



**UvA-DARE (Digital Academic Repository)**

**Contributions to latent variable modeling in educational measurement**

Zwitser, R.J.

[Link to publication](#)

*Citation for published version (APA):*

Zwitser, R. J. (2015). Contributions to latent variable modeling in educational measurement.

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## Chapter 3

# Ordering Individuals with Sum Scores: the Introduction of the Nonparametric Rasch Model

### Summary

When a simple sum or number correct score is used to evaluate the ability of individual testees, then, from an accountability perspective, the inferences based on the sum score should be the same as inferences based on the complete response pattern. This requirement is fulfilled if the sum score is a sufficient statistic for the parameter of a unidimensional model. However, the models for which this does hold, are known as being restrictive. It is shown that the less restrictive (non)parametric models could result in an ordering of persons that is different compared to an ordering based on the sum score. To arrive at a fair evaluation of ability with a simple number correct score, ordinal sufficiency is defined as a minimum condition for scoring. The Monotone Homogeneity Model, together with the property of ordinal sufficiency of the sum score, is introduced as the nonparametric Rasch Model (npRM). A basic outline for testable hypotheses about ordinal sufficiency, as well as illustrations with real data, are provided.

---

This chapter has been conditionally accepted for publication as: Zwitser, R.J. & Maris, G. (submitted). Ordering Individuals with Sum Scores: the Introduction of the Nonparametric Rasch Model. *Psychometrika*.

### 3.1 Introduction

One of the elementary questions in psychological and educational measurement is: how to score a test? Usually, tests consists of multiple items about the same topic. One of the issues is whether the scores on the individual items could fairly be summarized with only one total score, or whether multiple sub scores are needed. The answer to this question could be justified with the use of item response theory (IRT) models. If a unidimensional model fits the data, then it is defensible to report only one total score per person.

Assume that we have a unidimensional test, then the next question is: how should the total score be computed? One approach could be to estimate the person parameter, and report these to the testees. However, although this approach might be intuitively clear for those who have a basic knowledge of statistics, for the general public a more acceptable approach to communicating test results might be via an observed score, specifically the sum score (Sijtsma & Hemker, 2000).

But if someone wants to report sum scores instead of parameter estimates, then, from an accountability perspective, the following question does arise: are inferences based on the sum score the same as inferences based on the parameter estimate? In case of the Rasch model (RM, Rasch, 1960; Fischer, 1974; Hessen, 2005; Maris, 2008) the answer is clearly *yes*, because in this model the sum score is a sufficient statistic for the person parameter. This property, as we will explain in section 3.3, implies that all *available* information in the data about the ordering of *individual testees* is in correspondence with the ordering of the sum scores. However, the RM is known as a restrictive model. One of the less restrictive alternatives is the nonparametric Monotone Homogeneity Model (MHM, Mokken, 1971; see also Sijtsma & Molenaar, 2002). A well-known property of this model is that the person parameters are stochastically ordered by the sum score (Mokken, 1971; Grayson, 1988; Huynh, 1994). This property is very useful for comparisons between *groups of persons*, because it implies that testees with a higher sum score have *on average* a higher value of the person parameter than testees with a lower sum score. However, it will be demonstrated in section 3.2.2 that this property is not satisfactory for making ordinal inferences about *individual testees*, because the ordering based on the sum score could be different compared to the ordering of the parameters based on the available item responses. To arrive at a less restrictive nonparametric

model that enables the ordering of individuals based on the sum score, we define the minimal condition in section 3.3: ordinal sufficiency. With this property we can introduce the nonparametric Rasch Model. In section 3.4 we derive some testable implications of ordinal sufficiency. This is illustrated with an example based on real data.

## 3.2 Some models under consideration

All IRT models considered in this paper are unidimensional monotone latent variable models for dichotomous responses, i.e., they all assume at least Unidimensionality (UD), Local Independence (LI) and Monotonicity (M). The score on item  $i$  is denoted by  $X_i$ :  $X_i = 1$  for a correct response and  $X_i = 0$  otherwise. Let the random vector  $\mathbf{X} = [X_1, X_2, \dots, X_p]$  be the total score pattern on a test with  $p$  items and let  $\mathbf{x}$  denote a realization of  $\mathbf{X}$ . The person parameter, sometimes referred to as ability parameter or latent trait, is denoted by  $\theta$ .

### 3.2.1 Parametric IRT models

Examples of parametric unidimensional monotone latent trait models are the Rasch Model (RM, Rasch, 1960),

$$P(X_i = 1|\theta) = P(x_i|\theta) = \frac{\exp(\theta - \delta_i)}{1 + \exp(\theta - \delta_i)},$$

and the Two-Parameter Logistic Model (2PLM, Birnbaum, 1968),

$$P(x_i|\theta) = \frac{\exp[\alpha_i(\theta - \delta_i)]}{1 + \exp[\alpha_i(\theta - \delta_i)]},$$

in which  $\alpha_i$  and  $\delta_i$  are parameters related to item  $i$ . Both models contain sufficient statistics for their parameters.

**Definition 1.** *A statistic  $H(\mathbf{X})$  is sufficient for parameter  $\theta$  if the conditional distribution of  $\mathbf{X}$ , given the statistic  $H(\mathbf{X})$ , does not depend on the parameter  $\theta$ , i.e.,*

$$P(\mathbf{X} = \mathbf{x}|H(\mathbf{X}) = a, \theta) = P(\mathbf{X} = \mathbf{x}|H(\mathbf{X}) = a).$$

Sufficiency implies that *all* statistical information in the data  $\mathbf{X}$  about the parameter  $\theta$  is kept by the statistic  $H(\mathbf{x})$ . It has already been mentioned that in the RM the sum score

$$\sum_i X_i = X_+$$

is a sufficient statistic for  $\theta$ . Another well-known example of a sufficient statistic is the weighted sum score

$$\sum_i \alpha_i X_i$$

in the 2PLM, if the weights are known. Therefore, we can easily demonstrate that in a case where the 2PLM fits the data well, inferences based on  $\theta$  could be different compared to inferences based on  $X_+$ .

In section 3.3.2 we will also consider the Normal Ogive Model (NOM, Lord & Novick, 1968),

$$P(x_i|\theta) = \int_{-\infty}^{\theta-\delta_i} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt = \Phi(\theta - \delta_i).$$

### 3.2.2 Nonparametric IRT models

A well-known nonparametric model is the Monotone Homogeneity Model (MHM, Mokken, 1971; see also Sijtsma & Molenaar, 2002). The MHM only assumes UD, LI and M. For the MHM, it has been shown that  $X_+$  has a *likelihood ratio ordering* in  $\theta$  (Grayson, 1988; Huynh, 1994), i.e.,

$$\forall a > b, \theta_2 > \theta_1 : \frac{P(X_+ = a|\theta_2)}{P(X_+ = b|\theta_2)} \geq \frac{P(X_+ = a|\theta_1)}{P(X_+ = b|\theta_1)}. \quad (3.1)$$

From (3.1) it can easily be derived that

$$P(\Theta > s|X_+ = a) \geq P(\Theta > s|X_+ = b), \quad (3.2)$$

for all  $s$ , and  $a > b$ . The property in (3.2) is called *stochastic ordering of the latent trait by  $X_+$*  (SOL; Hemker, Sijtsma, Molenaar, & Junker, 1997), also

denoted by<sup>1</sup>

$$(\Theta|X_+ = a) \underset{st}{\geq} (\Theta|X_+ = b), \quad \text{if } a > b,$$

which equals

$$\mathcal{E}(g(\Theta)|X_+ = a) \geq \mathcal{E}(g(\Theta)|X_+ = b)$$

for all  $a > b$ , and all bounded increasing functions  $g$  (Ross, 1996, prop. 9.1.2.). This implies that all statistics for central tendency of  $\Theta$ , e.g., the median, mode, or mean, are ordered by  $X_+$ .

The SOL property has been used as justification for ordering individuals with the sum score (see, e.g., Mokken, 1971, and Meijer et al., 1990). However, for making ordinal inferences about individuals (e.g., passing or failing an exam) this property might not be sufficient. Consider, for instance, a test with three items that satisfy the assumptions of the MHM. The first item is a Guttman item (Guttman, 1950), whereas the last two items have a constant probability of success, e.g.,  $P(x_i|\theta) = 0.5$ . Next, consider two persons. The first person answers the second and third item correct, while the second person only answers the first item correct. According to the SOL property, we would conclude that

$$(\Theta|X_+ = 2) \underset{st}{\geq} (\Theta|X_+ = 1).$$

However, the item characteristics above imply that

$$(\Theta|\mathbf{X} = [0, 1, 1]) \underset{st}{<} (\Theta|\mathbf{X} = [1, 0, 0]).$$

Recall that the models considered in this paper are all unidimensional models. This implies all available information in the data about individual differences can be summarized with only one score per subject. The accountability issue mentioned above can be rephrased into the question whether the ordering based on the sum score is the same as the ordering based on the complete response pattern. This example demonstrates that the answer to this question is *no* for

---

<sup>1</sup>In general,

$$X \underset{st}{\geq} Y \text{ denotes } P(X > a) \geq P(Y > a) \quad \text{for all } a,$$

$$X \underset{st}{>} Y \text{ denotes } P(X > a) > P(Y > a) \quad \text{for all } a,$$

$$X \underset{st}{=} Y \text{ denotes } P(X > a) = P(Y > a) \quad \text{for all } a.$$

the MHM.

So far, the only model that satisfies this condition is the RM. However, the RM is known as a restrictive model, which leads to the wish for less restrictive nonparametric alternatives (Meijer et al., 1990). This alternative will be considered at the end of the next section.

### 3.3 Sufficiency

Before we propose a nonparametric alternative that justifies the use of the sum score for the purpose of individual measurement (section 3.3.3), we first describe the property of sufficiency in more detail. In section 3.3.1 we describe the condition under which a sufficient statistic exists. This leads to another representation of sufficiency whereby we can easily propose ordinal sufficiency as a weaker form of sufficiency that still enables ordinal measurement with an observed score (section 3.3.2).

#### 3.3.1 The existence of a sufficient statistic

The derivations in this section are based on the work of Milgrom (1981). We start with three lemmas.

**Lemma 1.**  $X \underset{st}{\geq} Y$  if and only if  $\mathcal{E}(g(X)) \geq \mathcal{E}(g(Y))$  for all bounded non-decreasing functions  $g$ .

*Proof.* See Ross (1996, prop. 9.1.2). □

**Lemma 2.** The distribution of  $\mathbf{X}$  conditionally on  $\Theta$  has monotone likelihood ratio (MLR) if and only if

$$\forall \mathbf{x}_1, \mathbf{x}_2, \Theta : (\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{>} (\Theta | \mathbf{X} = \mathbf{x}_1)$$

or

$$(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{<} (\Theta | \mathbf{X} = \mathbf{x}_1)$$

or

$$(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{=} (\Theta | \mathbf{X} = \mathbf{x}_1).$$

*Proof.* (if)

$$\forall \mathbf{x} : P(\Theta \leq \theta | \mathbf{X} = \mathbf{x})$$

is a bounded non-decreasing function of  $\theta$ . Hence, if we assume for  $\mathbf{x}_1$  and  $\mathbf{x}_2$  that  $(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{>} (\Theta | \mathbf{X} = \mathbf{x}_1)$ , we infer that

$$\forall \mathbf{x}_3 : \mathcal{E}[P(\Theta | \mathbf{X} = \mathbf{x}_3) | \mathbf{X} = \mathbf{x}_2] > \mathcal{E}[P(\Theta | \mathbf{X} = \mathbf{x}_3) | \mathbf{X} = \mathbf{x}_1].$$

Since

$$\forall \mathbf{x} : \mathcal{E}[P(\Theta | \mathbf{X} = \mathbf{x}) | \mathbf{X} = \mathbf{x}] = 1/2,$$

we obtain that

$$\mathcal{E}[P(\Theta | \mathbf{X} = \mathbf{x}_1) | \mathbf{X} = \mathbf{x}_2] > \mathcal{E}[P(\Theta | \mathbf{X} = \mathbf{x}_2) | \mathbf{X} = \mathbf{x}_1],$$

or explicitly

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\theta} \frac{P(\mathbf{x}_1 | \theta^*) f(\theta^*)}{P(\mathbf{x}_1)} \frac{P(\mathbf{x}_2 | \theta) f(\theta)}{P(\mathbf{x}_2)} d\theta^* d\theta \\ & > \int_{-\infty}^{\infty} \int_{-\infty}^{\theta} \frac{P(\mathbf{x}_2 | \theta^*) f(\theta^*)}{P(\mathbf{x}_2)} \frac{P(\mathbf{x}_1 | \theta) f(\theta)}{P(\mathbf{x}_1)} d\theta^* d\theta, \end{aligned}$$

and hence

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\theta} [P(\mathbf{x}_1 | \theta^*) P(\mathbf{x}_2 | \theta) - P(\mathbf{x}_2 | \theta^*) P(\mathbf{x}_1 | \theta)] f(\theta^*) f(\theta) d\theta^* d\theta > 0. \quad (3.3)$$

Notice that Lemma 2 holds for all  $\Theta$ , which denotes the random variable (uppercase). This implies that (3.3) does hold for every prior  $f(\theta)$ . Therefore,

$$\forall \theta^* < \theta : P(\mathbf{x}_1 | \theta^*) P(\mathbf{x}_2 | \theta) > P(\mathbf{x}_2 | \theta^*) P(\mathbf{x}_1 | \theta),$$

which completes the first part of the proof.

(only if) This part of the proof is trivial as MLR implies stochastic ordering (see, e.g., Ross, 1996).  $\square$

**Lemma 3.** *If the distribution of  $\mathbf{X}$  conditionally on  $\Theta$  has monotone likelihood ratio (MLR), then there exists a function  $H$  such that both  $\mathbf{X} \perp\!\!\!\perp \theta | H(\mathbf{X})$  and  $H(\mathbf{X}) | \Theta$  has MLR.*



*Proof.* Let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be such that  $(\Theta|\mathbf{X} = \mathbf{x}_1) \stackrel{st}{=} (\Theta|\mathbf{X} = \mathbf{x}_2)$ , and let  $g$  be a non-decreasing bounded function such that

$$H(\mathbf{x}) = \mathcal{E}(g(\Theta)|\mathbf{X} = \mathbf{x})$$

then  $H(\mathbf{x}_1) = H(\mathbf{x}_2)$  and

$$P(\mathbf{x}_1|\theta) = \frac{P(\mathbf{x}_1)}{P(\mathbf{x}_2)}P(\mathbf{x}_2|\theta)$$

such that

$$\begin{aligned} P(\mathbf{X} = \mathbf{x}_2|H(\mathbf{X}) = H(\mathbf{x}_2), \theta) &= \frac{P(\mathbf{x}_2|\theta)}{\sum_{\mathbf{x}_1: H(\mathbf{x}_1)=H(\mathbf{x}_2)} P(\mathbf{x}_1|\theta)} \\ &= \frac{P(\mathbf{x}_2|\theta)}{\sum_{\mathbf{x}_1: H(\mathbf{x}_1)=H(\mathbf{x}_2)} \frac{P(\mathbf{x}_1)}{P(\mathbf{x}_2)} P(\mathbf{x}_2|\theta)} \\ &= \frac{P(\mathbf{x}_2)}{\sum_{\mathbf{x}_1: H(\mathbf{x}_1)=H(\mathbf{x}_2)} P(\mathbf{x}_1)}, \end{aligned}$$

which does not depend on  $\theta$ , and therefore completes the first part of the proof.

For the second part, let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be such that  $(\Theta|\mathbf{X} = \mathbf{x}_2) > (\Theta|\mathbf{X} = \mathbf{x}_1)$ . Then, obviously,  $H(\mathbf{x}_2) > H(\mathbf{x}_1)$ . Since  $(\Theta|\mathbf{X} = \mathbf{x}) \stackrel{st}{=} (\Theta|H(\mathbf{X}) = H(\mathbf{x}))$  we obtain that  $(\Theta|H(\mathbf{X}) = H(\mathbf{x}_2)) \stackrel{st}{>} (\Theta|H(\mathbf{X}) = H(\mathbf{x}_1))$ , and the conclusion follows from Lemma 2.  $\square$

With these lemmas we can now describe under which conditions a sufficient statistic does exist.

**Theorem 1.**

$$\exists H : \mathbf{X} \perp\!\!\!\perp \Theta|H(\mathbf{X})$$

if and only if

$$\forall \mathbf{x}_1, \mathbf{x}_2, \Theta : (\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{>} (\Theta | \mathbf{X} = \mathbf{x}_1),$$

or

$$(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{<} (\Theta | \mathbf{X} = \mathbf{x}_1),$$

or

$$(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{=} (\Theta | \mathbf{X} = \mathbf{x}_1).$$

*Proof.* Direct from Lemma 2 and 3. □

With this representation of sufficiency, we can introduce the minimal condition for ordering persons with an observed score.

### 3.3.2 Ordinal sufficiency

From Theorem 1 it follows that sufficiency of statistic  $H$  implies that

$$(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{>} (\Theta | \mathbf{X} = \mathbf{x}_1), \quad \text{if } H(\mathbf{x}_2) > H(\mathbf{x}_1), \quad (3.4)$$

and

$$(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{=} (\Theta | \mathbf{X} = \mathbf{x}_1), \quad \text{if } H(\mathbf{x}_2) = H(\mathbf{x}_1). \quad (3.5)$$

The core of this paper is the following: if the purpose of a test is to order subjects, then (3.4) is the only property of interest: if we order subjects based on an observed score, then their posterior distributions of  $\Theta$  should be stochastically ordered in the same direction. Therefore, we call the condition in (3.4) *ordinal sufficiency* (OS).

**Definition 2.** A statistic  $H(\mathbf{X})$  is *ordinally sufficient* for  $\Theta$  if  $H(\mathbf{x}_2) > H(\mathbf{x}_1)$  implies  $(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{>} (\Theta | \mathbf{X} = \mathbf{x}_1)$ .

OS allows the ordering based on  $H$  to be coarser than the ordering based on the response patterns. That is, the following can occur:

$$(\Theta | \mathbf{X} = \mathbf{x}_2) \underset{st}{>} (\Theta | \mathbf{X} = \mathbf{x}_1), \quad \text{for some } \mathbf{x}_2 \text{ and } \mathbf{x}_1, \text{ for which } H(\mathbf{x}_2) = H(\mathbf{x}_1).$$

In the next section, we consider for some specific IRT models whether the sum score is ordinal sufficient for  $\theta$ .

## Determining ordinal sufficiency of the sum score in a particular model

**Normal Ogive Model** The following counter example shows that for the NOM the sum score is not ordinal sufficient for  $\theta$ .

Consider a test with 9 items of which the item parameters are

$$\delta = [\delta_1, \delta_2, \dots, \delta_9] = [2, 2, 2, 2, 2, -2, -2, -2, -2].$$

Furthermore, assume that

$$\Theta \sim N(0, 1).$$

The  $\delta_i$  parameters indicate that the first 5 items are difficult and that the last 4 items are easy.

Consider the following two answer patterns:

- $\mathbf{x}_1 = [1, 1, 1, 1, 1, 0, 0, 0, 0]$ ;
- $\mathbf{x}_2 = [0, 0, 0, 0, 0, 1, 1, 1, 1]$ .

The posterior distributions of  $\Theta$  for these two answer patterns are displayed in Figure 3.1. From this counter example it can be seen that these posterior distributions are not stochastically ordered.

**2PL Model** Consider two response vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . These vectors can, after applying the same permutation of indices to both, be expressed as

$$\begin{aligned}\mathbf{x}_1 &= \mathbf{y} \cup (1 - \mathbf{z}), \\ \mathbf{x}_2 &= \mathbf{y} \cup \mathbf{z},\end{aligned}$$

in which  $\mathbf{y}$  is the common part of  $\mathbf{x}_1$  and  $\mathbf{x}_2$ .

It is derived in the appendix that  $X_+$  is ordinal sufficient for  $\theta$  if for an item response model  $P(x_i|\theta)$  it can be shown that

$$\sum_g z_g \log \left[ \frac{\left( \frac{P(z_g|\theta_2)}{1-P(z_g|\theta_2)} \right)}{\left( \frac{P(z_g|\theta_1)}{1-P(z_g|\theta_1)} \right)} \right] > \sum_g (1-z_g) \log \left[ \frac{\left( \frac{P(z_g|\theta_2)}{1-P(z_g|\theta_2)} \right)}{\left( \frac{P(z_g|\theta_1)}{1-P(z_g|\theta_1)} \right)} \right], \quad \theta_2 > \theta_1, \quad z_+ > \frac{n}{2}, \quad (3.6)$$

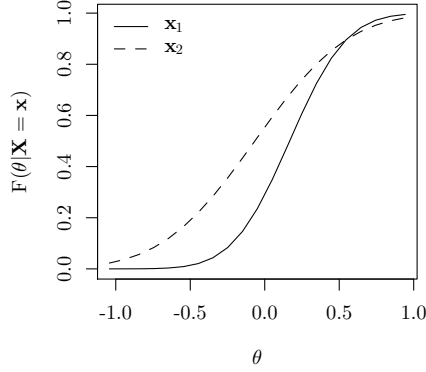


Figure 3.1: Posterior distributions of  $\theta$  for two response patterns under the NOM.

in which  $\mathbf{z}$  is the subset of  $n$  items to which is responded differently in  $\mathbf{x}_1$  and  $\mathbf{x}_2$ ,

$$z_+ = \sum_{g=1}^n z_g,$$

and where it is assumed that  $x_{+2} > x_{+1}$ .

For the 2PLM, (3.6) results in

$$\begin{aligned} \sum_g z_g \log \left[ \frac{\exp[\alpha_g(\theta_2 - \delta_g)]}{\exp[\alpha_g(\theta_1 - \delta_g)]} \right] &> \sum_g (1 - z_g) \log \left[ \frac{\exp[\alpha_g(\theta_2 - \delta_g)]}{\exp[\alpha_g(\theta_1 - \delta_g)]} \right] \\ &\Downarrow \\ \sum_g z_g \alpha_g (\theta_2 - \theta_1) &> \sum_g (1 - z_g) \alpha_g (\theta_2 - \theta_1), \quad \theta_2 > \theta_1 \\ &\Downarrow \\ \sum_{g:z=1} \alpha_g &> \sum_{g:z=0} \alpha_g, \quad z_+ > \frac{n}{2}. \end{aligned} \quad (3.7)$$

Here it can be seen that OS of  $X_+$  depends on the  $\alpha_i$  parameters.

Let  $\alpha$  be the vector of all parameters  $\alpha_i$ , i.e.,

$$\alpha = [\alpha_1, \alpha_2, \dots, \alpha_p]$$

and let  $\alpha'$  and  $\alpha^*$  be subsets of  $\alpha$  such that

$$\begin{aligned}\alpha' \cup \alpha^* &= \alpha, \\ \alpha' \cap \alpha^* &= \emptyset, \\ \dim(\alpha') &> \dim(\alpha^*).\end{aligned}$$

Following from (3.7),  $X_+$  is ordinal sufficient if

$$\forall \alpha', \alpha^* : \sum_i \alpha'_i > \sum_j \alpha^*_j. \quad (3.8)$$

It can be seen that (3.8) holds if, for  $p$  even, the sum of the smallest  $(\frac{p}{2} + 1)$  elements of  $\alpha$  is larger than the sum of the  $(\frac{p}{2} - 1)$  largest elements of  $\alpha$ . If  $p$  is odd, then the sum of the smallest  $(\frac{p}{2} + \frac{1}{2})$  elements of  $\alpha$  has be larger than the the sum of the largest  $(\frac{p}{2} - \frac{1}{2})$  elements of  $\alpha$ .

**Nonparametric IRT Models** In Section 3.2.2, it was shown that in the MHM the sum score is not ordinal sufficient. In this section, we consider two additional assumptions. The first is *invariant item ordering* (IIO):

$$P(x_1|\theta) \leq P(x_2|\theta) \leq \dots \leq P(x_p|\theta), \quad \text{for all } \theta.$$

The MHM model together with IIO, is known as the Double Monotonicity Model (DMM, Mokken, 1971; see also Sijtsma & Molenaar, 2002). The second assumption is *monotone traceline ratio* (MTR, Post, 1992):

$$\frac{P(x_i|\theta)}{P(x_j|\theta)} \text{ is a non-decreasing function of } \theta, \quad \text{for all } i < j.$$

In order to show that both the addition of IIO and MTR to the MHM do not result in a model with an ordinal sufficient sum score, we, once again, consider an example of a three-item test. The *item response functions* (IRFs) are as

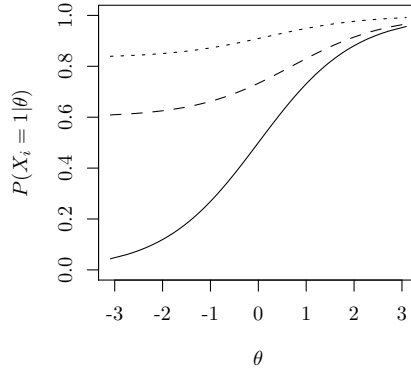


Figure 3.2: IRFs for the three items in (3.9):  $P(x_1|\theta)$  (solid),  $P(x_2|\theta)$  (dashed), and  $P(x_3|\theta)$  (dotted).

follows:

$$\begin{aligned}
 P(x_1|\theta) &= \frac{\exp(\theta)}{\exp(\theta) + 1} \\
 P(x_2|\theta) &= \frac{\exp(\theta) + 1.2}{\exp(\theta) + 2} \\
 P(x_3|\theta) &= \frac{\exp(\theta) + 1}{\exp(\theta) + 1.2}
 \end{aligned} \tag{3.9}$$

depicted in Figure 3.2. These IRFs satisfy IIO,

$$P(x_1|\theta) \leq P(x_2|\theta) \leq P(x_3|\theta), \quad \text{for all } \theta,$$

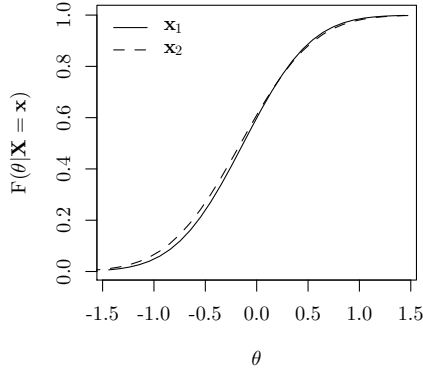


Figure 3.3: Posterior distributions of  $\theta$  for two response patterns under the DMM and MTR.

as well as MTR:

$$\frac{P(x_1|\theta)}{P(x_2|\theta)} \text{ increasing in } \theta;$$

$$\frac{P(x_1|\theta)}{P(x_3|\theta)} \text{ increasing in } \theta;$$

$$\frac{P(x_2|\theta)}{P(x_3|\theta)} \text{ increasing in } \theta.$$

Now consider two particular response patterns:

- $\mathbf{x}_1 = [1, 0, 0]$ ;
- $\mathbf{x}_2 = [0, 1, 1]$ .

The posterior distributions of  $\theta$  given  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , based on a standard normal distribution of  $\Theta$ , are displayed in Figure 3.3. This counter example shows that the additional assumptions of IIO and MTR to the MHM does not lead to a model with an ordinal sufficient sum score.

### 3.3.3 Nonparametric Rasch model

None of the nonparametric models described above have an ordinal sufficient sum score for  $\theta$ . Therefore, we propose a new nonparametric model. This model assumes UD, LI, M, and OS. Since it is known that the RM is the only model for which the sum score is a sufficient statistic for the latent trait (Fischer, 1974; Hessen, 2005; Maris, 2008), we call the this new model the *nonparametric Rasch Model* (npRM).

The difference between the MHM and npRM can also be displayed in the following way. Under the MHM, the ordering based on some score patterns is in accordance with the ordering based on the sum score. Specifically, if  $\mathbf{z}$ , which is the uncommon part of  $\mathbf{x}_1$  and  $\mathbf{x}_2$  (see Section 3.3.2), is such that it contains only ones, then

$$(\Theta|\mathbf{X} = \mathbf{x}_2) \underset{st}{>} (\Theta|\mathbf{X} = \mathbf{x}_1).$$

This can easily be demonstrated with two score patterns that differ on only one item (e.g. in a case of 3 items  $[1,1,0]$  and  $[0,1,0]$ ). The likelihood ratio of these two score patterns is

$$\begin{aligned} \frac{P[1, 1, 0|\theta]}{P[0, 1, 0|\theta]} &= \frac{P(x_1|\theta)P(x_2|\theta)[1 - P(x_3|\theta)]}{[1 - P(x_1|\theta)]P(x_2|\theta)[1 - P(x_3|\theta)]} \\ &= \frac{P(x_1|\theta)}{1 - P(x_1|\theta)}. \end{aligned}$$

Because  $P(x_1|\theta)$  is a monotone increasing function in  $\theta$  (model assumption),

$$\frac{P(x_1|\theta)}{1 - P(x_1|\theta)}$$

is also increasing in  $\theta$ , as well as the likelihood ratio of these two score patterns. This likelihood ratio ordering implies stochastic ordering (Ross, 1996). For the case of a three-item test, all pairs of response patterns for which this does hold, are displayed in Figure 3.4. However, in order to conclude that the sum score is ordinal sufficient, also score patterns for which  $\mathbf{z}$  contains zeros and ones, have to lead to a stochastic ordering of  $\theta$  that is in accordance with the ordering based on the sum score. In other words, not only the patterns in Figure 3.4 should have a stochastic ordering of the



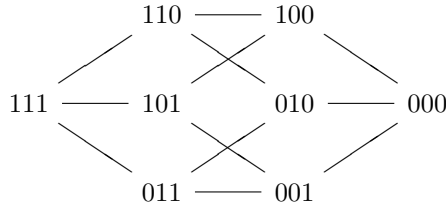


Figure 3.4: Partial likelihood ratio order for a three-item test under the MHM.

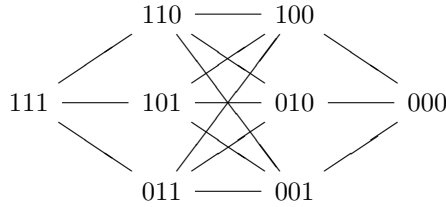


Figure 3.5: Partial likelihood ratio order for a three-item test under the npRM.

posterior distributions of  $\theta$ , but all patterns that are displayed in Figure 3.5 should meet this condition. For instance, it must hold that

$$(\Theta|\mathbf{X} = [0, 1, 1]) \underset{st}{>} (\Theta|\mathbf{X} = [1, 0, 0]).$$

The question whether this ordering does hold or not, could be verified with a statistical test. This test will be introduced in the next section.

### 3.4 Testable implications of ordinal sufficiency

In order to determine whether the sum score is ordinal sufficient, we introduce some testable implications of OS. First two lemma's are provided.

**Lemma 4.** *OS of the sum score for a set of items implies OS of the sum score for any subset of items.*

*Proof.* Let  $\mathbf{x}_k$  denote a realization of  $\mathbf{X}$ ,  $x_{ik}$  denote the response on item  $i$  in  $\mathbf{x}_k$ ,  $\mathbf{x}_k^{[i]}$  denote the subset of  $\mathbf{x}_k$  without item  $i$ ,  $x_{+k}$  denote the sum score of responses in  $\mathbf{x}_k$ , and  $x_{+k}^{[i]}$  denote the sum score of responses in  $\mathbf{x}_k^{[i]}$ .

It needs to be proven that if  $x_{+1} > x_{+2}$ , and

$$\frac{f(\theta|\mathbf{X} = \mathbf{x}_1)}{f(\theta|\mathbf{X} = \mathbf{x}_2)}$$

is increasing in  $\theta$ , then

$$\frac{f(\theta|\mathbf{x}_1^{[i]})}{f(\theta|\mathbf{x}_2^{[i]})}$$

is also increasing in  $\theta$  if  $x_{+1}^{[i]} > x_{+2}^{[i]}$ .

Since we assume local independence,

$$f(\theta|\mathbf{x}) = \frac{P(x_i|\theta)P(\mathbf{x}^{[i]}|\theta)f(\theta)}{P(x_i|\mathbf{x}^{[i]})P(\mathbf{x}^{[i]})},$$

and

$$\frac{f(\theta|\mathbf{x}_1)}{f(\theta|\mathbf{x}_2)} = \frac{P(x_{i1}|\theta)f(\theta|\mathbf{x}_1^{[i]})}{P(x_{i2}|\theta)f(\theta|\mathbf{x}_2^{[i]})} \frac{P(x_{i2}|\mathbf{x}_2^{[i]})}{P(x_{i1}|\mathbf{x}_1^{[i]})}.$$

Let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be such that  $x_{+1}^{[i]} > x_{+2}^{[i]}$ . Following from local independence, we are free to assume  $x_{i1} = x_{i2} = x_i$ . Then, we find using the above relation that

$$\frac{f(\theta|\mathbf{x}_1^{[i]})}{f(\theta|\mathbf{x}_2^{[i]})} \propto \frac{f(\theta|\mathbf{x}_1)}{f(\theta|\mathbf{x}_2)}.$$

Since the right hand side is increasing in  $\theta$  if  $x_{+1} > x_{+2}$ , and  $x_{+k} = x_{+k}^{[i]} + x_i$ , the result follows.  $\square$

**Lemma 5.** *If OS of the sum score holds in all subsets of  $(p-1)$  items, then it also holds for all  $p$  items, provided  $p$  is even.*

*Proof.* Any pair of response patterns  $\mathbf{x}_1$  and  $\mathbf{x}_2$  for which  $x_{+1} > x_{+2}$  can, after applying the same permutation of indices to both, be expressed as

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{y} \cup \mathbf{z} \\ \mathbf{x}_2 &= \mathbf{y} \cup (1 - \mathbf{z}) \end{aligned}$$

where  $z_+ > (1 - z)_+$ . We consider two cases.

The first case is when  $\mathbf{y} \neq \emptyset$ . Then

$$\frac{f(\theta|\mathbf{z})}{f(\theta|1-\mathbf{z})} \propto \frac{f(\theta|\mathbf{x}_1)}{f(\theta|\mathbf{x}_2)}.$$

Since the left hand side is increasing in  $\theta$  if  $z_+ > (1-z)_+$ , and  $x_{+k} = z_+ + y_+$ , it follows that the right hand side is also increasing in  $\theta$ , and that  $x_{+1} > x_{+2}$ , which completes the first part of the proof.

The second case is when  $\mathbf{y} = \emptyset$ . Then any pair of response patterns  $\mathbf{x}_1$  and  $\mathbf{x}_2$  with  $x_{+1} > x_{+2}$  and with an even number of items, can, after applying the same permutation of indices to both, be expressed as

$$\begin{aligned} \mathbf{x}_1 &= (x_i = 1) \cup \mathbf{z} \\ \mathbf{x}_2 &= (x_i = 0) \cup (1 - \mathbf{z}) \end{aligned}$$

where  $\mathbf{z} = \mathbf{x}^{[i]}$  and  $z_+ > (1-z)_+$ .

Observe that

$$\frac{f(\theta|\mathbf{z})}{f(\theta|1-\mathbf{z})} \frac{P(x_i = 1|\theta)}{P(x_i = 0|\theta)} \propto \frac{f(\theta|\mathbf{x}_1)}{f(\theta|\mathbf{x}_2)}.$$

Since both parts of the left hand side are increasing in  $\theta$  if  $z_+ > (1-z)_+$ , it follows that the right hand side is also increasing in  $\theta$  if  $x_{+1} > x_{+2}$ , which completes the second part of the proof. □

From these lemmas it follows that if (3.4) holds for  $\mathbf{X}$ , it also has to hold for all subsets of  $\mathbf{X}$ .

Let  $I$  denote the set of all  $p$  item indices, i.e.,  $I = \{1, 2, 3, \dots, i, \dots, p\}$ , let  $S$  denote a subset of item indices, i.e.,  $S \subset I$ , let  $\mathbf{S}$  denote the vector of responses on the items in  $S$ , and let  $X_+^{[S]}$  denote the sum score of the items of  $\mathbf{X}$  that are not in subset  $S$ . Following from local independence,

$$P(X_+^{[S]}|\mathbf{s}) = \int P(X_+^{[S]}|\theta)f(\theta|\mathbf{s})d\theta.$$

It is already mentioned that under the MHM,  $\theta$  has monotone likelihood ratio (MLR) in  $X_+$  (Grayson, 1988; Huynh, 1994). A well-known property of the MLR is that also the reverse is true, i.e.  $X_+$  has monotone likelihood ratio in

$\theta$ . If the likelihood ratio increases in  $\theta$ , then it follows that if

$$(X_+^{[S]}|\mathbf{s}_2)_{st} > (X_+^{[S]}|\mathbf{s}_1),$$

then

$$(\Theta|\mathbf{s}_2)_{st} > (\Theta|\mathbf{s}_1).$$

This leads to the following analogy for testing ordinal sufficiency: the null hypothesis ( $H_0$ ) is that  $H(\mathbf{X})$  is ordinal sufficient for  $\theta$ . If  $H_0$  is true, then

$$\forall \mathbf{s}_1, \mathbf{s}_2 : (\Theta|\mathbf{s}_2)_{st} > (\Theta|\mathbf{s}_1), \quad \text{if } H(\mathbf{s}_2) > H(\mathbf{s}_1).$$

These multiple sub-hypotheses can be tested by determining whether

$$\forall \mathbf{s}_1, \mathbf{s}_2 : (X_+^{[S]}|\mathbf{s}_2)_{st} > (X_+^{[S]}|\mathbf{s}_1), \quad \text{if } H(\mathbf{s}_2) > H(\mathbf{s}_1).$$

If one of these sub-hypotheses is rejected, then  $H(\mathbf{X})$  is not ordinal sufficient for  $\theta$ . For testing stochastic ordering, we refer to the literature about the Kolmogorov and Smirnov theorems (see, e.g., Doob, 1949; Conover, 1999b).

### 3.4.1 Example

This procedure will briefly be demonstrated with an example. The examples are based on data from the Dutch *Entrance Test* (in Dutch: *Entreetoets*), which consists of multiple parts that are administered annually to 125,000 grade 5 pupils. One of the parts is a test with 120 math items. To gain insight into the item characteristics, we first analyzed a sample of 30,000 examinees<sup>2</sup> with the *One-Parameter Logistic Model* (OPLM, Verhelst & Glas, 1995; Verhelst et al., 1993). The OPLM with integer  $\alpha_i$  parameters did not fit the data well,  $R_{1c} = 5,956$ ,  $df = 357$ ,  $p < 0.001$ , however, the item parameter estimates can be informative for the selection of subsets of items for this illustration. The parameters of a selection of six items are displayed in Table 3.1.

The smallest subsets that can be tested on ordinal sufficiency are subsets of three items. According to the rule in Section 3.3.2 the subset that contains

---

<sup>2</sup>A sample had to be drawn because of limitations of the OPLM software package w.r.t. the maximum number of observations.

Table 3.1: Estimated OPLM parameters of six items from the example data set.

item	$\alpha_i$	$\delta_i$
1	2	0.275
2	4	-0.156
3	3	-0.296
6	2	-0.460
8	4	0.104
10	5	-0.029

item 1, 2, and 3 has an ordinal sufficient sum score. This is confirmed by the empirical cumulative distributions (ecds) of  $(X_+^{[S]}|[0, 0, 1])$  and  $(X_+^{[S]}|[1, 1, 0])$  in Figure 3.6a. In contrast, the subset with item 1, 6, and 10 does not have an ordinal sufficient sum score. Figure 3.6b displays the ecds of  $(X_+^{[S]}|[0, 0, 1])$  and  $(X_+^{[S]}|[1, 1, 0])$ . A third example are the ecds of  $(X_+^{[S]}|[0, 0, 1])$  and  $(X_+^{[S]}|[1, 1, 0])$ , based on the subset with item 1, 6 and 8. According to the  $\alpha_i$ -parameters, the  $(X_+^{[S]}|[0, 0, 1])$  and  $(X_+^{[S]}|[1, 1, 0])$  are not expected to be stochastically ordered. This expectation is confirmed by Figure 3.6c.

These three cases can also be tested with the one-sided Kolmogorov-Smirnov (KS) test (Conover, 1999a). The corresponding hypotheses are

$$H_0 : (X_+^{[S]}|[0, 0, 1]) \underset{st}{\leq} (X_+^{[S]}|[1, 1, 0]);$$

$$H_A : (X_+^{[S]}|[0, 0, 1]) \underset{st}{>} (X_+^{[S]}|[1, 1, 0])$$

The KS-test was performed with the `ks.test`<sup>3</sup> function in R (R Development Core Team, 2013). The test statistics are  $D^- = 0.0002$ ,  $p = .9998$ ;  $D^- = 0.0829$ ,  $p < .001$ ; and  $D^- = 0.0016$ ,  $p = .9792$ , respectively.

### 3.5 Discussion

In the present study, the minimal conditions for ordinal inferences about individuals are considered. It was shown that common nonparametric models, which are known for their ordering properties (i.e., SOL), are not fully satisfactory for the purpose of measurement at the level of individuals. The

<sup>3</sup>This function computes the classical KS-test for continuous distributions, and therefore does not allow for ties. However, alternative analyses with the `ks.boot` function from the `Matching` package (Sekhon, 2011), a function that allows for ties, show similar results.

reason was that the ordering based on the sum score is not always in accordance with the ordering based on the complete response pattern. In order to guarantee this accordance, OS has been defined as a minimal condition for fair scoring of individuals.

This aspect of fairness should be distinguished from measurement error and the asymptotic behavior of a statistic. It could be that the ordinal inferences change if the test administration is extended or repeated. However, these changes are then based on additional information. OS refers to inferences based on all available information.

OS is a property that can hold for any scoring rule, but this study only focused on the sum score. It has been shown that OS of the sum score need not hold for the NOM, but that it does for the RM as well as for 2PLMs with a relatively homogeneous set of discrimination parameters. The latter case implies that ignoring the weights in the scoring rule need not have an effect on ordinal inferences.

In Section 3.3.2, it was shown that the MHM, as well as the extensions with IIO or/and MTR, does not imply OS of the sum score. However, this does not mean that these models are useless. The property that the latent trait is stochastically ordered by the sum score is, for instance, very useful in survey applications. It implies that the means (or other statistics of central tendency) of the posterior distributions of  $\theta$  are ordered in accordance with the sum score. This says that people with a higher sum score have *on average* a larger ability compared to people with a lower sum score, and therefore groups of people can be ordered based on the sum score.

The introduction of OS and the npRM leaves some topics for further research. The first is about model fit. It was shown in section 3.4 that OS has testable implications. However, the proposed procedure contains many pairwise subtests. For instance, for a test with ten items, 29,002 subtests (!) have to be performed on the same data. Maybe, the procedure could be reduced to those subtests that provide the most information about the null-hypothesis that the sum score is ordinal sufficient. This topic needs further study.

The second is how to equate two tests that both have an ordinal sufficient score. In other words, how do these score distributions relate to each other?

The final comment is about deriving other OS test statistics. Ordinal

sufficiency is a property that can hold for any scoring rule. And for any scoring rule provided, the test described above can be used in order to determine whether that scoring rule is ordinal sufficient or not. However, this approach can also be used in the reverse direction, i.e., it can be used in order to find the scoring rule that is ordinal sufficient for a particular test. For monotone latent variable models, there always exists an ordinal sufficient statistic for the latent trait. For instance, the statistic that assigns the value 0 to those who made all items incorrect, the value 1 to those who made some items incorrect and some items correct, and the value 2 to those who made all items correct. This example is practically of limited value, however, it demonstrates that one can look for a statistic that assigns examinees to categories, such that the ordering between categories is ordinal sufficient. This also demonstrates that OS is as condition a good deal weaker than sufficiency. Whereas most IRT models do not allow for a sufficient statistic, they all admit of (at least one) OS statistic.

## Appendix

Let  $n$  be the number of items in  $\mathbf{z}$ , and define

$$\begin{aligned} X_+ &= \sum_i X_i \\ Y_+ &= \sum_h Y_h \\ Z_+ &= \sum_g Z_g \end{aligned}$$

with realizations  $x_+$ ,  $y_+$  and  $z_+$ , respectively.

Following from the partitioning of  $\mathbf{x}_1$  and  $\mathbf{x}_2$ ,

$$\begin{aligned} x_{+2} &= y_+ + z_+, \\ x_{+1} &= y_+ + (n - z_+). \end{aligned}$$

In cases where

$$x_{+2} > x_{+1},$$

we obtain

$$\begin{aligned} z_+ &> (n - z_+), \\ z_+ &> \frac{n}{2}. \end{aligned}$$

Now we consider the posterior distribution

$$f(\theta|\mathbf{X} = \mathbf{x}) = \frac{\prod_i P(x_i|\theta)^{x_i} [1 - P(x_i|\theta)]^{1-x_i} f(\theta)}{P(\mathbf{X} = \mathbf{x})}.$$



The likelihood ratio can be written as

$$\begin{aligned}
\frac{f(\theta|\mathbf{X} = \mathbf{x}_2)}{f(\theta|\mathbf{X} = \mathbf{x}_1)} &= \frac{\prod_i P(x_{i2}|\theta)^{x_{i2}} [1 - P(x_{i2}|\theta)]^{1-x_{i2}} f(\theta)}{P(\mathbf{X} = \mathbf{x}_2)} \\
&= \frac{\prod_i P(x_{i1}|\theta)^{x_{i1}} [1 - P(x_{i1}|\theta)]^{1-x_{i1}} f(\theta)}{P(\mathbf{X} = \mathbf{x}_1)} \\
&= \frac{\prod_i P(x_{i2}|\theta)^{x_{i2}} [1 - P(x_{i2}|\theta)]^{1-x_{i2}}}{P(\mathbf{X} = \mathbf{x}_1)} \frac{P(\mathbf{X} = \mathbf{x}_2)}{\prod_i P(x_{i1}|\theta)^{x_{i1}} [1 - P(x_{i1}|\theta)]^{1-x_{i1}}} \\
&= \prod_i \frac{P(x_{i2}|\theta)^{x_{i2}} [1 - P(x_{i2}|\theta)]^{1-x_{i2}}}{P(x_{i1}|\theta)^{x_{i1}} [1 - P(x_{i1}|\theta)]^{1-x_{i1}}} \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \\
&= \prod_h \frac{P(y_h|\theta)^{y_h} [1 - P(y_h|\theta)]^{1-y_h}}{P(y_h|\theta)^{y_h} [1 - P(y_h|\theta)]^{1-y_h}} \prod_g \frac{P(z_g|\theta)^{z_g} [1 - P(z_g|\theta)]^{1-z_g}}{P(z_g|\theta)^{1-z_g} [1 - P(z_g|\theta)]^{z_g}} \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \\
&= \prod_g \left( \frac{P(z_g|\theta)}{1 - P(z_g|\theta)} \right)^{z_g} \left( \frac{1 - P(z_g|\theta)}{P(z_g|\theta)} \right)^{1-z_g} \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)}.
\end{aligned}$$

The natural logarithm of likelihood ratio is

$$\begin{aligned}
\log \left( \frac{f(\theta|\mathbf{X} = \mathbf{x}_2)}{f(\theta|\mathbf{X} = \mathbf{x}_1)} \right) &= \log \left( \prod_g \left( \frac{P(z_g|\theta)}{1 - P(z_g|\theta)} \right)^{z_g} \left( \frac{1 - P(z_g|\theta)}{P(z_g|\theta)} \right)^{1-z_g} \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \right) \\
&= \log \left( \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \right) + \sum_g z_g \log \left( \frac{P(z_g|\theta)}{1 - P(z_g|\theta)} \right) + \sum_g (1 - z_g) \log \left( \frac{1 - P(z_g|\theta)}{P(z_g|\theta)} \right) \\
&= \log \left( \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \right) + \sum_g z_g \log \left( \frac{P(z_g|\theta)}{1 - P_g(z_g|\theta)} \right) + \sum_g (z_g - 1) \log \left( \frac{P(z_g|\theta)}{1 - P(z_g|\theta)} \right) \\
&= \log \left( \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \right) + \sum_g (2z_g - 1) \log \left( \frac{P(z_g|\theta)}{1 - P(z_g|\theta)} \right). \tag{3.10}
\end{aligned}$$

It is generally known that if

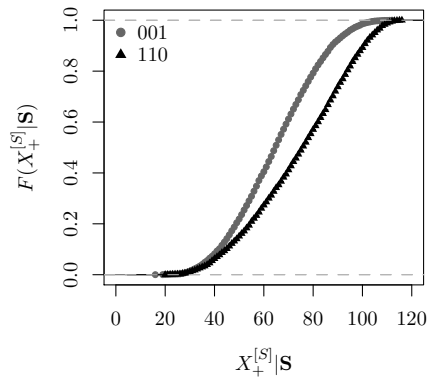
$$\frac{f(\theta_2|\mathbf{X} = \mathbf{x}_2)}{f(\theta_2|\mathbf{X} = \mathbf{x}_1)} > \frac{f(\theta_1|\mathbf{X} = \mathbf{x}_2)}{f(\theta_1|\mathbf{X} = \mathbf{x}_1)},$$

then

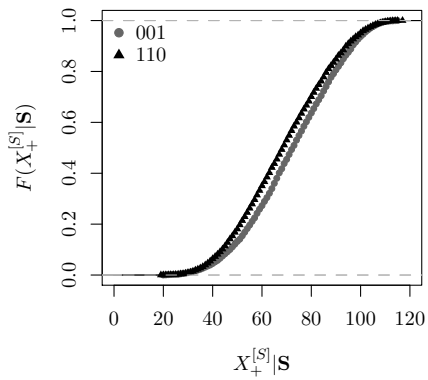
$$\log \left( \frac{f(\theta_2|\mathbf{X} = \mathbf{x}_2)}{f(\theta_2|\mathbf{X} = \mathbf{x}_1)} \right) > \log \left( \frac{f(\theta_1|\mathbf{X} = \mathbf{x}_2)}{f(\theta_1|\mathbf{X} = \mathbf{x}_1)} \right).$$

Now, following from (3.10), the likelihood ratio can be written as

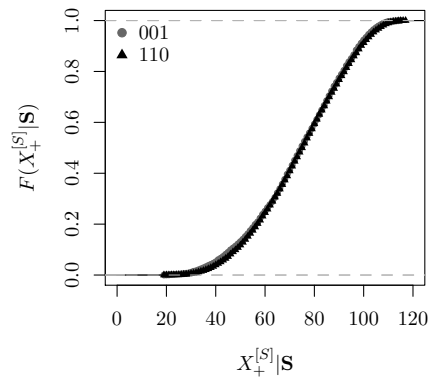
$$\begin{aligned}
& \log \left( \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \right) + \sum_g (2z_g - 1) \log \left( \frac{P(z_g | \theta_2)}{1 - P(z_g | \theta_2)} \right) > \\
& \log \left( \frac{P(\mathbf{X} = \mathbf{x}_1)}{P(\mathbf{X} = \mathbf{x}_2)} \right) + \sum_g (2z_g - 1) \log \left( \frac{P(z_g | \theta_1)}{1 - P(z_g | \theta_1)} \right) \\
& \quad \downarrow \\
& \sum_g (2z_g - 1) \log \left( \frac{P(z_g | \theta_2)}{1 - P(z_g | \theta_2)} \right) > \sum_g (2z_g - 1) \log \left( \frac{P(z_g | \theta_1)}{1 - P(z_g | \theta_1)} \right) \\
& \quad \downarrow \\
& \sum_g (2z_g - 1) \log \left( \frac{P(z_g | \theta_2)}{1 - P(z_g | \theta_2)} \right) - \sum_g (2z_g - 1) \log \left( \frac{P(z_g | \theta_1)}{1 - P(z_g | \theta_1)} \right) > 0 \\
& \quad \downarrow \\
& \sum_g (2z_g - 1) \left[ \log \left( \frac{P(z_g | \theta_2)}{1 - P(z_g | \theta_2)} \right) - \log \left( \frac{P(z_g | \theta_1)}{1 - P(z_g | \theta_1)} \right) \right] > 0 \\
& \quad \downarrow \\
& \sum_g (2z_g - 1) \log \left[ \frac{\left( \frac{P(z_g | \theta_2)}{1 - P(z_g | \theta_2)} \right)}{\left( \frac{P(z_g | \theta_1)}{1 - P(z_g | \theta_1)} \right)} \right] > 0 \\
& \quad \downarrow \\
& \sum_g z_g \log \left[ \frac{\left( \frac{P(z_g | \theta_2)}{1 - P(z_g | \theta_2)} \right)}{\left( \frac{P(z_g | \theta_1)}{1 - P(z_g | \theta_1)} \right)} \right] > \sum_g (1 - z_g) \log \left[ \frac{\left( \frac{P(z_g | \theta_2)}{1 - P(z_g | \theta_2)} \right)}{\left( \frac{P(z_g | \theta_1)}{1 - P(z_g | \theta_1)} \right)} \right], \\
& \quad \theta_2 > \theta_1, \quad z_+ > \frac{n}{2}.
\end{aligned}$$



(a)



(b)



(c)

Figure 3.6: The ecdfs of  $(X_+^{[S]}|[0, 0, 1])$  and  $(X_+^{[S]}|[1, 1, 0])$ .  $\mathbf{S}$  contains: item 1, 2, and 3 (a); 1, 6, and 10 (b); 1, 6, and 8 (c).