BAYESIAN STATISTICS 7, pp. 000–000
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

Rasch Models with Exchangeable Rows and Columns

STEFFEN L. LAURITZEN
Aalborg Universitet, Denmark
steffen@math.auc.dk

SUMMARY

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.

Keywords: 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

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

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

$$P_{\alpha\beta}\left\{(X_{ij}=x_{ij})_{i=1,\dots,m;j=1,\dots,n}\right\}=\prod_{i=1}^{m}\prod_{j=1}^{n}\frac{(\alpha_i\beta_j)^{x_{ij}}}{1+\alpha_i\beta_j}=\frac{\prod_{i=1}^{m}\alpha_i^{r_i}\prod_{j=1}^{n}\beta_j^{c_j}}{\prod_{i=1}^{m}\prod_{j=1}^{n}(1+\alpha_i\beta_j)}, \qquad (2)$$

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 $X_1, \dots, X_n, \dots$ is said to be *exchangeable* if for all n

$$P\left\{(X_i = x_i)_{i=1,\dots,n}\right\} = P\left\{(X_i = x_{\pi(i)})_{i=1,\dots,n}\right\} \text{ for all } \pi \in S(n),$$

i.e. if its distribution is invariant under finite permutations. Clearly, if $X_1, \dots, X_n, \dots$ 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, $X_1, \dots, X_n, \dots$ is exchangeable if and only if for all n the sum $t_n(x) = \sum_i x_i$ is summarizing for $p(x_1, \dots, x_n)$:

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

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 $X_1, \dots, X_n, \dots$ is exchangeable if and only if there exists a distribution function F on $[0,1]$ such that for all n*

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

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

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

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

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

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

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

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

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

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. de Finetti'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

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

We denote the corresponding group of transformations 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 $\mathcal{P}_{\mathrm{RCE}}$;*
(ii) *$\mathcal{T}$ is P-trivial;*
(iii) *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 $U = (U_i)_{i=1,\dots}$ and $V = (V_i)_{i=1,\dots}$ 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 $\mathcal{F} = \sigma(U_i, V_i, i = 1, \dots)$ 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 $\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} y_{ij}$ for all i and $\sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{m} y_{ij}$ 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 for 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

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

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 $P(\cdot \,|\, \mathcal{S}) \in \mathcal{P}_{\mathrm{RCS}}$.*

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

Note that $P(\cdot \,|\, \mathcal{S})$ 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, $P(\cdot \,|\, \mathcal{F})$ 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

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

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

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

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 of 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

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

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 $v \ge u$ 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

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

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 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

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

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

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

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.

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 $t_{mn}(x) = (\sum_{i=1}^m \delta_{r_i}, \sum_{j=1}^n \delta_{c_j})$ 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 (1963) 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:* $a \preceq b \Leftrightarrow \sum_{i=1}^k a_i \leq \sum_{j=1}^k b_j$ 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

$$\mathrm{E}(W \mid Z) = Z; \quad \mathrm{E}(W \mid Y) = Y; \quad 0 \leq W \leq 1. \tag{11}$$

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 (11) 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

$$\int_s^1 x\, F(dx) + \int_t^1 y\, G(dy) \leq \int_0^1 x\, F(dx) + \{1 - F(s)\}\{1 - G(t)\}. \tag{12}$$

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 $0 \leq \psi \leq 1$, (11) may also be written as

$$\int_0^1 \psi(x, y)\, F(dx) = y \text{ a.e. (F)}, \qquad \int_0^1 \psi(x, y)\, G(dy) = x \text{ a.e. (G)}. \tag{13}$$

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

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

Proposition 3. *Let F and G be two distribution functions on $[0, 1]$. Then there exists a function ψ satisfying (13) 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

$$\int_0^s F(x)\,dx \le sF(s) + tG(t) - F(s) - G(t) + F(s)G(t) + \int_t^1 G(x)\,dx. \qquad (14)$$

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 (14) yields that (11) holds for all s, t if and only if it holds for all s, u that

$$\int_0^s F(x)\,dx \le (s-u)F(s) + \int_0^u G^*(x)\,dx. \qquad (15)$$

If we assume (15), we may let $u = s$ and deduce that $F \preceq G^*$. Conversely, if we assume $F \preceq G^*$, we have

$$\begin{aligned}(s-u)F(s) + \int_0^u G^*(x)\,dx &\ge (s-u)F(s) + \int_0^u F(x)\,dx \qquad (16)\\ &\ge \int_0^s F(x)\,dx\end{aligned}$$

because differentiation w.r.t. u shows that the right-hand side of (16) is at minimum for $u = s$. Thus we have shown (15), as needed. □

The proposition on p. 1793 of Gutmann *et al.* (1991) can now 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 (13) if and only if $F = G^*$.*

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

$$F_{mn}(x) = \frac{1}{m}\sum_1^m \delta_{r_i/n}([0,x]), \quad G_{mn}(y) = \frac{1}{n}\sum_1^n \delta_{c_j/m}([0,y]) \qquad (17)$$

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

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

where $\check{B}$ is what we choose to call the *Rasch transform* defined as

$$\check{B}(x) = \int_0^\infty \frac{xy}{1+xy}\,B(dy).$$

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

$$F(x) = P(\bar{X}_{i\infty} \leq x) = P(\check{B}(\alpha_i) \leq x) = A(\check{B}^{-1}(x)) \tag{18}$$

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 (17) of empirical distributions of the row- and column averages as the summarizing statistic for $\mathcal{P}_{\text{RCES}}$. In analogy with (3) of de Finetti's theorem, we then have that for any $P \in \mathcal{P}_{\text{RCES}}$ this pair converges to a pair (F, G) of distributions satisfying $F \preceq G^*$ and the mixing measure μ_P on $\mathcal{E}_{\text{RCES}}$ 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 $\mathcal{E}_{\text{RCES}}$ can be identified with the set of subconjugate pairs.

So consider a pair (F, G). From (18) 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

$$F(\check{B}(x)) = A(x) \text{ and } G(\check{A}(y)) = B(y) \text{ for all } x \text{ and } y. \tag{19}$$

In the case where (F, G) are empirical distributions of the form (17), (19) is easily seen to be equivalent to the equation system

$$\frac{r_i}{n} = \frac{1}{n}\sum_j \frac{\alpha_i\beta_j}{1+\alpha_i\beta_j} \text{ and } \frac{c_j}{m} = \frac{1}{m}\sum_i \frac{\alpha_i\beta_j}{1+\alpha_i\beta_j}, \tag{20}$$

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 (20) 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. *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. *J. 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), 1–198. Lecture Notes in Mathematics 1117. Berlin: Springer.

Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British J. Math. Statist. Psychology* **26**, 31–44.

Besag, J. and Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* **76**, 633–642.

Dawid, A. P. (1982). Intersubjective statistical models. In *Exchangeability in Probability and Statistics*, (G. Koch and F. Spizzichino, eds.), 217–232. Amsterdam: North-Holland.

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**, 251–299.

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthése* **36**, 271–281.

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

Diaconis, P. and Freedman, D. (1981). On the statistics of vision: the Julesz conjecture. *J. Math. Psych.* **24**, 112–138.

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), 15–38.

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

Freedman, D. (1962). Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Statist.* **33**, 916–923.

Gale, D. (1957). A theorem on flows in networks. *Pacific J. Math.* **7**, 1073–1082.

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

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470–501.

Holst, C. (1995). *Item Response Theory*. Ph.D. Thesis, University of Copenhagen, Denmark.

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

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. *Phil. Trans. Roy. Soc. London B* **294**, 83–94.

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

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

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

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

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models* (2nd edn.). London: Chapman and Hall.

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 A* **58**, 225–42.

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

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

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*, (M. Blegvad, ed.), 58–94. Copenhagen: Munksgaard.

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

Ressel, P. (1988). Integral representations for distributions of symmetric stochastic processes. *Prob. Th. Rel. Fields* **79**, 451–67.

Ressel, P. (1994). Non-homogeneous de Finetti-type theorems. *J. Theor. Prob.* **7**, 469–82.

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

Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, 423–39.

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

DISCUSSION

MICHAEL GOLDSTEIN *(University of Durham, England)*

This paper contains results which are both deep and elegant. Are they important? I am reminded of the following quote from the introduction to the collection “Studies in subjective probability” (1964), in which Kyburg and Smokler write

> In some ways, the most important concept of the subjectivistic theory is that of exchangeable events. Until this notion was introduced by de Finetti in 1931, the subjectivistic theory of probability remained pretty much of a philosophical curiosity. None of those for whom probability theory was a means of livelihood or knowledge paid much attention to it.

Why is exchangeability so important? It will be helpful to have a story to hang this discussion on, so let’s suppose that a new television program is created, called First Kiss. In this program, a group of men and women compete as follows. Each man and each woman kiss exactly once. Each kiss is determined to be either “good” or “bad”. This determination is made by a strictly objective method which, due to space limitations, I will not be able to describe here. Each player is scored by the number of good kisses that they achieve, and the winners go on to compete in further stages of the competition.

Suppose that we want to carry out a statistical analysis of a round of the game. We have a table with rows corresponding to men, columns to women. The $(i, j)^{th}$ entry, X_{ij}, is 1 for a good kiss, and zero for a bad kiss. While there are many ways that we might choose to analyse such a table, a simple and fairly standard approach would be to apply a linear log-odds model, representing the process generating the table as

$$L_{ij} = \log \frac{\mathrm{P(X_{ij} = 1)}}{\mathrm{P(X_{ij} = 0)}} = r_i + c_j \tag{1}$$

where r_i, c_j are row and column constants representing the ability of each contestant. We might then fit (1) to the table of data using our favourite Bayes or likelihood approach, and then carry out some form of diagnostic check for model fit. However, this sidesteps the fundamental question as to why we should entertain a model such as (1), in the first place.

One of de Finetti's fundamental contributions was to show how beliefs about underlying and unobservable parameters could be inferred strictly from beliefs expressed over observable quantities. In our problem, the results of this paper assure us of the following. Suppose that we can view our individual matrix as a sub-matrix of a (hypothetical) infinite matrix (i) with exchangeable rows and exchangeable columns, and (ii) for which row sums and column sums for any sub-matrix are sufficient statistics for that sub-matrix. Then the Rasch representation theorem tells us that our beliefs over our matrix must be exactly as if we believed (i) that each row has a true value r_i and each column has a true value c_j, satisfying (1) for all i, j; (ii) we don't know what values r_i and c_j are, but we believe that the sequence $r_1, r_2, ...$ is iid with probability distribution $\mathrm{P_R}$, and the sequence $c_1, c_2, ...$ is iid with probability distribution $\mathrm{P_C}$; (iii) we don't know what $\mathrm{P_R}, \mathrm{P_C}$ are, but we have a prior distribution P_{RC} over possible choices of $\mathrm{P_R}, \mathrm{P_C}$. Therefore, we see that the Bayesian analysis over (1) is indeed a necessary consequence of certain beliefs over the observables. Further the diagnostic analysis of the model that we might carry out is precisely that which critically scrutinises the generalised constraints on our beliefs which we require in order to apply the Rasch representation.

This is an important and useful result, partly in giving meaning to our analysis and partly in directing us to the diagnostic testing which is appropriate to use of the model. However, there is a further consideration which I believe that we must apply before we can claim that this paper offers genuine insights for the subjectivistic theory. The representation theorem argues that our beliefs must be as if there were true underlying probability distributions generating true underlying parameter values. But what is it about our beliefs over the kisses which compels us to believe in these underlying parameter distributions? The central result of this paper is a deep one, whose proof winds its way through various other deep results from a variety of sources. Therefore, it is difficult to see whether the representation is based on natural, finite considerations, or whether at some point in the development a step has been introduced which only makes sense within an infinite collection and which has no meaningful finite counterpart.

I shall now suggest that the result is indeed a consequence of natural and finite considerations. First, let us recall how de Finetti's representation theorem works for coin tosses. If we judge that coin tosses are exchangeable, then we may consider the outcomes of a large, but finite, collection of tosses. We may imagine filling a bucket with tokens, where the $\mathrm{i^{th}}$ token is marked heads or tails depending on the result of the $\mathrm{i^{th}}$ toss. Suppose that the proportion of heads in the bucket is p. As the tosses are exchangeable, our beliefs, given p, about the outcome of tossing the coin k times is exactly as though we were to make k independent selections of tokens from the bucket without replacement. If the number of tokens in the bucket is large compared to k, then we may view the selections as almost independent, each with probability p for heads. Of course, we do not know what the value of p will be, and therefore we have a prior distribution over this value. Thus, our beliefs will be exactly as described by de Finetti's representation theorem, up to the approximation arising from the finite nature of the bucket. Thus, there is a final book-keeping step of allowing the size of the bucket to tend to infinity, and showing that the limit is smoothly and consistently achieved, but this argument is sufficient to show that the representation is really concerned with our beliefs over large finite collections of tosses.

For the Rasch representation, the argument is more complicated, but similar finite arguments show why the representation holds. To simplify the discussion, suppose that we consider that there are three levels of ability for the men, namely Superb (S), Acceptable (A) and Terrible (T), and similarly for the women. We do not know a priori how many people fall into each category, nor do we know the quantitative differences between the groups and nor do we know which category each individual should fall into.

However, now suppose that we envisage a large, but finite, array of outcomes of the game. The row sums allow us to classify the men into their appropriate groups to an arbitrary level of accuracy, as the array size increases. Similarly, the column sums allow us to classify the women. Therefore, we may consider that we have nine buckets filled with kisses. In each bucket, some kisses are good and some are bad. Let p_{mw} be the proportion of good kisses among group mw, where each ability of the men, m, and of the women, w, is one of S, A, T. Our probability that an individual pair i, j of people have a good kiss, conditional on the row and column sums for the layout, comes from using the i^{th} row sum and the j^{th} column sum to allocate the pair to the appropriate groups m_i and w_j and then, from the row and column exchangeability, viewing the probability that the couple have a good kiss as $p_{m_i w_j}$ independently of all other kisses.

Given the nine values p_{mw}, we now fit the Rasch model

$$P(X_{ij} = 1) = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j}, i, j = S, A, T \tag{2}$$

As it stands, the model is non-identifiable, so we nominate an individual to be the standard against which all others are judged. Suppose that we assign β_S, the score for superb women, to be one. This then fixes the scores for all men as $p_{mS} = (\alpha_m/[1 + \alpha_m]), m = S, A, T$. This now fixes each of the remaining scores for women, for example looking at the groups with $m = S$ gives $p_{Sw} = (\alpha_S \beta_w/[1 + \alpha_S \beta_w]), w = A, T$. (This argument breaks down if any of the p_{mw} values are zero, which is why a separate argument is required in the general statement of the theorem for the non-regular case.) We have now fixed all of the values α_i, β_j and we must check that (2) is indeed satisfied over all subgroups. This follows as all the information that we have used is based on conditioning on row and column sums. Such conditioning preserves row-column summarisability (as, conditional on the row and column sums, all configurations with these row and column sums have the same probability, so that any two sub-matrices with the same row and column sums must have the same conditional probability as each can be embedded in exactly the same number of configurations for the full matrix with the given row and column sums). Therefore, consider, for example, our assessment for $P(X_{TT} = 1)$. This is uniquely determined, as by row-column summarisability, we must assign the same probability to each of the events $[X_{TT} = 1, X_{TS} = 0, X_{ST} = 0, X_{SS} = 1]$ and $[X_{TT} = 0, X_{TS} = 1, X_{ST} = 1, X_{SS} = 0]$. Equivalently, we must assign

$$\frac{\alpha_S}{1 + \alpha_S} \frac{1}{1 + \alpha_S \beta_T} \frac{1}{1 + \alpha_T} p_{TT} = \frac{1}{1 + \alpha_S} \frac{\alpha_S \beta_T}{1 + \alpha_S \beta_T} \frac{\alpha_T}{1 + \alpha_T} (1 - p_{TT})$$

from which $p_{TT} = (\alpha_T \beta_T/[1 + \alpha_T \beta_T])$ as required. We therefore see that the exchangeability construction for Rasch matrices corresponds in this case to a mixture of our uncertainties as to the relative proportions of each of the groups and our beliefs over the magnitudes of the effects within each group, as expressed by our beliefs over the nine probability values p_{mw} constrained by the Rasch relations (2). The book-keeping that is required to produce the general result is far more detailed than for the classic de Finetti representation, as we not only have to let the size of each bucket tend to infinity but we also need to let the number of buckets tend to infinity, reversing the argument that I gave where we started by knowing the number of groups and instead defining the groups through observed similarity of row or column sums.

The technical difficulties in such an explicit construction are considerable. However, the above argument is, I hope, sufficient to suggest that the reason that the Rasch representation, so expertly presented in this paper, does indeed offer powerful, practical insights into the treatment of binary layouts is that it is a genuinely subjectivistic result which is based on intuitive and finite considerations.

REPLY TO THE DISCUSSION

First I would like to thank Michael Goldstein for his positive reaction to this paper. Although mating of salamanders is a potential application of the Rasch model, I admit that the First Kiss program is much more fascinating! The description of the nature and genesis of the Rasch model given by Michael Goldstein is both very illuminating, accurate, and on the point.

Indeed it was appropriate to mention that the non-degenerate Rasch model is nothing but an additive model for the log-odds, a model which is more familiar to statisticians today than it was in 1960, when Rasch introduced it.

It would be very valuable to have a derivation of the random Rasch model from finite considerations, as suggested by the discussant. Diaconis and Freedman (1980) give finite versions of de Finetti's classical theorem, with an explicit bound on the distance in total variation between the distribution of the first k of a sequence of exchangeable variables with a given finite length n, and the closest mixture of Bernouilli distributions. The bound, $4k/n$, originates from approximating the hypergeometric distribution with the binomial. Generalizations of this type of argument has e.g. been made by Diaconis, Eaton and Lauritzen (1992), and the corresponding infinite versions of de Finetti type theorems then usually follow by a simple limit argument.

The problem here is that the bookkeeping associated with deriving such bounds and controlling their asymptotic behaviour in the case of binary matrices is particularly difficult. Whereas there are efficient and well-known asymptotic results for the number $\binom{n}{x}$ of binary sequences of length n with sum x, it seems to be extremely hard to control $N(r, c) = |\mathcal{M}(r, c)|$, the number of binary matrices with row-sums equal to $r = (r_1, \ldots, r_m)$ and column-sums equal to $c = (c_1, \ldots, c_n)$. The combinatorial literature has only sporadic results; see for example O'Neil (1969), Békéssy, Békéssy and Komlós (1972), Bender (1974), Mineev and Pavlov (1976), and McKay (1984, 1985). The structure of the degenerate RCE matrices of Rasch type also indicates that the situation is quite complex.

ADDITIONAL REFERENCES IN THE DISCUSSION

Békéssy, A., Békéssy, P., and Komlós, J. (1972). Asymptotic enumeration of regular matrices. *Studia Scientiarum Mathematicarum Hungaria* **7**, 343–353.

Bender, E. A. (1974). The asymptotic number of non-negative integer matrices with given row and column sums. *Discrete Mathematics* **10**, 217-233.

Diaconis, P., Eaton, M. L. and Lauritzen, S. L. (1992). Finite forms of de Finetti theorems in linear models and multivariate analysis. *Scandinavian J. Statist.* **19**, 289–315.

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. *Ann. Prob.* **8**, 745–764.

McKay, B. (1984). Asymptotics for 0-1 Matrices with prescribed line sums. In *Enumeration and Design* (D. M. Jackson and S. A. Vanstone, eds.). New York: Academic Press, 225–228.

McKay, B. (1985). Asymptotics for 0-1 Matrices with prescribed row sums. *Ars Combinatoria* **19A**, 15–25.

Mineev, M. P. and Pavlov, A. I. (1976). On the number of (0-1) matrices with prescribed sums of rows and columns. *Dokl. Akad. Nauk SSSR* **230**, 1286-1290. Russian original has page numbers 271–274.

O'Neil, P. E. (1969). Asymptotics and random matrices with row-sum and column-sum restrictions. *Bull. Amer. Math. Soc.* **75**, 1276–1282.