

ELENA A. EROSHEVA  
STEPHEN E. FIENBERG  
BRIAN W. JUNKER

**Alternative statistical models and representations for  
large sparse multi-dimensional contingency tables**

*Annales de la faculté des sciences de Toulouse 6<sup>e</sup> série*, tome 11,  
n° 4 (2002), p. 485-505

[http://www.numdam.org/item?id=AFST\\_2002\\_6\\_11\\_4\\_485\\_0](http://www.numdam.org/item?id=AFST_2002_6_11_4_485_0)

© Université Paul Sabatier, 2002, tous droits réservés.

L'accès aux archives de la revue « Annales de la faculté des sciences de Toulouse » (<http://picard.ups-tlse.fr/~annales/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## Alternative statistical models and representations for large sparse multi-dimensional contingency tables<sup>(\*)</sup>

ELENA A. EROSHOVA<sup>(1)</sup>, STEPHEN E. FIENBERG<sup>(2)</sup>,  
BRIAN W. JUNKER<sup>(1)</sup>

---

**RÉSUMÉ.** — Une variété de stratégies permet d'approcher l'analyse de larges tableaux creux à l'aide de modèles présentant un nombre modeste de paramètres. Cet article présente quelques unes de ces stratégies, insistant sur le rôle des modèles log-linéaires et le concept d'échangeabilité : des généralisations multidimensionnelles du modèle de quasi-symétrie dû à Henri Caussinus, des versions bayésiennes du modèle de Rasch en théorie des questionnaires et ses généralisations, et le modèle de degré d'appartenance.

**ABSTRACT.** — A variety of strategies allow one to approach the analysis of large sparse contingency tables using models with a modest number of parameters. This paper gives an overview of some of these strategies, emphasizing the role of log-linear models and exchangeability: multi-dimensional generalizations of the model of quasi-symmetry due to Henri Caussinus, Bayesian versions of the Rasch model from item response theory and its generalizations, and the Grade of Membership model.

---

---

(\*) Reçu le 18 septembre 2001, accepté le 18 septembre 2002

The preparation of this paper was supported in part by Grant No. 1R03 AG18986-01, from the National Institute on Aging and by National Science Foundation Grant No. EIA-9876619 to the National Institute of Statistical Sciences.

<sup>(1)</sup> Department of Statistics, Carnegie Mellon University, Pittsburgh PA 15213-3890, U.S.A. E-mail: elenae@stat.cmu.edu and brian@stat.cmu.edu

<sup>(2)</sup> Department of Statistics and Center for Automated Learning & Discovery, Carnegie Mellon University, Pittsburgh PA 15213-3890, U.S.A. E-mail: fienberg@stat.cmu.edu

## 1. Introduction

The literature on log-linear models for the analysis of multi-dimensional contingency tables which arose in the 1960s and 1970s built on fundamental contributions by Birch, Darroch, Good, Goodman, and others (see the overview in Fienberg [29]). It was spurred on by theoretical developments (e.g., see Haberman [36]), practical applications involving large sparse tables (e.g., the National Halothane Study in the U.S. [13]), and advances in computing that made practical the use of approaches such as Deming-Stephan iterative proportional fitting algorithm, and approaches for the Generalized Linear model (e.g., [53]). From the outset, the log-linear model literature placed heavy emphasis on the fitting of parsimonious models, e.g., by setting as many higher-order interaction terms equal to zero as possible. Another way to induce parsimony is by invoking forms of symmetry or exchangeability (see for example, Bishop, Fienberg, and Holland [10], Bhapkar and Darroch [8], Ten Have and Becker [58], and Fienberg, Johnson, and Junker [30]), which is one of the focuses of this paper. Without such parsimonious representations for large sparse tables, the likelihood gets maximized on the boundary of the parameter space and many of the cell estimates are zero (or minus infinity in the logarithmic scale).

Henri Caussinus's landmark 1966 paper [15], while focused largely on two-way tables and the models of quasi-independence and quasi-symmetry, fits well with those theoretical developments and stimulated multidimensional generalizations that appeared initially in Bishop, Fienberg, and Holland [10], Chapter 8. Caussinus' insights on the links between quasi-independence and quasi-symmetry led to the reformulation of quasi-symmetry as a log-linear model in Bishop, Fienberg, and Holland [10]. This in turn allowed the adaptation of iterative proportional fitting algorithm (IPF) as an alternative to the related algorithm for estimation under quasi-symmetry presented in Caussinus [15]. The next steps described in that chapter then followed naturally: (1) re-expressing IPF for quasi-symmetry in terms of the standard IPF algorithm for no 2nd-order interaction in a three-way table constructed from the original two-way table, and (2) various multivariate generalizations of quasi-symmetry. The former awaited the subsequent theoretical work of Meyer [52] for a firm foundation, and the latter remained somewhat of a hidden curiosity until its link to other statistical models was recognized in the 1980s and 1990s.

In the late 1970s, Otis Dudley Duncan, a distinguished sociologist, discovered the Rasch model from the educational testing literature and was applying it in innovative ways to survey analysis (e.g., see Duncan [26]). But he was stymied when it came to understanding the theoretical basis

for the approach he had intuitively adopted. Fortunately, work by Tjur [60] provided the key to his problem and, in giving it a proper contingency table representation in Fienberg and Meyer [31], we were able to tie it directly to one of the multivariate generalizations of quasi-symmetry! These same ideas surfaced again in the 1990s in work on models for heterogeneity in a multiple-recapture setting by Darroch, Fienberg, Glonek, and Junker [23] and Fienberg, Johnson and Junker [30], and by Agresti [2] and Coull and Agresti [19]. For related discussion and links to other log-linear and related models, see Cox and Wermuth [20] and the papers by Agresti [3] and Goodman [35] in this issue.

Despite the 35 years that have passed since the emergence of the log-linear model literature, the interest in succinct parametric models for large sparse contingency tables remains, and quasi-symmetry ideas remain one of the essential tools that are being refined and used. In this paper, we review three different approaches to the modeling of large sparse contingency tables, standard log-linear models, including those with quasi-symmetric interactions (Section 2), the related class of Rasch model extensions (Sections 3 and 4), and the class of Grade of Membership (GoM) models which are of special interest in many of the same settings as are the Rasch and quasi-symmetry approaches (Section 5).

## 2. Log-linear models and quasi-symmetry

Let  $\mathbf{n} = (n_1, n_2, \dots, n_t)$  be a vector of observed counts for  $t$  cells, structured in the form of a cross-classification for  $n = \sum_{c=1}^t n_c$  observations. Now let  $\mathbf{m} = E[\mathbf{n}] = (m_1, m_2, \dots, m_t)$  be the vector of expected values that are assumed to be functions of unknown parameters  $\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots, \theta_s)$ , where  $s < t$ .

Let  $\mathcal{M}$  denote the log-linear model specified by  $\mathbf{m} = \mathbf{m}(\boldsymbol{\theta})$ . When the  $t$  cells form a  $J$ -dimensional cross-classification corresponding to  $J$  categorical variables, then we can rewrite the most general version of model  $\mathbf{m} = \mathbf{m}(\boldsymbol{\theta})$  in its more recognizable saturated log-linear form as

$$\begin{aligned} \log m_{i_1 i_2 \dots i_J} = u &+ \sum_{j=1}^J u_{j(i_j)} + \sum_{j_1 \neq j_2} u_{j_1 j_2 (i_{j_1} i_{j_2})} \\ &+ \sum_{j_1 \neq j_2 \neq j_3} u_{j_1 j_2 j_3 (i_{j_1} i_{j_2} i_{j_3})} \\ &+ \dots + u_{j_1 j_2 \dots j_J (i_{j_1} i_{j_2} \dots i_{j_J})}, \end{aligned} \tag{1}$$

To make model (1) identifiable, we require each subscripted  $u$ -term to sum to zero over any subscript following Bishop, Fienberg, and Holland [10], or we set selected  $u$ -terms equal to zero as in GLIM or S-Plus. By setting some of the interaction terms in (1) equal to zero, we get a reduced or unsaturated log-linear model.

For the general log-linear model,  $\mathcal{M}$ , the minimal sufficient statistics (MSSs) are given by the projection of  $\mathbf{n}$  onto  $\mathcal{M}$ ,  $P_{\mathcal{M}}\mathbf{n}$  (e.g., see Haberman [36]). In the case of model (1), the MSSs consist of the marginal totals corresponding to the highest order terms in the model. For example if we set all 3-factor and higher-order interactions, then model (1) reduces to

$$\log m_{i_1 i_2 \dots i_J} = u + \sum_{j=1}^J u_{j(i_j)} + \sum_{j_1 \neq j_2} u_{j_1 j_2(i_{j_1} i_{j_2})}, \quad (2)$$

and the MSSs are the two-way marginal totals, corresponding to the two-factor terms in (2).

The quasi-symmetry strategy for reducing the number of parameters in (1) sets all of the higher-order terms of the same order equal to one another, e.g.,

$$\begin{aligned} u_{j_1 j_2 j_3(i_{j_1} i_{j_2} i_{j_3})} &= u_{123(\text{perm}[i_{j_1} i_{j_2} i_{j_3}])} \quad \text{for all } j_1 \neq j_2 \neq j_3, \\ &\vdots \\ u_{j_1 j_2 \dots j_J(i_{j_1} i_{j_2} \dots i_{j_J})} &= u_{12 \dots J(\text{perm}[i_{j_1} i_{j_2} \dots i_{j_J}])}. \end{aligned} \quad (3)$$

where  $\text{perm}[\cdot]$  denotes any permutation of the sequence of indices in the argument (e.g., see Bishop, Fienberg and Holland [10], Darroch [22]). As Ten Have and Becker [58] note, quasi-symmetry in this sense involves a form of conditional exchangeability of expected cell values which combines both class parameter invariance (equivalence of  $u$ -terms of similar order) and class parameter symmetry of any interaction involving all variables of a class. Clearly, we can also separate out the various components of this model and utilize them in a hierarchical log-linear fashion.

Because the model which combines (1) with (3) is log-linear, we can read off the MSSs directly. They are the two-way marginals which correspond to the the first-order interaction parameters in (1) which are not in quasi-symmetrization of (3), and the sums of counts corresponding to the symmetry or equivalence sets defined by (3). Had we also symmetrized the first-order interaction terms,  $u_{j_1 j_2(i_{j_1} i_{j_2})}$ , we would replace the two-way marginals by the corresponding sums of counts and include the one-way marginals in the MSSs.

We can also combine the two strategies of modeling by setting some higher-order terms equal to zero and assuming class parameter symmetry and/or invariance for others. Computing maximum likelihood estimates under any of these models is straightforward using any generalized linear model computer program, such as the ones in S-Plus or SAS.

### 3. Variations on the Rasch model

In education and the social sciences, the  $J$  categorical variables whose cross-classification leads to the table  $\mathbf{n} = (n_1, n_2, \dots, n_t)$  modelled in equation (1) often arises as coded subject responses to a set of items (questions, statements, or tasks) on survey forms, self-report inventories, and mental tests. Subject  $i$  responds to item  $j$ , for  $j = 1, 2, \dots, J$ , coded as  $X_{ij} = 1$  or 0, indicating the agreement (1) or disagreement (0) of subject  $i$  with proposition  $j$ , success (1) or failure (0) at performing task  $j$ , presence (1) or absence (0) of a symptom or feature  $j$ , etc. The table  $\mathbf{n}$  is then the table of counts  $n_{\mathbf{x}}$  of the unique binary response patterns  $\mathbf{x} = (x_1, x_2, \dots, x_J)$ . Since  $t = 2^J$  grows exponentially with  $J$ ,  $\mathbf{n}$  is typically large and sparse.

It is often sensible to separate subject (row) effects from item (column) effects in the matrix of responses  $\mathcal{X} = \{X_{ij} : i = 1, 2, \dots, n; j = 1, 2, \dots, J\}$ . One of the earliest models to do this was developed by Georg Rasch [55], and specifies row and column effects in an additive logistic model for  $\mathcal{X}$ :

$$P_j(\theta_i) \equiv P[X_{ij} = 1 | \theta_i, \beta_j] = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}, \quad (4)$$

where  $\theta_i$  represents subject  $i$ 's propensity to respond positively to any item relative to other subjects; and  $\beta_j$  represents the difficulty of responding positively to item  $j$ , relative to other items. As  $\theta_i$  increases, the probability of responding 1 to item  $j$  increases, and as the difficulty parameter  $\beta_j$  increases, the probability of responding 1 to item  $j$  decreases. The Rasch model, together with similar models using different forms for  $P_j(\theta_i)$  developed by Loevinger [48], Lord [49], and others, became the core of what is now item response theory (e.g., see Fischer and Molenaar [32] and van der Linden and Hambleton [62]). Clearly, the number of parameters  $\theta_i$  increase linearly with the sample size  $n$ ; the  $\theta_i$  are typically treated as latent variables in the model (conditioned or integrated away when estimation of the  $\beta_j$  is desired). In the psychometric literature it is common to talk about a single latent variable for ability but that assumes a relationship among the  $\theta_i$  such as that they come from a common distribution (see Section 3.1 below).

We further assume ‘‘local independence’’ among item responses, that is, we assume all  $X_{ij}$  are conditionally independent, given the  $\theta$ 's and  $\beta$ 's.

Letting  $x_{ij}$  denote the observed values of  $X_{ij}$ , we may write the likelihood for the Rasch model as

$$P[X_{ij} = x_{ij}, 1 \leq i \leq n, 1 \leq j \leq J \mid \theta_1, \theta_2, \dots, \theta_n; \beta_1, \beta_2, \dots, \beta_J] = \prod_{i=1}^n \prod_{j=1}^J P_j(\theta_i)^{x_{ij}} [1 - P_j(\theta_i)]^{1-x_{ij}}. \quad (5)$$

The Rasch model in equation (4) has the interesting invariance property that the odds ratio comparing two items  $j$  and  $k$  is

$$\frac{P_j(\theta_i)}{1 - P_j(\theta_i)} \cdot \frac{1 - P_k(\theta_i)}{P_k(\theta_i)} = \exp(\beta_k - \beta_j), \quad (6)$$

which is independent of  $\theta_i$ . Hence any two items may be compared, in principle, using any convenient sample of subjects regardless of their propensities to respond positively (in practice, of course, subjects still contribute at least “small sample bias” to the estimated  $\beta$ 's). Similarly, the odds ratio comparing two subjects is independent of the item difficulty parameters. These two invariance properties are instances of “specific objectivity,” a property Rasch considered important in defining a measurement model, and from which equation (4) can be derived (c.f. Fischer and Molenaar [32]).

The Rasch model is also closely related to the Bradley-Terry [12] model of paired comparisons (see Fienberg and Meyer [31]). Consider two items  $j$  and  $k$  for which  $\beta_j < \beta_k$ . For any subject  $i$  responding positively to only one of these two items, it follows from equation (4) that

$$P[X_{ij} = 1 \mid X_{ij} + X_{ik} = 1, \theta_i, \beta_j, \beta_k] = \frac{\exp(\beta_k - \beta_j)}{1 + \exp(\beta_k - \beta_j)}, \quad (7)$$

independent of  $\theta_i$ ; this is exactly the Bradley-Terry model, which says that the log odds of choosing item  $j$  from the pair  $j$  and  $k$  depends only on a ratio of positive parameters of the form:  $\phi_j/\phi_k$ . We can use the the lack of dependence of the right hand side of equation (7) on  $\theta_i$  to construct empirical tests of the Rasch model, since we can estimate the probability in equation (7) without assuming the Rasch model and this should be the same in any subpopulation of subjects in which the Rasch model applies.

### 3.1. The Rasch model and quasi-symmetry for the $2^J$ table

Two estimation methods are traditionally associated with the Rasch model, joint maximum likelihood (JML) and conditional maximum likelihood (CML). Work of Erling Andersen in the 1970's summarized in Andersen [4], shows that JML estimators for  $\beta_j$  obtained by maximizing equation

(5) jointly in the  $\theta$ 's and  $\beta$ 's, are inconsistent, that is, asymptotically biased as  $n$  grows and  $J$  remains fixed. On the other hand, the CML approach, conditioning on row totals in  $\mathcal{X}$  to eliminate the  $\theta_i$ 's, lead to consistent estimators of  $\beta_j$ . Estimates of both the  $\theta$ 's and the  $\beta$ 's in the joint likelihood in equation (5) can be made to be consistent if both  $n$  and  $J$  are allowed to grow at controlled rates (Haberman [37]).

A more widespread approach to solving the inconsistency problem with JML has been to assume the scalars  $\theta_i$  are independent random effects following a common (discrete or continuous) distribution  $dF(\theta|\eta)$  with hyperparameters  $\eta$ . Integrating over  $\theta_i$  for each subject yields the marginal likelihood

$$\prod_{i=1}^n \int \prod_{j=1}^J P_j(\theta_i)^{x_{ij}} [1 - P_j(\theta_i)]^{1-x_{ij}} dF(\theta_i|\eta). \quad (8)$$

Maximizing equation (8) with respect to  $\beta_j$ 's and  $\eta$  (using an E-M-like algorithm, for example, as in Bock and Aitken [11]) yields what are called maximum marginal likelihood (MML) estimates  $\hat{\beta}_j$  and  $\hat{\eta}$ . These MML estimates are also consistent (asymptotically unbiased). The method based on equation (8) can be interpreted as an empirical Bayes method; it also links the Rasch model directly with other latent variable approaches such as factor analysis, where the  $\theta_i$ 's are treated as unobserved random variables, or, equivalently, as missing data.

Equation (8) also provides an important link to log-linear models. Under i.i.d. sampling of subjects with observed response patterns  $\mathbf{x} = (x_1, x_2, \dots, x_J)$  and missing latent variables  $\theta$ , the integral in equation (8) gives the cell probabilities for the unique binary response patterns  $\mathbf{x} = (x_1, x_2, \dots, x_J)$ . Indeed, since equation (4) implies that  $\log P_j(\theta)/[1 - P_j(\theta)] = \theta - \beta_j$ , it follows that these cell probabilities are

$$\pi_{\mathbf{x}} = \int \exp \left\{ \sum_{j=1}^J x_j (\theta - \beta_j) \right\} \prod_{j=1}^J \frac{1}{1 + e^{\theta - \beta_j}} dF(\theta|\eta). \quad (9)$$

This readily simplifies to a log-linear model of the form

$$\log(\pi_{\mathbf{x}}) = \alpha + b_1 x_1 + \dots + b_J x_J + \gamma(x_+), \quad (10)$$

where

$$b_j = -\beta_j \quad \forall j, \quad x_+ = \sum_{j=1}^J x_j, \quad \text{and} \quad \gamma(s) = \log E [e^{s\theta} | \mathbf{x} = \mathbf{0}], \quad (11)$$



(for details, see Tjur [60], Cressie and Holland, [21]; Fienberg and Meyer [31], Holland [39], Darroch, Fienberg, Glonek and Junker [23]). The log-linear model in equation (10) is equivalent to the full quasi-symmetry model of equation (3) for the  $2^J$  table, subject to moment-like order constraints on the  $\gamma(s)$  implied by equation (11) and detailed in Cressie and Holland [21]. Lindsay, Clogg and Grego [47] explore in detail connections between Rasch models and log-linear quasi-symmetry models, and consequences for the identifiability of the distribution  $dF(\theta|\eta)$  in a semiparametric formulation of equation (8).

### 3.2. The Rasch model and quasi-symmetry for the $K^J$ table

The Rasch model generalizes naturally to multinomial logit models depending on  $\beta$  and  $\theta$ , when there are more than two levels of response per measure and the categories are ordered. In particular, suppose each  $X_{ij}$  takes values in the discrete set  $\{0, 1, \dots, K-1\}$ . The polytomous Rasch model, or “partial credit model” (Masters [51]) specifies linear adjacent category logits as

$$\text{logit } P[X_{ij} = k+1 | X_{ij} \in \{k, k+1\}, \theta_i, \delta_{jk}] = \theta_i - \delta_{jk}, \quad (12)$$

or, equivalently, in a somewhat overparametrized form,

$$P_{jk}(\theta_i) \equiv P[X_{ij} = k | \theta_i, \delta_{j0}, \delta_{j1}, \dots, \delta_{j(K-1)}] = \frac{\exp\left\{\sum_{\ell=1}^k (\theta_i - \delta_{j\ell})\right\}}{\sum_{h=0}^{K-1} \exp\left\{\sum_{\ell=1}^h (\theta_i - \delta_{j\ell})\right\}}, \quad (13)$$

where sums with indices from 1 to 0 are defined to be zero. Under this model, the response categories are stochastically ordered by  $\theta$ :  $P[X_{ij} > c | \theta_i]$  is a non-decreasing function of  $\theta_i$  for each  $c$ . Agresti [1] considers related models, e.g., when  $\delta_{jk}$  in equation (12) is replaced by two terms additive in  $j$  and  $k$ .

Again, if we assume local independence and integrate out the latent variables  $\theta_i$  with respect to a common distribution  $dF(\theta|\eta)$ , we can write the cell probability for the pattern of polytomous responses  $\mathbf{x} = (x_1, x_2, \dots, x_J)$  in the  $K^J$  table as the integrated product multinomial

$$\pi_{\mathbf{x}} = \int \prod_{j=1}^J \prod_{k=0}^{K-1} P_{jk}(\theta)^{y_{jk}} dF(\theta|\eta),$$

where

$$y_{jk} = \begin{cases} 1, & \text{if } x_j = k, \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

Using the definition in equation (13) we see that

$$\begin{aligned} \pi_{\mathbf{x}} &= \int \prod_{j=1}^J \frac{\exp \sum_{k=0}^{K-1} \left[ y_{jk} \sum_{\ell=1}^k (\theta_i - \delta_{j\ell}) \right]}{\sum_{h=0}^{K-1} \exp \sum_{\ell=1}^h (\theta_i - \delta_{j\ell})} dF(\theta|\eta) \\ &= \int \frac{\exp \left[ x_+ \theta_i - \sum_{j=1}^J \sum_{\ell=1}^{x_j} \delta_{j\ell} \right]}{\sum_{h=0}^{K-1} \exp \sum_{\ell=1}^h (\theta_i - \delta_{j\ell})} dF(\theta|\eta), \end{aligned} \quad (15)$$

where, again,  $x_+ = \sum_{j=1}^J x_j$ . A simplification analogous to that leading from equation (9) to equation (10) produces the log-linear model

$$\log p(\mathbf{y}) = \alpha + \sum_{j=1}^J \sum_{l=0}^{K-1} y_{jk} b_{jk} + \gamma(x_+), \quad (16)$$

where  $b_{jk} = -\sum_{\ell=1}^k \delta_{j\ell}$  and  $\gamma(s) = E[e^{s\theta} | \mathbf{x} = \mathbf{0}]$ . Again, this is a quasi-symmetry representation for the  $K^J$  table, subject to moment constraints on  $\gamma(s)$ . Unlike the  $2^J$  case, however, this is not the most general  $K^J$  quasi-symmetry model expressed in equation (3), since any two interaction terms with the same sum  $x_+$  are constrained to be equal in equation (16). For ordinal tables with “equally spaced scores,” the implied log-linear model has a simpler main effects form (c.f., Agresti’s [3] “ordinal quasi-symmetry” model).

Variants of this approach have been outlined for social survey data by Sobel, Hout, and Duncan, [57], Thissen and Mooney [59], and Goodman [34]. Alternative forms of the model in equation (15) and equation (16) may be obtained by changing the form of the linear logits in equation (12) (e.g., see Huguenard, Lerch, Junker, Patz and Kass [41]).

#### 4. Semi-exchangeability variations on the Rasch model

Due to the stochastic ordering property, that  $P[X_{ij} > c | \theta_i]$  is nondecreasing in  $\theta_i$  for all  $i, j$  and  $c$ , the dichotomous and polytomous Rasch models exhibit strong positive dependence in the structure they induce for  $2^J$  and  $K^J$  tables. Holland and Rosenbaum [40] explore this dependence in detail, and Junker and Ellis [44] use related properties to characterize a broad class of item response models. However, in many applications, ranging from multiple-recapture census enumeration to epidemiology, biostatistics and even cognitive testing, this strong positive dependence structure is too restrictive for practical modeling.

One way to weaken this structure has been to generalize the quasi-symmetry model, especially in the case of dichotomous response data, for example by adding two-way interactions

$$\begin{aligned} \log(\pi_{\mathbf{x}}) &= \alpha + b_1 x_1 + \cdots + b_J x_J + \gamma(x_+) + \sum_{j_1 \neq j_2} \sum b_{j_1 j_2} x_{j_1} x_{j_2} \\ &\quad + \sum_{\ell} \gamma'(x_{\ell}, x_+). \end{aligned} \tag{17}$$

Kelderman [46] considered these so-called “generalized log-linear Rasch models” to develop hierarchically nested alternatives to the null hypothesis that the data follow the log-linear Rasch model of equation (10). Cormack [18] provided an independent, alternative development of these ideas. These ideas have also proven useful in extending the log-linear Rasch model to accommodate dependence in the table of counts  $n_{\mathbf{x}}$  that is Rasch-like but more general than the “exchangeable higher moments” structure of the Rasch model (see Darroch and McCloud, [24], Carriquiry and Fienberg [14], Biggeri *et al.* [9] and Bartolucci and Forcina [7]). In particular, the terms  $b_{j_1 j_2} x_{j_1} x_{j_2}$  allow for negative dependence between some pairs of items that are excluded by the basic Rasch model’s assertion of equal positive associations among all the items. It is also worth noting that models of the form equation (17) may contain interactions of all orders, but they may be severely constrained both by the form of the  $\gamma'(\cdot)$  terms in equation (17) and by moment inequality constraints of the type discussed after equation (11) above.

To relate log-linear models like equation (17) to the Rasch model of equation (4), we begin by rewriting the likelihood for the conditional  $2^J$  table of counts  $n_{\mathbf{x}|\theta}$ , given  $\theta$ , as

$$\begin{aligned} \pi_{\mathbf{x}|\theta} &= \prod_{j=1}^J P_j(\theta)^{x_j} [1 - P_j(\theta)]^{1-x_j} = \prod_{j=1}^J \left[ \frac{P_j(\theta)}{1 - P_j(\theta)} \right]^{x_j} \prod_{j=1}^J [1 - P_j(\theta)] \\ &= \exp\left\{ \sum_{j=1}^J \lambda_j(\theta) x_j \right\} \prod_{j=1}^J [1 - P_j(\theta)], \end{aligned}$$

where  $\lambda_j(\theta) = \log\{P_j(\theta)/[1 - P_j(\theta)]\}$ . Hence

$$\log \pi_{\mathbf{x}|\theta} = \alpha(\theta) + \sum_{j=1}^J \lambda_j(\theta) x_j. \tag{18}$$

where  $\alpha(\theta) = \log \prod_{j=1}^J [1 - P_j(\theta)]$ , and for the Rasch model in particular,  $\lambda_j(\theta) = \theta + b_j$  ( $b_j = -\beta_j$  in equation (4)).

Darroch, *et al.* [23] note that interdependencies in the observed  $2^J$  table may either be the result of collapsing over  $\theta$  (a version of Simpson's paradox, see Holland and Rosenbaum [40] or Kadane, Meyer and Tukey [45]) or they may be a consequence of the items being truly interdependent even when the data are disaggregated to the person or object level. Examples can be found in diverse applications, e.g., two web search engines may draw from the same pool of web pages (perhaps because of a common indexing strategy), so that a web page may be more likely to show up in one, given than it is in the other, quite apart from the visibility of the page to search engines in general. In multiple-recapture census work based on administrative lists, lists that by their nature penetrate nearly disjoint subpopulations induce a tendency toward negative dependence in the marginal distribution  $\pi_{\mathbf{x}}$  (Asher and Fienberg [6]).

To illustrate models for item-by-item dependencies that are not artifacts of aggregating over  $\theta$ , we consider adding to equation (18) the two-way interactions in the conditional (fixed  $\theta$ ) model:

$$\log \pi_{\mathbf{x}|\theta} = \alpha(\theta) + \sum_j \lambda_j(\theta)x_j + \sum_{j_1 \neq j_2} \lambda_{j_1 j_2}(\theta)x_{j_1}x_{j_2}, \quad (19)$$

where  $\alpha(\theta)$  is simply the usual log-linear model normalizing constant (sum of model terms over all binary patterns  $x_1, x_2, \dots, x_J$ ). We assume, as before, that  $\lambda_j(\theta) = \theta + b_j$ , and now we also assume that  $\lambda_{j_1 j_2}(\theta) = \theta + b_{j_1 j_2}$ . This leads to the form

$$\log \pi_{\mathbf{x}|\theta} = \alpha(\theta) + \sum_j b_j x_j + \theta x_+ + \sum_{j_1 \neq j_2} b_{j_1 j_2} x_{j_1} x_{j_2} + \theta \sum_{j_1 \neq j_2} x_{j_1} x_{j_2}, \quad (20)$$

(c.f. Jannarone [42] and Jannarone, Yu and Laughlin [43]).

Exponentiating, integrating with respect to the distribution of the random catchability effects,  $\theta$ , and taking the logarithm again, Fienberg, Johnson and Junker [30] obtained the log-linear model

$$\log(\pi_{\mathbf{x}}) = \alpha + b_1 x_1 + \dots + b_J x_J + \gamma(x_+) + \sum_{j_1 \neq j_2} b_{j_1 j_2} x_{j_1} x_{j_2}, \quad (21)$$

a submodel of equation (17) [subject to moment constraints analogous to equation (11)].

A different development generates terms like  $\gamma'(x_\ell, x_+)$  in equation (17) and interprets them in terms of heterogeneity across subjects that cannot be described by the simple Rasch model. Let us begin again with the basic likelihood in equation (18), and suppose now that  $\theta$  is multidimensional, i.e.,

$\theta = (\theta_1, \theta_2, \dots, \theta_q)$ . Moreover, suppose that different items depend on different  $\theta$ 's through the Rasch model. For example, suppose that  $\theta = (\theta_1, \theta_2)$  and we can partition the items into  $I$  items that depend only on  $\theta_1$  and  $J - I$  items that depend only on  $\theta_2$ . Then, after permuting item indices, the likelihood given  $\theta$  becomes

$$\pi_{\mathbf{x}|\theta_1\theta_2} = \prod_{j=1}^I P_j(\theta_1)^{x_j} [1 - P_j(\theta_1)]^{1-x_j} \prod_{j=I+1}^J P_j(\theta_2)^{x_j} [1 - P_j(\theta_2)]^{1-x_j}. \quad (22)$$

If, as would usually seem reasonable, the density of  $\theta = (\theta_1, \theta_2)$  does not factor, then a derivation which is similar to that leading from equation (20) to equation (21) above now leads us to

$$\log(\pi_{\mathbf{x}}) = \alpha + b_1 x_1 + \dots + b_J x_J + \gamma(x_+^{(1)}, x_+^{(2)}), \quad (23)$$

where  $x_+^{(1)} = \sum_{j=1}^I x_j$ , and  $x_+^{(2)} = \sum_{j=I+1}^J x_j$ . Equation (23) is a partial quasi-symmetry model in which two sets of items participate in separate quasi-symmetry relationships. This sort of structure was employed by Darroch, *et al.* [23] in triple-system enumeration, to model the differing visibility (modelled by  $\theta_1$  and  $\theta_2$ ) of persons in administrative lists, in U.S. Census lists and in a post-enumeration survey also conducted by the U.S. Census Bureau.

Clearly, the same construction can be used to add terms  $\gamma(k_1, k_+ - k_1)$  to the basic Rasch quasi-symmetry model in equation (10). This is equivalent to adding the term  $\gamma(k_1, k_+)$  to the model, since there is a 1-1 correspondence between levels of the pair  $(k_1, k_+ - k_1)$  and levels of the pair  $(k_1, k_+)$  (due to the constraint  $k_+ = k_1 + \dots + k_J$ ). Of course, we may also add item-by-item interactions to this model as in equation (19), in the end obtaining the full generality of equation (17). In addition to providing links between the Rasch model and hierarchically structured log-linear models, this approach is useful in relaxing the strong positive association constraints of the basic Rasch model and exhibiting some meaningful stochastic ordering properties, while retaining an interpretation in terms of latent variable models.

## 5. The Grade of Membership model

Log-linear models, as we introduced them in Section 2, focus on dependencies among variables and they assume that the units (subjects) are homogeneous. Latent structure models originated from searching for a phenomenon that would describe substantive differences among subjects in the sample. In particular, latent class models (Goodman [33]) assume that a

heterogeneous population is composed of a few subpopulations that are homogeneous in their responses with respect to the problem of interest, and that, given the latent class, the responses to the items are considered independent (a form of the “local independence” assumption). When this assumption is reasonable, by representing the likelihood as a mixture of latent classes, we can obtain a good fit to multi-way contingency table data that would otherwise require a number of not so easy to explain higher-order interactions in a log-linear model.

Rasch models involve parameters for two sets of objects, e.g. for items and subjects. The property of “specific objectivity” (Rost [56]) assures that the estimation of parameters for one set of objects is independent of the other set. Rasch showed that when only two response categories are considered, the dichotomous Rasch model is the only parametric model that fulfills the requirement of specific objectivity (see Andersen and Olsen [5]). By assuming random subject effects for the Rasch model, we can think in terms of disaggregating the  $2^J$  contingency table into independence tables by the value of subject parameter. It is common to assume a unimodal distribution, such as the normal, for subject parameters, and then integrate them out in order to estimate the item parameters. Unimodal distributions are consistent with the assumption, plausible in many cases, that a population can be represented on a latent continuum in such a way that a few individuals would have very high/low values of the subject parameters and the rest of the population would have values concentrated near the average. But such unimodality does not hold universally, and we often need alternatives to the Rasch models for modeling purposes.

We often observe another kind of population heterogeneity in medical classification problems. Examining whether each of  $J$  symptoms is present or absent for patients, clinicians can classify a good proportion of a population as either being healthy or having a disease, but the diagnosis for the rest of the population can be less certain. The uncertainty in the diagnosis can be incorporated into a subject parameter where the value of the parameter would indicate how close each subject is to having a disease. This idea gave rise to formulation of the Grade of Membership (GoM) model proposed originally in the 1970s by Max Woodbury and described in detail by Manton, Woodbury, and Tolley [50]. The GoM model regards the property of convexity in the response probabilities in a population as worthy of modeling. Intuitively, convexity is a two-fold phenomenon: first, extremal individual cases or extreme profiles must exist, at least theoretically, and, second, all other individual cases must be convex combinations of extremal cases. Extreme profiles are defined by the conditional probabilities of response for “certain diagnosis” cases. These are the item parameters as well.

Subject level parameters are components of the  $K$ -dimensional vector of membership scores. Each component represents how close an individual is to the respective extreme profile. In the GoM model we again invoke the local independence assumption: given the membership scores,  $J$  responses are considered independent. Therefore, the model disaggregates observed  $2^J$  table according to the vector of membership scores.

As before, we consider data arising in the form of a vector of  $J$  dichotomous manifest variables,  $\mathbf{X} = (X_1, X_2, \dots, X_J)$ , and we let  $\mathbf{x} = (x_1, x_2, \dots, x_J)$  denote a binary response pattern, where  $x_j = 0$  or  $1$  is a response to  $j$ th manifest variable, for  $j = 1, 2, \dots, J$ . We denote by  $\mathbf{G} = (G_1, G_2, \dots, G_K)$  a vector of membership scores with distribution  $H(g)$  such that  $G_k \geq 0$ , for  $k = 1, 2, \dots, K$ , and  $\sum_{k=1}^K G_k = 1$ . Denote the conditional probability of positive response for each extreme profile by

$$\lambda_{kj} = P(X_j = 1 | G_k = 1). \quad (24)$$

The probability of negative response,  $X_j = 0$ , is one minus the probability of positive response.

The marginal probability of response for the manifest variable  $X_j$ , given the membership scores, is a convex combination of the probabilities that correspond to the extremal cases:

$$\begin{aligned} P(X_j = x_j | \mathbf{G} = \mathbf{g}) &= \sum_{k=1}^K g_k \cdot P(X_j = x_j | G_k = 1) \\ &= \sum_{k=1}^K g_k \cdot \lambda_{kj}^{x_j} (1 - \lambda_{kj})^{1-x_j}, \end{aligned} \quad (25)$$

where  $g = (g_1, g_2, \dots, g_K) \in (0, 1)^K$ .

The GoM local independence assumption states that manifest variables are conditionally independent, given the latent variables. Thus, for the conditional probability of observing a response pattern  $x$  is

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} | \mathbf{G} = \mathbf{g}) &= \prod_{j=1}^J P(X_j = x_j | \mathbf{G} = \mathbf{g}) \\ &= \prod_{j=1}^J \sum_{k=1}^K g_k \cdot \lambda_{kj}^{x_j} (1 - \lambda_{kj})^{1-x_j}. \end{aligned} \quad (26)$$

This particular form of conditional probability function results in a property of the GoM model that distinguishes it from other continuous latent

variable models. Thus, for the Rasch model, Ramsay [54] has pointed out that the value of the ability parameter in the Rasch model is only indirectly related to its position within the manifold that corresponds to the model in the probability space. In contrast, for the GoM model, Erosheva [27] has shown that the values of the membership scores represent the distances along the model manifold from corresponding extreme profiles. Thus, the membership scores have the property of being intrinsic to the model manifold, which provides a natural characterization of the latent continuum.

Integrating the conditional probability in equation (26) with respect to the distribution of the membership scores, we obtain the marginal distribution of the data for the GoM model

$$P(\mathbf{X} = \mathbf{x}) = \int \prod_{j=1}^J \sum_{k=1}^K g_k \cdot \lambda_{kj}^{x_j} (1 - \lambda_{kj})^{1-x_j} dH(g), \quad (27)$$

where the integration is over  $(0, 1)^K$ .

Following a suggestion of Haberman [38], Erosheva [28] shows that we can derive the same distribution by treating the data as arising from a constrained latent class model with  $J$  latent variables. Let  $Z_1, Z_2, \dots, Z_J$  be multinomial exchangeable latent variables with the values  $z_j \in \{1, 2, \dots, K\}$ . Let the probability of observing a latent vector  $z = (z_1, z_2, \dots, z_J)$  be the expected value of the  $J$ -fold product of corresponding membership scores:

$$\pi_z = P(z_1, z_2, \dots, z_J) = E_H \left( \prod_{j=1}^J g_{z_j} \right). \quad (28)$$

By assuming that the conditional distribution of the  $j$ th manifest variable depends only on the  $j$ th component of the latent vector and setting

$$P(X_j = x_j | z_j = k) = \lambda_{kj}^{x_j} (1 - \lambda_{kj})^{1-x_j}, \quad (29)$$

Erosheva [28] has shown that the probability to observe response pattern  $x$  under this constrained latent class model coincides with the corresponding probability under the GoM model.

We can consider the classical notion of item independence in the context of the GoM model. Assume that items  $j_1$  and  $j_2$  are independent if they are independent in the  $2^J$  contingency table aggregated over the distribution of individual parameters. In this case, the observed  $2^J$  table would show independence or near independence of  $j_1$  and  $j_2$ . From the geometric representation of the GoM model, it can be seen that two items are independent only if the model manifold is a straight line on the surface of independence



in the subspace of the parameters of the corresponding  $2 \times 2$  table (Erosheva [27]). If the distribution of the GoM scores is not degenerate, this can happen only when one of the items, say  $j_1$ , has the same probability of response for all extreme profiles. Since the items in the GoM model are fully characterized by their extreme profile probabilities, this implies that item  $j_1$  is in fact independent from other  $J - 2$  items as well. In other words, item  $j_1$  is irrelevant for explaining population heterogeneity in the GoM model. This result is similar to the one for IRT models: item responses can be independent in the aggregated  $2^J$  contingency table only if each subject in the population has the same ability; otherwise, a positive correlation among items is to be expected. While the Rasch model can be used to model positive item dependencies, the GoM model can be employed to model more general item interactions in  $2^J$  table.

By using the latent class representation of the GoM model described above, we can give the GoM model a log-linear random effects form. Consider the  $(K^J \times 2^J)$ -dimensional table. Let  $m_{\mathbf{z}x_1x_2\dots x_J}$  be the expected count in the cell  $(\mathbf{z}, x_1, x_2, \dots, x_J)$ , where  $\mathbf{z} = (z_1, z_2, \dots, z_J)$ . Because the model assumes that the items are conditionally independent, given the value of the latent vector  $\mathbf{z}$ , the log-linear model will only have the main item effects and the interaction effects between each of the manifest variables and the latent vector-valued variable  $\mathbf{z}$ , i.e.,

$$\log m_{\mathbf{z}x_1x_2\dots x_J} = u + \sum_{j=1}^J \left( u_{x_j}^j + u_{\mathbf{z}x_j}^{\mathbf{z}j} \right). \quad (30)$$

The log-linear parameters are related to conditional probabilities in the GoM model as

$$P(x_j|\mathbf{z}) = \frac{\exp(u_{x_j}^j + u_{\mathbf{z}x_j}^{\mathbf{z}j})}{\sum_{x_j \in \{0,1\}} \exp(u_{x_j}^j + u_{\mathbf{z}x_j}^{\mathbf{z}j})}, \quad (31)$$

or, equivalently, as

$$\log \left[ \frac{P(x_j = 1|\mathbf{z})}{P(x_j = 0|\mathbf{z})} \right] = u_1^j + u_{\mathbf{z}1}^{\mathbf{z}j}. \quad (32)$$

By using the conditional probability from the GoM model and imposing the usual zero-sum constraints, we can write the log-linear main effect and interaction parameters as

$$u_1^j = \frac{1}{K} \sum_{k=1}^K \log \left[ \frac{\lambda_{kj}}{1 - \lambda_{kj}} \right], \quad (33)$$

$$u_{\mathbf{z}1}^{\mathbf{z}j} = \log \left[ \frac{\lambda_{z_j j}}{1 - \lambda_{z_j j}} \right] - \frac{1}{K} \sum_{k=1}^K \log \left[ \frac{\lambda_{kj}}{1 - \lambda_{kj}} \right]. \quad (34)$$

This random effects log-linear formulation has an underlying partial-symmetry structure that results from exchangeability in the components of the latent vectors: if  $\mathbf{z}$  and  $\mathbf{z}^*$  are such that  $z_j = z_j^*$ , then  $u_{\mathbf{z}1}^{z_j} = u_{\mathbf{z}^*1}^{z_j^*}$ . This partial-symmetry structure, however, does not resemble quasi-symmetry as proposed originally by Caussinus.

There has long been a fascination in the statistical literature about the development of additive as opposed to log-additive models for multi-dimensional contingency tables (e.g., see Darroch and Speed [25]). Intuitively, it is clear that additive models on the probability scale can not be additive on the log probability scale. There might be exceptions to these rule, e.g., if the  $g_k$  parameters of the GoM model were independent, then we can rewrite equation (30) as a constant plus  $J$  terms, each depending on  $\lambda_{jk}$  and the expected values of the membership scores. It is not clear whether there might exist an alternative parameterization that would provide a log-linear form for the GoM model.

Finally, we note the similarities of the GoM model and its representations to the latent budget model of van der Heijden, Mooijaart and de Leeuw [61], which is also described in terms of extreme profiles (“latent budgets”), and a simultaneous latent class model proposed by Clogg and Goodman [16] [17].

## 6. Summary

Henri Caussinus introduced the notion of quasi-symmetry in the context of two-way contingency tables in 1965. The subsequent generalizations of this quasi-symmetry model have proved to be especially useful in the modeling of multi-way contingency tables. In particular, they represent an interesting point of departure for alternatives to standard log-linear models for large sparse contingency tables. Such models offer a way to simplify parameter structures while at the same time allowing for interactions in the models.

In this paper, we have described some alternative models for large sparse tables that are based on latent structures, and we have described some of their relationships to quasi-symmetry and log-linear structures. These latent variable models scale to high dimensional settings and retain the simplicity of interpretation associated with the traditional local independence assumption.

Cox and Wermuth [20] suggest a somewhat different strategy for developing models for multivariate binary tables based on dichotomized Gaussian models. Their approach also uses a latent representation and essentially

replaces the logistic model of equation (4) by a multivariate Gaussian structure. They also consider a quadratic binary exponential model which is in essence a log-linear model without 2nd- and higher-order effects.

This area of categorical data analysis remains a fertile one for the development of new methodology.

## Bibliography

- [1] AGRESTI (A.). — Computing conditional maximum likelihood estimates for generalized Rasch models using simple loglinear models with diagonal parameters, *Scandinavian Journal of Statistics*, 20 (1993), 63–71.
- [2] AGRESTI (A.). — Simple capture-recapture models permitting unequal catchability and variable sampling effort, *Biometrics*, 50 (1994), 494–500.
- [3] AGRESTI (A.). — Links between binary and multi-category logit item response theory models and quasi-symmetric loglinear models, *Annales de la Faculté des Sciences de l'Université de Toulouse Mathématiques*, 11 (2002), This issue.
- [4] ANDERSEN (E. B.). — *Discrete Statistical Models with Social Science Applications*, North-Holland, Amsterdam, 1980.
- [5] ANDERSEN (E. B.), OLSEN (L. W.). — The Life of Georg Rasch as a Mathematician and as a Statistician, in Anne Boomsma, Marijtje A. J. van Duijn and Tom A. B. Snijders, editors, *Essays on Item Response Theory*, volume 157 of *Lecture Notes in Statistics*, pages 3-24, Springer-Verlag, New York, 2001.
- [6] ASHER (J.), FIENBERG (S. E.). — Statistical variations on an administrative records census, in *Proceedings of the Section on Social Statistics*, Alexandria, VA, 2001, American Statistical Association.
- [7] BARTOLUCCI (F.), FORCINA (A.). — The analysis of capture-recapture data with a Rasch-type model allowing for conditional dependence and multidimensionality, *Biometrics*, 57, (2001), 714–719.
- [8] BHAPKAR (V. P.), DARROCH (J. N.). — Marginal symmetry and quasi symmetry of general order, *Journal of Multivariate Analysis*, 34 (1990), 173–184.
- [9] BIGGERI (A.), STANGHELLINI (E.), MERLETTI (F.) and MARCHI (M.). — Latent class models for varying catchability and correlation among sources in capture-recapture estimation of the size of a human population, *Statistica Applicata*, 11 (1999), 563–576.
- [10] BISHOP (Y. M. M.), FIENBERG (S. E.) and HOLLAND (P. W.). — *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, MA, 1975.
- [11] BOCK (R. D.), AITKIN (M.). — Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm, *Psychometrika*, 46 (1981), 443–459.
- [12] BRADLEY (R. A.), TERRY (M. E.). — Rank analysis of incomplete block designs I. The method of paired comparisons, *Biometrika* (1952), 324–345.
- [13] BUNKER (J. P.), FORREST JR. (W. H.), MOSTELLER (F.) and VANDAM (L. D.), editors. — *The National Halothane Study*, Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1969. Report of the Subcommittee of the National Halothane Study, of the Committee on Anesthesia, Division of Medical sciences, National Academy of Sciences–National Research Council.

- [14] CARRIQUIRY (A.), FIENBERG (S. E.). — Rasch models, in *Encyclopedia of Biostatistics*, volume 5, 3724–3730, Wiley, New York, 1998.
- [15] CAUSSINUS (H.). — Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de l'Université de Toulouse*, 29 (année 1965), (1966), 77–183.
- [16] CLOGG (C. C.), GOODMAN (L. A.). — Latent structure analysis of a set of multidimensional contingency tables, *Journal of the American Statistical Association*, 79 (1984), 762–771.
- [17] CLOGG (C. C.), GOODMAN (L. A.). — Simultaneous latent structure analysis in several groups, *Sociological Methodology* (1985), 81–110.
- [18] CORMACK (R. M.). — Log-linear models for capture-recapture, *Biometrics*, 45 (1989), 395–413.
- [19] COULL (B. A.), AGRESTI (A.). — The Use of mixed logit models to reflect heterogeneity in capture-recapture studies, *Biometrics*, 55 (1999), 294–301.
- [20] COX (D. R.), WERMUTH (N.). — On some models for multivariate binary variables parallel in complexity with the multivariate gaussian distribution, *Biometrika*, 89 (2002), 462–469.
- [21] CRESSIE (N.), HOLLAND (P. W.). — Characterizing the manifest probabilities of latent trait models, *Psychometrika*, 48 (1983), 129–141.
- [22] DARROCH (J. N.). — Quasi-symmetry, in Samuel Kotz, Norman L. Johnson and Campbell B. Read, editors, *Encyclopedia of Statistical Sciences*, volume 7, 469–473, Wiley, New York, 1986.
- [23] DARROCH (J. N.), FIENBERG (S. E.), GLONEK (G. F.V.) and JUNKER (B. W.). — A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability, *Journal of the American Statistical Association*, 88 (1993), 1137–1148.
- [24] DARROCH (J. N.), MC CLOUD (P. I.). — Separating two sources of dependence in repeated influenza outbreaks, *Biometrika*, 77 (1990), 237–243.
- [25] DARROCH (J. N.), SPEED (T. P.). — Additive and multiplicative models and interactions, *Annals of Statistics*, 11 (1983), 724–738.
- [26] DUNCAN (O. D.). — Rasch Measurement: Further examples and discussion, in Charles F. Turner and Elizabeth Martin, editors, *Surveying Subjective Phenomena*, volume 2, chapter 12, 367–403, Russell Sage Foundation, New York, 1984.
- [27] EROSHEVA (E. A.). — Population Heterogeneity Representations of Log-Linear, Latent Class, Rasch, and Grade of Membership Models, Technical Report 731, Department of Statistics, Carnegie Mellon University, 2001.
- [28] EROSHEVA (E. A.). — *Grade of Membership and latent structure models with application to disability survey data*, PhD thesis, Carnegie Mellon University, Department of Statistics, 2002.
- [29] FIENBERG (S. E.). — Contingency tables and log-linear models: basic results and new developments, *Journal of the American Statistical Association*, 95 (2000), 643–647.
- [30] FIENBERG (S. E.), JOHNSON (M. S.), JUNKER (B. W.). — Classical multilevel and bayesian approaches to population size estimation using multiple lists, *Journal of the Royal Statistical Society, Series A*, 162 (1999), 383–405.
- [31] FIENBERG (S. E.), MEYER (M. M.). — Loglinear models and categorical data analysis with psychometric and econometric applications, *Journal of Econometrics*, 22 (1983), 191–214.
- [32] FISCHER (G. H.), MOLENAAR (I. W.), editors. — *Rasch Models: Foundations, Recent Developments, and Applications*, Springer-Verlag, New York, 1995.

- [33] GOODMAN (L. A.). — Exploratory latent structure analysis using both identifiable and unidentifiable models, *Biometrika*, 61 (1974), 215–231.
- [34] GOODMAN (L. A.). — Total-score models and Rasch-type models for the analysis of a multidimensional contingency table, or a set of multidimensional contingency tables, with specified and/or unspecified order for response categories, *Sociological Methodology*, 20 (1990), 249–294.
- [35] GOODMAN (L. A.). — Contributions to the statistical analysis of contingency tables: Notes on quasi-symmetry, quasi-independence, log-linear models, log-bilinear models, and correspondence analysis models, *Annales de la Faculté des Sciences de l'Université de Toulouse, Mathématiques*, 2002, This issue.
- [36] HABERMAN (S. J.). — *The Analysis of Frequency Data*, University of Chicago Press, Chicago, IL, 1974.
- [37] HABERMAN (S. J.). — Maximum likelihood estimates in exponential response models, *Annals of Statistics*, 5 (1977), 815–841.
- [38] HABERMAN (S. J.). — Review of “Statistical Applications Using Fuzzy Sets,” by Kenneth G. Manton, Max A. Woodbury, and H. Dennis Tolley, *Journal of the American Statistical Association*, 90 (1995), 1131–1133.
- [39] HOLLAND (P. W.). — The Dutch identity-A new tool for the study of item response models, *Psychometrika*, 55 (1990), 5–18.
- [40] HOLLAND (P. W.), ROSENBAUM (P. R.). — Conditional association and unidimensionality in monotone latent variable models, *Annals of Statistics*, 14 (1986), 1523–1543.
- [41] HUGENARD (B. R.), LERCH (F. J.), JUNKER (B. W.), PATZ (R. J.) and KASS (R. E.). — Working memory failure in phone-based interaction, *ACM Transactions on Computer-Human Interaction*, 4 (1997), 67–102.
- [42] JANNARONE (R. J.). — Conjunctive item response theory kernels, *Psychometrika*, 51 (1986), 357–373.
- [43] JANNARONE (R. J.), YU (K. F.) and LAUGHLIN (J. E.). — Easy Bayes estimation for Rasch-type models, *Psychometrika*, 55 (1990), 449–460.
- [44] JUNKER (B. W.), ELLIS (J. L.). — A characterization of monotone unidimensional latent variable models, *Annals of Statistics*, 25 (1997), 1327–1343.
- [45] KADANE (J. B.), MEYER (M. M.), TUKEY (J. W.). — Yule’s association paradox and ignored stratum heterogeneity in capture-recapture studies, *Journal of the American Statistical Association*, 94 (1999), 855–859.
- [46] KELDERMAN (H.). — Loglinear Rasch model tests, *Psychometrika*, 49 (1984), 223–245.
- [47] LINDSAY (B.), CLOGG (C.), GREGO (J.). — Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis, *Journal of the American Statistical Association*, 86 (1991), 96–107.
- [48] LOEVINGER (J.). — The technique of homogeneous tests compared with some aspects of “scale analysis” and factor analysis, *Psychological Bulletin*, 45 (1948), 507–530.
- [49] LORD (F. M.). — A Theory of Test Scores. *Psychometric Monographs*, 7 (1952).
- [50] MANTON (K. G.), WOODBURY (M. A.), TOLLEY (H. D.). — *Statistical Applications Using Fuzzy Sets*, Wiley, New York, 1994.
- [51] MASTERS (G. N.). — A Rasch model for partial credit scoring, *Psychometrika*, 47 (1982), 149–174.
- [52] MEYER (M. M.). — Transforming contingency tables, *Annals of Statistics*, 10 (1982), 1172–1181.

## Alternative statistical models

- [53] NELDER (J. A.), WEDDERBURN (R. W. M.). — Generalized linear models, *Journal of the Royal Statistical Society, Series A*, 135 (1972), 370–384.
- [54] RAMSAY (J. O.). — A geometrical approach to item response theory, *Behavior-metrika*, 23 (1996), 3–17.
- [55] RASCH (G.). — *Probabilistic Models for Some Intelligence and Attainment Tests*, Niesen and Lydiche, Copenhagen, expanded English edition (1980), University of Chicago Press edition, 1960.
- [56] ROST (J.). — The Growing Family of Rasch models, in Anne Boomsma, Marijtje A. J. van Duijn and Tom A. B. Snijders, editors, *Essays on Item Response Theory*, volume 157 of *Lecture Notes in Statistics*, pages 25–42, Springer-Verlag, New York, 2001.
- [57] SOBEL (M. E.), HOUT (M.), DUNCAN (O. D.). — Exchange, structure, and symmetry in occupational-mobility, *American Journal of Sociology*, 91 (1985), 359–372.
- [58] TEN HAVE (T. R.), BECKER (M. P.). — Multivariate Contingency Tables and the Analysis of Exchangeability, *Biometrics*, 51 (1995), 1001–1016.
- [59] THISSEN (D.), MOONEY (J. A.). — Loglinear item response models, with applications to data from social surveys, *Sociological Methodology*, 19 (1989), 299–330.
- [60] TJUR (T.). — A connection between Rasch's item analysis model and a multiplicative Poisson model, *scandinavian journal of statistics*, 9 (1982), 23–30.
- [61] VAN DER HEIJDEN (P. G. M.), MOOIJAART (A.), DE LEEUW (J.). — Constrained latent budget analysis, *Sociological Methodology*, 22 (1992), 279–320.
- [62] VAN DER LINDEN (W. J.), HAMBLETON (R. K.). — *Handbook of Modern Item Response Theory*, Springer-Verlag, New York, 1997.