Chapter 28
Non-Rasch Measurement Models for Ordered Response Categories

This chapter summarizes the most common non-Rasch models considered for analysing ordered response category items. These models fall into two distinct classes. The models of the first class have a structure consistent with the PRM but with a greater number of parameters. The models of the second class are structurally different from the PRM but can have the same or more parameters than the PRM. Models from both classes do not have the sufficient statistic properties of the PRM. The application of these models arises from the Item Response Theory (IRT) paradigm in which the main criterion for the choice of the model is that of statistical fit of the responses to the model. These models are chosen to describe or summarize the data, and do not arise from any fundamental principles that are independent of the data. The full class of models, and their connection to the respective paradigms, are summarized in Andrich (2011).

For efficiency of exposition, we begin with the class of models which specializes to the PRM.

The Nominal Response Model

Bock (1972) presented the model he called the *nominal response mode* (NRM), equivalent in form and notation to

$$\Pr\{x; \beta, (\psi), (\varphi)\} = e^{\psi_x + \varphi_x \beta}/\gamma, \;\; x = 0, 1, 2, \ldots, m. \tag{28.1}$$

Again, because the response $x; \;\; x = 0, 1, 2, \ldots, m$ is of a single person to a single item, we do not subscript the person and item parameters β and δ, nor the two vectors $(\kappa), \;\; (\varphi)$ which characterize the categories of the item. Here, the response variable $x; \;\; x = 0, 1, 2, \ldots, m$ is simply the ordinal count of the category of the response, beginning with the first category and γ is again the normalizing factor which is the sum of the numerators of Eq. (28.1). In the development of the Rasch model,

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8_28

this same equation appeared earlier (Rasch, 1961), which was developed further by Andersen (1977), and then interpreted in terms of thresholds and discrimination at the thresholds in Andrich (1978). In these publications, $\kappa_x,\ \varphi_x,\ x = 0, 1, 2, \ldots, m$ are called, respectively, the category coefficient and the scoring function and we use these terms in this chapter. In order to connect this model to the PRM, and better understand it, we now summarize the original derivation of the PRM.

Relationship Between the PRM and the NRM

This section follows the derivation of the threshold form of the PRM shown in Chap. 27. However, there is one important difference. Instead of specifying the dichotomous Rasch model for the latent dichotomous responses at the thresholds in the full space Ω, the 2PL model (Birnbaum, 1968) we encountered in Chap. 18 was specified. This specification appeared in the original derivation of the threshold form of the PRM in Andrich (1978).

Thus, instead of applying the dichotomous Rasch model of Eq. (27.1) of the previous chapter as the probability of a dichotomous response at the thresholds, $x = 1, 2, 3$, the equation applied was

$$\Pr\{y_x = 1; \beta, \delta_x | \Omega\} = e^{\alpha_x(\beta - \delta_x)}/\gamma, \tag{28.2}$$

where α_x is the discrimination at threshold x of item i. In the dichotomous Rasch model, and in terms of Eq. (28.2), it will be recalled that $\alpha_x = 1$.

Table 28.1 reproduces the essential elements of Table 27.4 for responses within the Guttman subspace Ω^G, but with Eq. (28.2) as the latent response probability at each threshold and again immediately notated by successive integers, $x = 1, 2, 3$.

Following the division of the probabilities in the last column of Table 28.1 by Γ, which ensures the probabilities sum to 1, the model takes the general form

Table 28.1 Probabilities of responses in the Guttman subspace Ω^G when the dichotomous response at threshold x is the 2PL model

$\Pr\{y_1, y_2, y_3\}$	$= \Pr\{y_1\|\Omega\}\Pr\{y_2\|\Omega\}\Pr\{y_3\|\Omega\}$	
$\Pr\{x = 0\}$	$= 1.1.1/\gamma_1\gamma_2\gamma_3$	$= e^{0\beta}/\gamma_1\gamma_2\gamma_3$
$\Pr\{x = 1\}$	$= e^{\alpha_1\beta - \alpha_1\delta_1}.1.1/\gamma_1\gamma_2\gamma_3$	$= e^{\alpha_1\beta - \alpha_1\delta_1}/\gamma_1\gamma_2\gamma_3$
$\Pr\{x = 2\}$	$= e^{\alpha_1\beta - \alpha_1\delta_1}e^{\alpha_2\beta - \alpha_2\delta_2}.1/\gamma_1\gamma_2\gamma_3$	$= e^{(\alpha_1+\alpha_2)\beta - \alpha_1\delta_1 - \alpha_2\delta_2}/\gamma_1\gamma_2\gamma_3$
$\Pr\{x = 3\}$	$= e^{\alpha_1\beta - \alpha_1\delta_1}e^{\alpha_2\beta - \alpha_2\delta_2}e^{\alpha_3\beta - \alpha_3\delta_3}/\gamma_1\gamma_2\gamma_3$	$= e^{(\alpha_1+\alpha_2+\alpha_3)\beta - \alpha_1\delta_1 - \alpha_2\delta_2 - \alpha_3\delta_3}/\gamma_1\gamma_2\gamma_3$
$\sum_{\Omega^G} \Pr\{(y_1, y_2, y_3)\|\Omega = \Pr\{y_1\|\Omega\}\Pr\{y_2\|\Omega\}\Pr\{y_3\|\Omega\} = \Gamma < 1.\}$		

$$\Pr\{x; \beta, (\alpha), (\delta)|\Omega^G\} = e^{(\alpha_1+\alpha_2+\ldots+\alpha_x)\beta-(\alpha_1\delta_1+\alpha_2\delta_2+\ldots+\alpha_x\delta_x)}/\gamma \tag{28.3}$$

where $x = 0, 1, 2, \ldots, m$.

Now, define

$$\varphi_0 = 0; \ \varphi_x = \alpha_1 + \alpha_2 + \ldots \alpha_x; \ x = 1, 2, \ldots, m, \tag{28.4}$$

$$\psi_0 = 0; \ \psi_x = -(\alpha_1\delta_1 + \alpha_2\delta_2 + \ldots \alpha_x\delta_x); \ x = 1, 2, \ldots, m, \tag{28.5}$$

to give the model

$$\Pr(x; \beta, (\psi), (\varphi)) = e^{\psi_x+\varphi_x\beta}/\gamma, \ x = 0, 1, 2, \ldots, m. \tag{28.6}$$

where we now take for granted the subspace Ω^G and drop its specification.

We see that Eq. (28.6) is the form of the NRM of Eq. (28.1).

With the constraints $\varphi_0 = 0; \ \psi_0 = 0$ on the categories of each item, the number of independent parameters for each item are effectively $2m$. Although they are not typically viewed in this way, the parameters embody a location (difficulty) and discrimination at each threshold, a generalization of the 2PL. Where the model is applied, the parameters $\varphi_x, \ \psi_x$ are attempted to be estimated without consideration of what these parameters might characterize. It is evident from Eqs. (28.4) and (28.5) that φ_x is the sum of discriminations of all thresholds up to threshold x in the required order, and that ψ_x is of the same cumulative structure but with the location and discrimination parameters at the thresholds entangled. With only one response in one of the $m + 1$ categories, this model is not easy to implement and is not used routinely in major assessments.

To see the way the NRM is a generalization of the PRM, suppose, as in the dichotomous Rasch model, that the discriminations α_x are identical. Let $\alpha_x = \alpha, \ x = 1, 2, \ldots, m$. Then, from Eq. (28.4),

$$\varphi_0 = 0; \ \varphi_x = (\alpha + \alpha + \ldots \alpha) = x\alpha; \ x = 1, 2, \ldots, m, \tag{28.7}$$

and

$$\psi_0 = 0; \ \psi_x = -\alpha(\delta_1 + \delta_2 + \ldots + \delta_x); \ x = 1, 2, \ldots, m. \tag{28.8}$$

Then, defining $\delta_0 = 0$ for convenience, the NRM of Eq. (28.6) takes the form

$$\Pr\{x; \beta, (\delta)\} = e^{-\alpha(\delta_0+\delta_1+\delta_2+\ldots+\delta_x)+x\alpha\beta}/\gamma; \ x = 0, 1, 2, \ldots, m. \tag{28.9}$$

Absorbing the common discrimination α into $\beta, \ (\delta)$, or simply defining $\alpha = 1$, gives the PRM in the form

$$\Pr\{x; \beta, (\delta)\} = e^{-(\delta_0+\delta_1+\delta_2+\ldots+\delta_x)+x\beta}/\gamma; \ x = 0, 1, 2, \ldots, m. \tag{28.10}$$

Thus, the PRM is an algebraic specialization of the NRM expressed in the form of threshold locations and discriminations at these thresholds with the discriminations at the thresholds all constant. The equal discriminations at the thresholds give the integer scoring function.

However, the uniform discriminations at the thresholds go beyond simply the discriminations at the thresholds within each item, they are uniform across all items. Including now an item and a person subscript, Eq. (28.10) takes the form

$$\Pr\{x;\beta_n,(\delta_i)\} = e^{-(\delta_{i0}+\delta_{i1}+\delta_{i2}+\ldots+\delta_{ix})+x\beta_n}/\gamma_{ni}$$
$$= e^{-\sum\limits_{k=0}^{x}\delta_{ik}+x\beta_n}/\gamma_{ni};\;\; x = 0,1,2,\ldots,m_i \qquad (28.11)$$

Equation (28.11) is the partial credit parameterization of the PRM which we encountered in Eq. (21.6) in Chap. 21. The equal discriminations at the thresholds among all items give the total score of a person across all items, an integer, as the sufficient statistic for the person parameter. With different discriminations at the thresholds, the NRM does not have a sufficient statistic in the sense that the person and item parameters can be separated in the estimation as in the PRM.

The Generalized Partial Credit Model

The generalized partial credit model is also a special case of the NRM, but not to the degree that the PRM is specialized (Muraki, 1992; Muraki & Muraki, 2016). Although it retains the condition that all thresholds within an item have the same discrimination, it permits variable discrimination α_i among the items. This gives the model, with subscripts present,

$$\Pr\{x;\beta_n,(\delta_i),(\alpha_i)\} = \mathrm{e}^{\alpha_i\left(-\sum_{k=0}^{x}\delta_{ik}+x\beta_n\right)}/\gamma_{ni};\;\; x = 0,1,2,\ldots,m_i. \qquad (28.12)$$

The generalized partial credit model also does not have sufficient statistics of the form of the PRM, but because it has a smaller number of parameters than the NRM, it is more tractable than the NRM. As indicated earlier, it is applied from the perspective of the IRT paradigm.

We now turn to the second class of models which is structurally different from the PRM.

The Graded Response Model

The model now called the graded response model (GRM) for the analysis of ordered response categories has its origins in the work of Thurstone. The possibility of collecting data in the form which implied the model was mentioned at the end of Thurstone (1928) and then further developed in Edwards and Thurstone (1952). In modern psychometric form, it is presented in Samejima (1969, 2016), and in a contingency table context, where the dependent variable is in the form of ordered response categories, it is presented in Bock (1975). The GRM was the standard model for the analysis of ordered response categories before the advent of the PRM.

In the PRM, there is a distinct latent response process at each threshold which is then constrained by the category order. In contrast, in the GRM there is only one response process across the continuum and the outcome of this process is portioned into categories.

To show the structure of the GRM, let

$$P_x = \Pr\{x; \beta, (\alpha), (\delta)\}, \;\; x = 0, 1, 2, \ldots, m, \tag{28.13}$$

be the probability of a response in category x, using the same notation as in the PRM. Again, we do not subscript the person parameter β and the vectors of item parameters $(\alpha), \; (\delta)$, the response being that of a single person responding to a single item. Although using the same notation as in the PRM, the item parameters are different in the two models.

Now, define the cumulative probability π_x for category x and above as follows:

$$\pi_x = \mathrm{P}_x + \mathrm{P}_{x+1} + \mathrm{P}_{x+2} \ldots . + \mathrm{P}_m; \;\; \pi_0 = 1, \;\; \pi_m = \mathrm{P}_m. \tag{28.14}$$

By definition, the cumulative probabilities π_x decrease with x. Figure 28.1 shows the response process of the GRM as a cumulative probability. The categories are bounded by adjacent thresholds $\delta_x, \;\; x = 1, 2, \ldots, m$ which are different from the thresholds of the PRM.

The curve of Fig. 28.1 is defined in terms of the 2PL model (Birnbaum, 1968), that is, for a fixed person location β

$$\pi = e^{\alpha(\beta-\delta)}/\gamma, \tag{28.15}$$

where again the γ is the normalizing factor.

Then, the specific response in category x *or greater* is given by

$$\pi_x = e^{\alpha(\beta-\delta_x)}/\gamma. \tag{28.16}$$

The probability of a response in category x is then given by

$$\mathrm{P}_x = \pi_x - \pi_{x+1} = e^{\alpha(\beta-\delta_x)}/\gamma - e^{\alpha(\beta-\delta_{x+1})}/\gamma. \tag{28.17}$$

It is possible to specialize the GRM so that the discriminations, α, are the same across items. Then, the GRM and the PRM have the same number of parameters. However, the scale of the GRM is different from PRM, though in any data set, the

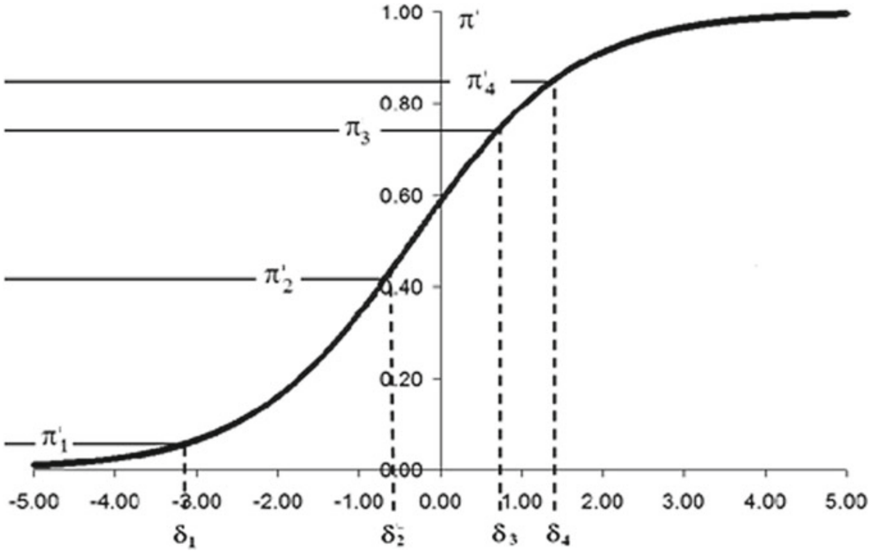

Fig. 28.1 The cumulative response structure of the graded response model

estimates of the person parameters will be highly correlated—that is a property of the data.

The structure of the GRM ensures that its thresholds, which are different from the thresholds in the PRM, are necessarily in order. This results from the feature that $\pi_x < \pi_{x-1}$. This means that those using the GRM tend not to focus on evidence that categories might not be operating as intended. However, the points of intersection of the adjacent categories in category characteristic curves of the GRM may still show reversals—they will do so if an analysis with the PRM shows reversals. An example of a data set with respective threshold estimates from the PRM and the GRM is shown in Andrich (2011).

Estimation of Parameters in the Non-Rasch Models

We saw in Chap. 7 how the person parameter can be eliminated in the dichotomous Rasch model and then the item parameters can be estimated independently of the person parameters. This is because the Rasch model has sufficient statistics for its parameters. Because the non-Rasch models do not have such sufficient statistics, it is not possible to separate the estimation of the item and person parameters in the same way. Therefore, some other assumptions or constraints are required. One approach is to assume a distribution of the person parameters, such as normal, and impose it as a constraint in the estimation. Another approach is to place a constraint on the observed distribution of total scores. In any case, these methods involve first estimating a set of item parameters, then estimating a set of person parameters given the estimates of the item parameters, and then returning to the estimates of the person parameters, and so on, until the estimates converge. In many cases, all estimates do not converge and some upper limit on an estimate of an item difficulty parameter or discrimination parameter may be imposed.

These methods of estimation may also be used with the Rasch model and are used in many Rasch model software packages. RUMM2030 uses a particular kind of conditional estimation which does eliminate the person parameters in the process of estimating the item parameters. In this method, the conditional responses to pairs of items are essential elements of the estimation. The method is described in more detail in Andrich and Luo (2003).

Exercises

Describe, in one paragraph each, two differences between the Rasch and non-Rasch models used for analysing items with ordered categories.

References

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika, 42,* 69–81.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–574.

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*(5), 571–585.

Andrich, D., & Luo, G. (2003). Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. *Journal of Applied Measurement, 4*(3), 205–221.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–545). Reading, Massachusetts: Addison-Wesley.

Bock, R. D. (1972). Estimating item parameters and latent ability when response are scored in two or more nominal categories. *Psychometrika, 37,* 29–51.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Edwards, A. L., & Thurstone, L. L. (1952). An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika, 17,* 169–180.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–176.

Muraki, E. & Muraki, M. (2016). Generalized partial credit model. In W. J. van der Linden (Ed.), *Handbook of item response theory: Models* (Vol. 1, Chapter 8, pp. 127–137). Boca Raton, Florida: Taylor and Francis.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 4, pp. 321–333). Berkeley, California: University of California Press. Reprinted in Bartholomew, D. J. (Ed.) (2006). *Measurement: Sage benchmarks in social research methods* (Vol. I, pp. 319–334). London: Sage Publications.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monographs, 34*(2, No.17).
Samejima, F. (2016). Graded response models. In W. J. van der Linden (Ed.), *Handbook of item response theory: Models* (Vol. 1, Chapter 6, pp. 95–108). Boca Raton, Florida: Taylor and Francis.
Thurstone, L. L. (1928). The measurement of opinion. *Journal of Abnormal and Social Psychology, 22,* 415–430.

Further Reading

Andrich, D. (1995). Distinctive and incompatible properties of two common classes of IRT models for graded responses. *Applied Psychological Measurement, 19*(1), 101–119.
Nering, M., & Ostini, R. (Eds.). (2010). *Handbook of polytomous item response theory models: Developments and applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.
Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51,* 567–577.