Rasch Models with Exchangeable Rows and Columns*

Steffen L. Lauritzen†
Aalborg University

April 3, 2002

Abstract

The article studies distributions of doubly infinite binary matrices with exchangeable rows and columns which satisfy the further property that the probability of any $m \times n$ submatrix is a function of the row- and column sums of that matrix. We show that any such distribution is a (unique) mixture of random Rasch distributions. The non-degenerate elements of these distributions were introduced by Rasch (1960). We investigate the relationship between these random Rasch distributions and a problem in visual perception, the characters of a certain Abelian semigroup, and the problem of existence of measures with given marginals.

Key words: de Finetti's theorem, extreme point models, intelligence testing, majorization, marginal problem, partial exhangeability,

1 Introduction

This article is concerned with the dichotomous Rasch model for item analysis (Rasch 1960). The model was developed to describe outcomes of psychological testing experiments. An item (question, problem) labeled i is presented to a person labeled j and a binary response X_{ij} is recorded. The Rasch model asserts that responses are independent, and that there are parameters α_i ('easinesses') characteristic for the items and parameters β_j ('abilities') characteristic for the persons so that

$$P(X_{ij} = 1 \mid \alpha_i, \beta_j) = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j}. \tag{1}$$

*This is Research Report R-02-2005, Department of Mathematical Sciences, Aalborg University.

†Address for correspondence: Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark. email: steffen@math.auc.dk.

The model and its variants has been the subject of intensive study and interest in the psychometric literature. But the Rasch model is also of fundamental interest in many other contexts and in itself. For example, X_{ij} could indicate whether or not a batter i is getting a hit when matched with a pitcher j (Gutmann *et al.* 1991), X_{ij} could indicate presence or absence of a given species i of bird on island j (Wilson 1987), or X_{ij} could denote the success or failure of mating when a female salamander i is paired with male salamander j (McCullagh and Nelder 1989, page 439 ff.). An overview of literature related to the Rasch model can be found in Fischer and Molenaar (1995).

The Rasch model can in some sense be seen as the fundamental model of randomness for (0,1)-matrices and much effort has also been devoted to derivations of the Rasch model as the unique model satisfying certain fundamental principles. Rasch favoured deriving it from his principle of *specific objectivity* (Rasch 1967, 1977), but, for example, he also derived it from sufficiency arguments, using the basic assumption that the probability of any binary $m \times n$ matrix should only depend on the row- and column sums of this matrix (Rasch 1971), a property clearly satisfied by the Rasch model, since

$$
\begin{aligned}
P_{\alpha\beta}\left(X_{ij} = x_{ij}, i = 1, \ldots, m; j = 1, \ldots, n\right) &= \prod_{i=1}^{m}\prod_{j=1}^{n} \frac{(\alpha_i\beta_j)^{x_{ij}}}{1+\alpha_i\beta_j} \\
&= \frac{\prod_{i=1}^{m} \alpha_i^{r_i} \prod_{j=1}^{n} \beta_j^{c_j}}{\prod_{i=1}^{m}\prod_{j=1}^{n}(1+\alpha_i\beta_j)}, \quad (2)
\end{aligned}
$$

where $r_i = \sum_j x_{ij}$ and $c_j = \sum_i x_{ij}$. Other derivations (Andersen 1973) assume sufficiency of the column sums c_j when item parameters α_i are arbitrary but known and show that this leads to the Rasch model. The derivations usually have an implicit or explicit assumption of *independence* and *regularity* ($0 < P(X_{ij} = 1) < 1$); see Fischer (1995) for a survey. We note that the proof and theorem supplied by Rasch (1971) is inaccurate as it stands, but can be modified to become correct.

In the present paper we attempt to identify Rasch models as *extreme point models* (Lauritzen 1988). More precisely, we replace the assumption of independence with the *exchangeability* of rows and columns and show (Theorem 2) that (randomized) Rasch models are extreme points in the simplex of row-column exchangeable binary matrices with distributions only depending on row- and column sums. This yields an extension of de Finetti's theorem for binary sequences which supplements the results of Aldous (1981), see also Dawid (1982).

The article is composed as follows. In Section 2 we recapitulate some basic concepts and results on exchangeability and convex sets of measures. Section 3 reviews and extends the main results about random binary matrices, and Section 4 places the results in a slightly wider perspective.

2 Preliminaries

2.1 Exchangeable sequences and summarization

We begin by rephrasing some classical results. A sequence $X_1, \ldots, X_n, \ldots$ is said to be *exchangeable* if for all n

$$X_1, \ldots, X_n \stackrel{\mathcal{D}}{=} X_{\pi(1)}, \ldots, X_{\pi(n)} \text{ for all } \pi \in S(n),$$

i.e. if its distribution is invariant under finite permutations. Clearly, if $X_1, \ldots, X_n, \ldots$ are independent and identically distributed, they are exchangeable, but not conversely.

A statistic $t(x)$ is *summarizing* for a discrete probability distribution P (Freedman 1962) if $P(X = x) = p(x) = h(t(x))$ for some function h. Note that if $t(x)$ is summarizing for $P \in \mathcal{P}$, it is sufficient for $\mathcal{P}$ and for all $P \in \mathcal{P}$, $p(x \mid t)$ is uniform on $\{x : t(x) = t\}$.

For binary variables, $X_1, \ldots, X_n, \ldots$ is exchangeable if and only if for all n the sum $t_n(x) = \sum_i x_i$ is summarizing for $p(x_1, \ldots, x_n)$:

$$P(X_1 = x_1, \ldots, X_n = x_n) = h_n(\textstyle\sum_i x_i)$$

because the group of permutations $S(n)$ *acts transitively on binary n-vectors with fixed sum,* i.e. if x and y are two such vectors, there is a permutation which sends x into y, and thus t_n is a *maximal invariant.*

In general, a statistic t is summarizing for P if and only if P is invariant under the group of transformations that leaves t unchanged. Thus the term *partial exchangeability* has often been used for this more general concept.

de Finetti (1931) shows that all exchangeable sequences are mixtures of Bernoulli sequences:

Theorem 1 (de Finetti) *A binary sequence $X_1, \ldots, X_n, \ldots$ is exchangeable if and only if there exists a distribution function F on $[0, 1]$ such that for all n*

$$p(x_1, \ldots, x_n) = \int_0^1 \theta^{t_n(x)} (1-\theta)^{n-t_n(x)} \, dF(\theta).$$

It further holds that F is the distribution function of the limiting frequency:

$$Y = \lim_{n \to \infty} \frac{1}{n} \sum_i X_i, \quad P(Y \leq y) = F(y) \tag{3}$$

and the Bernoulli distribution is obtained by conditioning with $Y = \theta$:

$$P(X_1 = x_1, \ldots, X_n = x_n \,|\, Y = \theta) = \theta^{t_n}(1-\theta)^{n-t_n}.$$

de Finetti's theorem, with its generalisations (Hewitt and Savage 1955) and variants (Diaconis 1977; Diaconis and Freedman 1980), has received much attention in the literature on probability and mathematical statistics (Kingman 1978; Aldous 1985).

2.2 Convex sets, mixtures and extreme points

Theorem 1 gives an *integral representation* of exchangeable measures. To pursue this perspective on de Finetti's Theorem, we need some basic facts about convex sets of measures.

In the following $\mathcal{P}$ denotes a set of probability measures on a space $\mathcal{X}$ which is the countable product of finite sets. Such a set $\mathcal{P}$ is *convex* if

$$P_1, P_2 \in \mathcal{P} \text{ and } 0 < \alpha < 1 \implies \alpha P_1 + (1-\alpha)P_2 \in \mathcal{P}$$

and Q is an *extreme point* of $\mathcal{P}$ if

$$Q = \alpha P_1 + (1-\alpha)P_2 \text{ with } 0 < \alpha < 1, P_1, P_2 \in \mathcal{P} \implies Q = P_1 = P_2.$$

If $\mathcal{P}$ is equipped with the weak topology and $\mathcal{A}$ is a Borel subset of $\mathcal{P}$, P is a *mixture* of elements in $\mathcal{A}$ if there is a probability measure μ on $\mathcal{A}$ such that for all Borel subsets B of $\mathcal{X}$

$$P(B) = \int_{\mathcal{A}} A(B)\mu(dA).$$

A fundamental result is Choquet's Theorem: *If $\mathcal{P}$ is compact, the set of extreme points $\mathcal{E}$ of $\mathcal{P}$ is a non-empty Borel subset of $\mathcal{P}$, and any element of $\mathcal{P}$ is a mixture of the extreme points :*

$$P(B) = \int_{\mathcal{E}} E(B)\mu_P(dE).$$

A convex set $\mathcal{P}$ is a *simplex* if the mixing measure μ_P is uniquely determined by P.

deFinetti's Theorem can alternatively be formulated as: *The exchangeable meaures $\mathcal{P}_{\mathrm{E}}$ form a compact simplex, with Bernoulli measures as extreme points.*

A compact simplex is a *Bauer simplex* if the extreme points $\mathcal{E}$ form a closed (and therefore compact) subset of $\mathcal{P}$. The simplex $\mathcal{P}_{\mathrm{E}}$ is a Bauer simplex since the extreme points can be identified with the interval $[0, 1]$.

3 Partially exchangeable binary matrices

This section first reviews results about binary matrices with exchangeable rows and columns, then binary matrices where the row- and column sums are summarizing statistics, and finally gives new results about matrices with both properties.

3.1 Row-column exchangeable matrices

The distribution P is said to be *row–column exchangeable* (RCE) if for all permutations $\pi \in S(m)$ and $\rho \in S(n)$ we have

$$P\left(X_{ij} = x_{ij}, i = 1, \ldots, m; j = 1, \ldots, n\right) = \\ P\left(X_{ij} = x_{\pi(i)\rho(j)}, i = 1, \ldots, m; j = 1, \ldots, n\right).$$

The corresponding group of transformations of binary matrices is denoted by $\mathcal{G}_{RC}(m, n)$ and as before $\mathcal{G}_{RC} = \cup_{m,n}\mathcal{G}_{RC}(m, n)$ denotes the similar group acting on infinite matrices.

The set $\mathcal{P}_{\mathrm{RCE}}$ of RCE distributions was e.g. studied by Aldous (1981), Diaconis and Freedman (1981), Hoover (1982) and Lynch (1984), and the main results have been collected and extended in Aldous (1985).

There seems to be no simple expression for a statistic which is maximal invariant under the action of $\mathcal{G}_{RC}(m, n)$ and thus there is no simple description of the sufficient (and summarizing) statistic.

Two σ-fields are particularly important for the study of $\mathcal{P}_{\mathrm{RCE}}$ and $\mathcal{P}_{\mathrm{RCS}}$. These are the *tail-field* $\mathcal{T}$ and *shell-field* $\mathcal{S}$ where

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\{X_{ij}, \min(i, j) \geq n\}, \quad \mathcal{S} = \bigcap_{n=1}^{\infty} \sigma\{X_{ij}, \max(i, j) \geq n\}.$$

We exploit the following which is Proposition 14.8 of Aldous (1985):

Proposition 1 *If X has distribution $P \in \mathcal{P}_{\mathrm{RCE}}$ then the following are equivalent:*

(i) *P is extreme in $\mathcal{P}_{\mathrm{RCE}}$;*

(i) *$\mathcal{T}$ is P-trivial*

(i) *X is P-dissociated.*

Here a σ-field $\mathcal{A}$ is said to be *P-trivial* if $P(A) \in \{0,1\}$ for all $A \in \mathcal{A}$ and X is *P-dissociated* if for all A_1, A_2, B_1, B_2 with $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$,

$$\{X_{ij}\}_{i \in A_1, j \in B_1} \text{ and } \{X_{ij}\}_{i \in A_2, j \in B_2} \text{ are independent w.r.t. } P.$$

Diaconis and Freedman (1981) introduce the notion of a *ϕ-matrix*. This is constructed from a measurable function $\phi : [0,1]^2 \to [0,1]$ and independent sequences $U = (U_i)_{i=1,\ldots}$ and $V = (V_i)_{i=1,\ldots}$ of independent random variables, uniformly distributed on the unit interval $[0,1]$, by letting X_{ij} be conditionally independent given U and V and

$$P(X_{ij} = 1 \mid \mathcal{F}) = \phi(u_i, v_j),$$

where $\mathcal{F} = \sigma(U_i, V_i, i = 1, \ldots)$ is the *effect field* of the ϕ-matrix. Clearly ϕ-matrices are necessarily dissociated. Following Lynch (1984), a ϕ-matrix X is said to be *canonical* if

$$X \perp\!\!\!\perp \mathcal{S} \mid \mathcal{F},$$

i.e. if $\mathcal{F}$ captures the whole effect of $\mathcal{S}$ on X. Proposition 1 in combination with Corollary 2.4 of Lynch (1984) now yields:

Proposition 2 *P is extreme in $\mathcal{P}_{\mathrm{RCE}}$ if and only if P is the distribution of a canonical ϕ-matrix.*

Although this proposition gives a relatively simple description of $\mathcal{E}_{\mathrm{RCE}}$, it is still too implicit to be useful for statistical purposes. In particular, it is difficult to get a handle on the ambiguity of the function ϕ as many different ϕ-functions yield the same distribution of its ϕ-matrix.

3.2 Summarized matrices

We consider distributions P of doubly infinite matrices $X = \{X_{ij}\}_{i,j=1,2,\ldots}$ of binary random variables. P is said to be *row-column summarized* (RCS) if the probability of any $m \times n$ (initial) submatrix, depends only on the row- and column sums, i.e. if for all m and n

$$P\left(X_{ij} = x_{ij}, i = 1, \ldots, m; j = 1, \ldots, n\right) = \\ P\left(X_{ij} = y_{ij}, i = 1, \ldots, m; j = 1, \ldots, n\right)$$

whenever

$$\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} y_{ij} \text{ for all } i \text{ and } \sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{m} y_{ij} \text{ for all } j.$$

For $r = (r_1, \ldots, r_m)$ and $c = (c_1, \ldots, c_n)$, we let $\mathcal{M}(r, c)$ denote the set of $m \times n$-matrices with row- and column sums equal to (r, c). We then introduce the group $\mathcal{G}_S(m, n)$ of *switches*. This is the group of one-to-one transformations of $\mathcal{M}(r, c)$ generated by *simple switches,* where a simple switch changes a specified 2×2-submatrix of $x \in \mathcal{M}(r, c)$ as

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and otherwise leave the entries of x invariant.

Theorem 3.1 of Ryser (1957) says that if x and y are matrices in $\mathcal{M}(r, c)$, there exists a switch $g \in \mathcal{G}_S(m, n)$ such that $x = gy$. In other words:

Lemma 1 *The group $\mathcal{G}_S(m, n)$ of switches acts transitively on $\mathcal{M}(r, c)$.*

It thus follows that (r, c) is a maximal invariant under the action of $\mathcal{G}_S(m, n)$. If we let $\mathcal{G}_S = \cup_{m,n} \mathcal{G}_S(m, n)$ denote the similar group acting on infinite matrices we thus have:

Corollary 1 *A distribution P on the set of doubly infinite binary matrices is row–column summarized if and only if it is $\mathcal{G}_S$-invariant.*

Lemma 1 and its corollary was also exploited by Besag and Clifford (1989) to construct a Markov chain Monte Carlo algorithm for simulating from the uniform distribution on $\mathcal{M}(r, c)$, see also Holst (1995), Rao *et al.* (1996) and Ponocny (2001).

We let $\mathcal{P}_{\mathrm{RCS}}$ denote the set of RCS-distributions. Lauritzen (1988) showed that $\mathcal{P}_{\mathrm{RCS}}$ is a convex simplex and found partially the extreme points $\mathcal{E}_{\mathrm{RCS}}$ of this simplex, in particular that $\mathcal{E}_{\mathrm{RCS}} \neq \mathcal{P}_{\mathrm{R}}$, where $\mathcal{P}_{\mathrm{R}}$ is the set of *Rasch distributions* given by (2). More precisely it was shown (Propositions 9.2 and 9.3, page 250) that $P_{\alpha,\beta} \in \mathcal{E}_{\mathrm{RCS}}$ if

$$\sum_{i=1}^{\infty} \frac{\alpha_i \beta_i}{(1 + \alpha_i)(1 + \beta_i)(1 + \alpha_i \beta_i)} = \infty \tag{4}$$

and $P_{\alpha,\beta} \notin \mathcal{E}_{\mathrm{RCS}}$ unless

$$\sum_{i=1}^{\infty} \frac{\alpha_i}{(1 + \alpha_i)^2} = \sum_{i=1}^{\infty} \frac{\beta_i}{(1 + \beta_i)^2} = \infty. \tag{5}$$

Roughly speaking, the conditions are preventing that the α and β sequences vary too rapidly with i. It is therefore natural to expect simpler results when extra symmetry, such as exchangeability, is assumed. The condition (4) implies (5), but it is not known whether any of these two conditions are both necessary and sufficient for $P_{\alpha,\beta}$ to be in $\mathcal{E}_{\mathrm{RCS}}$.

Note that $\mathcal{P}_{\mathrm{RCS}}$ is not a Bauer simplex. To see this, define $P^n_{\alpha,\beta}$ by

$$\alpha_i = \beta_i = \begin{cases} 2^{-i} & \text{if } i < n \\ 1 & \text{otherwise.} \end{cases}$$

Then $P^n_{\alpha,\beta} \in \mathcal{E}_{\mathrm{RCS}}$ for all n as it satisfies (4), but as n tends to infinity it converges to a measure which violates (5), hence $\mathcal{E}_{\mathrm{RCS}}$ is not closed.

3.3 Summarized and exchangeable matrices

This section deals with the set of distributions $\mathcal{P}_{\mathrm{RCES}} = \mathcal{P}_{\mathrm{RCE}} \cap \mathcal{P}_{\mathrm{RCS}}$ which are both RCE and RCS. We let $\mathcal{G}_{RCS}(m,n)$ denote the group of transformations generated by row-column permutations $\mathcal{G}_{RC}(m,n)$ and switches $\mathcal{G}_S(m,n)$ and $\mathcal{G}_{RCS}$ the corresponding group of transformations on infinite matrices.

Lemma 2 *The pair of empirical measures induced by the row- and column sums $t_{mn}(x) = (\sum_{i=1}^m \delta_{r_i}, \sum_{j=1}^n \delta_{c_j})$ is a maximal invariant under the action of $\mathcal{G}_{RCS}(m,n)$.*

Proof Clearly, t_{mn} is invariant so we just have to show that $\mathcal{G}_{RCS}(m,n)$ acts transitively on the set of matrices with a given value of t_{mn}. So assume $t_{mn}(x) = t_{mn}(y) = t$. We first permute the rows and columns of x and y to form $g_1 x$ and $g_2 x$ with increasing row- and column sums using $g_1, g_2 \in \mathcal{G}_{RC}(m,n)$. But then $g_1 x$ and $g_2 y$ have identical row- and column sums and Lemma 1 yields the existence of $g \in \mathcal{G}_S(m,n)$ so that $g g_1 x = g_2 y$. Then $g^* = g_2^{-1} g g_1 \in \mathcal{G}_{RCS}(m,n)$ has $g^* x = y$ as desired. □

The set $\mathcal{P}_{\mathrm{RCES}}$ is a proper subset of $\mathcal{P}_{\mathrm{RCE}}$. This is because the group $\mathcal{G}_{RC}(m,n)$ does not act transitively on sets of matrices x with a fixed value $t_{mn}(x) = t$. For example, if we let

$$x = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

then $\det x = 1$ and $\det y = 0$. Since the absolute value of the determinant is invariant under permutation of rows and columns, there is no element $g \in \mathcal{G}_{RC}(3,3)$ so that $x = gy$. Indeed, for a ϕ-matrix with $\phi(u,v) = uv$, we have $p(x) = 665/2985984$ but $p(y) = 1/4096$, so this distribution is in $\mathcal{P}_{\mathrm{RCE}} \setminus \mathcal{P}_{\mathrm{RCS}}$.

General results (Lauritzen 1988) imply that $\mathcal{P}_{\mathrm{RCES}}$ is a simplex. To identify the extreme points $\mathcal{E}_{\mathrm{RCES}}$ of $\mathcal{P}_{\mathrm{RCES}}$ we need the following lemma:

Lemma 3 *If $P \in \mathcal{P}_{\mathrm{RCES}}$ and $E \in \mathcal{T}$ with $P(E) > 0$, then $P(\cdot \mid E) \in \mathcal{P}_{\mathrm{RCES}}$.*

Proof Let $E \in \mathcal{T}$ with $P(E) > 0$. By Lemma 2 we must just show that $P(\cdot \mid E)$ is $\mathcal{G}_{RCS}$-invariant. So let $g \in \mathcal{G}_{RCS}$. When $E \in \mathcal{T}$ it is clearly invariant both under switches, row- and column permutations so $gE = E$. Let D be an arbitrary measurable subset of infinite binary matrices. We thus have

$$\begin{aligned} P(gX \in D \mid X \in E) &= \frac{P(gX \in D \wedge X \in E)}{P(X \in E)} \\ &= \frac{P(gX \in (D \cap gE))}{P(X \in E)} \\ &= \frac{P(X \in D \cap E)}{P(X \in E)} \\ &= P(X \in D \mid X \in E) \end{aligned}$$

and thus $P(\cdot \mid E) \in \mathcal{P}_{\mathrm{RCES}}$. □

Further, we have

Lemma 4 $\mathcal{E}_{\mathrm{RCES}} = \mathcal{P}_{\mathrm{RCES}} \cap \mathcal{E}_{\mathrm{RCE}}$.

Proof The inclusion $\mathcal{E}_{\mathrm{RCES}} \supseteq \mathcal{P}_{\mathrm{RCES}} \cap \mathcal{E}_{\mathrm{RCE}}$ is obvious. To show the reverse inclusion, we assume that $P \in \mathcal{P}_{\mathrm{RCES}} \subseteq \mathcal{E}_{\mathrm{RCE}}$ and show that $P \notin \mathcal{E}_{\mathrm{RCES}}$. Since $P \notin \mathcal{E}_{\mathrm{RCE}}$, $\mathcal{T}$ is not P-trivial. Thus there exists $E \in \mathcal{T}$ with $0 < P(E) < 1$. We may now write

$$P(\cdot) = P(\cdot \mid X \in E)P(E) + P(\cdot \mid X \notin E)(1 - P(E)).$$

By Lemma 3 this expresses P as a non-trivial convex combination of two elements of $\mathcal{P}_{\mathrm{RCES}}$ implying that $P \notin \mathcal{E}_{\mathrm{RCES}}$. □

Next we must realize that $\mathcal{P}_{\mathrm{RCS}}$ is stable under shell-conditioning:

Lemma 5 *If $P \in \mathcal{P}_{\mathrm{RCS}}$ then $P(\cdot \mid \mathcal{S}) \in \mathcal{P}_{\mathrm{RCS}}$.*

Proof Using Corollary 1 this follows as in the proof of Lemma 3 because G_S leaves sets in $\mathcal{S}$ invariant. □

Note that $P(\cdot \mid \mathcal{S})$ is typically *not* RCE. Next we define ϕ to be of *Rasch type* if it satisfies the functional equation:

$$\phi(u,v)\bar{\phi}(u,v^*)\bar{\phi}(u^*,v)\phi(u^*,v^*) = \bar{\phi}(u,v)\phi(u,v^*)\phi(u^*,v)\bar{\phi}(u^*,v^*), \tag{6}$$

where we have let $\bar{\phi} = 1 - \phi$. If P is the distribution of a ϕ-matrix with ϕ of Rasch type, $P(\cdot \mid \mathcal{F})$ is $\mathcal{G}_S$ invariant. This therefore also holds for its unconditional distribution P, implying that $P \in \mathcal{P}_{\mathrm{RCES}}$. In fact, below we show that such ϕ-matrices exactly correspond to the extreme points $\mathcal{E}_{\mathrm{RCES}}$ of $\mathcal{P}_{\mathrm{RCES}}$.

Theorem 2 *If $P \in \mathcal{P}_{\mathrm{RCES}}$, it is in $\mathcal{E}_{\mathrm{RCES}}$ if and only if P is the distribution of a ϕ-matrix of Rasch type.*

Proof If X is a ϕ-matrix, its distribution P is extreme in $\mathcal{P}_{\mathrm{RCE}}$ and *a forteriori* extreme in $\mathcal{P}_{\mathrm{RCES}} \subset \mathcal{P}_{\mathrm{RCE}}$. Thus we only need to show the converse.

Assume P is an extreme point of $\mathcal{P}_{\mathrm{RCES}}$. From Lemma 4 and Proposition 2 we get that P is the distribution of some canonical ϕ-matrix X. Lemma 5 implies

$$P\left(\left\{\begin{array}{ll} X_{ij}=1 & X_{ij^*}=0 \\ X_{i^*j}=0 & X_{i^*j^*}=1 \end{array}\right\} \middle|\, \mathcal{S}\right) = P\left(\left\{\begin{array}{ll} X_{ij}=0 & X_{ij^*}=1 \\ X_{i^*j}=1 & X_{i^*j^*}=0 \end{array}\right\} \middle|\, \mathcal{S}\right).$$

Since X is canonical, $P(X \in A \mid \mathcal{S}) = P(X \in A \mid \mathcal{F})$ and thus

$$\begin{aligned} &\phi(U_i,V_j)\{1-\phi(U_i,V_{j^*})\}\{1-\phi(U_{i^*},V_j)\}\phi(U_{i^*},V_{j^*}) = \\ &\quad \{1-\phi(U_i,V_j)\}\phi(U_i,V_{j^*})\phi(U_{i^*},V_j)\{1-\phi(U_{i^*},V_{j^*})\}, \end{aligned}$$

i.e. it is a ϕ-matrix of Rasch type. □

If $0 < \phi < 1$ the solutions of (6) all have the special form

$$\phi(u,v) = \frac{a(u)b(v)}{1+a(u)b(v)} \tag{7}$$

where a and b map the unit inverval into the positive half-line. This is seen by letting $\rho = \phi/\bar{\phi}$ and fixing (u^*,v^*) whereby (6) can be manipulated to

$$\rho(u,v) = \frac{\rho(u,v^*)\rho(u^*,v)}{\rho(u^*,v^*)}$$

so that we may let $a(u) = \rho(u, v^*)/\rho(u^*, v^*)$ and $b(v) = \rho(u^*, v)$ to satisfy (7). Indeed we may without loss of generality assume that a and b are determined from distribution functions A and B on the positive halfline as $a = A^{-1}$ and $b = B^{-1}$ so that A and B are the distributions of $\alpha_i = a(U_i)$ and $\beta_j = b(V_j)$ respectively.

A *regular random Rasch distribution* is now defined to be the distribution of a ϕ-matrix with ϕ having the form (7) and the set of such distributions is denoted by $\mathcal{P}_{\mathrm{RR}}$.

Note that there is ambiguity between a and b in the sense that one can be transformed by multiplication with a positive constant and the other with division without changing the ϕ-matrix. But modulo this, the pairs (A, B) are in a one-to-one correspondence with the elements of $\mathcal{P}_{\mathrm{RR}}$.

It follows that the extreme points of $\mathcal{P}_{\mathrm{RCES}}$ which are non-degenerate, in the sense that they correspond to truly non-deterministic matrices, are regular random Rasch distributions. More precisely, if we say that P is *regular* when

$$0 < P(X_{ij} = 1 \,|\, \mathcal{S}) < 1 \text{ for all } i, j,$$

we have:

Corollary 2 *P is a regular extreme point of $\mathcal{P}_{\mathrm{RCES}}$ if and only if $P \in \mathcal{P}_{\mathrm{RR}}$.*

There are many non-regular extreme points of $\mathcal{P}_{\mathrm{RCES}}$, essentially correponding to all non-regular solutions of the functional equation (6). An example of such a solution is

$$\phi(u, v) = \chi_{\{u \leq v\}} = \begin{cases} 1 & \text{if } u \leq v \\ 0 & \text{otherwise.} \end{cases}$$

This is an example of a ϕ-matrix of Rasch type which is 'deterministic' in the sense that it is $\mathcal{S}$-measurable. In the context of intelligence tests, the interpretation of this model is that a person with ability v solves a problem of difficulty u with certainty if $v \geq u$ but never if $u > v$. Variants of the model appear for h and k being monotone functions from the unit interval to itself and then

$$\phi^*(u, v) = \chi_{\{h(u) \leq k(v)\}}. \tag{8}$$

It seems plausible that these are the only ϕ-matrices of Rasch type which are essentially $\mathcal{S}$-measurable. Proposition 3.6 of Aldous (1981) implies that this is true if and only if all solutions to the functional equation (6) with $\phi \in \{0, 1\}$ had the form (8).

There are many more ϕ-matrices of Rasch type. Consider for example

$$\phi(u,v) = \begin{cases} \dfrac{a(u)b(v)}{1+a(u)b(v)} & \text{if } 1/3 < u,v < 2/3 \\ \chi_{\{u \le v\}} & \text{otherwise} \end{cases}$$

which can be seen as dividing difficulties and abilities into three classes of equal size: *low, medium, high*, so that an ordinary Rasch model prevails when persons of medium ability solve questions of medium difficulty whereas other combinations yield a deterministic outcome.

There are similar models with more than three groups. For example, one can keep cutting out middle thirds of the unit interval as above to get

$$\phi(u,v) = \begin{cases} \dfrac{a(u)b(v)}{1+a(u)b(v)} & \text{if } 1/9 < u,v < 2/9 \\ \dfrac{a(u)b(v)}{1+a(u)b(v)} & \text{if } 1/3 < u,v < 2/3 \\ \dfrac{a(u)b(v)}{1+a(u)b(v)} & \text{if } 7/9 < u,v < 8/9 \\ \chi_{\{u \le v\}} & \text{otherwise} \end{cases},$$

and so on. Since the simplex $\mathcal{P}_{\text{RCES}}$ is a Bauer simplex (Ressel 2002, personal communication), the set of its extreme points is closed. Thus the sequence of distributions of ϕ-matrices defined by this procedure will converge to what could be termed a *Cantor–Rasch distribution* with an infinite number of groups.

Although these non-regular Rasch models are unusual, they are by no means counterintuitive.

4 Other perspectives

4.1 The Julesz conjecture

Diaconis and Freedman (1981) discuss a conjecture of Julesz (1975, 1980) in visual perception saying that two 'random patterns' (i.e. binary matrices) with the same first- and second-order statistics (joint distributions of singletons and pairs) cannot be visually distinguished. Indeed they give several examples of ϕ-matrices with the same first- and second-order statistics as a purely random 'coin tossing' matrix which are visually very different from such a matrix.

Here we show that such counterexamples cannot be of Rasch type. This implies that certain types of deviations from the Rasch model may indeed be visually detected from inspecting large matrices.

Say a binary matrix is *purely random* if X_{ij} are all independent and $P(X_{ij}=1)=1/2$ for all i,j.

Theorem 3 *Let X be a ϕ-matrix of Rasch type with the same first- and second-order statistics as a purely random matrix. Then it is a purely random matrix.*

Proof Theorem (3.8) of Diaconis and Freedman (1981) says that a ϕ-matrix has the same first- and second-order statistics as a purely random matrix if

$$\int \phi(u,v)\,du = 1/2 \text{ a.e. } (u) \text{ and } \int \phi(u,v)\,dv = 1/2 \text{ a.e. } (v), \tag{9}$$

so this is what we assume. We will show that if ϕ is of Rasch type, it is a.e. constant and thus equal to $1/2$.

Since ϕ is of Rasch type it satisfies (6). If we expand and reduce the terms in this equation we find that it is equivalent to

$$\begin{aligned}&\phi(u,v)\phi(u^*,v^*)+\phi(u,v)\phi(u,v^*)\phi(u^*,v)+\phi(u,v^*)\phi(u^*,v)\phi(u^*,v^*)=\\&\quad\phi(u,v^*)\phi(u^*,v)+\phi(u,v)\phi(u,v^*)\phi(u^*,v^*)+\phi(u,v)\phi(u^*,v)\phi(u^*,v^*).\end{aligned}$$

Integrating this equation with respect to u and using (9) yields that for almost all v and almost all v^*

$$\begin{aligned}&\phi(u^*,v^*)/2+I(v,v^*)\phi(u^*,v)+\phi(u^*,v)\phi(u^*,v^*)/2=\\&\quad\phi(u^*,v)/2+I(v,v^*)\phi(u^*,v^*)+\phi(u^*,v)\phi(u^*,v^*)/2,\end{aligned}$$

where $I(v,v^*)=\int\phi(u,v)\phi(u,v^*)\,du$. Reducing and rearranging terms leads to

$$\{\phi(u^*,v^*)-\phi(u^*,v)\}\{I(v,v^*)-1/2\}=0. \tag{10}$$

Next let $A=\{v \mid I(v,v^*)=1/2\}$, then for almost all $v\notin A$ we have $\phi(u,v)=\phi(u,v^*)$ and hence also $I(v,v^*)=I(v^*,v^*)$ for almost all $v\notin A$, whereby

$$\int I(v,v^*)\,dv=\int_A I(v,v^*)\,dv+\int_{A^c} I(v,v^*)\,dv=\lambda(A)/2+I(v^*,v^*)\{1-\lambda(A)\},$$

where λ is Lebesgue measure. Using now that

$$I(v^*,v^*)=\int\phi(u,v^*)^2\,du\geq\left\{\int\phi(u,v^*)\,du\right\}^2=1/4$$

and

$$\int I(v,v^*)\,dv=\int\int\phi(u,v)\phi(u,v^*)\,du\,dv=\int\int\phi(u,v)\phi(u,v^*)\,dv\,du=1/4,$$

we find

$$1/4 \geq \lambda(A)/2 + \{1 - \lambda(A)\}/4 = 1/4 + \lambda(A)/4$$

whereby $\lambda(A) = 0$. Hence $\phi(u, \cdot)$ is constant almost everywhere. By symmetry, $\phi(\cdot, v)$ is also constant so ϕ must be constant and equal to $1/2$. □

Note that we have not proved the somewhat stronger statement saying that two ϕ-matrices of Rasch type which have identical first- and second order statistics, have identical distributions, and indeed this does not hold in general.

To elaborate a little on this issue in the regular case, let A and B denote the distributions of the row- and column-effects $a(U_i)$ and $b(V_j)$ as in the previous section and let ϕ be given by (7). Introduce further what we choose to call the *Rasch transform* $\breve{A}$ of A as

$$\breve{A}(x) = \int_0^\infty \frac{xy}{1+xy} A(dy). \tag{11}$$

Clearly, $\breve{A}$ will be strictly increasing and differentiable with $\breve{A}(0) = 0$ and $\lim_{x\to\infty} \breve{A}(x) = 1$, so $\breve{A}$ is the distribution function of an absolutely continuous probability distribution on $(0, \infty)$.

The first-order statistics of a binary random Rasch matrix is

$$\begin{aligned} P(X_{11} = 1) &= \int\int \phi(u, v)\, du\, dv = \int_0^\infty \int_0^\infty \frac{xy}{1+xy} A(dx)\, B(dy) \\ &= \int_0^\infty \breve{A}(y)\, B(dy) = \int_0^\infty \breve{B}(x)\, A(dx). \end{aligned}$$

The interesting second-order statistics are similarly

$$\begin{aligned} P(X_{11} = 1, X_{12} = 1) &= \int_0^\infty \int_0^\infty \int_0^\infty \frac{xy}{1+xy} \frac{xz}{1+xz} A(dx)\, B(dy)\, B(dz) \\ &= \int_0^\infty \breve{B}(x)^2\, A(dx) \end{aligned}$$

and

$$P(X_{11} = 1, X_{21} = 1) = \int_0^\infty \breve{A}(y)^2\, B(dy).$$

Therefore, if the first and second-order statistics correspond to independence

$$P(X_{11} = 1, X_{21} = 1) = P(X_{11} = 1)P(X_{21} = 1)$$

we must have $\breve{A}(y)$ almost surely constant with respect to B and thus B must be degenerate. Symmetry yields A degenerate. But in general the first- and second-order statistics will not identify A and B.

4.2 Analytic properties of $\mathcal{P}_{\mathrm{RCES}}$

In several papers, Ressel (1985, 1988, 1994) has studied convex sets of measures with symmetry properties from an analytic point of view. For example, he has considered simplices of probability distributions which are summarized by additive statistics with values in Abelian semigroups. The case of $\mathcal{P}_{\mathrm{RCES}}$ is such an example, where the semigroup S is the subsemigroup of pairs of measures on the non-negative integers generated by the summarizing statistics

$$t_{mn}(x) = \left(\sum_{i=1}^{m} \delta_{r_i}, \sum_{j=1}^{n} \delta_{c_j}\right)$$

for $m, n = 1, 2, \ldots$. This family of statistics can be shown (Ressel 2002, personal communication) to be 'strongly almost additive' and thus 'strongly positivity forcing' (Ressel 1994), which implies that $\mathcal{P}_{\mathrm{RCES}}$ is a Bauer simplex and the extreme points $\mathcal{E}_{\mathrm{RCES}}$ are determined by normalized characters $\sigma \in \hat{S}$ so that for a binary $m \times n$ matrix x it holds

$$p_\sigma(x) = \sigma(t_{mn}(x)),$$

where $\sigma(t \oplus s) = \sigma(t)\sigma(s)$ for all $s, t \in S$. Theorem 2 thus identifies the characters of this semigroup in terms of solutions to the functional equation (6), albeit in a rather implicit fashion.

To describe the characters in more detail, we may represent the elements of the semigroup S by vectors (r, c) with elements ordered so that $r_1 \leq \cdots \leq r_m$ and $c_1 \leq \cdots \leq c_n$. Then $(r, c) \in S$ if and only if the set of matrices $\mathcal{M}(r, c)$ with row sums r and column sums c is non-empty. Gale (1957) and Ryser (1957) have shown that $\mathcal{M}(r, c) \neq \emptyset$ *if and only if* $r \preceq c^*$, where c^* is the *conjugate* sequence of c

$$c_j^* = |\{l \mid c_l \geq j\}|$$

and $\preceq$ denotes *majorization*

$$a \preceq b \Leftrightarrow \sum_{i=1}^{k} a_i \leq \sum_{j=1}^{k} b_j \text{ for all } k = 1, \ldots, m.$$

The value of the character ρ_ϕ where ϕ satisfies (6) is then given as

$$\rho_\phi(r, c) = \int \cdots \int \prod_i \prod_j \phi(u_i, v_j)^{x_{ij}} \{1 - \phi(u_i, v_j)\}^{1 - x_{ij}}\, du_i dv_j,$$

where x is an arbitrary element of $\mathcal{M}(r, c)$. The description is still somewhat implicit since many choices of ϕ lead to the same character ρ_ϕ.

4.3 Marginal problems

A problem related to the Rasch model was investigated by Gutmann *et al.* (1991). Simulation models for baseball were considered in which a random batter of batting average Y was confronted with a random pitcher of pitching average Z. If we let $W = \psi(Y, Z)$ denote the probability of a hit, we must have

$$\mathbf{E}(W \mid Z) = Z; \quad \mathbf{E}(W \mid Y) = Y; \quad 0 \leq W \leq 1. \tag{12}$$

Dawid *et al.* (1995) discuss the related problem of coherent combination of experts' opinions. Here Y and Z are experts' opinons in the form of their subjective probabilities for some event A. Then $W = \psi(Y, Z)$ is a *coherent combination* of the experts' opinions if and only if (12) holds.

If F and G are the distribution functions of Y and Z such a function ψ exists if and only if it holds for all $s, t \in [0, 1]$ that

$$\int_s^1 x\, F(dx) + \int_t^1 y\, G(dy) \leq \int_0^1 x\, F(dx) + \{1 - F(s)\}\{1 - G(t)\}. \tag{13}$$

This was shown as Theorem 4 of Gutmann *et al.* (1991), using classical results of Kellerer (1961) and Strassen (1965). Gutmann *et al.* (1991) also show that if this condition is met, ψ can be chosen to be increasing in each of its arguments, and ψ can also be chosen to be the indicator of a set, although not always both simultaneously, see Proposition 4 below. Note that for $0 \leq \psi \leq 1$, (12) may also be written as

$$\int_0^1 \psi(x, y)\, F(dx) = y \text{ a.e. (F)}, \quad \int_0^1 \psi(x, y)\, G(dy) = x \text{ a.e. (G)}. \tag{14}$$

Clearly, if F and G are such a pair and we let

$$\phi(u, v) = \psi(F^{-1}(u), G^{-1}(v)),$$

we obtain a ϕ-matrix of batting outcomes.

The results of Gutmann *et al.* (1991) can be seen as a continuous analogue of the Gale–Ryser theorem. To make this more precise, we define the *conjugate* F^* of a distribution function F on the unit interval by

$$F^*(x) = 1 - F^{-1}(1 - x),$$

where F^{-1} is the left-continuous inverse of F:

$$F^{-1}(x) = \sup\{y \mid F(y) \leq x\}.$$

As in the discrete case we say that G *majorizes* F and write $F \preceq G$ if

$$\int_0^s F(x)\,dx \le \int_0^s G(x)\,dx \text{ for all } s \in [0,1].$$

Proposition 3 *Let F and G be two distribution functions on $[0,1]$. Then there exists a function ψ satisfying (14) if and only if $F \preceq G^*$.*

Proof We simply show that (12) holds if and only if $F \preceq G^*$. Partial integration in (12) yields

$$\int_0^s F(x)\,dx \le sF(s) + tG(t) - F(s) - G(t) + F(s)G(t) + \int_t^1 G(x)\,dx. \quad (15)$$

A small picture makes it clear that

$$\int_t^1 G(x)\,dx = \int_0^{1-G(t)} G^*(x)dx + (1-t)G(t).$$

Letting $u = 1 - G(t)$ and inserting the above into (15) yields that (12) holds for all s, t if and only if it holds for all s, u that

$$\int_0^s F(x)\,dx \le (s-u)F(s) + \int_0^u G^*(x)\,dx. \quad (16)$$

If we assume (16), we may let $u = s$ and deduce that $F \preceq G^*$. Conversely, if we assume $F \preceq G^*$, we have

$$\begin{aligned} (s-u)F(s) + \int_0^u G^*(x)\,dx &\ge (s-u)F(s) + \int_0^u F(x)\,dx \qquad (17) \\ &\ge \int_0^s F(x)\,dx \end{aligned}$$

because differentiation w.r.t. u shows that the right-hand side of (17) is at minimum for $u = s$. Thus we have shown (16), as needed. □

The proposition on p. 1793 of Gutmann *et al.* (1991) can be rephrased as

Proposition 4 *If F and G are continuous, there exists $\psi \in \{0,1\}$ which is increasing in each of its arguments and satisfies (14) if and only if $F = G^*$.*

The analogy with the Gale–Ryser theorem becomes clearer if we let

$$F_{mn}(x) = \frac{1}{m}\sum_1^m \delta_{r_i/n}([0,x]), \quad G_{mn}(y) = \frac{1}{n}\sum_1^n \delta_{c_j/m}([0,y]) \quad (18)$$

whereby some manipulation shows that

$$r \preceq c^* \iff F_{mn} \preceq G^*_{mn}.$$

If we consider a random Rasch ϕ-matrix, given by distributions (A, B) of row- and column sums, we get for the infinite row- and column averages

$$\bar{X}_{i\infty} \quad = \quad \mathbf{E}(\bar{X}_{i\infty} \,|\, \mathcal{S}) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{\alpha_i \beta_j}{1+\alpha_i\beta_j} = \int_0^\infty \frac{\alpha_i\beta}{1+\alpha_i\beta} B(d\beta) = \check{B}(\alpha_i),$$

where $\check{B}$ is the Rasch transform defined in (11). Similarly we get $\bar{X}_{\infty j} = \check{A}(\beta_j)$. Thus if we let $F(x)$ denote the distribution function of the row average $\bar{X}_{i\infty}$, we have

$$F(x) = P(\bar{X}_{i\infty} \leq x) = P(\check{B}(\alpha_i) \leq x) = A(\check{B}^{-1}(x)) \tag{19}$$

and similarly $G(x) = B(\check{A}^{-1}(x))$ where G is the distribution function of $\bar{X}_{\infty j}$.

Clearly, we may consider the pair $t^*_{mn} = (F_{mn}, G_{mn})$ in (18) of empirical distributions of the row- and column averages as the summarizing statistic for $\mathcal{P}_{\mathrm{RCES}}$. In analogy with (3) of de Finetti's theorem, we then have that for any $P \in \mathcal{P}_{\mathrm{RCES}}$ this pair converges to a pair (F, G) of distributions satisfying $F \preceq G^*$ and the mixing measure μ_P on $\mathcal{E}_{\mathrm{RCES}}$ is the distribution of this pair; we refrain from giving the details of the argument.

An obvious question to ask next is whether to any given *subconjugate pair* (F, G) of distributions, i.e. pair of distributions satisfying $F \preceq G^*$, one can find a ϕ-matrix of Rasch type, so that $\mathcal{E}_{\mathrm{RCES}}$ can be identified with the set of subconjugate pairs.

So consider a pair (F, G). From (19) it follows that these are the distributions of row- and column averages of a regular random Rasch model if and only if there exist distributions A and B on $(0, \infty)$ so that

$$F(\check{B}(x)) = A(x) \text{ and } G(\check{A}(y)) = B(y) \text{ for all } x \text{ and } y. \tag{20}$$

In the case where (F, G) are empirical distributions of the form (18), (20) is easily seen to be equivalent to the equation system

$$\frac{r_i}{n} = \frac{1}{n} \sum_j \frac{\alpha_i\beta_j}{1+\alpha_i\beta_j} \text{ and } \frac{c_j}{m} = \frac{1}{m} \sum_i \frac{\alpha_i\beta_j}{1+\alpha_i\beta_j}, \tag{21}$$

where then A and B are the empirical distributions of $\{\alpha_i\}$ and $\{\beta_j\}$. This fact is most directly seen when row sums and column sums are all different and ordered to be increasing, since then

$$F_{mn}(\check{B}(\alpha_i)) = F_{mn}(r_i/n) = i/n = A(\alpha_i)$$

and similarly with β_j.

The equation system (21) is exactly the maximum likelihood equations for estimation of the parameters in the Rasch model and these are known to have a solution (Fischer 1981) if and only if $r \prec s^*$ where $\prec$ denotes strict majorization

$$a \prec b \Leftrightarrow \sum_{i=1}^{k} a_i < \sum_{j=1}^{k} b_j \text{ for all } k = 1, \ldots, m,$$

and the solution is unique up to multiplication of α_i with a positive constant c and division of β_j with the same constant. Thus, if we say that (F, G) are *strictly subconjugate* if $F \prec G^*$, where $\prec$ means strict majorization:

$$F \prec G \Longleftrightarrow \int_0^s F(x)\,dx < \int_0^s G(x)\,dx \text{ for all } 0 < s < 1$$

it seems natural to conjecture:

Conjecture 1 *Let (F, G) be a pair of distribution functions on $[0, 1]$. Then there exists a ϕ-matrix of Rasch type with distributions of asymptotic marginal row- and column averages given by F and G if and only if $F \preceq G^*$. Moreover, the distribution of the ϕ-matrix is injectively parametrized by (F, G) and the corresponding ϕ-matrix is regular if and only if $F \prec G^*$.*

However, at present it is not clear to the author how to prove this, although part of the conjecture should follow from a suitable limiting argument, using the result about existence and uniqueness of the maximum likelihood estimates. Note that the case $F = G^*$ of Proposition 4 indeed corresponds to the non-regular Rasch-matrix determined by (8).

Acknowledgements

The results in the present paper have their origin in my visit to London in 1979 and I am indebted to inspiration from discussions with Phil Dawid during that visit, where I first learned about subconjugate distributions and majorization and the results later published in Aldous (1981). Ever since, I have from time to time worked on the issues in the present paper. Significant progress was made in 1997 while I was a Fellow at the Center for

Advanced Study in Behavioral Sciences at Stanford, USA, and I gratefully acknowledge financial support from NSF grant # SBR–9022192. Finally a visit to Klitgaarden — a refuge in Skagen, Denmark, for scientists and artists — provided the necessary peace to complete the paper and consequently I am deeply indebted to this institution.

References

Aldous, D. (1981). Representations for partially exchangeable random variables. *Journal of Multivariate Analysis*, **11**, 581–98.

Aldous, D. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint–Flour XIII — 1983*, (ed. P. Hennequin), pp. 1–198. Springer-Verlag, Heidelberg. Lecture Notes in Mathematics 1117.

Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, **26**, 31–44.

Besag, J. and Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika*, **76**, 633–42.

Dawid, A. P. (1982). Intersubjective statistical models. In *Exchangeability in Probability and Statistics*, (ed. G. Koch and F. Spizzichino), pp. 217–32. North-Holland Publishing Company, Amsterdam.

Dawid, A. P., DeGroot, M. H., and Mortera, J. (1995). Coherent combination of experts' opinions (with discussion). *Test*, **4**, 263–313.

de Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale*, **4**, 251–99.

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthése*, **36**, 271–81.

Diaconis, P. and Freedman, D. (1980). de Finetti's theorem for Markov chains. *Annals of Probability*, **8**, 745–64.

Diaconis, P. and Freedman, D. (1981). On the statistics of vision: the Julesz conjecture. *Journal of Mathematical Psychology*, **24**, 112–38.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, **46**, 59–77.

Fischer, G. H. (1995). Derivations of the Rasch model. In Fischer and Molenaar (1995), pp. 15–38.

Fischer, G. H. and Molenaar, I. W. (ed.) (1995). *Rasch Models. Foundations, Recent Developments, and Applications.* Springer-Verlag, New York.

Freedman, D. (1962). Invariants under mixing which generalize de Finetti's theorem. *Annals of Mathematical Statistics*, **33**, 916–23.

Gale, D. (1957). A theorem on flows in networks. *Pacific Journal of Mathematics*, **7**, 1073–82.

Gutmann, S., Kemperman, J. H. B., Reeds, J. A., and Shepp, L. A. (1991). Existence of probability measures with given marginals. *Annals of Probability*, **19**, 1781–97.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, **80**, 470–501.

Holst, C. (1995). *Item Response Theory.* PhD thesis, University of Copenhagen.

Hoover, D. N. (1982). Row–column exchangeability and a general model for exchangeability. In *Echangeability in Probability and Statistics*, (ed. G. Koch and F. Spizzichino), pp. 281–91. North-Holland, Amsterdam.

Julesz, B. (1975). Experiments in the visual perception of textures. *Scientific American*, **232**, 34–43.

Julesz, B. (1980). Spatial nonlinearities in the instantaneous perception of textures with identical power spectra. *Philosophical Transactions of the Royal Society of London, Series B*, **294**, 83–94.

Kellerer, H. G. (1961). Funktionen auf Produkträumen mit vorgegebenen Marginal-Funktionen. *Mathematische Annalen*, **144**, 323–44.

Kingman, J. F. C. (1978). Uses of exchangeability. *Annals of Probability*, **6**, 183–97.

Lauritzen, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics.* Springer-Verlag, Heidelberg. Lecture Notes in Statistics 49.

Lynch, J. (1984). Canonical row–column-exchangeable arrays. *Journal of Multivariate Analysis*, **15**, 135–40.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Monographs on statistics and applied probability, (2nd edn). Chapman and Hall, London.

Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the Rasch model. *Psychometrika*, **66**, 437–60.

Rao, A. R., Jana, R., and Bandyopadhyay, S. (1996). A Markov chain Monte Carlo method for generating random (0,1)-matrices with given marginals. *Sankhyā, Series A*, **58**, 225–42.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*, Studies in mathematical psychology, Vol. 1. Danmarks Pædagogiske Institut, Copenhagen.

Rasch, G. (1967). An informal report on a theory of objectivity in comparisons. In *Psychological Measurement Theory*, (ed. L. J. T. van der Kamp and C. A. J. Vlek), pp. 1–19. University of Leyden, Leyden.

Rasch, G. (1971). Proof that the necessary condition for the validity of the multiplicative dichotomic model is also sufficient. Unpublished manuscript.

Rasch, G. (1977). On specific objectivity. an attempt at formalizing the request for generality and validity of scientific statements. In *The Danish Yearbook of Philosophy*, (ed. M. Blegvad), pp. 58–94. Munksgaard, Copenhagen.

Ressel, P. (1985). de Finetti-type theorems: An analytical approach. *Annals of Probability*, **13**, 898–922.

Ressel, P. (1988). Integral representations for distributions of symmetric stochastic processes. *Probability Theory and Related Fields*, **79**, 451–67.

Ressel, P. (1994). Non-homogeneous de Finetti-type theorems. *Journal of Theoretical Probability*, **7**, 469–82.

Ryser, H. J. (1957). Combinatorial properties of matrices of zeros and ones. *Canadian Journal of Mathematics*, **9**, 371–7.

Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, **36**, 423–39.

Wilson, J. B. (1987). Methods for detecting non-randomness in species cooccurrences: A contribution. *Oecologica*, **73**, 579–82.