Nonparametric unfolding models for dichotomous data

Wendy J. Post and Tom A. B. Snijders

Netherlands Interdisciplinary Demographic Institute and Department of Statistic and Measurement Theory, University of Groningen

Abstract: What are essential requirements, formulated in terms of item response theory, for unidimensional unfolding models for dichotomous data, if one does not wish to make specific assumptions concerning the form of the tracelines and of the population distribution of latent trait values? Tracelines should be unimodal, of course, but this requirement is not sufficient to derive empirically testable consequences. Two basic postulates are formulated concerning population-independent inference about subjects' latent trait values on the basis of observed responses to items. These postulates are proven to be equivalent to total positivity of orders 2 and 3 for the traceline family. Given these postulates, unimodality of the tracelines leads to some empirically testable results. These are formulated as properties of the conditional adjacency matrix and of the correlation matrix.

Key words: Unfolding, item response theory, unimodal response models, total positivity, unidimensional scaling, measurement theory.

1. Introduction

Unfolding models for dichotomous data are regarded in this paper as unidimensional latent trait models, where subjects and stimuli are represented on the same scale, and where the probability of a positive response of a subject to a given stimulus will be higher when the position of the subject on the latent trait is closer to the position of the stimulus. To express this description without using the term "closeness," or "distance," one can say that as subjects occupy more rightist (or leftist) positions on the latent dimension, the items to which they tend to answer positively are more rightist (or leftist, respectively); while all subjects agree about the scale order (from left to right) of the items. The basic issue confronted in this paper is the relation between the latent trait and the observable stochastic indicators of this trait. We shall present a mathematical model (based on unimodal tracelines in the framework of item response theory; see below) that expresses the conceptual idea of unfolding, and specify this model further by requiring certain properties for the inference from the observed indicators to the latent trait. These properties allow the derivation of certain statistical associations between the item responses that can be used to check the appropriateness of a dichotomous unfolding model on the basis of observed dichotomous data.

Author's address: Tom A. B. Snijders, Dept. of Statistics and Measurement Theory, University of Groningen, Grote Kruisstraat 2/1, NL — 9712 TS Groningen, The Netherlands; email T.A.B.SNIJDERS@PPSW.RUG.NL

The set of stimuli that together constitute the scale is supposed to be given. The position of a subject on the latent trait can be identified with his or her (hypothetical) ideal stimulus. This paper treats probabilistic models for dichotomous "pick any" data: each subject is confronted with the same set of stimuli, or items, and the response to each stimulus can be scored as positive (1) or negative (0). For example, subjects can be asked whether they agree or disagree with each of a number of statements. The "pick any" assumption means that the subject responds independently to the various stimuli; the number of positive responses is not fixed in advance.

The founding father of unfolding analysis was Coombs (1950, 1964). His approach was in terms of preference functions of subjects rather than tracelines of stimuli. In this paper, we propose a model in terms of item response theory (IRT) and point items (an early reference about point items in IRT is Torgerson, 1958, p. 312); we shall indicate below that an important subclass of our class of models can also be interpreted as being based on preference functions. Coombs proposed parallelogram analysis, a unidimensional deterministic unfolding model for dichotomous data. In parallelogram analysis, it is assumed that all subjects pick those stimuli that are close to their positions, and no others. This assumption is hardly ever satisfied in empirical data sets, a common deficiency of deterministic models. Probabilistic models are more realistic. Several authors have proposed different versions. Parametric models were presented by Andrich (1988) and Hoijtink (1990). These authors proposed models based on single-peaked (or unimodal) functions involving the positions of each subject and stimulus. Cliff et al. (1988) proposed a nonparametric scaling method, implicitly based on the assumption that there are equal distances between the stimuli positions. Formann (1988) used a latent class approach for modeling non-monotone dichotomous items. Unfolding models have also been proposed for other types of data (rankings, paired comparisons, etc.) and other data collection designs (e.g., pick k/n designs, where the number of choices is fixed at k). There exists a large collection of literature on probabilistic unfolding models for paired comparisons data, initiated by Coombs, Greenberg, and Zinnes (1961), and followed up by, e.g., Bechtel (1968), Sixtl (1973), and Zinnes and Griggs (1974). Bossuyt (1990) studied several probabilistic unfolding theories for paired comparisons data.

A nonparametric method for unidimensional unfolding for dichotomous data was developed by Van Schuur (1984, 1988). The associated computer program is called MUDFOLD, for Multiple UniDimensional unFOLDing (Van Schuur & Wierstra, 1987; Van Schuur & Post, 1990). This method constructs a scale of items which minimizes, in a certain sense, the deviations ("errors") from the deterministic model; the scale is (in first instance) considered acceptable if the number of errors is less than expected under the null model of statistically independent items. Many of the ideas on which MUDFOLD is based were derived from Mokken's nonparametric item-response model for *cumulative* items (Mokken, 1971; Mokken & Lewis, 1982), which may be considered to lie between two "extreme models" defined by the deterministic Guttman scale on one hand and the

null model of independent responses on the other hand. Mokken (1971) formulates his nonparametric probabilistic item-response model by imposing some order requirements on the family of tracelines. In this paper we formulate a nonparametric *unfolding* model for dichotomous data. The aim of this unfolding model is that it should embody the concept of unfolding in an item-response model with a minimal set of assumptions (hence the nonparametric approach), while allowing the derivation of statistical associations between the item responses on which diagnostic checks for the model assumption can be based.

This approach along the lines of item response theory deviates from Coombs' approach which is based on single-peaked preference functions. However, if the preference is a monotone function of the distance between the position of the subject and the position of the stimulus, and if this function is the same for all subjects, models formulated in terms of single-peaked preference functions can be translated into models formulated in terms of unimodal tracelines.

In section 2 some basic requirements of unidimensional unfolding models are formulated in an IRT framework. In addition to the assumptions of local independence and unimodal tracelines, these assumptions include two properties concerning the measurement model. These express a form of population-independent inference with respect to subjects' latent trait values. The first measurement-related assumption expresses that for a randomly drawn subject from any population, a positive answer to a given item i should indicate higher latent trait values accordingly as item i has itself a higher position. The second is analogous but more complicated, and specifically related to unimodality. In section 3, it is proven that these properties can be reformulated in terms of the mathematical properties of total positivity of order 2 or 3, respectively, for the family of tracelines. These requirements make it possible to deduce, in section 4, intuitively appealing properties for the probabilities of joint choices of pairs of items, which hold *irrespective* of the population distribution of the latent trait. These properties can be used for empirically checking the unfolding model. Under some special conditions, the assumptions imply that the choice probabilities of the items are also unimodal, as is demonstrated in section 5. In section 6 an empirical dataset of Andrich (1988) is analyzed, in which the properties of the diagnostic statistics are utilized. The paper ends with a discussion in section 7.

2. A latent trait and measurement model for unfolding

Our unfolding model is along the lines of item response theory. As to the data structure, it is assumed that there are n items, indicated by the numbers 1 through n; these numbers also give the *scale order* of the items. The response of a subject on item i is considered to be a random variable X_i; $X_i = 1$ if the subject responds positively to item i, and 0 otherwise. The vector of responses of a subject is the realization of the random vector $X = (X_1, \ldots, X_n)$. We make the usual assump-

tions of the existence of a latent trait and local stochastic independence, reflecting the "pick any" data collection design and the unidimensionality of the set of stimuli:

A1. There exists a unidimensional latent trait, θ, such that every subject has a position on this latent trait; the probability that a subject with trait value θ responds positively to item i is denoted $p_i(\theta)$. This probability, regarded as a function of θ, is the *traceline* of item i.

A2. The responses of a subject on the various stimuli, given the latent trait value, are independent:

$$P(X = x|\theta = \theta_0) = \prod_{i=1}^{n} p_i(\theta_0)^{x_i} (1-p_i(\theta_0))^{1-x_i}.$$

The further assumptions, A3 to A5, are assumptions regarding the family of tracelines $p_1(\theta)$ to $p_n(\theta)$. For cumulative models, the tracelines would be required to be non-decreasing in θ, see Mokken and Lewis (1982). For unfolding models, in contrast, the tracelines are required to be weakly unimodal. This property is defined as follows; note that if in the definition x equals $-\infty$ ($+\infty$), the function f(.) must be non-increasing (non-decreasing).

Definition. A function f: $X \rightarrow R$, where $X \subset R$, is *weakly unimodal* if there exists an x ε $[-\infty, +\infty]$ such that f(y) is non-decreasing for $y \leq x$ and non-increasing for $y \geq x$.

A3. For every item i, the traceline $p_i(\theta)$ is a weakly unimodal function of θ. Let σ_i be a value where $p_i(\theta)$ is maximal. (It is possible that there is an interval of values where $p_i(\theta)$ is maximal, in which case σ_i is not unique; it is also possible that σ_i is $\pm \infty$.) It is possible to choose the σ_i so that they are non-decreasing as a function of i: $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_n$.

The value σ_i is regarded as the scale position of item i. Since in A3, the tracelines are not required to be strictly unimodal, this value is not necessarily unique. Usually, however, tracelines will be strictly unimodal in the sense that the maximum is assumed in exactly one value σ_i and that $p_i(\theta)$ is increasing left of σ_i and decreasing right of σ_i. When some of the σ_i are $-\infty$ and $+\infty$, the traceline family consists of one or more non-decreasing functions, followed by some unimodal functions, followed again by some non-decreasing functions.

The basic characteristic of unfolding models in an item-response context is unimodality of tracelines. In Coombs' preference function model for unfolding, the subject picks the stimuli closest to his or her position. If the critical distance is the same for all subjects, say, δ, this preference function model can also be formulated

as a traceline model: the tracelines are $p_i(\theta) = f(\theta-\sigma_i)$ where $f(d) = 1$ for $|d| \leq \delta$ and $f(d) = 0$ for $|d| > \delta$. Just as in the Coombsian model *subjects* may have different critical distances within which they find the stimuli acceptable, so *items* may in our IRT model have different widths δ_i; if the additional requirement is made that items are deterministic (i.e., $p_i(\theta)$ is either 0 or 1), this leads to tracelines

(1) $p_i(\theta) = 1$ if $|\theta-\sigma_i| \leq \delta_i$ and 0 otherwise.

For tracelines (1), however, an additional requirement is necessary in order that subjects at all possible values of θ agree on the order (from 1 to n) of the stimuli on the latent dimension: both the left boundaries $\sigma_i-\delta_i$ and the right boundaries $\sigma_i+\delta_i$ should be non-decreasing functions of i. If this does not hold, then there are subject positions θ where the set of items picked by the subject is not a series of adjacent items. It can be concluded that the Coombsian model can be accommodated as a deterministic model in the framework of assumptions A1—A3, except that not all differences in the "width" of a subject's preferences can be modeled. The model with tracelines (1) where $\sigma_i-\delta_i$ as well as $\sigma_i+\delta_i$ are non-decreasing functions of i, will be indicated in this paper as the *deterministic unfolding* model.

Assumption A3 is a statement about the conditional distributions of the X_i, given values of θ. To have testable properties for the unfolding model, however, we need statements about the joint distribution of the observable data $X = (X_1, \ldots, X_n)$ without conditioning on the nonobservable value of θ, but under the assumption that θ has been drawn at random from a population of θ-values. In IRT terms, this is a "marginal" approach. It turns out that assumptions A1 through A3 do not suffice to derive testable properties of the joint distribution of the vector X. In the discussion following the definition of A3 we already saw that additional assumptions are needed to derive some intuitively acceptable properties; at least, in the deterministic version of the model. The additional assumptions proposed below are measurement related and refer to the relation between the *order of stimuli* and the order of subjects.

It is assumed in this paper that subjects are drawn at random from a population of subjects; the cumulative distribution function (cdf) of latent scale values in this population is denoted $G(\theta)$. This subject distribution is arbitrary, it may be continuous or discrete. By $g(\theta)$ shall be denoted the probability density function or, in the discrete case, the probability function corresponding to G. Expectations of functions $f(\theta)$ will be written as integrals $\int f(\theta)g(\theta)d\theta$ (with assumed integration limits $-\infty$ and ∞); for discrete distributions, this can be regarded as a sum, while in the general case it may be regarded as an abstract integral $\int f(\theta)dG(\theta)$. The marginal popularity of item i, i.e., the probability that a random subject from a population with cdf G picks item i, is denoted by

(2) $P_G(i) = \int p_i(\theta)g(\theta)d\theta.$

An important role will be played by the *conditional distribution* of θ, given that the subject has responded positively to a given item. The conditional cdf of θ, given that a subject responds positively to item i, is denoted $G\{. \mid X_i = 1\}$ and given by

(3) $G\{\theta_0 \mid X_i = 1\} = P_G\{\theta \leq \theta_0 \mid X_i = 1\} = \int_{-\infty}^{\theta_0} p_i(\theta)g(\theta)d\theta / P_G(i).$

The corresponding conditional probability density function (for continuous θ) or probability function (for discrete θ) is

(4) $g(\theta \mid X_i = 1) = p_i(\theta)g(\theta)/P_G(i).$

Note that (3) and (4) are undefined if the popularities are zero.

Since we intend to propose a nonparametric model with a minimal set of assumptions concerning the latent distribution in the population, the measurement properties expressed below in A4 and A5 are required to hold for any cdf $G(\theta)$ for the latent trait. This implies that A4 and A5 are assumptions only on the family of tracelines. For a given unordered set of tracelines, there is at most one order in which they satisfy A4 and A5; this order is not specific for a given subject distribution. The motivation for requiring assumptions A4 and A5 is twofold: they are appealing as measurement properties in their own right; and they are regularity conditions that enable us to prove, in Section 4, properties of the joint distribution of $(X_1, \ldots, X_n)$ that can be used as diagnostic checks for the unfolding model defined by A1—A3.

The first basic relation between the given order of the items and the latent trait, expressed formally in assumption A4, is the following: when $i < j$, then the information $\{X_j = 1\}$ should lead to a higher estimate of the subject's latent trait value than the information $\{X_i = 1\}$. This is a desirable property of item response models in a wide variety of situations: e.g., in ability testing, knowing that a subject has correctly answered a difficult item points more towards a high ability value than knowing that the subject has correctly answered an easy item; for attitude items on a political left-right scale, knowing that a subject endorses a rightist item points more towards a rightist scale value than knowing that the subject endorses a more leftist item. This property is formulated here without reference to any particular estimator of the subject's scale value. Instead, we use the conditional distribution of the subject's latent trait value θ given that the item has been answered positively. This conditional distribution is more fundamental than a particular estimator of θ as a function of X_i.

A4. For any probability distribution, $G(\theta)$, of latent trait values, and any value θ_0 on the latent trait, the following holds. When a subject is chosen at random from a population with distribution $G(\theta)$, then the conditional probability

$$P_G\{\theta > \theta_0 \mid X_i = 1\}$$

is a non-decreasing function of i for $P_G(i) > 0$.

A4 can be formulated in terms of stochastic ordering of probability distributions (see, e.g., Lehmann, 1986), which is defined as follows: distribution P is stochastically larger than distribution Q if

$$P\{\theta > \theta_0\} \geq Q\{\theta > \theta_0\} \text{ for every } \theta_0.$$

Thus, A4 states that for any subject distribution, the conditional distribution of the subject's latent trait value θ, given that $X_i = 1$, is a stochastically increasing function of the item number i. The stochastic order of latent trait values was also considered by Rosenbaum (1985); he, however, studied the order between two given subpopulations, whereas we study the order between conditional distributions of latent trait values given $\{X_i = 1\}$, for different items i. Note that we are not comparing the information $\{X_i = 0, X_j = 1\}$ with $\{X_i = 1, X_j = 0\}$, but $\{X_j = 1\}$ with $\{X_i = 1\}$; the former comparison is treated in Rosenbaum (1987), but is less relevant to unfolding theory.

Assumption A4 is a relevant condition for cumulative (monotone) as well as unfolding (unimodal) models. It can be proven that the Guttman deterministic cumulative scaling model as well as the deterministic unfolding model defined above satisfy A4 (hint: for these models the joint distribution of (θ, X_i) has the property $\{X_i = 1$ if and only if $\theta_{l,i} \leq \theta \leq \theta_{r,i}\}$ for certain $\theta_{l,i}$ and $\theta_{r,i}$, where $\theta_{r,i} = \infty$ for Guttman while $\theta_{r,i}$ is usually finite for unfolding).

We shall also need a somewhat more complicated analogue of A4, which is relevant for unfolding models and not for cumulative models. Assumption A4 is concerned with conditional probabilities of events of the form $\{\theta > \theta_0\}$. What should one expect with respect to events of the form $\{\theta_1 \leq \theta \leq \theta_2\}$? Consider, e.g., a scale for a left-right dimension, containing extreme leftist (low i) through centrist to extreme rightist (high i) statements; subjects are asked whether they endorse the statements. A subject is chosen at random from a population with cdf G. The event $\{\theta_1 \leq \theta \leq \theta_2\}$ corresponds to a "central" subject's position on the scale. The conditional probability

(5) $P_G\{\theta_1 \leq \theta \leq \theta_2 | X_i = 1\}$

indicates the probability that the subject occupies such a central position, given that he or she endorses statement i. The following property for (5) seems intuitively appealing: for the very leftist statement i=1 this probability will be low; as i becomes higher, (5) will increase as a function of i up to a certain maximum, assumed for a "centrally scaled" item; as i becomes higher still, which means that the item becomes more and more rightist, (5) will decrease again. In other words, (5) is a weakly unimodal function of the item number i.

Assumption A5 expresses the unimodality of the conditional probability (5).

A5. For any cumulative probability function, $G(\theta)$, of latent trait values, and for all $\theta_1 < \theta_2$, the conditional probability (5) is a weakly unimodal function of i.

Assumptions A4 and A5 can be regarded as regularity conditions concerning the measurement properties of the unfolding model that allow to translate the unimodality of the non-observable tracelines into similar but probabilistic properties for observable quantities, properties holding irrespective of the (equally non-observable) population distribution of the latent trait.

3. Equivalent conditions in terms of total positivity

Conditions A4 and A5 have been formulated in Section 2 as measurement properties. In this section it will be demonstrated that they are equivalent to mathematical properties of the traceline family $\{p_i(\theta)|i = 1, \ldots, n\}$, namely, total positivity of order 2 and 3, respectively. This makes available the mathematical machinery concerning totally positive functions, developed by Karlin (1968) and others. However, the properties of totally positive functions of order 3 used in this paper are deduced from principles in the appendix. With the aid of these properties it will be possible to derive testable properties of the joint distribution of $(X_1, \ldots, X_n)$. The available theory for totally positive functions also provides a number of examples of traceline families satisfying A4 and A5. The present section is an interlude presenting the connection between assumptions A4 and A5 and the theory of total positivity, and presenting the unimodality preserving property which is used in the proofs in Sections 4 and 5. Readers who are not interested in the mathematical proofs may skip sections 3.1 and 3.3.

Karlin (1968) provides an extensive treatment of the theory of total positivity; Marshall and Olkin (1977) give an introduction. Lehmann (1986) presents applications of total positivity of order 2 to one-sided statistical testing problems, and of order 3 to two-sided testing problems, while Brown, Johnstone, and MacGibbon (1981) give a review of statistical applications. The definitions and properties that play a role in this paper are discussed in the following subsection.

3.1 Total positivity

Definition

Let $K(x,y)$ be a function defined for $x \in A$, $y \in B$, where A and B are sets of real numbers. The function K is totally positive of order r (TP$_r$) if, for every natural number $m = 1, \ldots, r$ and for all ordered sets $x_1 < x_2 < \ldots < x_m$ and $y_1 < y_2 < \ldots < y_m$ with $x_i \in A$, $y_j \in B$, we have the inequality

$$(6) \quad \det (K(x_i, y_j))_{\substack{i=1,\ldots,m \\ j=1,\ldots,m}} = \begin{vmatrix} K(x_1, y_1) & K(x_1, y_2) & \ldots & K(x_1, y_m) \\ K(x_2, y_1) & K(x_2, y_2) & \ldots & K(x_2, y_m) \\ \vdots & \vdots & & \vdots \\ K(x_m, y_1) & K(x_m, y_2) & \ldots & K(x_m, y_m) \end{vmatrix} \geq 0$$

The function K is totally positive (TP) if it is TP_r for all positive integers r.

This property is symmetric in x and y. Note that total positivity of order r (also abbreviated to TP_r) implies non-negativity of K (take $m = 1$). For $m=2$, one obtains

$$(7) \quad \det (K(x_i, y_j))_{\substack{i=1,2 \\ j=1,2}} = K(x_1, y_1)K(x_2, y_2) - K(x_1, y_2)K(x_2, y_1) \geq 0.$$

This shows that TP_2 can be reformulated as follows: $K(x,y) \geq 0$, and for all $x_1 < x_2$, the ratio $K(x_2,y)/K(x_1,y)$ is a non-decreasing function of y. For a family of probability density functions $f(x;\mu)$ parametered by a real parameter μ, the TP_2 property (where μ plays the role of y) is known as monotone likelihood ratio (MLR) (e.g., Lehmann, 1986). The TP_2 property can be applied to traceline families, where θ and i will then play the role of x and y: the traceline family $\{p_i(\theta)|i = 1, \ldots, n\}$ is TP_2 if and only if for all $j > i$ the ratio $p_j(\theta)/p_i(\theta)$ is a nondecreasing function of θ. We shall call TP_2 of tracelines monotone traceline ratio (MTR).

A basic property which follows immediately from the definition, and which is used in some proofs, is the following:

Property 1.

If $K(x,y)$ is TP_r, and g(x) and h(y) are non-negative functions, then the function $g(x)h(y)K(x,y)$ is TP_r.

According to the so-called basic composition formula (Karlin, 1968, p. 17) we have the following property for integrals (and, for discrete variables z, for sums):

Property 2.

Suppose that $K_1(x,z)$ and $K_2(z,y)$ both are TP_r functions for $x \in A$, $y \in B$, $z \in C$. Then the integral

$$K(x,y) = \int_C K_1(x,z)K_2(z,y)dz$$

is also TP_r.

We only use this property for $r=2$. For this case, a proof is given in the appendix.

3.2 Examples

It will be demonstrated in Section 3.3 that the TP_3 property is appealing in principle for unfolding models, but are there any examples of traceline families that satisfy it? In this section, we discuss some examples. Checking TP_3 usually is not too hard, but there seem to be no straightforward rules for checking the TP_3 property. For the examples of TP_3 traceline families given below, we rely on Karlin (1968), who provides a host of examples, using various methods for proving total positivity. These methods are not specific for TP_3 but prove total positivity of all orders (TP).

Distance models

An important class of families of tracelines is the class of location families, where

$$p_i(\theta) = p(\theta - \sigma_i)$$

for a certain function $p(\theta)$ and item parameters σ_i. For location families all tracelines have the same form, but they are shifted with respect to one another. If the function $p(\theta)$ is even, i.e., $p(\theta) = p(-\theta)$ for all θ, then the IRT model is a distance model, for probabilities are functions of the distance $|\theta - \sigma_i|$. Distance models are of special interest, because in most parametric unfolding models in the literature it is assumed that the preference is a monotone non-increasing function of the distance between subject and stimuli.

For location families, TP_2 or MTR is equivalent to log-concavity of the tracelines (Karlin, 1968, p. 32; Lehmann, 1986, p. 509), i.e., nonpositivity of the second order derivative of $\log(p(\theta))$ if $p(\theta)$ is twice differentiable. A log-concave function is necessarily monotone or unimodal, so that distance models with the MTR property must be either cumulative models or unfolding models in the sense of Assumption A3.

Some examples of symmetric location families which are TP are the following three (these tracelines are not probability densities, since they are normed to have a maximum equal to 1 instead of to integrate to 1; nevertheless, we indicate them by the name of the corresponding probability densities).
The normal traceline (Karlin, 1968, p. 19):

$$p(\theta) = \exp(-\theta^2).$$

The logistic density traceline (Karlin, 1968, p. 19):

$$p(\theta) = \frac{4\exp(-\theta)}{(1 + \exp(-\theta))^2}.$$

Molenaar and Münnich (1993) used the hyperbolic secant traceline (Karlin, 1968, p. 349; a linear transformation of the argument has been applied to obtain a simpler formula):

$$p(\theta) = \frac{2}{\exp(\theta) + \exp(-\theta)} .$$

In contrast to the mentioned families, the PARELLA traceline proposed by Hoijtink (1990), which is the location family with the Cauchy form

$$p(\theta) = \frac{1}{1 + \theta^2} ,$$

is not TP_3; this function is not log-concave, so it cannot even have MTR. (Hoijtink also considers this function with θ^2 raised to a positive power g; this function also is not log-concave). An example of the counter-intuitive consequences is given in Section 3.3.

The tracelines of the deterministic unfolding model mentioned in Section 2 are not TP_3; see Post (1992).

Several TP families of functions that are not location families can be found in Karlin (1968). For example, the family of gamma densities with a fixed scale parameter is TP in the shape parameter, and the family of non-central t-densities with a fixed number of degrees of freedom is TP in the non-centrality parameter. It should be noted that from a given family of TP_3 densities, one can construct other families by using property 1 and also by monotone transformations of the latent parameter θ and of the probabilities.

3.3 Total positivity of traceline families

The importance of total positivity in our context is based on the so-called variation-diminishing property of totally positive functions, which is crucial in the mathematical theory of total positivity and discussed in Karlin (1968, pp. 21 & 233). For total positivity of order 3, this property amounts to a unimodality-preserving property which is equivalent to assumption A5. This equivalence is expressed in the next theorem. The theorem refers for a latent trait cdf $G(\theta)$ with density function $g(\theta)$ and an arbitrary real-valued function $f(\theta)$, to the conditional expectation

$$E_G\{f(\theta)|X_i = 1\} = \frac{\int f(\theta)p_i(\theta)g(\theta)d\theta}{\int p_i(\theta)g(\theta)d\theta}$$

as a function of i.

Theorem 1.

Assumption A5 is equivalent to the following property: For every cdf $G(\theta)$ and every weakly unimodal function $f(\theta)$ for which this conditional expectation exists, $E_G\{f(\theta)|X_i=1\}$ is a weakly unimodal function of i.

The proof is given in the appendix. Theorem 1 is crucial for the derivation of testable properties in Section 4.

We now formulate the theorem on the equivalence of assumption A4 to monotone traceline ratio.

Theorem 2.

Assume A1 and A2. The following properties are equivalent: a) Assumption A4; b) the traceline family $p_i(\theta)$ has MTR; and c) for all $G(\theta)$, the conditional density functions $g(\theta|x_i=1)$ have MLR as a function of i, for those i for which the popularities $P_G(i)$ defined in (2) are positive.

Note that A4 is concerned with stochastic ordering, which is by itself a weaker property than MLR (Lehmann, 1986, p. 86). This might seem to imply that MTR (the same as MLR but in another context) is stronger than A4. However, the equivalence between (a) and (b) in Theorem 2 is valid because of the additional specification in A4 that the stochastic ordering holds for all latent trait distributions G. The proof of Theorem 2 is given in the appendix.

Theorem 3 gives a similar equivalence as Theorem 2, now for TP_3 traceline families. This theorem also is proven in the appendix.

Theorem 3.

(a) If the traceline family $\{p_i(\theta)|i = 1, \ldots, n\}$ is TP_3, then it satisfies assumptions A4 and A5. (b) Conversely, if the traceline family satisfies A4 and A5 while $p_i(\theta) > 0$ for all i and all θ, then the traceline family is TP_3.

The argumentation in Section 2 for assumptions A4 and A5 implies that, because of these theorems, it is desirable for traceline families in unfolding models to satisfy TP_3.

We noted in section 3.2 that the Cauchy traceline family based on the function $p(\theta) = 1/(1+\theta^2)$ does not have MTR. According to Theorem 2, MTR is equivalent to A4, which requires the stochastic ordering, for every subject distribution, of the conditional subject distributions $G\{\theta_0|X_i=1\}$ for $i = 1, \ldots, n$. So for the Cauchy traceline there must exist subject distributions, discrete as well as continuous, for which this stochastic ordering does not hold. We shall give a numerical example. Consider the two tracelines

$$p_1(\theta) = 1/\{1+\theta^2\}, \; p_2(\theta) = 1/\{1+(\theta-1)^2\}.$$

Let $G(\theta)$ be a two point distribution with probabilities $g(\theta_1)=g(\theta_2)=1/2$ for the values $\theta_1=1.5$, $\theta_2=3$. Then the conditional probabilities are

$$g(\theta_1|X_1=1) = \frac{g(\theta_1)p_1(\theta_1)}{g(\theta_1)p_1(\theta_1)+g(\theta_2)p_1(\theta_2)} = 40/53$$

and $g(\theta_2|X_1=1) = 13/53$. Similarly, $g(\theta_1|X_2=1) = 4/5$ and $g(\theta_2|X_2=1) = 1/5$. Of the two possible values θ_1 and θ_2, the smaller one, θ_1, has a higher probability conditional on $X_2=1$ than on $X_1=1$. So $G(\theta|X_1=1)$ is stochastically larger, instead of smaller, than $G(\theta|X_2=1)$. This demonstrates the counter-intuitive property of the Cauchy traceline family as an example of a non-MTR family: knowing that the subject has chosen a rightist stimulus does not necessarily imply that we have more ground to believe he has a rightist latent trait value than knowing that he has chosen a leftist stimulus.

4. Diagnostic statistics: The conditional adjacency matrix

How should one check in empirical data whether the items form an unfolding scale? We focus on one intuitively appealing property and refer the reader to Van Schuur (1984, 1988) and Post (1992) for checks based on other principles, notably those based on the number of error patterns, which are defined as response triples $(X_i, X_j, X_k) = (1, 0, 1)$ where $i < j < k$. To check the unfolding model, one would like to check unimodality of tracelines; but these are not directly observable, as one cannot directly observe the subjects' latent trait values. A way to get around this difficulty is the following. The subsample of subjects choosing item j will have, on the average, smaller θ values than the (partially overlapping) subsample of subjects choosing item $j+1$; this is the verbal expression of the stochastic ordering of the distributions $G(\theta|X_j=1)$ and $G(\theta|X_{j+1}=1)$. Consider an item i of which the unimodality of the traceline is to be checked. The probability of choosing item i in the subsample of subjects choosing item j, denoted $P(i|j)$, is an average of the traceline $p_i(\theta)$ with respect to this conditional distribution $G(\theta|X_j=1)$. Since these conditional distributions are stochastically increasing as a function of j it seems likely that, if the traceline $p_i(\theta)$ is unimodal, the conditional probability $P(i|j)$ is a unimodal function of j for $j \neq i$. In this section, it will be proven that this result is correct provided that the additional requirement of TP_3 tracelines is made.

We assume that a random sample of subjects from a population is available, and by X_{ih} we denote the indicator variable for the choice of item i by subject h similarly as in Section 2. The marginal probabilities of joint choices of pairs of items are

$$(8) \qquad P(i,j) = P\{X_{ih} = X_{jh} = 1\} = \int p_i(\theta)p_j(\theta)g(\theta)d\theta, \quad (i \neq j)$$

where $g(\theta)$ is the probability density of the subject distribution. (The fact that these probabilities depend on g is left implicit in the notation.) The number of subjects in the sample is denoted by N; the number of subjects in the sample choosing both item i and item j is denoted by $N(i,j)$. The joint choice probabilities are estimated by

$$\hat{P}(i,j) = \frac{N(i,j)}{N} = \frac{1}{N} \sum_{j=1}^{N} X_{ih} X_{jh}. \qquad\qquad (i \neq j)$$

The matrix $\hat{P}(i,j)$, called the adjacency matrix, is used as a diagnostic statistic in the MUDFOLD method; see Van Schuur (1984). As a diagnostic statistic it is based on the deterministic Coombsian unfolding model. In this model subjects choose those stimuli, and only those, which are close to their ideal point. This implies that every subject choosing two given items also has to choose all items in between, so that $P(i,j) \geq P(i,k)$ for all $i < j < k$. For an adjacency matrix of a perfect and correctly ordered Coombsian unfolding scale, the highest values therefore are on the diagonal, and the entries are monotone decreasing both in the rows and in the columns as one moves away from the diagonal.

This unimodality pattern does not always hold for the population value of the adjacency matrix under the probabilistic unfolding model defined by assumptions A1—A3 or A1—A5. To illustrate this, consider tracelines of stimuli i, j, k with item positions $\sigma_i \leq \sigma_j \leq \sigma_k$, as shown in Figure 1. (Whether or not these tracelines satisfy A4 and A5 is immaterial.)
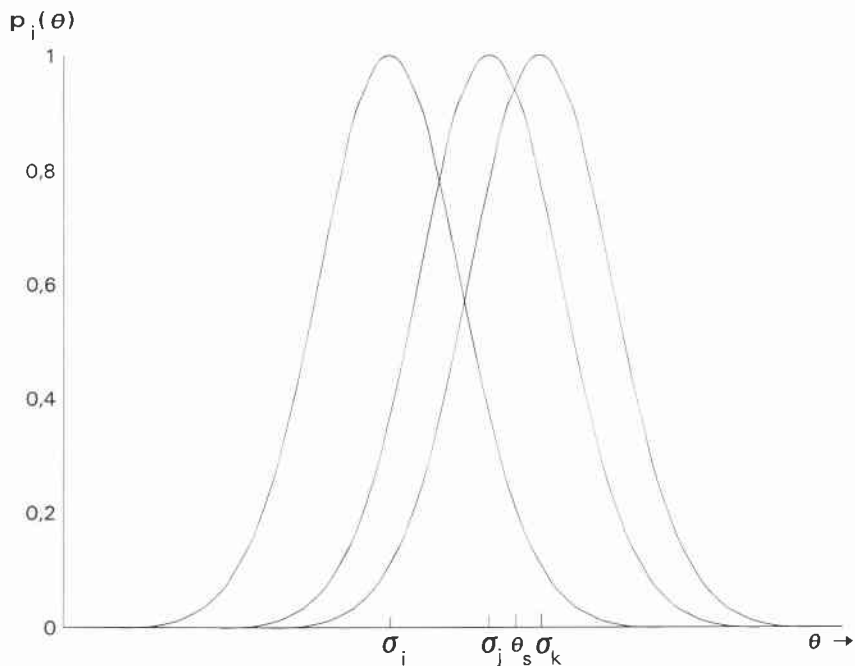

Figure 1: Three tracelines satisfying assumptions A1, A2, and A3.

The monotonicity property $P(i,j) \geq P(i,k)$ is equivalent to

$$\int p_i(\theta)g(\theta)\{p_j(\theta)-p_k(\theta)\}d\theta \geq 0.$$

Define θ_s as the value for which $p_j(\theta_s) = p_k(\theta_s)$. For $\theta > \theta_s$ the term between braces is negative. For any subject distribution concentrated on values larger than θ_s, i.e., satisfying $P\{\theta \leq \theta_s\}=0$, this implies

$$\int p_i(\theta)g(\theta)\{p_j(\theta)-p_k(\theta)\}d\theta < 0,$$

contradicting the proposed monotonicity.

However, it follows directly from properties 1 and 2 of Section 3.1 that if assumptions A1, A2, and A4 are satisfied, the expected values $P(i,j)$ of the adjacency matrix do satisfy the TP$_2$ property. Further it appears that for another matrix closely related to the adjacency matrix, another unimodality property can be derived in accordance with the intuitive reasoning given at the beginning of this section. To introduce this matrix, we note that the conditional probability of choosing item i, given that item j ($i \neq j$) is chosen, is given by

$$P(i|j) = P\{X_{ih} = 1 \mid X_{jh} = 1\} = P(i,j)/P(j) = \int p_i(\theta)g(\theta|X_j=1)d\theta.$$

This conditional probability can of course be estimated by

$$N(i|j) = N(i,j)/N(j) \qquad (i \neq j)$$

The matrix with entries $N(i|j)$ will be called the conditional adjacency matrix. The following theorem demonstrates a certain unimodality pattern of the rows of its population version $P(i|j)$. Note that the diagonal values of the conditional adjacency matrix are undefined, because independent replications of a single item for a single subject are usually not observed. This leads to a complication in the formulation of the theorem.

In the theorem, column $j^*(i)$ is defined as the column where the entries in row i assume their maximum:

$$P(i|j^*(i)) = \max_{j;j \neq i} P(i|j).$$

If the maximum is not unique, then the smallest $j \neq i$ for which the maximum is assumed may be chosen for $j^*(i)$.

Theorem 4. $P(i|j)$

Assume that assumptions A1 to A5 hold. Then (a) $P(i|j)$ is a weakly unimodal function of j for every fixed i (where $j \neq i$); (b) $j^*(i)$ is a non-decreasing function of i, except possibly for inversions at the diagonal, i.e., pairs of adjacent rows i, $i+1$ for which $j^*(i) = i+1$ while $j^*(i+1) = i$.

Proof: (a) $P(i|j)$ is the conditional expectation of the weakly unimodal function $p_i(\theta)$ under the condition $X_j=1$. Since A5 holds, applying Theorem 1 to $f(\theta)$

$= p_i(\theta)$ yields that for fixed i, $P(i|j)$ is weakly unimodal as a function of j. (b) Define $P(i,i)$ as the integral in (8) for $i = j$. The traceline family satisfies A4 and hence (by Theorem 2) is TP_2; it follows from Property 2 that the matrix $P(i,j)$ is TP_2. Property 1 implies then that the conditional adjacency matrix is also TP_2. Define $j^+(i)$ as the column where $P(i|j)$ is maximal among all j, including $i = j$. We have

$$P(i+1|j) = \frac{P(i+1|j)}{P(i|j)} P(i|j).$$

The ratio $P(i+1,j)/P(i,j)$ is, due to the TP_2 property, a nondecreasing function of j. Part (a) implies that for row i, $P(i|j)$ is nondecreasing in j for $1 \le j \le j^+(i)$. Hence $P(i+1|j) = P(i,j)[P(i+1,j)/P(i,j)]$ is also nondecreasing for $1 \le j \le j^+(i)$, so that $j^+(i+1) \ge j^+(i)$. This implies that $j^*(i+1) \ge j^*(i)$, except possibly for inversions at the diagonal which may occur if $j^+(i+1) = i+1$ or $j^+(i) = i$. QED.

The order pattern for $P(i|j)$ is observed, except for chance fluctuations, in the conditional adjacency matrix. Therefore, one can check the unfolding model by checking whether the conditional adjacency matrix does not deviate too strongly from the pattern indicated in Theorem 4. The question as to the statistical significance of an observed deviation will be treated in another paper.

In addition to the conditional adjacency matrix, one could study the matrix with entries

$$(9) \qquad N(i|j,j+1) = \frac{N(i,j,j+1)}{N(j,j+1)} \qquad j \ne i-1, i$$

where $N(i,j,j+1)$ is the number of people simultaneously choosing items i, j, and $j+1$. This is relevant only if the sample size is large enough that the numerator and denominator of (9), divided by N, are stable estimators of the corresponding probabilities. Similarly as above, it can be proven that the conditional subject distributions, given that $X_j = X_{j+1} = 1$, are stochastically increasing as a function of j. Since these conditional distributions will be sharper (have smaller variance) than the conditional distributions given that $X_j = 1$, matrix (9) will permit a more sensitive examination of the unimodality of the tracelines than the conditional adjacency matrix provided that the sample size is large enough.

Last but not least something can be said about the patterns of the correlation matrix of the items. In the cumulative Mokken model, all correlations are necessarily nonnegative. For the unfolding model, however, negative correlations are possible. If two items are close together on the latent trait, subjects tend to either choose them both or not choose them both, resulting in a positive correlation. If two items i and j are widely separated in the sense that $p_i(\theta)$ is low where $p_j(\theta)$ is high and vice versa, then subjects will tend to respond positively either to item i or to item j but not to both, resulting in a negative correlation. Davison (1977) derived for a metric unidimensional unfolding model that the correlation matrix exhibits a simplex pattern (Guttman, 1954), i.e., the correlations are non-decreas-

　　　　　　　Wendy J. Post and Tom A. B. Snijders

ing from the first column towards the diagonal and non-increasing from the diago-
nal towards the last column. A weaker property than the simplex pattern is derived
for our model; this is formulated in the following corollary.

Corollary.

If A1 to A5 hold, each row and column of the correlation matrix has at most two
sign changes. If the number of sign changes in a row or column is 2, then the first
sign pattern is negative. In other words, the correlation matrix has the sign struc-
ture of Figure 2.

$$
\begin{pmatrix}
+ & + & + & - & - & - \\
+ & + & + & + & - & - \\
+ & + & + & + & + & - \\
+ & + & + & + & + & + \\
- & + & + & + & + & + \\
- & - & + & + & + & + \\
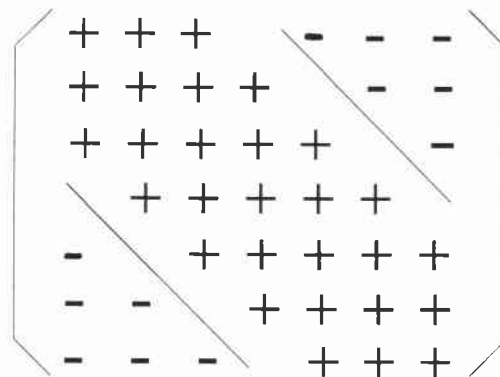- & - & - & + & + & +
\end{pmatrix}
$$

Figure 2: The sign pattern of the correlation matrix if assumptions
A1 to A5 hold.

Proof: We prove the equivalent assertion for the covariance matrix. Define $P(i,i)$
and $P(i|i)$ as the appropriate integrals, corresponding to independent replications
of the single item i. Consider any fixed number c and any fixed item i. Since $P(i|j)$
is a unimodal function of j, $P(i|j) - c$ has at most two sign changes (defined as
in the proof of Theorem 1 in the Appendix, which contains a more elaborate dis-
cussion of the relation between sign changes and unimodality); and if it has two,
then the unimodality implies that the sign pattern is $- + -$. Therefore, also
$P(i,j) - cP(j) = P(j)\{P(i|j) - c\}$ has at most two sign changes, with sign pattern
$- + -$ if the number of changes is two. For $c = P(i)$ and $i \neq j$ the covariances
are obtained. For $i = j$, the values are positive $P(i,i) - P^2(i)$; these are not equal
to the variances but they do have the same sign, so the sign pattern of the matrix
$P(i,j) - P(i)P(j)$ is the same as that of the covariance matrix. This shows that the
covariance matrix has at most two sign changes per row and therefore also per
column. QED.

Summarizing this section, the following diagnostic properties have been pro-
posed: The TP_2 property of the (conditional) adjacency matrix can be used to
check model assumptions A1, A2, and A4. The unimodality per row of the con-

ditional adjacency matrix with the non-decreasing positions of the maxima, as well as the sign structure of the correlation matrix, can be used to check the entire model (assumptions A1 to A5). Of course, these properties are deduced from the nonparametric unfolding model: they are necessary but not sufficient.

5. Unimodal popularities

A question of some interest, albeit of secondary importance, is whether the popularities $P(i)$ are unimodal, i.e.,

$$P(1) \leq P(2) \leq \ldots \leq P(i^*) \geq P(i^*+1) \geq \ldots \geq P(n), \text{ for some } i^*.$$

In our unfolding model, it is not necessary that the popularities are unimodal: It is easy to construct examples of bimodal popularities in a model satisfying A1 to A5 by using a bimodal subject density function $g(\theta)$. However, under the additional requirements that the tracelines are a location family and the subject density is unimodal, the popularities must indeed be unimodal, as is expressed in the following theorem.

Theorem 5.
Suppose that A1—A5 are satisfied, and that the subject distribution has a unimodal probability density $g_u(\theta)$. Then the popularities $P(i)$ are a weakly unimodal function of i.

Proof: Schoenberg (1951) showed that if the unimodal function $p(\theta)$ defines a TP_3 location family (this function is therefore called a Polya frequency function of order 3), then it has a finite integral. Let $p(\theta-\sigma_i)$ be the traceline of item i; then the assumptions of the theorem and Schoenberg's result imply that

$$\int p(\theta-\sigma_i)d\theta < \infty \text{ for every i.}$$

This integral does not depend on i because the tracelines constitute a location family. Applying Theorem 1 in its slightly stronger form (see the Appendix) with $g(\theta) = 1$ and $f(\theta) = g_u(\theta)$ and Theorem 3 immediately yields the weak unimodality of the popularities.

6. An example

A dataset of Andrich (1988) concerning statements about the attitude towards capital punishment is used to illustrate the utilization of the conditional adjacency matrix and correlation matrix for model checking. Graduate students, taking an introductory course in educational measurement and statistics, were asked to respond to statements by agreeing or disagreeing with them. Eight statements were

answered by 54 students. In Table 1 the statements are presented with their popularities.

Table 1: Statements about capital punishment and their popularity

	Statement	Popularity
A.	Capital punishment is one of the most hideous practices of our time.	0.44
B.	The state cannot teach the sacredness of human life by destroying it.	0.65
C.	Capital punishment is not an effective deterrent to crime.	0.67
D.	I don't believe in capital punishment but I am not sure it isn't necessary.	0.46
E.	I think capital punishment is necessary but I wish it were not.	0.48
F.	Until we find a more civilized way to prevent crime we must have capital punishment.	0.44
G.	Capital punishment is justified because it does act as a deterrent to crime.	0.35
H.	Capital punishment gives the criminal what he deserves.	0.35

In the MUDFOLD program an ordering of stimuli can be supplied by the user for which several statistics are then calculated. The conditional adjacency and correlation matrices are incorporated in its newest version (Van Schuur & Post, 1990). Both statistics are also incorporated in the newest version of PARELLA, the program for parametric probabilistic unfolding for dichotomous data based on Cauchy tracelines (see Hoijtink, Molenaar, & Post, 1992). The order of the items as presented in Table 1 was obtained by Andrich. This scale can be interpreted as ranging from 'strongly against' to 'strongly in favor of' capital punishment. The associated conditional adjacency matrix and correlation matrix are presented in Table 2 and Table 3. The row maxima in the conditional adjacency matrix are printed in bold. According to Theorem 3 the conditional adjacency matrix should have two properties, namely, each row should exhibit a unimodal pattern and the maxima of the rows should be situated in positions that move to the right as one moves downwards in the matrix, except possibly for inversions around the diagonal.

Inspecting the unimodality condition of each row, we see some small disturbances. For instance, in row B we have to exchange columns F and G to obtain a unimodal pattern. In row H the columns A, B, and C cause disturbances of the unimodal pattern. These disturbances, however, are small and do not occur structurally in the same columns. The largest disturbance occurs in row F in columns G and H. We could say that if there are inversions, they occur either within the first three columns or within the last two columns. Checking the condition for the maxima shows that, this condition is satisfied and there are indeed some

inversions around the diagonal. For instance row C has a maximum observed in cell (C,B), while it is plausible that the maximum occurs on the non-observable diagonal element. The same argument holds for the maximum of row H.

The sign pattern of the correlation matrix shows no disturbances. Each row (column) has exactly one sign change. Note that the simplex pattern does not hold.

Table 2: The conditional adjacency matrix for the order of stimuli obtained by Andrich.

	A	B	C	D	E	F	G	H
A	—	0.57	**0.58**[1]	0.48	0.23	0.21	0.16	0.16
B	0.83	—	**0.86**	0.68	0.35	0.29	0.32	0.11
C	0.87	**0.89**		0.76	0.38	0.33	0.16	0.21
D	0.50	0.49	**0.53**	—	0.46	0.42	0.42	0.42
E	0.25	0.26	0.28	0.48	—	**0.96**	0.84	0.89
F	0.21	0.20	0.22	0.40	0.88	—	0.79	**0.89**
G	0.12	0.17	0.08	0.32	0.62	0.62	—	**0.74**
H	0.12	0.06	0.11	0.32	0.65	0.71	**0.74**	—

[1] The bold elements are the observed maxima of the rows.

Table 3: The correlation matrix for the ordering obtained by Andrich.

	A	B	C	D	E	F	G	H
A	1	0.35	0.40	0.07	−0.41	−0.43	−0.42	−0.42
B	0.35	1	0.63	0.06	−0.61	−0.67	−0.51	−0.84
C	0.40	0.63	1	0.18	−0.58	−0.63	−0.80	−0.71
D	0.07	0.06	0.18	1	0.00	−0.08	−0.06	−0.06
E	−0.41	−0.61	−0.58	0.00	1	0.85	0.53	0.61
F	−0.43	−0.67	−0.63	−0.08	0.85	1	0.51	0.67
G	−0.42	−0.51	−0.80	−0.06	0.53	0.51	1	0.59
H	−0.42	−0.84	−0.71	−0.06	0.61	0.67	0.59	1

The small disturbances in the conditional adjacency matrix could indicate some doubt about the order of the first three items and also about the order of the last four items. This is emphasized by the results of the search procedure incorporated in MUDFOLD, which finds another unfolding scale, in the order BCADEFHG. We see that the first three items are permuted and also the last two items. The criteria of this procedure have been documented in Van Schuur (1984) and Van Schuur and Wierstra (1987). The order obtained by MUDFOLD can also be interpreted as ranging from 'against' to 'in favor of' capital punishment. The sign pattern of the correlation matrix is the same for both scales, but there are some changes in the conditional adjacency matrix as can be seen in Table 4.

Again small disturbances are apparent in the rows of the matrix, but their number is smaller. The maxima conform to the requirements: with increasing row number, they move from left to right except for some inversions at the diagonal.

Wendy J. Post and Tom A. B. Snijders

Table 4: The conditional adjacency matrix for the order obtained by MUDFOLD.

	B	C	A	D	E	F	H	G
B	—	**0.86**[1]	0.83	0.68	0.35	0.29	0.11	0.32
C	**0.89**	—	0.87	0.76	0.38	0.33	0.21	0.16
A	0.57	**0.58**	—	0.48	0.23	0.21	0.16	0.16
D	0.49	**0.53**	0.50	—	0.46	0.42	0.42	0.42
E	0.26	0.28	0.25	0.48	—	**0.96**	0.89	0.84
F	0.20	0.22	0.21	0.40	0.88	—	**0.89**	0.79
H	0.06	0.11	0.12	0.32	0.65	0.71	—	**0.74**
G	0.17	0.08	0.12	0.32	0.62	0.62	**0.74**	—

[1] The bold elements are the observed maxima of the rows.

The differences of the statistics for the two orders, however, are too small to conclude which order is best. One should only conclude that the items A, B, and C are close together on the scale as are items G and H, because permutations do not provide evidently better results. This is supported by the rather large conceptual equivalence of the relevant items.

7. Discussion

In this paper nonparametric conditions for unfolding models for dichotomous pick any/n data were introduced. These conditions were formulated in terms of item response theory with unimodal tracelines. The purpose here was to construct an unfolding model with a minimal set of assumptions, where the unfolding nature of the scale and the order of the items do not depend on the population of subjects. Unimodality of the tracelines proved not sufficient to derive empirically testable consequences. Special assumptions for the family of tracelines are required, namely, total positivity of orders 2 and 3. These rather technical assumptions were proven to be equivalent to sensible properties concerning population-independent inference about latent trait values; these properties were expressed in assumptions A4 and A5.

From a measurement point of view, assumptions A4 and A5 are desirable properties for an unfolding model. These assumptions led us to the conditional adjacency matrix as a diagnostic tool. This matrix is intuitively appealing, since its rows may be regarded as observable proxies for the tracelines of the items: the expected value for the element of the conditional adjacency matrix,

$$P(i|j) = \int p_i(\theta) g(\theta | X_j = 1) d\theta$$

is an average of the traceline $p_i(\theta)$ with respect to the conditional distribution of the latent trait, given a positive answer to item j; as j increases, this conditional distribution increases stochastically. We could say that when proceeding from left

to right through the row of the conditional adjacency matrix, the traceline appears through a blurred (by averaging) and error-prone (by sampling errors) window, where the window also moves from left to right (stochastic increasingness). Properties of the conditional adjacency matrix and correlation matrix under the non-parametric unfolding model were derived, which permit their use as diagnostic statistics.

However, the TP_3 property poses certain restrictions on the form of the tracelines. This is illustrated in that for location families, TP_2 tracelines are log-concave, and therefore tend to zero exponentially fast for low and for high values of the latent trait. For empirical applications, this seems rather restrictive. The Cauchy tracelines of Hoijtink (1990), although not being TP_2, have the advantage of tending to zero more slowly when $|\theta|$ gets large. More research will be needed to receive a more definite answer to the empirical applicability of TP_3 traceline families in attitudinal scaling.

Another problem, but of a less serious nature, is the statistical testing of the properties of the diagnostics. If the unimodality patterns in the rows of the conditional adjacency matrix are disturbed, or if the maxima do not have the required pattern, one may ask whether these deviations are significant or whether they are due to sample fluctuations. This problem leads to a rather complex statistical testing problem under order restrictions. A more detailed discussion about this problem will be given in another paper.

8. Appendix: Proofs of properties in Section 3

Proof of Property 2 for $r=2$.

The definition of $K(x,y)$ in terms of K_1 and K_2 implies that

$$K(x_1,y_1)K(x_2,y_2) - K(x_1,y_2)K(x_2,y_1) =$$
$$= \iint K_1(x_1,z_1)K_2(z_1,y_1)K_1(x_2,z_2)K_2(z_2,y_2)dz_1dz_2$$
$$- \iint_{z_1<z_2} K_1(x_1,z_1)K_2(z_1,y_2)K_1(x_2,z_2)K_2(z_2,y_1)dz_1dz_2$$
$$- \iint_{z_1\leq z_2} K_1(x_1,z_2)K_2(z_2,y_2)K_1(x_2,z_1)K_2(z_1,y_1)dz_1dz_2 =$$
$$= \iint_{z_1<z_2} K_1(x_1,z_1)K_1(x_2,z_2)\{K_2(z_1,y_1)K_2(z_2,y_2) - K_2(z_1,y_2)K_2(z_2,y_1)\}dz_1dz_2$$
$$+ \iint_{z_1\leq z_2} K_2(z_1,y_1)K_2(z_2,y_2)\{K_1(x_1,z_1)K_1(x_2,z_2) - K_1(x_1,z_2)K_1(x_2,z_1)\}dz_1dz_2.$$

In both terms of the last expression, the integrands are non-negative because of the TP_2 property of K_1 and K_2. This establishes TP_2 for the function K. QED.

Theorem 1 will be proven in a slightly stronger form, indicated as Theorem 1[*]; this is necessary for the proof of Theorem 5, and Theorem 1[*] is not harder to prove than Theorem 1. The difference is that in Theorem 1[*] the function $g(\theta)$ does not need to be integrable (recall that $g(\theta)$ is a probability density function in the

main part of the paper). The function $F_g(\theta_1,\theta_2,i)$ in the formulation below is to be identified with $P_G\{\theta_1 \leq \theta \leq \theta_2 | X_i = 1\}$ in Theorem 1, and $H_g(f,i)$ with $E_G\{f(\theta)|X_i = 1\}$.

Theorem 1[*].

Let $\{p_i(\theta)|i = 1, \ldots, n\}$ be a family of non-negative functions. For $\theta_1 < \theta_2$ and non-negative functions $g(\theta)$, define

$$F_g(\theta_1, \theta_2, i) = \frac{\int_{\theta_1}^{\theta_2} p_i(\theta)g(\theta)d\theta}{\int p_i(\theta)g(\theta)d\theta}$$

$$H_g(f, i) = \frac{\int f(\theta)p_i(\theta)g(\theta)d\theta}{\int p_i(\theta)g(\theta)d\theta}.$$

The following poperties are equivalent: a) (identical to A5) For all $\theta_1 < \theta_2$ and all non-negative functions $g(\theta)$ for which the integrals defining $F_g(\theta_1,\theta_2,i)$ are positive and finite, $F_g(\theta_1,\theta_2,i)$ is a weakly unimodal function of i. b) For all non-negative functions $g(\theta)$ and all weakly unimodal functions $f(\theta)$ for which the integrals defining $H_g(f,i)$ are positive and finite, $H_g(f,i)$ is a weakly unimodal function of i.

Proof.

Step 1. (b) $\rightarrow$ (a):

For given θ_1 and θ_2, define the function $s(\theta) = 1$ for $\theta_1 \leq \theta \leq \theta_2$ and $s(\theta) = -1$ for $\theta < \theta_1$ and for $\theta > \theta_2$. It follows from (b) that $H_g(s,i)$ is a weakly unimodal function of i. Since $2F_g(\theta_1,\theta_2,i) = H_g(s,i) + 1$, it follows that $F_g(\theta_1,\theta_2,i)$ is also a weakly unimodal function of i.

Step 2. (a) $\rightarrow$ (b):

In this proof, the number of sign changes of a sequence x(i) is used. This is the number of sign changes of x(i) as i traverses its domain of possible values from low to high, without considering changes to zero values: for example, the sequence $x(i) = (i-3)^2$ for $i = 1$ to 5 has no sign changes, and the sequence $x(i) = (i-3)^2 - 1$ has two. The number of sign changes is related in the following way to unimodality. The sequence x(i) is weakly unimodal if and only if, for every real number c, the sequence $x(i) - c$ has at most two sign changes, and if it has two, the sign pattern is $- + -$.

Let $f(\theta)$ be a weakly unimodal and $g(\theta)$ a non-negative function. We shall prove unimodality of $H_g(f,i)$ by demonstrating that, for every c, $H_g(f,i) - c$ has at most two sign changes as a function of i.

If $c \leq \inf_\theta f(\theta)$ or $c \geq \sup_\theta f(\theta)$, it can be concluded immediately that $H_g(f,i) - c$ has no sign changes. Therefore, assume further that $\inf_\theta f(\theta) < c < \sup_\theta f(\theta)$.

Since $f(\theta)$ is unimodal, there exist θ_1 and θ_2 (possibly $\pm\infty$) with $\theta_1 < \theta_2$ such that $f(\theta)-c \geq 0$ for $\theta_1 \leq \theta \leq \theta_2$ and ≤ 0 elsewhere. Define the function $s(\theta)$ as in Step 1. Then $h(\theta) = s(\theta)(f(\theta)-c)g(\theta)$ is non-negative, and integrable because $f(\theta)$ is bounded. Define $P_g(i) = \int p_i(\theta)g(\theta)d\theta$ and $P_h(i)$ likewise. Using $s^2(\theta) = 1$ it can be concluded that

$$2F_h(\theta_1,\theta_2,i)-1 = H_h(s,i) = \{P_g(i)/P_h(i)\}\{H_g(f,i)-c\}.$$

It follows from (a) that $F_h(\theta_1,\theta_2,i)$ is a weakly unimodal function of i; hence $2F_h(\theta_1,\theta_2,i)-1$ has at most two sign changes; if it has two, the sign pattern is $- + -$. (It can be demonstrated by a continuity argument, that θ_1 or $\theta_2 = \pm\infty$ does not present a problem.) Since $P_g(i)/P_h(i) > 0$ this implies that $H_g(f,i)-c$ has at most two sign changes, and if it has two, then the sign pattern is $- + -$. QED.

Proof of Theorem 2.
Step 1. a) $\rightarrow$ b):
   Note that

$$P_G\{\theta > \theta_0|X_i=1\} = 1-G(\theta_0|X_i=1),$$

where $G(\theta_0|X_i=1)$ is the conditional cdf defined in (3). This shows that A4 states that for any G, for all θ_0, and for all $i < j$,

(10) $$G(\theta_0|X_j=1) \leq G(\theta_0|X_i=1).$$

Now let θ_1, θ_2, i, j be arbitrary with $\theta_1 < \theta_2$ and $i < j$. To prove (b) we use a suitably chosen cdf $G(\theta)$ depending on i, j, θ_1, and θ_2, namely, the cdf corresponding to the two-point discrete distribution with probabilities $g(\theta_1) = g(\theta_2) = 1/2$. Since (a) implies that (10) holds for this G, it can be concluded that

$$\frac{p_j(\theta_1)}{p_j(\theta_1) + p_j(\theta_2)} = G(\theta_1|X_j=1) \leq G(\theta_1|X_i=1) = \frac{p_i(\theta_1)}{p_j(\theta_1) + p_j(\theta_2)}$$

implying

$$p_i(\theta_2)/p_i(\theta_1) \leq p_j(\theta_2)/p_j(\theta_1).$$

Since this holds for all $\theta_1 < \theta_2$, the traceline family has MTR (cf. (7)).

Step 2. b) → c):

Formula (4) gives the conditional density function $g(\theta|X_i=1)$. Since $g(\theta)$ and $1/P_G(i)$ are non-negative functions, it follows from Property 1 that if $p_i(\theta)$ is TP_2 (i.e., has MTR), then also $g(\theta|X_i=1)$ is TP_2 (has MLR).

Step 3. c) → a):

If a family of probability densities, or probability functions, has MLR as a function of an index i, then the corresponding family of probability distributions is stochastically non-decreasing as a function of i. A proof can be found in Lehmann (1986, p. 85). This shows that (c) implies (a). QED.

Proof of Theorem 3.
In this proof the identity

(11) $(A_1B_2-A_2B_1)(B_2C_3-B_3C_2) - (A_2B_3-A_3B_2)(B_1C_2-B_2C_1)$

$$= B_2 \begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix}$$

can be proven by working out both expressions.

Proof of part (a). It follows from Theorem 2 that A4 is satisfied. To prove A5, denote $F(i) = P_G\{\theta_1 \leq \theta \leq \theta_2|X_i=1\}$; if $F(i)$ is not weakly unimodal, then there must exist some $i < n-1$, for which $F(i+1) < \min\{F(i), F(i+2)\}$. This must be proven impossible. In order not to make notation too cluttered, we prove that $F(i+1) \geq \min\{F(i), F(i+2)\}$ for $i=1$. For other i the proof is identical. Define

$A(\theta) = g(\theta)$ if $\theta < \theta_1$ and 0 elsewhere;
$A_i = P\{\theta < \theta_1 \text{ and } X_i = 1\}$;
$B(\theta) = g(\theta)$ if $\theta_1 \leq \theta \leq \theta_2$ and 0 elsewhere;
$B_i = P_G\{\theta_1 \leq \theta \leq \theta_2 \text{ and } X_i = 1\}$;
$C(\theta) = g(\theta)$ if $\theta > \theta_2$ and 0 elsewhere;
$C_i = P_G\{\theta > \theta_2 \text{ and } X_i = 1\}$.

The definition of the determinant implies that

$$\iiint A(\theta_a)B(\theta_b)C(\theta_c)g(\theta_a)g(\theta_b)g(\theta_c) \begin{vmatrix} p_1(\theta_a) & p_1(\theta_b) & p_1(\theta_c) \\ p_2(\theta_a) & p_2(\theta_b) & p_2(\theta_c) \\ p_3(\theta_a) & p_3(\theta_b) & p_3(\theta_c) \end{vmatrix} d\theta_a d\theta_b d\theta_c$$

$$= \iiint \begin{vmatrix} A(\theta_a)g(\theta_a)p_1(\theta_a) & B(\theta_b)g(\theta_b)p_1(\theta_b) & C(\theta_c)g(\theta_c)p_1(\theta_c) \\ A(\theta_a)g(\theta_a)p_2(\theta_a) & B(\theta_b)g(\theta_b)p_2(\theta_b) & C(\theta_c)g(\theta_c)p_2(\theta_c) \\ A(\theta_a)g(\theta_a)p_3(\theta_a) & B(\theta_b)g(\theta_b)p_3(\theta_b) & C(\theta_c)g(\theta_c)p_3(\theta_c) \end{vmatrix} d\theta_a d\theta_b d\theta_c$$

(12)
$$= \begin{vmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \\ A_3 & B_3 & C_3 \end{vmatrix}.$$

Because of the definition of the functions A, B, and C, the triple integral in the first expression has non-zero contributions only for $\theta_a < \theta_b < \theta_c$. With the TP_3 property, this implies that (12) is non-negative. Similarly, the TP_2 property implies that all principal minors of (12) such as $A_1B_2 - A_2B_1$ and $B_1C_3 - B_3C_1$ are non-negative. We wish to prove that $F(2) \geq \min\{F(1), F(3)\}$. Given $F(i) = B_i/(A_i + B_i + C_i)$; simple calculations show that $F(2) < \min\{F(1), F(3)\}$ is equivalent with $\{A_1B_2 - A_2B_1 < B_1C_2 - B_2C_1$ and $B_2C_3 - B_3C_2 < A_2B_3 - A_3B_2\}$. It follows from (a) together with the non-negativity of all these minors that this is incompatible with the positivity of (12).

Proof of part (b). Assume that the family of positive tracelines does satisfy A4, but not TP_3; we shall prove that it then cannot satisfy A5. Because of Theorem 2, the traceline family is TP_2. Since TP_3 does not hold, there exist $\theta_a < \theta_b < \theta_c$ and $i < j < k$ such that

$$\begin{vmatrix} p_i(\theta_a) & p_i(\theta_b) & p_i(\theta_c) \\ p_j(\theta_a) & p_j(\theta_b) & p_j(\theta_c) \\ p_k(\theta_a) & p_k(\theta_b) & p_k(\theta_c) \end{vmatrix} < 0.$$

Define $\theta_1 = (\theta_a + \theta_b)/2$ and $\theta_2 = (\theta_b + \theta_c)/2$. Consider a discrete probability distribution concentrated on the points θ_a, θ_b, and θ_c with positive probabilities $g(\theta_a)$, $g(\theta_b)$, $g(\theta_c)$; use the same notation with A_1, etc., as in part (a) of this proof, with the modification that the distribution of θ now is discrete rather than continuous and item numbers 1, 2, and 3 are replaced by i, j, and k. Then the determinant (12) is negative. With (11) and the positivity of $p_j(\theta_b)$, this implies

$$(A_1B_2 - A_2B_1)(B_2C_3 - B_3C_2) < (A_2B_3 - A_3B_2)(B_1C_2 - B_2C_1).$$

Note that $g(\theta_a)$ is a proportionality factor for the A_i, and similarly for $g(\theta_c)$. By taking $g(\theta_a):g(\theta_c)$ suitably, it can be ensured that $A_1B_2 - A_2B_1 = B_1C_2 - B_2C_1$. This implies $B_2C_3 - B_3C_2 < A_2B_3 - A_3B_2$. Taking $g(\theta_a)$ slightly smaller will not affect the validity of the latter inequality, but will make $A_1B_2 - A_2B_1 < B_1C_2 - B_2C_1$. This means for the F(i) as defined above that $F(j) < \min\{F(i), F(k)\}$, contradicting assumption A5. QED.

9. Acknowledgements

10. References

Andrich, D. (1988) The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement, 12*, 33—51.

Bechtel, G. G. (1968). Folded and unfolded scaling of preferential pair comparisons. *Journal of Mathematical Psychology, 5*, 333—357.

Bossuyt, P. (1990). *A comparison of probabilistic unfolding theories for paired comparisons data.* Berlin: Springer.

Brown, L. D., Johnstone, I. M. & MacGibbon, K. B. (1981). Variation diminishing transformations: A direct approach to total positivity and its statistical applications. *Journal of the American Statistical Association, 76*, 824—832.

Cliff, N., Collins, L. M., Zatkin, J., Gallipeau, D., & McCormick, D. J. (1988). An ordinal scaling method for questionnaire and other ordinal data. *Applied Psychological Measurement, 12*, 83—97.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review, 57*, 145—158.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Coombs, C. H., Greenberg, M. G., & Zinnes, J. L. (1961). A double law of comparative judgment for the analysis of preference choice and similarity data. *Psychometrika, 26*, 165—171.

Davison, M. L. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. *Psychometrika, 42*, 523—548.

Formann, A. K. (1988). Latent class models for nonmonotone dichotomous items. *Psychometrika, 53*, 45—62.

Guttman, L. A. (1954). A new approach to factor analysis: The radix. In Lazarsfeld, P. F. (Ed.) *Mathematical thinking in the social sciences.* New York: Columbia University Press.

Hoijtink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika, 55*, 641—656.

Hoijtink, H., Molenaar I. W., & Post, W. J. (1992). *PARELLA User's Manual.* Groningen, The Netherlands: iec ProGAMMA.

Karlin, S. (1968). *Total positivity.* Stanford, CA: Stanford University Press.

Lehmann, E. L. (1986). *Testing statistical hypotheses*, (2nd ed.). New York: Wiley.

Marshall, A. W. & Olkin, I. (1977). *Inequalities: Theory of majorization and its applications.* New York: Academic Press.

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* New York-Berlin: de Gruyter (Mouton).

Mokken, R. J. & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 10*, 279—285.

Molenaar, I. W. & Münnich, A. (1993). A relation between point items and cumulative items. In R. F. Steyer, K. F. Wender, & K. F. Widdman (eds.) *Psychometric Methodology.* Stuttgart: Gustav Fischer Verlag.

Post, W. J. (1992). *Nonparametric unfolding models: a latent structure approach.* Leiden: DSWO Press.

Rosenbaum, P. R. (1985). Comparing distributions of item responses for two groups. *British Journal of Mathematical and Statistical Psychology, 38*, 206—215.

Rosenbaum, P. R. (1987). Comparing item characteristic curves. *Psychometrika, 52*, 217—233.

Schoenberg, I. J. (1951). On Polya frequency functions I. *Journal d'Analyse Mathématique, 1*, 331—374.

Sixtl, F. (1973). Probabilistic unfolding. *Psychometrika, 38*, 235—248.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Van Schuur, W. H. (1984). *Structure in political beliefs, a new model for stochastic unfolding with application to European party activists.* Amsterdam: CT Press.

Van Schuur, W. H. (1988). Stochastic unfolding. Chapter 9 (p. 137—157) in W.E. Saris & I.N. Gallhofer (eds.), *Sociometric research*, vol.1: Data collection and scaling. London: MacMillan.

Van Schuur, W. H. & Wierstra, T. (1987). *MUDFOLD Manual*, (1st ed.). Groningen: iecProGAMMA.

Van Schuur, W. H. & Post, W. J. (1990). *MUDFOLD Manual*, (2nd. ed.). Groningen: iecProGAMMA.

Zinnes, J. L. & Griggs, R. A. (1974). Probabilistic multidimensional unfolding analysis. *Psychometrika, 39*, 327—350.