Springer Proceedings in Mathematics & Statistics

Roger E. Millsap
Daniel M. Bolt
L. Andries van der Ark
Wen-Chung Wang *Editors*

Quantitative Psychology Research

The 78th Annual Meeting of the Psychometric Society

Springer

Springer Proceedings in Mathematics & Statistics

Volume 89

More information about this series at http://www.springer.com/series/10533

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of select contributions from workshops and conferences in all areas of current research in mathematics and statistics, including OR and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Roger E. Millsap • Daniel M. Bolt
L. Andries van der Ark • Wen-Chung Wang
Editors

Quantitative Psychology Research

The 78th Annual Meeting of the Psychometric Society

Springer

Editors
Roger E. Millsap
Department of Psychology
Arizona State University
Tempe, AZ, USA

Daniel M. Bolt
Department of Educational Psychology
University of Wisconsin
Madison, USA

L. Andries van der Ark
University of Amsterdam
Amsterdam, The Netherlands

Wen-Chung Wang
Department of Psychological Studies
The Hong Kong Institute
of Education, Hong Kong
Hong Kong SAR

ISSN 2194-1009 ISSN 2194-1017 (electronic)
ISBN 978-3-319-07502-0 ISBN 978-3-319-07503-7 (eBook)
DOI 10.1007/978-3-319-07503-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014954131

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume represents presentations given at the 78th annual meeting of the Psychometric Society, organized by Cito and held at the Musis Sacrum in Arnhem, the Netherlands, during July 22–26, 2013. The meeting attracted 334 participants from 28 countries, with 242 papers being presented, along with 49 poster presentations, five pre-conference workshops, three keynote presentations, six invited presentations, six state-of-the-art lecturers, and three invited symposia. We thank the local organizer Anton Béguin and his staff and students for hosting this very successful conference.

After the 77th meeting in Lincoln, Nebraska, the idea was presented to publish a proceedings volume from the conference so as to allow presenters to quickly make their ideas available to the wider research community, while still undergoing a thorough review process. Because the first volume was received successfully, it was suggested that we publish proceedings more regularly. Hence, this is the second volume, and a third volume following the 79th meeting in Madison, Wisconsin, is expected.

We asked authors to use their presentation at the meeting as the basis of their chapters, possibly extended with new ideas or additional information. The result is a selection of 29 state-of-the-art chapters addressing a diverse set of topics, including classical test theory, item response theory, factor analysis, measurement invariance, test equating and linking, mediation analysis, cognitive diagnostic models, marginal models, and multi-level models.

The joy of editing these proceedings was overshadowed by the tragic news that Roger E. Millsap had passed away suddenly on May 9, 2014. As editor of Psychometrika and former president, Roger played an important role in the Psychometric Society. He was also the initiator and principal editor of the proceedings. He passed away shortly after finalizing these proceedings. We will always remember him fondly as the driving force of this project, and we will miss the friendly, helpful, and competent advice of this well-seasoned editor. May you rest in peace Roger.

Amsterdam, The Netherlands — L. Andries van der Ark
Madison, WI, USA — Daniel M. Bolt
Hong Kong — Wen-Chung Wang

Contents

Chapter 1
What Do You Mean by a Difficult Item? On the Interpretation of the Difficulty Parameter in a Rasch Model

Ernesto San Martín and Paul De Boeck

Abstract Three versions of the Rasch model are considered: the fixed-effects model, the random-effects model with normal distribution, and the random-effects model with unspecified distribution. For each of the three, we discuss the meaning of the difficulty parameter starting each time from the corresponding likelihood and the resulting identified parameters. Because the likelihood and the identified parameters are different depending on the model, the identification of the parameter of interest is also different, with consequences for the meaning of the item difficulty. Finally, for all the three models, the item difficulties are monotonically related to the marginal probabilities of a correct response.

1.1 Introduction

In the Rasch model, the probability of success in an item is defined on the basis of a contribution from the part of the person (person ability) and a contribution from the part of the item (item difficulty), while the contribution from the part of the persons does not depend on the item and neither does the effect of the items depend on the person. The Rasch model is, therefore, a main-effect model. The basic formula is the following:

$$Y_{pi} \sim \text{Bern}\left[\Psi(\theta_p - \beta_i)\right], \tag{1.1}$$

E. San Martín (✉)
Faculty of Mathematics and Faculty of Education, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile

Center for Operations Research and Econometrics CORE, 30 Voie du Roman Pays, Louvain-la-Neuve, Belgium
e-mail: esanmart@mat.puc.cl

P. De Boeck
Department of Psychology, The Ohio State University,
1835 Neil Avenue Columbus, Columbus, OH 43210, USA
e-mail: deboeck.2@osu.edu

R.E. Millsap et al. (eds.), *Quantitative Psychology Research*, Springer Proceedings in Mathematics & Statistics 89, DOI 10.1007/978-3-319-07503-7_1

where $\Psi(x) = \exp(x)/(1+\exp(x))$, θ_p is the effect of the person on the probability, also called ability, and β_i is the effect of the item on the probability, also called the difficulty.

Different choices can be made for how the effects of the persons are considered. Either the persons are modeled with fixed-effects (FE) or with random-effects (RE), and for these random-effects one can either specify the distribution, for example, the normal distribution (RE-N), or one can leave the distribution unspecified (RE-U). The three models have led to three different likelihood functions (or sampling probabilities) and, accordingly, to three different ways to estimate the corresponding parameters: joint maximum likelihood (JML) in the case of the FE model, parametric marginal maximum likelihood (MML) in the case of the RE-N model, and semi-parametric MML in the case of the RE-U model.

It is our purpose to infer the consequences these choices have for the meaning of the other parameter of the model, β_i, the item difficulty. We will make this inference from the likelihood for each of the three types of models. This approach is justified by the fact that the likelihood function is supposed to generate the responses patterns and, therefore, it provides the statistical meaning of the parameters indexing it; for details, see Bamber and Van Santen (2000), Cox and Hinkley (1974), Fisher (1922), and McCullagh (2002). Consequently, the inference proceeds in two steps:

1. One step for the *identified parameterization*, which is the parameterization as far as possible on the basis of the likelihood.
2. Another step for the further identification of the *parameter of interest*, which will require to establish an injective relationship (under constraints, if necessary) between the parameter of interest and the identified parameterization.

It will be shown that the β_i parameters have a different meaning in the three Rasch models. The identified parameterization differs and also the identification of the parameter of interest is different. Furthermore, we will also discuss a condition based on marginal probabilities under which the difficulty of an item i is larger than the difficulty of an item j; this empirical condition can also be related to the empirical difficulty of an item (that is, the proportion of persons correctly answering the item).

In order to differentiate between the three models, item difficulty parameters of the three models will be denoted with different symbols: β_i^{FE}, $\beta_i^{\mathrm{RE\text{-}N}}$, and $\beta_i^{\mathrm{RE\text{-}U}}$. It will be shown that the meaning of these three parameters depends on the choice that is made for how to treat the effects from the part of the persons and the assumptions one is making regarding these effects.

1.2 Fixed-Effects Specification

1.2.1 Assumptions

The specification of a Rasch model rests on the following two assumptions:

Assumption 1: $\{Y_{pi} : p = 1,\ldots,N; i = 1,\ldots,I\}$ are mutually independent.

Assumption 2: For each person p and each item i, $Y_{pi} \sim \text{Bern}(\pi_{pi})$, where $\pi_{pi} = \Psi(\theta_p - \beta_i^{\text{FE}})$ and $\Psi(x) = \exp(x)/(1+\exp(x))$.

1.2.2 Likelihood and Identified Parameters

These assumptions induce the following likelihood function:

$$P^{(\theta_{1:N},\beta_{1:I}^{\text{FE}})}(\mathbf{Y}_1 = \mathbf{y}_1 \ldots \mathbf{Y}_N = \mathbf{y_N}) = \prod_{p=1}^{N}\prod_{i=1}^{I} \pi_{pi}^{y_{pi}}(1-\pi_{pi})^{1-y_{pi}}$$
$$= \prod_{p=1}^{N}\prod_{i=1}^{I} \frac{\exp[y_{pi}(\theta_p - \beta_i)]}{1+\exp(\theta_p - \beta_i)},$$

where $\mathbf{Y}_p = (Y_{p1},\ldots,Y_{pI})^\top \in \{0,1\}^I$, $\theta_{1:N} = (\theta_1,\ldots,\theta_N)$, and similarly for $\beta_{1:I}$.

The parameter of a Bernoulli distribution is identified. This fact, together with Assumption 1, implies that the identified parameters are $\{\pi_{pi} : p = 1,\ldots,P; i = 1,\ldots,I\}$.

1.2.3 Parameters of Interest

The problem now is to identify the parameter of interest $(\theta_{1:N}, \beta_{1:I}^{\text{FE}})$, which means to write them as functions of the identified parameters. From Assumption 2, it follows that

$$\theta_p - \beta_i^{\text{FE}} = \ln\left[\frac{\pi_{pi}}{1-\pi_{pi}}\right], \quad p = 1,\ldots,N; i = 1,\ldots,I;$$
$$\beta_i - \beta_j = \ln\left[\frac{1-\pi_{pi}}{\pi_{pi}}\,\frac{\pi_{pj}}{1-\pi_{pj}}\right], \quad \text{for all person } p \text{ and } i \neq j.$$

These relationships show that $\{\theta_p - \beta_i^{\text{FE}} : p = 1,\ldots,N; i = 1,,\ldots,I\}$ as well as $\{\beta_i^{\text{FE}} - \beta_j^{\text{FE}} : i = 1,\ldots,I,\ j = i,\ldots,I\}$ are identified since they are written as functions of identified parameters. Therefore, the parameters of interest $(\theta_{1:N}, \beta_{1:I}^{\text{FE}})$ are identified if one identification restriction is imposed. Two possibilities can be considered:

1. To restrict one person parameter, namely $\theta_1 = 0$. Under this restriction, the difficulty parameter becomes

$$\beta_i^{\mathrm{FE}} = \ln\left(\frac{1-\pi_{1i}}{\pi_{1i}}\right),$$

that is, the logarithm of the ratio between the probability that person 1 incorrectly answers item i and the probability that person 1 correctly answers the item.

2. To restrict one item parameter, namely $\beta_1^{\mathrm{FE}} = 0$. Under this restriction, the difficulty parameter becomes

$$\beta_i^{\mathrm{FE}} = \ln\left(\frac{1-\pi_{pi}}{\pi_{pi}}\,\frac{\pi_{p1}}{1-\pi_{p1}}\right), \tag{1.2}$$

that is, the logarithm of the odd ratio between item 1 and item i for each person p.

The first restriction depends on a specific person who is present in one application of the test. Therefore, the second identification restriction is more convenient since we may apply the same test to various sets of persons; see Andersen (1980).

1.2.4 Relationship of Item Difficulty with Empirical Difficulty

Regarding the relationship between item difficulty and empirical difficulty, from (1.2) it follows that

$$\beta_i^{\mathrm{FE}} > \beta_j^{\mathrm{FE}} \iff P^{(\theta_p,\beta_{1:I}^{\mathrm{FE}})}(Y_{pi}=1) < P^{(\theta_p,\beta_{1:I}^{\mathrm{FE}})}(Y_{pj}=1) \quad \text{for all persons } p. \tag{1.3}$$

Thus, item i is more difficult than item j if the probability that the person correctly answers item i is less than the probability that a person correctly answers item j.

1.2.5 Comments

The previous considerations lead to the following comments:

1. The fixed-effects specification of the Rasch model is a rather easy model from the perspective of identification, easier than the other two specifications, and it is therefore often implicitly used to interpret the parameters of the Rasch model, even when one is interesting in is the random-effects specification; for more discussion, see San Martín and Rolin (2013).
2. On the other hand, for an estimation of the model, mostly the assumption of a random-effects model is made, because the maximum likelihood estimator of the difficulty parameters is inconsistent due to the presence of the incidental parameters. For details, see Andersen (1980), Ghosh (1995), and Lancaster (2000).

1.3 Random-Effects Specification

The random-effects assumption for the persons leads to consider the ability parameters as realizations of an iid process. Using the statistical jargon, the person's abilities are now considered as random-effects.

1.3.1 Assumptions

A random-effects specification rests on the following assumptions:

Assumption 1: $\{\mathbf{Y}_p : p = 1,\ldots,N\}$ are mutually independent conditionally on $\theta_{1:N}$.

Assumption 2: For each person p, the conditional distribution of $\mathbf{Y}_p$ given $\theta_{1:N}$ only depends on θ_p and it is parameterized by $\beta_{1:I}^{\text{RE-N}}$.

Assumption 3: For each person p, $\{Y_{pi} : i = 1,\ldots,I\}$ are mutually independent conditionally on θ_p. This is the so-called axiom of local independence.

Assumption 4: For each item i, $(Y_{pi} \mid \theta_p) \sim \text{Bern}\,[\Psi(\theta_p - \beta_i^{\text{RE-N}})]$.

Assumption 5: θ_p's are mutually independent and identically distributed, with a common distribution $\mathcal{N}(0,\sigma^2)$.

1.3.2 Likelihood and Identified Parameters

These assumptions imply that the response patterns $\mathbf{Y}_p$'s are mutually independent and identically distributed. To describe the likelihood function, it is enough to describe the probability of each of the 2^I response patterns, namely

$$
\begin{aligned}
q_{12\cdots I} &= P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{p1}=1, Y_{p2}=1,\ldots,Y_{p,I-1}=1, Y_{pI}=1),\\
q_{12\cdots \bar{I}} &= P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{p1}=1, Y_{p2}=1,\ldots,Y_{pI,}-1=1, Y_{pI}=0),\\
\vdots\quad & \quad \vdots \\
q_{\bar{1}\bar{2}\cdots \bar{I}} &= P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{p1}=0, Y_{p2}=0,\ldots,Y_{pI,}-1=0, Y_{pI}=0),
\end{aligned}
\tag{1.4}
$$

and

$$
P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(\mathbf{Y}_p = \mathbf{y}) = \int_{-\infty}^{\infty} \prod_{i=1}^{I} \frac{\exp[y_i\,(\sigma\theta - \beta_i^{\text{RE-N}})]}{1+\exp(\sigma\theta - \beta_i^{\text{RE-N}})}\,\phi(\theta)\,d\theta,
$$

with $\phi(\cdot)$ as the density of a standard normal distribution. Therefore, the likelihood function corresponds to a multinomial distribution $\text{Mult}(2^I, \mathbf{q})$, where $\mathbf{q} = (q_{12\cdots I}, q_{12\cdots I-1,\bar{I}}, \cdots, q_{\bar{1}\bar{2}\cdots\bar{I}})$. Consequently, the parameter $\mathbf{q}$ is the identified parameter.

1.3.3 Parameters of Interest

It is possible to prove that $\beta_{1:I}^{RE-N}$ and σ can be written in terms of the identified parameter $\mathbf{q}$ without restrictions on the item parameters. The proof rests on the following arguments:

1. Let

$$\alpha_i \doteq P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{pi}=1) = \int_{-\infty}^{\infty} \Psi(\sigma\theta - \beta_i^{\text{RE-N}})\phi(\theta)\,d\theta \doteq p(\sigma, \beta_i^{\text{RE-N}}).$$

   The parameter α_i is an identified parameter because it is a function of $\mathbf{q}$.
2. The function $p(\sigma, \beta_i^{\text{RE-N}})$ is a strictly decreasing continuous function of $\beta_i^{\text{RE-N}}$. Therefore it is invertible and consequently

$$\beta_i^{\text{RE-N}} = p^{-1}(\sigma, \alpha_i). \tag{1.5}$$

3. For $i \neq j$, let

$$\alpha_{ij} \doteq P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{pi}=1, Y_{pj}=1) = \int_{-\infty}^{\infty} \Psi(\sigma\theta - \beta_i^{\text{RE-N}})\,\Psi(\sigma\theta - \beta_j^{\text{RE-N}})\,\phi(\theta)d\theta.$$

   The parameter α_{ij} is also an identified parameter because it is a function of $\mathbf{q}$. Using (1.5), it follows that

$$\begin{aligned}\alpha_{ij} &= \int_{-\infty}^{\infty} \Psi(\sigma\theta - p^{-1}(\sigma,\alpha_i))\,\Psi(\sigma\theta - p^{-1}(\sigma,\alpha_j))\,\phi(\theta)\,d\theta.\\ &\doteq r(\sigma, \alpha_i,, \alpha_j).\end{aligned}$$

   It can be shown that $r(\sigma, \alpha_i, \alpha_j)$ is a strictly increasing continuous function of σ; for details, see San Martín and Rolin (2013). It follows that

$$\sigma = r^{-1}(\alpha_{ij}, \alpha_i, \alpha_j). \tag{1.6}$$

1.3.4 Relationship of Item Difficulty with Empirical Difficulty

Regarding the relationship between item difficulty and empirical difficulty, from (1.5) it follows that

$$\beta_i^{\text{RE-N}} > \beta_j^{\text{RE-N}} \Longleftrightarrow P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{pi}=1) < P^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{pj}=1) \quad \text{for all person } p. \tag{1.7}$$

Thus, item i is more difficult than item j if probability that a person correctly answers item i is less than the probability that the person correctly answers item j.

1.3.5 Comments

The previous considerations lead to the following comments:

1. For each person p, the responses are positively correlated, that is,

$$cov^{(\beta_{1:I}^{\text{RE-N}},\sigma)}(Y_{pi},Y_{pj}) > 0$$

   for $i \neq j$. This is a marginal correlation and it follows from both Assumption 3 and the strict monotonicity of $\Psi(\theta_p - \beta_i^{\text{RE-N}})$ as a function of θ_p for all $\beta_i^{\text{RE-N}}$.
2. According to equality (1.6), σ represents the dependency between items i and j induced by both the marginal probabilities α_i and α_j and the joint marginal probability α_{ij}. Furthermore, this dependency is the same for all pairs of items since equality (1.6) is valid for all pairs of items i and j.
3. The item difficulty $\beta_i^{\text{RE-N}}$ is not only a function of the marginal probability α_i of correctly answering the item i, but also of terms of the common dependency.
4. The previous identification analysis is valid in the case $\pi_{pi} = \Phi(\theta_p - \beta_i^{\text{RE-N}})$, where Φ is the distribution function of a standard normal distribution; see San Martín and Rolin (2013). In this case, it is possible to show that

$$\alpha_i = \Phi\left(-\frac{\beta_i^{\text{RE-N}}}{\sqrt{1+\sigma^2}}\right).$$

   Therefore, the difficulty parameter $\beta_i^{\text{RE-N}}$ can be written as

$$\beta_i^{\text{RE-N}} = -\sqrt{1+\sigma^2}\,\Phi^{-1}(\alpha_i). \tag{1.8}$$

   It follows that the difficulty parameter $\beta_i^{\text{RE-N}}$ is decreasing with σ. In other words, the larger the individual differences, the more extreme the difficulty parameters become.
5. There is not an explicit function as (1.8) for the logistic model, but approximately the same equation applies with σ^2 premultiplied by $16\sqrt{3}/(15\pi)$; see Molenberghs et al. (2010), Zeger et al. (1988).
6. The distribution of θ_p can be specified as a $\mathcal{N}(\mu,\sigma^2)$. In this case, the identified parameters are $(\widetilde{\beta}_{1:I}^{\text{RE-N}},\sigma)$, where $\widetilde{\beta}_i^{\text{RE-N}} \doteq \beta_i^{\text{RE-N}} - \mu$. In order to identify the difficulty parameters $\beta_{1:I}^{\text{RE-N}}$ and the scale parameter μ, it is enough to introduce a linear restriction on the item parameters $\beta_{1:I}^{\text{RE-N}}$.

1.4 Semiparametric Specification

As pointed out by Woods and Thissen (2006) and Woods (2006), there exist specific fields, such as personality and psychopathology, in which the normality assumption of θ_p is not realistic. For instance, psychopathology and personality latent variables are likely to be positively skewed, because most persons in the general population have low pathology, and fewer persons have severe pathology. However, the distribution G of θ_p is unobservable and, consequently, though a researcher may hypothesize about it, it is not known in advance of an analysis. Therefore, any a priori parametric restriction on the shape of the distribution G could be considered as a mis-specification.

1.4.1 Assumptions

These considerations lead to extend parametric Rasch models by considering the distribution G as a parameter of interest, and thus specifying semi-parametric Rasch models. These models rest on the following assumptions:
Assumptions 1–4 as in the random-effects specification.

Assumption 5: θ_p's are mutually independent and identically distributed, with a common unspecified distribution G.

1.4.2 Likelihood and Identified Parameters

As in the random-effects specification, these assumptions imply that the response patterns $\mathbf{Y}_p$'s are mutually independent and identically distributed, with a common multinomial distribution Mult$\,(2^I,\mathbf{q})$, with $\mathbf{q} = (q_{12\cdots I}, q_{12\cdots I-1,\bar{I}}, \cdots, q_{\bar{1}\bar{2}\cdots\bar{I}})$, where

$$\begin{aligned}
q_{12\cdots I} &= P^{(\beta^{\text{RE-U}}_{1:I},G)}(Y_{p1}=1, Y_{p2}=1, \ldots, Y_{p,I-1}=1, Y_{pI}=1),\\
q_{12\cdots \bar{I}} &= P^{(\beta^{\text{RE-U}}_{1:I},G)}(Y_{p1}=1, Y_{p2}=1, \ldots, Y_{pI,}-1=1, Y_{pI}=0),\\
\vdots\quad & \vdots\\
q_{\bar{1}\bar{2}\cdots\bar{I}} &= P^{(\beta^{\text{RE-U}}_{1:I},G)}(Y_{p1}=0, Y_{p2}=0, \ldots, Y_{pI,}-1=0, Y_{pI}=0),
\end{aligned}$$

and

$$P^{(\beta^{\text{RE-U}}_{1:I},G)}(\mathbf{Y}_p=\mathbf{y}) = \int \prod_{i=1}^{I} \frac{\exp[y_i\,(\theta-\beta^{\text{RE-U}}_i)]}{1+\exp(\theta-\beta^{\text{RE-U}}_i)}\,G(d\theta). \tag{1.9}$$

Therefore, the likelihood function is parametrized by **q**, which corresponds to the identified parameter.

Following San Martín et al. (2011), Equation (1.9) can be rewritten as follows: for all $\mathscr{J} \subset \{1,\ldots,I\} \setminus \emptyset$,

$$P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\bigcap_{j\in\mathscr{J}^c}\{Y_{pj}=0\}\right)=$$

$$=\exp\left(-\sum_{j\in\mathscr{J}}\beta^{\text{RE-U}}_j\right)\times\int_{-\infty}^{\infty}\frac{e^{|\mathscr{J}|\theta}}{\prod\limits_{1\le i\le I}(1+e^{\theta-\beta^{\text{RE-U}}_i})}G(d\theta). \tag{1.10}$$

By taking (1.10) with $\mathscr{J}=\{1\}$ and after (1.10) with $\mathscr{J}=\{i\}$, we identify $(\beta^{\text{RE-U}}_2-\beta^{\text{RE-U}}_1,\ldots,\beta^{\text{RE-U}}_I-\beta^{\text{RE-U}}_1)$ because

$$\beta^{\text{RE-U}}_j-\beta^{\text{RE-U}}_1=\ln\left[\frac{P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=1\}\cap\bigcap\limits_{2\le i\le I}\{Y_{pi}=0\}\right)}{P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{pj}=1\}\cap\bigcap\limits_{i\neq j}\{Y_{pi}=0\}\right)}\right]. \tag{1.11}$$

Not only the item differences can be identified, but also some characteristics of the distribution G. As a matter of fact, working with the identified parameters $\beta^{\text{RE-U}}_j-\beta^{\text{RE-U}}_1$ leads to a shift of θ which we express with $u=\theta+\beta^{\text{RE-U}}_1$. Thus, for all $\mathscr{J}\subset\{1,\ldots,I\}$, (1.9) can be rewritten as

$$P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\bigcap_{j\in J^c}\{Y_{pj}=0\}\right)=$$

$$=e^{\left[-\sum_{j\in\mathscr{J}}(\beta^{\text{RE-U}}_j-\beta^{\text{RE-U}}_1)\right]}\int_{-\infty}^{\infty}\frac{e^{|\mathscr{J}|u}}{\prod\limits_{1\le i\le I}\left[1+e^{u-(\beta^{\text{RE-U}}_i-\beta^{\text{RE-U}}_1)}\right]}G_{\beta^{\text{RE-U}}_1}(du),$$

where $G_{\beta^{\text{RE-U}}_1}((-\infty,x])\doteq G((-\infty,x+\beta^{\text{RE-U}}_1])$. Therefore, the functionals

$$m_{G_{\beta^{\text{RE-U}}_1}}(k)=\int_{-\infty}^{\infty}\frac{e^{ku}}{\prod\limits_{1\le i\le I}\left[1+e^{u-(\beta^{\text{RE-U}}_i-\beta^{\text{RE-U}}_1)}\right]}G_{\beta^{\text{RE-U}}_1}(du),$$

for $k=0,1,\ldots,I$, are identified. Note that $m_{G_{\beta^{\text{RE-U}}_1}}(0)=m_{G_{\beta^{\text{RE-U}}_1}}(|\emptyset|)$ corresponds to $P^{(\beta^{\text{RE-U}}_{1:I},G)}(Y_{p1}=0,\ldots,Y_{pI}=0)$. Summarizing, the following $I+1$ relationships follow: For all $\mathscr{J}\subset\{1,\ldots,I\}$ such that $|\mathscr{J}|=k$,

$$m_{G_{\beta_1^{\text{RE-U}}}}(k) =$$

$$= P^{(\beta_{1:I}^{\text{RE-U}},G)}\left(\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\bigcap_{j\in J^c}\{Y_{pj}=0\}\right)\times e^{\left[\sum_{j\in\mathscr{J}}(\beta_j^{\text{RE-U}}-\beta_1^{\text{RE-U}})\right]} \quad (1.12)$$

for $k = 0, 1, \ldots, I$. These $I+1$ identified parameters will be used for an alternative way to identify the difficulties and to derive an interesting difficulty ratio.

1.4.3 Parameters of Interest

In order to identify the item parameters $\beta_{1:I}^{\text{RE-U}}$, the previous equalities suggest to introduce an identification restriction, namely $\beta_1^{\text{RE-U}} = 0$. Under this restriction, the difficulty parameters $\beta_i^{\text{RE-U}}$ are given by Eq. (1.11) with $\beta_1^{\text{RE-U}} = 0$. Moreover, using equalities (1.11) and (1.12) with $\beta_1^{\text{RE-U}} = 0$, the following relations follow:

$$\beta_j^{\text{RE-U}} = \ln\left[\frac{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{p1}=1, Y_{pj}=0)}{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{p1}=0, Y_{pj}=1)}\right] \quad \text{for all persons } p; \quad (1.13)$$

$$\frac{\beta_i^{\text{RE-U}}}{\beta_j^{\text{RE-U}}} = \ln\left[\frac{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pi}=0, Y_{pj}=1)}{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pi}=1, Y_{pj}=0)}\right] \quad \text{for all persons } p. \quad (1.14)$$

For a proof, see Appendix.

1.4.4 Relationship of Item Difficulty with Empirical Difficulty

Regarding the relationship between item difficulty and empirical difficulty, from (1.14) it is possible to prove that

$$\beta_i^{\text{RE-U}} > \beta_j^{\text{RE-U}} \iff P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pi}=1) < P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pj}=1) \quad \text{for all person } p. \quad (1.15)$$

Thus, item i is more difficult than item j if probability that a person correctly answers item i is less than the probability that the person correctly answers item j.

1.4.5 Comments

The previous considerations lead to the following comments:

1. Equalities (1.13) and (1.14) apply independent of the distribution G.
2. Equality (1.13) shows that the difficulty of an item j essentially corresponds to a ratio of probabilities involving two items: the item j itself and the item 1. This ratio can be interpreted as a *mirror property* between items 1 and j.
3. Equality (1.14) can also be interpreted as a *mirror property* between items i and j.

1.5 Discussion

In the random-effects specification of the Rasch model, it is not possible to make a distinction between a Rasch model with abilities distributed according to a $\mathscr{N}(0,\sigma^2)$ and a 2PL model with equal discriminations and abilities distributed according to a $\mathscr{N}(0,1)$. Both models are identified and, therefore, this is an example of equivalent models: the distribution generating the response patterns is not enough to distinguish between these two equivalent models. Let us remark that for the 2PL model with different discrimination parameters, the situation is different; for details and a first interpretation of the parameters of interest, see San Martín et al. (2013, Appendix B).

Relations (1.3), (1.7), and (1.15) suggest that the comparison between item difficulties can empirically be interpreted in terms of the proportion of persons answering correctly one or other item. This also suggests that the estimations of the difficulty parameters in the three models will be (almost) perfectly correlated. However, the *meaning* of these estimators is quite different. For the fixed-effects specification, *item difficulty* is interpreted in terms of odd ratio [see equality (1.2)]; for the random-effects specification, *item difficulty* is interpreted as a function of both the marginal probability of correctly answering the item and the dependency common to all pairs of items [see equalities (1.5) and (1.6)]; and for the semi-parametric specification, *item difficulty* is interpreted in terms of the mirror property (1.13).

Appendix

Proof of Equality (1.13)

Consider the reparameterization $\eta_i = \exp(\beta_i)$ and let

$$\mathscr{A}_{\mathscr{J}} = \bigcap_{j\in\mathscr{J}} \{Y_{pj}=1\} \ \cap \bigcap_{j\in\mathscr{J}^c\setminus\{1,i\}} \{Y_{pj}=0\}.$$

Let $\mathscr{J} \subset \{2,\ldots,I\}$ and $i \notin \mathscr{J}$. Using (1.12), it follows that

$$
\begin{aligned}
m_G(|\mathscr{J}|+1) &= P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{pi}=1\}\cap\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\bigcap_{j\in(\mathscr{J}\cup\{i\})^c}\{Y_{pj}=0\}\right)\times\prod_{j\in\mathscr{J}}\eta_j\times\eta_i\\
&= P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=1\}\cap\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\bigcap_{j\in\mathscr{J}^c\setminus\{1\}}\{Y_{pj}=0\}\right)\times\prod_{j\in\mathscr{J}}\eta_j.
\end{aligned}
$$

It follows that

$$
\eta_i=\frac{P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=1,Y_{pi}=0\}\cap\mathscr{A}_{\mathscr{J}}\}\right)}{P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=0,Y_{pi}=1\}\cap\mathscr{A}_{\mathscr{J}}\}\right)}. \tag{1.16}
$$

for all $\mathscr{J}\subset\{2,\ldots,I\}$ and $i\notin\mathscr{J}$. Therefore, using (1.16),

$$
\begin{aligned}
\frac{P^{(\beta^{\text{RE-U}}_{1:I},G)}(Y_{pi}=0,Y_{pj}=1)}{P^{(\beta^{\text{RE-U}}_{1:I},G)}(Y_{pi}=1,Y_{pj}=0)} &= \frac{\sum_{\{\mathscr{J}\subset\{2,\ldots,I\}:i\notin\mathscr{J}\}} P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=1,Y_{pi}=0\}\cap\mathscr{A}_{\mathscr{J}}\right)}{\sum_{\{\mathscr{J}\subset\{2,\ldots,I\}:i\notin\mathscr{J}\}} P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=0,Y_{pi}=1\}\cap\mathscr{A}_{\mathscr{J}}\right)}\\
&= \frac{\sum_{\{\mathscr{J}\subset\{2,\ldots,I\}:i\notin\mathscr{J}\}} \eta_i\, P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=0,Y_{pi}=1\}\cap\mathscr{A}_{\mathscr{J}}\right)}{\sum_{\{\mathscr{J}\subset\{2,\ldots,I\}:i\notin\mathscr{J}\}} P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{p1}=0,Y_{pi}=1\}\cap\mathscr{A}_{\mathscr{J}}\right)}\\
&= \eta_i.
\end{aligned}
$$

Proof of Equality (1.14)

Let $\mathscr{J}$ such that $|\mathscr{J}|=I-2$ and denote the label of two items excluded from $\mathscr{J}$ as i and i'. Using (1.12), it follows that

$$
\begin{aligned}
m_G(|\mathscr{J}\cup\{i\}|) &= P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{pi}=1\}\cap\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\{Y_{pi'}=0\}\cap\bigcap_{j\in\mathscr{J}^c\setminus\{i'\}}\{Y_{pj}=0\}\right)\times\\
&\quad\times\prod_{j\in\mathscr{J}}\eta_j\times\eta_i,
\end{aligned}
$$

$$
\begin{aligned}
m_G(|\mathscr{J}\cup\{i'\}|) &= P^{(\beta^{\text{RE-U}}_{1:I},G)}\left(\{Y_{pi'}=1\}\cap\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\{Y_{pi}=0\}\cap\bigcap_{j\in\mathscr{J}^c\setminus\{i\}}\{Y_{pj}=0\}\right)\times\\
&\quad\times\prod_{j\in\mathscr{J}}\eta_j\times\eta_{i'}.
\end{aligned}
$$

Therefore,

$$\frac{\eta_i}{\eta_{i'}} = \frac{P^{(\beta_{1:I}^{\text{RE-U}},G)}\left(\{Y_{pi}=0,Y_{pi'}=1\}\cap\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\bigcap_{j\in\mathscr{J}^c\setminus\{i\}}\{Y_{pj}=0\}\right)}{P^{(\beta_{1:I}^{\text{RE-U}},G)}\left(\{Y_{pi}=1,Y_{pi'}=0\}\cap\bigcap_{j\in\mathscr{J}}\{Y_{pj}=1\}\cap\bigcap_{j\in\mathscr{J}^c\setminus\{i'\}}\{Y_{pj}=0\}\right)}. \tag{1.17}$$

Let $\mathscr{J}\subset\{1,\dots,I\}$ such that $|\mathscr{J}|=I-2$ and take $i,i'\notin\mathscr{J}$. Denote

$$\mathscr{A}_{\mathscr{J}} = \bigcap_{j\in\mathscr{J}}\{Y_{pi}=1\} \quad\cap\bigcap_{j\in\mathscr{J}^c\setminus\{i\}}\{Y_{pi}=0\},$$

$$\mathscr{B}_{\mathscr{J}} = \bigcap_{j\in\mathscr{J}}\{Y_{pi}=1\} \quad\cap\bigcap_{j\in\mathscr{J}^c\setminus\{i'\}}\{Y_{pi}=0\}.$$

Then, using (1.17),

$$\begin{aligned}\frac{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pi}{=}0,Y_{pi'}{=}1)}{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pi{=}1},Y_{pi'}{=}0)} &= \frac{\sum_{\{\mathscr{J}\subset\{1,\dots,I\}:|\mathscr{J}=I-2,i,i'\notin\mathscr{J}\}} P^{(\beta_{1:I}^{\text{RE-U}},G)}(\{Y_{pi}{=}0,Y_{pi'}{=}1\}\cap\mathscr{A}_{\mathscr{J}})}{\sum_{\{\mathscr{J}\subset\{1,\dots,I\}:|\mathscr{J}=I-2,i,i'\notin\mathscr{J}\}} P^{(\beta_{1:I}^{\text{RE-U}},G)}(\{Y_{pi}{=}1,Y_{pi'}{=}0\}\cap\mathscr{B}_{\mathscr{J}})}\\ &= \frac{\eta_i}{\eta_{i'}}\frac{\sum_{\{\mathscr{J}\subset\{1,\dots,I\}:|\mathscr{J}=I-2,i,i'\notin\mathscr{J}\}} P^{(\beta_{1:I}^{\text{RE-U}},G)}(\{Y_{pi}{=}1,Y_{pi'}{=}0\}\cap\mathscr{B}_{\mathscr{J}})}{\sum_{\{\mathscr{J}\subset\{1,\dots,I\}:|\mathscr{J}=I-2,i,i'\notin\mathscr{J}\}} P^{(\beta_{1:I}^{\text{RE-U}},G)}(\{Y_{pi}{=}1,Y_{pi'}{=}0\}\cap\mathscr{B}_{\mathscr{J}})}\\ &= \frac{\eta_i}{\eta_{i'}}.\end{aligned}$$

Proof of Relation (1.15)

Using the same notation introduced above and the ratio $\eta_i'/\eta_{i'}$, it follows that

$$\begin{aligned}\frac{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pi'}=1)}{P^{(\beta_{1:I}^{\text{RE-U}},G)}(Y_{pi}=1)} &= \frac{\sum_{\{\mathscr{J}:|\mathscr{J}|=I-2,i,i'\notin\mathscr{J}\}} P^{(\beta_{1:I}^{\text{RE-U}},G)}(\{Y_{pi'}=1,Y_{pi}=0\}\cap\mathscr{A}_{\mathscr{J}})}{\sum_{\{\mathscr{J}:|\mathscr{J}|=I-2,i,i'\notin\mathscr{J}\}} P^{(\beta_{1:I}^{\text{RE-U}},G)}(\{Y_{pi'}=0,Y_{pi}=1\}\cap\mathscr{B}_{\mathscr{J}})}\\ &= \frac{\eta_i}{\eta_{i'}}.\end{aligned}$$

Acknowledgements This research was partially funded by the ANILLO Project SOC1107 *Statistics for Public Policy in Education* from the Chilean Government.

References

Andersen EB (1980) Discrete statistical models with social science applications. North Holland, Amsterdam

Bamber D, Van Santen JPH (2000) How to asses a model's testability and identifiability. J Math Psychol 44:20–40

Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman and Hall, London

Fisher RA (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc Lond A 222:309–368

Ghosh M (1995) Inconsistent maximum likelihood estimators for the Rasch model. Stat Probab Lett 23:165–170

Lancaster T (2000) The incidental parameter problem since 1948. J Econom 95:391–413

McCullagh P (2002) What is a statistical model? (with Discussion). Ann Stat 30:1225–1310

Molenberghs G, Verbeke G, Demetrio CGB, Vieira A (2010) A family of generalized linear models for repeated measures with normal and cojugate random effects. Stat Sci 25:325–347

San Martín E, Rolin J-M (2013) Identification of parametric Rasch-type models. J Stat Plan Inference 143:116–130

San Martín E, Jara A, Rolin J-M, Mouchart M (2011) On the Bayesian nonparametric generalization of IRT-types models. Psychometrika 76:385–409

San Martín E, Rolin J-M, Castro M (2013) Identification of the 1PL model with guessing parameter: parametric and semi-parametric results. Psychometrika 78:341–379

Woods CM (2006) Ramsay-curve item response theory (RC-IRT) to detect and correct for nonnormal latent variables. Psychol Methods 11:235–270

Woods CM, Thissen D (2006) Item response theory with estimation of the latent population distribution using spline-based densities. Psychometrika 71:281–301

Zeger SL, Liang K-Y, Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. Biometrics 44:1049–1060

Chapter 2
Thurstonian Item Response Theory and an Application to Attitude Items

Edward H. Ip

Abstract The assessment of attitudes has a long history dating back at least to the work of Thurstone. The Thurstonian approach had its "golden days," but today it is seldom used, partly because judges are needed to assess the location of an item, but also because of the emergence of contemporary tools such as the IRT. The current work is motivated by a study that assesses medical students' attitudes toward obese patients. During the item-development phase, the study team discovered that there were items on which the team members could not agree with regard to whether they represented positive or negative attitudes. Subsequently, a panel of $n = 201$ judges from the medical profession were recruited to rate the items, and the responses to the items were collected from a sample of $n = 103$ medical students. In the current work, a new methodology is proposed to extend the IRT for scoring student responses, and an affine transformation maps the judges' scale onto the IRT scale. The model also takes into account measurement errors in the judges' ratings. It is demonstrated that the linear logistic test model can be used to implement the proposed Thurstonian IRT approach.

Keywords Item response theory • Likert scaling • Linear logistic test model • Attitudes toward obese persons • Equal-appearing interval scaling

2.1 Introduction

Together with the Guttman scale, Thurstone and Likert scaling are perhaps the most prominently featured and researched scaling techniques in the history of psychological measurement, especially in the assessment of attitudes. Historically, Thurstone was one of the first quantitative psychologists to set his sights on the development of a theory for psychological scaling (Thurstone 1925, 1928). Thurstone's pioneer work on conception of attitude was based on the assessment of subjective attitudinal responses. The covert responses—or a sample of them—are

E.H. Ip (✉)
Department of Biostatistical Sciences, Wake Forest School of Medicine,
Medical Center Blvd., WC23, Winston-Salem, NC 27157, USA
e-mail: eip@wakehealth.edu

R.E. Millsap et al. (eds.), *Quantitative Psychology Research*, Springer Proceedings
in Mathematics & Statistics 89, DOI 10.1007/978-3-319-07503-7_2

linguistically represented in the form of opinion statements, which can then be located on a single evaluative dimension (Ostram 1989). Based on the principle of comparative judgment, Thurstone developed several scaling methods, of which the best known is the equal-appearing interval scale (Thurstone and Chave 1929). Given a collection of items, each of which contains a statement concerning the psychological construct of interest, the technique consists of two steps.

First, a panel of judges is recruited to rate the items in terms of their favorability to the construct of interest. Thurstone suggested using integral values of 1–11 for the rating scale. The 11-point scale then becomes the psychological continuum on which the statements have been judged, and the distribution of judgments obtained is used to calculate a typical value, which can then be taken as the scale-value of the statement on the 11-point psychological continuum. The value could be the median or the mean of the judgment distribution, and descriptive statistics such as standard deviations and the interquartile range are then used to eliminate questions that have overly dispersed judgment scores. Ideally, the equal-appearing interval scale is established by a final collection of items with small dispersions so that the scale-value of the statements on the psychological continuum are relatively equally spaced. In the second step, the statements are presented to subjects with instructions to indicate those with which they are willing to agree and those with which they disagree. The attitude score for a subject is based on the mean or the median of the scale-values of the statements agreed with. In other words, if the responses are dichotomously coded as 1 for Agree and 0 for Disagree, then the attitude score is an average of a weighted combination of the response categories, of which the weights are the scale-score.

One of the most fascinating aspects of Thurstone's scaling procedure is that the scale is determined by expert judges on a unidimensional continuum and that the operating characteristic of a Thurstone item may reflect either an underlying dominant-response process or an ideal-point process (Coombs 1964; Roberts and Laughlin 1996). In the most common form of the dominance mechanism, respondents and items are represented by positions on a latent trait, and the responses are determined by a comparison process: if the respondent's trait value is greater than the item-trait value, then the response to the item is positive; otherwise, the response is negative. The item-characteristic curve (ICC) of the item response for a dominant-response process is monotone and can be well captured by existing item response theory (IRT; Lord 1980) models. An example of a monotone ICC for equal-appearing interval scaling is the Sickness-Impact Profile (SIP; Bergner et al. 1981). Judges rated the SIP items on the severity of the dysfunction described in an item on an equal-interval 11-point scale. The end points were labeled "minimally dysfunctional" and "severely dysfunctional" to provide meaningful referents. An item concerning how sickness impacts work is: "I act irritable and impatient with myself—for example, talk badly about myself, swear at myself, blame myself for things that happen." A monotone ICC for this item implies that a respondent with a higher SIP trait value (more dysfunctional) is more likely to endorse this item than someone with a lower SIP trait value (less dysfunctional). For an empirical comparison between IRT scaling and Thurstone scaling in education, see Williams et al. (1998).

The Thurstone scaling procedure could also be used to describe an ideal point-response process, a model commonly used in attitudinal measurement of political and social views. Like the dominant-response process, the ideal-point process postulates that the individual's response also depends on the relative position of the person's trait value and the position of the item on the scale. However, a respondent in an ideal-point process is more likely to endorse statements that have trait values close to the respondent's. Thus, the ICC from an ideal-point process is not monotonic with respect to the trait and typically has a single peak at the location of the item. These models are often referred to as unfolding models in the IRT literature. An example of an unfolding item is a well-known General Social Survey (GSS) item on legalized abortion. The respondent in the GSS is asked when legalized abortion is allowed on a collection of seven conditions such as: "The family has a very low income and cannot afford any more children" and "The woman wants it for any reason." For respondents who hold a more centralist view about legalized abortion, the likelihood of endorsing the former statement would be higher than it would be for those who hold a liberal view about abortion as well as those who are strong anti-abortion.

In this paper, we only focus on Thurstone's equal-appearing scaling methods for items that do not fold—or items that are supposed to follow a dominant-response process so that their ICCs are monotonic. We argue that the equal-appearing scaling method is a way to set scales according to experts' views of the construct of interest and that it could be operationalized through IRT models in which the location parameter of an item can be obtained by careful scaling of the judges' ratings. The extent to which the judges disagree on the location of an item can be incorporated into the IRT model by assuming that the rating scores from the sample of judges are normally distributed with a mean m and a standard deviation σ, both of which could be directly estimated from the judges' data. As such, the proposed model can be viewed as an IRT implementation for equal-appearing scaling, which is distinct from the Thurstonian item response model proposed in Brown and Maydeu-Olivares (2012). We further demonstrate that the uncertainty associated with the judges' ratings would lead to an attenuation of the slope of the ICC, which, in modern IRT language, means that the information contained in the item is less than 1 at the same scale location but has a steeper slope. Also, we show that through a convolution technique the proposed Thurstonian IRT model can be solved using the estimation procedure for the linear logistic test model (LLTM; Fischer 1973).

The remainder of this paper is organized as follows: first, we describe the Thurston IRT model, then we illustrate the proposed model using a data set collected from a study of attitudes. Finally, we conclude with a discussion.

2.2 Thurstonian IRT: Method

We begin with a simple Rasch model:

$$P(Y_{ij}=1\Big|\theta_j)=\frac{\exp\left(\theta_j+b_i\right)}{1+\exp\left(\theta_j+b_i\right)}, \tag{1}$$

where Y_{ij} is the binary response of individual j to item i, with 1 indicating a correct or positive response; θ_j is the latent trait for individual j; and b_i is the intercept parameter for item i or the individual. We further assume that the intercept parameter b_i is a function of item attributes $\underset{\sim}{w}_i$ and the judge's rating, which has a mean m_i and variance σ_i^2. Specifically, we write:

$$b_i=\eta_1^T\underset{\sim}{w}_i+\eta_2\left(m_i+\varepsilon_i\right),\;\; \varepsilon_i\sim N\left(0,\sigma_i^2\right), \tag{2}$$

where η denotes regression coefficients.

$$P\left(Y_{ij}=1\Big|\theta_j,\varepsilon_i\right)=\frac{\exp\left[\theta_j+\eta_1^T\underset{\sim}{w}_i+\eta_2\left(m_i+\varepsilon_i\right)\right]}{1+\exp\left[\theta_j+\eta_1^T\underset{\sim}{w}_i+\eta_2\left(m_i+\varepsilon_i\right)\right]}, \tag{3}$$

$$\theta_j\sim N\left(0,1\right),\;\; \varepsilon_i\sim N\left(0,\sigma_i^2\right).$$

In other words, we have

$$P\left(Y_{ij}=1\Big|\theta_j,\varepsilon_i\right)=\frac{\exp\left[\theta_j+b_i'+\eta_2\varepsilon_i\right]}{1+\exp\left[\theta_j+b_i'+\eta_2\varepsilon_i\right]}, \tag{4}$$

where $b_i'=\eta_1^T\underset{\sim}{w}_i+\eta_2 m_i$.

By integrating out the error term $\eta_2\varepsilon_i$ through a convolution technique (Zeger et al. 1988; Caffo et al. 2007; Ip 2010), we now have

$$P\left(Y_{ij}=1\Big|\theta_j\right)=\frac{\exp\left[a_i^*\theta_j+b_i^*\right)\right]}{1+\exp\left[a_i^*\theta_j+b_i^*\right)\right]}, \tag{5}$$

where $a_i^*=\lambda_{\text{logit}}(a_{i1}+\frac{\eta_2\rho\sigma_i}{\sigma_1})$, $b_i^*=\lambda_{\text{logit}}b_i^{`}$, $\lambda_{\text{logit}}=[k^2\eta_2^2(1-\rho^2)\sigma_i^2+1]^{-1/2}$, and $k=16\sqrt{3}/(15\pi)=0.588$. The factor a_i^* represents an attenuation factor for the slope of θ, which is assumed to be 1.0 in a Rasch model, and ρ represents the correlation between ε and θ, which is set to zero.

Figure 2.1 shows the change in attenuation as a function of the standard deviation of the measurement error. Generally speaking, when the noise level (measurement error) increases, the attenuation factor becomes smaller and varies almost linearly from no attenuation (=1.0) to a value of 0.5. Notably, the graphs show that attenuation is approximately 0.8 when the noise level (SD = 1) reaches the level of the signal (SD = 1). We call the model specified by Eqs. (4) and (5) the Thurstonian LLTM model.

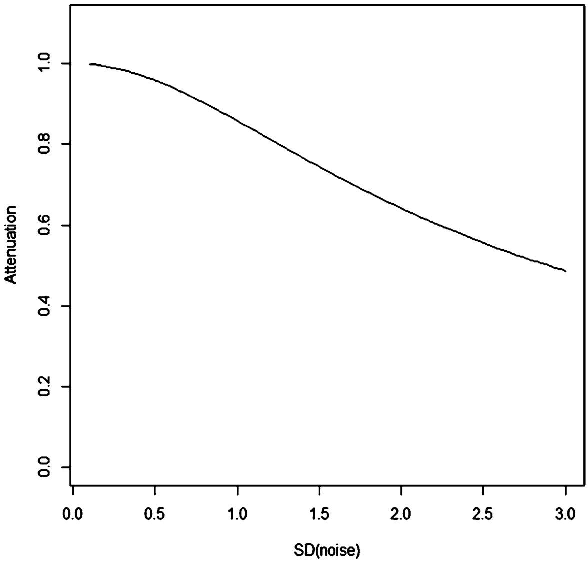

Fig. 2.1 Attenuation factor as a function of the standard deviation of the judges' ratings

2.3 Real Data Example

2.3.1 Data

The data were a subset of data collected from a recent study on the development of a curriculum for medical school students for counseling obese patients. The Nutrition, Exercise, and Weight Management (NEW) study collected attitude data using an instrument—the NEW Attitude Scale (Ip et al. 2013)—which comprises 31 items measuring attitudes across three domains: nutrition, exercise, and weight management. Examples of items include "I do feel a bit disgusted when treating a patient who is obese" (Item 23), and "The person and not the weight is the focus of weight-management counseling" (Item 25). In the item-development process, the study team had a consensus view for some items but divergent views for others. An example of a consensus item was "Overweight individuals tend to be lazy about exercise" (Item 13), which the team agreed represented an unfavorable attitude. An item that solicited divergent views was "Patients are likely to follow an agreed-upon plan to increase their exercise" (Item 10). Some tended to feel that an endorsement of the item suggested a favorable attitude because the physician sounded positive about the outcome, but others argued that the item should be viewed negatively because the physician might not appreciate the challenges that an obese person encountered when prescribed an exercise program. The study team decided to use the Thurstonian approach of soliciting judges' opinions about the positivity/negativity of the items. A total of 201 judges (approximately 50% clinically focused and the remaining research focused) rated the items. A sample of N = 103 medical students completed the instrument. Using the scores that were derived from traditional Thurstone scaling, the test–retest reliability of the instrument was 0.89 (N = 24). Pearson correlations between two other anti-obesity measures were the Anti-Fat Attitudes Questionnaire (AFA; Lewis et al. 1997) and the Beliefs About Obese Persons Scale (BAOP; Allison et al. 1991) were -0.47 and 0.23, respectively. This shows satisfactory convergent validity with existing measures of attitudes toward obese individuals. A full report about the validation of the instrument can be found in Ip et al. (2013).

To illustrate the range of concordance in judges' ratings across items, we used two items as examples. Figures 2.2 and 2.3 show, respectively, the distributions of ratings for Item 23 and Item 25. The former item has a relatively high level of consensus as being indicative of an unfavorable attitude, as demonstrated by the small standard deviation (SD = 0.8). In contrast, Item 25 exhibits high variance in the judges' ratings (SD = 2.2).

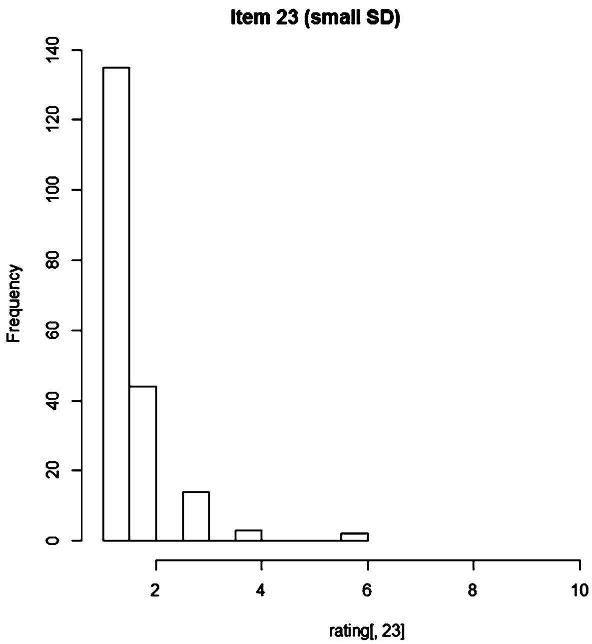

Fig. 2.2 Distribution of judges' ratings for Item 23