ijK})(\mu_{IjK} - \mu_{ijK})}, \tag{6.20}$$

$$\ln \psi_{13,ik} = \ln \frac{\mu_{iJk}(1 - \mu_{IJk} - \mu_{iJK} + \mu_{iJk})}{(\mu_{iJK} - \mu_{iJk})(\mu_{IJk} - \mu_{iJk})}, \tag{6.21}$$

$$\ln \psi_{23,jk} = \ln \frac{\mu_{Ijk}(1 - \mu_{IJk} - \mu_{IjK} + \mu_{Ijk})}{(\mu_{IJk} - \mu_{Ijk})(\mu_{IjK} - \mu_{Ijk})}, \tag{6.22}$$

together with a set of three-way associations (generalized cross-ratios):

$$\begin{aligned} \ln \psi_{123,ijk} = \\ \ln \Bigg[& \frac{\mu_{ijk}(\mu_{iJK} - \mu_{ijK} - \mu_{iJk} + \mu_{ijk})}{(\mu_{ijK} - \mu_{ijk})(\mu_{iJk} - \mu_{ijk})} \\ & \times \frac{(\mu_{IjK} - \mu_{ijK} - \mu_{Ijk} + \mu_{ijk})}{(\mu_{Ijk} - \mu_{ijk})} \\ & \times \frac{(\mu_{IJk} - \mu_{iJk} - \mu_{Ijk} + \mu_{ijk})}{(1 - \mu_{iJK} - \mu_{IjK} - \mu_{IJk} + \mu_{ijK} + \mu_{iJk} + \mu_{Ijk} - \mu_{ijk})} \Bigg]. \end{aligned} \tag{6.23}$$

Clearly, the link functions (6.17)–(6.23) are all expressed in terms of contrasts of log-probabilities and hence fit in (6.2). Molenberghs and Lesaffre

TABLE 6.14. *Fluvoxamine Trial. Cross-classification of therapeutic effect at the second, third, and fourth occasion. In parentheses, the fitted values. The first entry corresponds to Model 1, the second entry corresponds to Model 2, the third entry corresponds to the generalized RC model. Part I.*

		Side 4			
Side 2	Side 3	1	2	3	4
1	1	11	1	0	0
		(10.18)	(1.19)	(0.10)	(0.02)
		(13.75)	(2.07)	(0.18)	(0.05)
		(10.99)	(0.48)	(0.00)	(0.00)
	2	0	1	1	0
		(0.60)	(1.46)	(0.30)	(0.06)
		(0.89)	(1.88)	(0.40)	(0.10)
		(1.16)	(0.97)	(0.02)	(0.00)
	3	0	0	0	0
		(0.05)	(0.18)	(0.15)	(0.06)
		(0.08)	(0.21)	(0.13)	(0.05)
		(0.05)	(0.15)	(0.07)	(0.01)
	4	0	0	0	0
		(0.01)	(0.03)	(0.04)	(0.02)
		(0.02)	(0.05)	(0.04)	(0.02)
		(0.01)	(0.02)	(0.03)	(0.03)

(1994, 1999) describe ways to determine the joint probabilities μ_{ijk} from the links and to compute maximum likelihood estimates. Indeed, the key issue in a marginal model of this type is the construction of the joint probabilities. The univariate marginal probabilities μ_{iJK}, μ_{IjK}, and μ_{IJk} are easily determined from inverting (6.17)–(6.19), just as with (6.11) and (6.12). The pairwise marginal probabilities μ_{ijK}, μ_{iJk}, and μ_{Ijk}, can be written in analogy with (6.16), as links (6.20)–(6.22) have the same form as (6.13). Determining the third order cumulative probabilities μ_{ijk} is more difficult and details are given in Chapter 7, in particular in Sections 7.3 and 7.7.

To illustrate the model, let us analyze the three therapeutic effect measurements. Model 1 assumes the marginal logits (6.17)–(6.19) are independent of covariate effects, yielding 9 marginal parameters. Each of the association parameters ψ in (6.20)–(6.23) is assumed independent of covariate effects as well as of the category indicators i, j, and k, yielding three pairwise and one three-way association parameters. This brings the

TABLE 6.15. *Fluvoxamine Trial. Cross-classification of therapeutic effect at the second, third, and fourth occasion. In parentheses, the fitted values. The first entry corresponds to Model 1, the second entry corresponds to Model 2, the third entry corresponds to the generalized RC model. Part II.*

		Side 4			
Side 2	Side 3	1	2	3	4
2	1	33	2	0	0
		(36.27)	(2.92)	(0.33)	(0.08)
		(30.39)	(3.18)	(0.35)	(0.09)
		(32.34)	(2.76)	(0.00)	(0.00)
	2	13	23	2	0
		(13.80)	(18.28)	(2.14)	(0.39)
		(13.03)	(16.81)	(2.19)	(0.46)
		(16.88)	(17.72)	(0.73)	(0.00)
	3	1	2	3	0
		(0.47)	(1.84)	(1.87)	(0.62)
		(0.44)	(1.31)	(1.05)	(0.40)
		(1.48)	(4.48)	(2.39)	(0.24)
	4	0	1	1	1
		(0.10)	(0.30)	(0.34)	(0.21)
		(0.10)	(0.29)	(0.22)	(0.11)
		(0.20)	(0.83)	(0.99)	(0.95)

total number of parameters to 13. Marginal parameter estimates (standard errors in parentheses) are

$$\begin{array}{rclrclrcl} \widehat{\eta}_{11} &=& -2.76(0.27) & \widehat{\eta}_{21} &=& -1.04(0.14) & \widehat{\eta}_{31} &=& -0.21(0.13) \\ \widehat{\eta}_{12} &=& -0.45(0.13) & \widehat{\eta}_{22} &=& 0.75(0.13) & \widehat{\eta}_{32} &=& 1.58(0.17) \\ \widehat{\eta}_{13} &=& 1.00(0.15) & \widehat{\eta}_{23} &=& 2.40(0.22) & \widehat{\eta}_{33} &=& 3.12(0.32). \end{array}$$

The constant global cross-ratios are $\widehat{\psi}_{12} = \widehat{\psi}_{12,ij} = \exp(2.58) = 13.18(3.08)$ for the first and the second outcome, $\widehat{\psi}_{13} = \widehat{\psi}_{13,ik} = \exp(1.38) = 3.99(0.89)$ for the first and the third outcome, and $\widehat{\psi}_{23} = \widehat{\psi}_{23,jk} = \exp(3.08) = 21.76(5.74)$ for the second and the third outcome. The three-way interaction, $\widehat{\psi}_{123} = \widehat{\psi}_{123,ijk} = \exp(0.18) = 1.19(0.66)$, is not significantly different from 1. Fitted frequencies are given in Tables 6.14–6.17.

The overall deviance goodness-of-fit statistic is 37.13 on 50 degrees of freedom ($p = 0.9115$). Inspecting standardized residuals, 62 out of 64 are

TABLE 6.16. *Fluvoxamine Trial. Cross-classification of therapeutic effect at the second, third, and fourth occasion. In parentheses, the fitted values. The first entry corresponds to Model 1, the second entry corresponds to Model 2, the third entry corresponds to the generalized RC model. Part III.*

		Side 4			
Side 2	Side 3	1	2	3	4
3	1	12	1	0	0
		(8.39)	(1.04)	(0.16)	(0.04)
		(7.41)	(1.12)	(0.17)	(0.05)
		(12.59)	(1.70)	(0.00)	(0.00)
	2	25	25	1	1
		(22.86)	(24.33)	(1.45)	(0.27)
		(24.97)	(28.79)	(2.09)	(0.38)
		(19.74)	(24.31)	(1.48)	(0.00)
	3	1	8	5	1
		(1.15)	(10.76)	(7.30)	(1.56)
		(1.33)	(9.96)	(6.61)	(1.64)
		(2.75)	(8.61)	(4.98)	(0.63)
	4	0	3	0	0
		(0.22)	(0.83)	(1.16)	(0.97)
		(0.30)	(1.05)	(1.02)	(0.71)
		(0.43)	(1.78)	(2.07)	(1.91)

less than 2 in absolute value, the remaining ones being 2.24 and 2.39. Thus, model fit is acceptable, but one might want to simplify the model further. We will in turn simplify the marginal and association structures. First, the three sets of logits reveal an increase over time, suggesting an improving response to therapy. A simpler model would assume: $\eta_{1i} = \alpha_i$, $\eta_{2j} = \alpha_j + \pi_2$, and $\eta_{3k} = \alpha_k + \pi_3$ $(i, j, k = 1, 2, 3)$. We interpret α_1, α_2, and α_3 as cut-off points at the first occasion and π_2 and π_3 as "proportional" shift parameters at occasions 2 and 3 respectively. Second, one might argue that the association between outcomes is mainly a function of the time lag between the outcomes, but not so much of the measurement times themselves. This is supported by the fact that $\ln \psi_{12}$ and $\ln \psi_{23}$ are roughly the same (given their standard errors of about 0.24), with $\ln \psi_{13}$ approximately half of the other association. Should one grant belief to this assumption, then an association model of the form $\gamma = \ln \psi_{12} = 2 \ln \psi_{13} = \ln \psi_{23}$ might be considered. The multiplier 0.5 for $\ln \psi_{13}$ is suggested by the data and has

TABLE 6.17. *Fluvoxamine Trial. Cross-classification of therapeutic effect at the second, third, and fourth occasion. In parentheses, the fitted values. The first entry corresponds to Model 1, the second entry corresponds to Model 2, the third entry corresponds to the generalized RC model. Part IV.*

		Side 4			
Side 2	Side 3	1	2	3	4
4	1	1	0	0	0
		(1.96)	(0.42)	(0.07)	(0.02)
		(1.42)	(0.36)	(0.06)	(0.02)
		(0.08)	(0.05)	(0.00)	(0.00)
	2	5	6	0	0
		(8.87)	(5.58)	(0.46)	(0.11)
		(8.16)	(5.55)	(0.44)	(0.11)
		(5.50)	(11.75)	(2.70)	(0.03)
	3	7	18	9	1
		(3.07)	(19.88)	(7.74)	(0.99)
		(3.26)	(19.69)	(6.67)	(0.97)
		(3.82)	(13.39)	(10.15)	(2.78)
	4	0	2	8	6
		(0.51)	(3.13)	(7.48)	(4.78)
		(0.68)	(4.42)	(7.84)	(4.45)
		(0.98)	(3.98)	(4.37)	(3.41)

limited empirical or theoretical support. Alternatively, one could estimate this parameter from the data. Third, the three-way interaction can be set to zero. There are six parameters in total.

Parameter estimates (standard errors) for this model are estimated to be $\widehat{\alpha}_1 = -2.41(0.17)$, $\widehat{\alpha}_2 = -0.52$ (0.12), $\widehat{\alpha}_3 = 1.02(0.14)$ for the cut-off points, with time shifts $\widehat{\pi}_2 = 1.32(0.11)$ for the second period and $\widehat{\pi}_3 = 2.17(0.16)$ for the third period. The single association parameter is equal to $\widehat{\gamma} = 2.81(0.17)$, resulting in $\widehat{\psi}_{12} = \widehat{\psi}_{23} = 16.56(2.77)$ and $\widehat{\psi}_{13} = \sqrt{16.56} = 4.07(0.34)$. Fitted frequencies are given in Table 6.14–6.17. This model has a deviance of 43.67 on 57 degrees of freedom ($p = 0.9029$), and again only two standardized residuals are larger than 2 (being 2.07 and 2.70), showing that there is some support in the data for the assumed model. Finally, comparing Models 1 and 2, yields a deviance of 6.54 on 7 degrees of freedom ($p = 0.4782$), indicating that the first and more complex model is not necessary.

A similar model is obtained from the analysis of side effects at times 2, 3, and 4. Analyzing initial severity, side effects at time 2, and therapeutic effect at time 2, yields a satisfactory model with only constant association. No details on these models are included. These results are promising because they support the thesis that for a range of ordinal data applications, parsimonious marginal global cross-ratio models are sufficient to describe the data.

In case nominal data are to be analyzed, then the model can be adapted to cell probabilities μ^*_{ijk}. This would mean that (6.17)–(6.23) have to be changed in the spirit of (6.14). In particular, the global cross-ratios might have to be replaced by their local counterparts.

In addition to the extensions studied sofar, it is possible to extend the RC model to more than two dimensions. One option is to generalize Model (6.5) by defining

$$\mu^*_{ijk} = \alpha_i \beta_j \gamma_k e^{\phi \lambda_i \nu_j \omega_k} \tag{6.24}$$

with obvious notation. Of course, the marginal pairwise local odds ratio for a pair (i, j) has a very complicated form and (6.5) is not a submodel of (6.24) in the sense that the interpretation of the parameters will change in passing from a bivariate to a trivariate model. The conditional pairwise odds ratio on the other hand is $\ln \theta^*_{ij|k} = \phi \omega_k (\lambda_i - \lambda_{i+1})(\nu_j - \nu_{j+1})$, where ω_k can be considered an adjustment for the category conditioned upon. The three-way odds ratio is similar in structure to the two-way odds ratio of the bivariate model (6.5).

Fitting Model (6.24) to the trivariate therapeutic data of Tables 6.14–6.17 yields a deviance of 67.96 on 47 degrees of freedom ($p = 0.0243$), indicating that the fit is not satisfactory. Fitted frequencies are displayed in Table 6.14–6.17. One could consider more elaborate alternatives, such as trivariate versions of the R+C+RC model (6.9). However, as indicated earlier, for this kind of data, the marginal model defined in terms of cumulative probabilities seems to be more promising, as it yields very parsimonious descriptions of the association structure.

An alternative fashion to extend (6.5) would start from three pairwise marginal RC models:

$$\mu^*_{ij+} = \alpha^{(12)}_i \beta^{(12)}_j e^{\phi^{(12)} \lambda^{(12)}_i \nu^{(12)}_j}, \tag{6.25}$$

$$\mu^*_{i+k} = \alpha^{(13)}_i \gamma^{(13)}_k e^{\phi^{(13)} \lambda^{(13)}_i \tau^{(13)}_k}, \tag{6.26}$$

$$\mu^*_{+jk} = \beta^{(23)}_j \gamma^{(23)}_k e^{\phi^{(23)} \nu^{(23)}_j \tau^{(23)}_k}, \tag{6.27}$$

$(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K)$. For (6.25)–(6.27) to define a valid probability mass function μ^*_{ijk}, complicated restrictions must be satisfied: summing (6.25) over j and (6.26) over k yields I restrictions:

$$\alpha^{(12)}_i \sum_{j=1}^{J} \beta^{(12)}_j e^{\phi^{(12)} \lambda^{(12)}_i \nu^{(12)}_j} = \alpha^{(13)}_i \sum_{k=1}^{K} \gamma^{(13)}_k e^{\phi^{(13)} \lambda^{(13)}_i \tau^{(13)}_k},$$

$(i = 1, \ldots, I)$, with similarly J and K restrictions for the other two marginals.

6.7 Relation to Latent Continuous Densities

Several publications are devoted to the comparison of local and global association models. Important references are Goodman (1981b), Mardia (1970), Dale (1984), and Becker (1989). An argument, often used to claim superiority of local over global association models, is the close relationship between Goodman's UM model and discretizations of the bivariate normal distribution (Goodman 1981b, Becker 1989). Also, their close connection with log-linear modeling is brought forward.

Holland and Wang (1987) introduced the *local dependence function* (LDF) of a bivariate continuous density function f as an analog to the local cross-ratios for contingency tables (Yule and Kendall 1950). The probability of a rectangular cell around (x, y) with edges dx and dy is approximated by $f(x, y)dxdy$. For cells around (x, y), (x, v), (u, y) and (u, v), the log local cross-ratio is given by

$$\theta(x, y; u, v) = \ln \left[\frac{f(x, y)f(u, v)}{f(x, v)f(u, y)} \right].$$

The *local dependence function* (LDF) at (x, y) is defined as

$$\gamma_f(x, y) = \lim_{dx \to 0, dy \to 0} \frac{\theta(x, y; x + dx, y + dy)}{dx\, dy} = \frac{\partial^2}{\partial x \partial y} \ln f(x, y). \qquad (6.28)$$

Holland and Wang (1987) show that a bivariate density is characterized by its LDF and its two marginal densities. Further, a bivariate normal is characterized by a constant LDF and two normal marginal densities. Precise statements and proofs are found in Holland and Wang (1987).

The LDF of a normal density with correlation ρ is equal to $\phi = \rho/(1-\rho^2)$. Exactly this quantity, together with appropriately chosen scores α_i, β_j, λ_i and ν_j, are used by Becker (1989) to approximate the discretized normal by (6.5). Note that a special version of the RC model, i.e., the UM model, implies a constant local cross-ratio. It can be observed from Wang (1987), who provides an alternative way of computing normal probabilities, that the local association model introduced by (6.5) and the bivariate normal naturally go together. This explains why the local association models fit far better the discretized normal than do global cross-ratio models. In general, local association models correspond to bivariate densities via the LDF.

An analogous relationship holds between the Dale model and the Plackett distribution (Plackett 1965, Mardia 1970). If the global cross-ratio is constant (or in particular zero) throughout a contingency table, then it

corresponds to a bivariate Plackett distribution (constant "Yulean association"). This was the case for the global association models, selected in the case of Tables 6.4 and 6.5.

However, we observed that model construction is restricted to neither a constant local association, nor a constant global association. Within family (6.4), one can even consider non-linear association models. In particular, we considered various types of row and column effects, together with interactions. This suggests that the normal distribution and the Plackett distribution are not the only ones of interest as continuous distributions, underlying a contingency table. Different forms for the local and global cross-ratios correspond to different distributions.

The correspondence between contingency tables and distribution functions in the Dale model case is very easy. The definition of the distribution is found by the continuous version of (6.16), of which the explicit form is straightforward. A continuous version of (6.15) would include linear (and quadratic) terms in x and y, together with an interaction term.

Let us again turn attention to Goodman's R+C+RC model (6.9). To construct a continuous density having a similar association structure, we first select a local dependence function of the form

$$\gamma(\boldsymbol{\phi}; x, y) = \phi_1 + \phi_2 f_2(x) + \phi_3 g_3(y) + \phi_4 f_4(x) g_4(y), \tag{6.29}$$

where f_k and g_k are integrable functions. Molenberghs and Lesaffre (1999) show how the corresponding density can be approximated. The RC model is found by setting all terms, except those with subscript 4, equal to zero.

The models, fitted to Table 6.1, can be seen as extensions of both an underlying normal and an underlying Plackett distribution. The choice between different models should not be made on the ground of potential classes of underlying densities, but on the shape (structure) of associations. Figure 6.1 presents local and global cross-ratios found from the fitted values of both the RC+R+C model and the global cross-ratio model with row and column effects, as well as interactions. Obviously, there is little pattern in the local cross-ratios, whereas the global cross-ratios show a clear tendency: all associations are high, with an increase if the dichotomy is constructed closer to the categories with low labels, being highest between the variables I(Father's status ≤ 1) and I(Child's status ≤ 1). This implies that social mobility increases with increasing category. There is also slight evidence that the association surface is symmetric, which would then correspond to a global cross-ratio distribution with symmetric global cross-ratio function, such as a symmetric second-degree polynomial.

6.8 Conclusions and Perspective

In this chapter, we presented association models for cross-classified data that belong to the unified multivariate logistic framework, described by

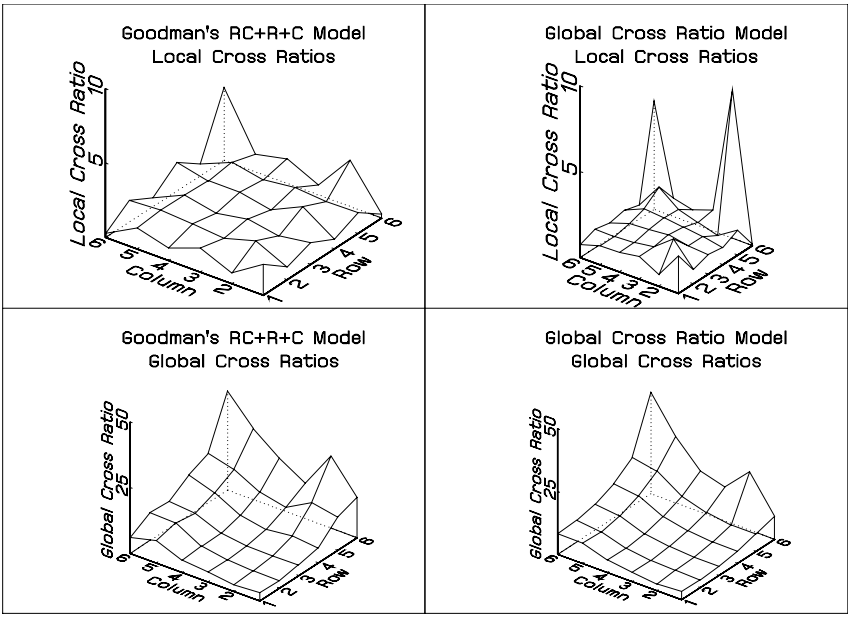

FIGURE 6.1. *Fluvoxamine Trial. Local and global cross-ratios, found from fitting the RC+R+C model and the global cross-ratio model.*

McCullagh and Nelder (1989) and Glonek and McCullagh (1995). This family provides a versatile way of exploring the association structure of cross-classified data. It encompasses both local and global measures of association, with emphasis on cross-ratios (odds ratios), as log cross-ratios can be written as contrasts of log-probabilities. Both fully marginal models, such as the Dale model and its multivariate extensions, as models with a conditional flavor, such as Goodman's (1981a) RC model, are members of this family. Further, linear as well as non-linear link functions (e.g., involving interactions between row and column effects) fit within this family.

We argue that, in spite of the close connection between an RC model and an underlying normal density and the absence of this connection with a fully marginal model, this last category of models provides a flexible toolkit to explore the association structure of cross-classified data, whether of nominal or of ordinal type. We infer from the examples that they often yield parsimonious descriptions of the association structure. Further, marginal association models are easily extended to marginal regression models to include covariate effects. Extensions to multi-way tables are possible, both with the RC as well as with the marginal family.

Both Dale (1984) and Anscombe (1981) suggest the use of global cross-ratios as soon as the outcomes are recorded as ordinal variables. We have shown that this choice is supported by a very good fit for this kind of model

to a range of applications. Further, we claim that the global cross-ratio can lead to interesting interpretations of the association structure itself, which we think is an often neglected aspect of data analysis.

An argument, in favor of RC models, is their computational simplicity. However, with the current state of high quality statistical software, fitting marginal global association models poses no problems.

The Dale model, being a marginal model, is a member of a wider class of marginal models encompassing, for example, the probit model, and allowing for the analysis of multivariate and longitudinal data, with or without covariates, and with measurements sequences of length longer than two. This is the topic of the next chapter.