Samejima Models

In: Polytomous Item Response Theory Models

By: Remo Ostini & Michael L. Nering

Pub. Date: 2011
Access Date: February 26, 2018
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780761930686
Online ISBN: 9781412985413
DOI: http://dx.doi.org/10.4135/9781412985413
Print pages: 61-85

Samejima Models

The only major approach to polytomous IRT modeling that is distinct from the Rasch-type models is the work of Samejima, which is built on the cumulative boundary measurement approach of Thurstone. Samejima (1969) initially developed two such models designed to allow researchers to estimate respondents' standing on a trait from their responses to a set of items with an ordered polytomous response format. Much of this work focused on maximum likelihood estimation for the person (trait) parameter. The two models had an identical form and differed only in that one employed a normal ogive to model responses, and ultimately to construct an item's ICRFs, whereas the second used a logistic function for this purpose. Samejima (1972) later expanded and formalized the basic framework of her earlier work to accommodate free-response data that could conceivably include a potentially unlimited number of unordered responses to a test item.

Samejima provides a limited number of practical measurement models for specific situations. Unfortunately, in the literature, there is often a failure to differentiate a specific practical model, by name, from the general class of models to which it belongs. This has led to some confusion in labeling and understanding Samejima's framework and the constituent, specific models.

Framework

Table 4.1 presents a way to think of Samejima's theoretical framework. Working from left to right, we begin with specific data types. These can be specialized from free-response data either at the outset by the format of the data collection, or by subsequent, valid ordering or categorization procedures. Of course, in many practical testing situations, the data will, in fact, be collected in a format corresponding to one of the more specific types.

Samejima (1972, 1997b) describes what are essentially two broad theoretical classes of models that can be used to model the specific data types listed in the first column. The two classes of models are referred to as the heterogeneous and the homogeneous cases, and the appropriate class for each data type is listed in the second column of the table. The most interesting feature to note here is that ordered, polytomous data can be modeled by either the heterogeneous or homogeneous classes of models. Nominal data, however, can be modeled only by the heterogeneous class of models (Samejima, 1972).

The final column lists specific models that have been developed over time, with respect to the general class of models to which they belong and, by extension, the specific data types to which they explicitly apply. Two other interesting points highlighted in this final column are that not all of the specific models have been developed by Samejima herself, and that in some

cases, a number of applicable models are potentially available for use.

TABLE 4.1 Samejima's General Framework for Polytomous Data

Theoretical Framework				
More specific polytomous data types	*which can be modeled by* →	General classes of models	*using* →	Specific models
Continuous nominal (e.g., ungrouped Rorschach data)		Heterogeneous (continuous)		None
Continuous ordered (e.g., graphic rating scales)		Heterogeneous (continuous) Homogeneous (continuous)		Continuous Rasch model (Müller, 1987) Continuous response model (Samejima, 1973)
Discrete nominal (e.g., multiple-choice test items)		Heterogeneous (categorical)		Nominal response model (Bock, 1972)
Discrete ordered (e.g., rating scale items)		Heterogeneous (categorical)		Acceleration model (Samejima, 1995); Polytomous Rasch models; Generalized partial credit models (per Muraki, 1992)
		Homogeneous (categorical)		Graded response models (Samejima, 1969)

The lack of a clear distinction between the theoretical framework and the specific models is probably due to the fact that the specific models developed by Samejima grow directly out of the underlying framework. That framework is tied directly to the response processes hypothesized to generate the various types of response data.

This is an important difference between Samejima's models and the polytomous Rasch models. As has been mentioned numerous times, the Rasch models are built on the requirement of model parameter separability, with the associated issues of sufficiency, specific objectivity, and fundamental measurement. Any hypothesized response process that may have generated the data to be modeled is essentially an afterthought.

In Samejima's framework, however, the plausibility of the response process thought to generate data is paramount, because for Samejima (1979a), the main role of a mathematical model in psychology is to plausibly represent psychological reality. Thus, the hypothetical response process is not ancillary speculation, as it tends to be for the Rasch models; rather, it is the foundation for her whole approach to modeling polytomous data. Any discussion of Samejima's work therefore must begin with her interpretation of a realistic underlying psychological reality and proceed by showing the bridge to specific models.

From Response Process to Specific Model

The response process hypothesized by Samejima (1972) to underlie the specific models in her

framework is built on the notion that each response category of an item exerts a level of attraction on individuals. By implication, this attraction varies across individuals who vary in trait level, and is a function of θ. This attraction toward a category is also defined contingent on the person having already been attracted by the previous category. The function describing the probability of being attracted to a particular category g, over and above an attraction to the previous category, $g - 1$, is designated $M_{ig}(\theta)$. Samejima (1995, 1997b) calls this the *processing function*.

In the context of an entire item, then, being attracted to a category must take all prior category attractions into account. This can be expressed as the serial product of the attractions to each successive category, over and above the attraction to the respective prior categories. Competing with this attraction for any given category is the further attraction of the next category. The probability of being attracted to the next category can be designated $M_{ig+1}(\theta)$. This can also be thought of as the probability of rejecting the previous category. In those terms, the probability is designated $U_{ig}(\theta)$. Naturally, this implies that the value of M_{ig} is also the probability of rejecting category $g - 1$.

Given this web of competing attractions (and rejections), the probability of responding in any given category is simply a combination of being attracted through all previous categories up to the given category, but then no further; that is, the probability of the serial attraction up to the category multiplied by the probability of not rejecting that category.

In the case of ordered categories, this process assumes that to respond in a particular category, a person must have passed through all preceding categories (Samejima, 1972). The psychological process being postulated is therefore a cumulative process of successively accepting and then rejecting categories, where rejecting a category is defined as being more attracted to the next category, until a category is reached where the probability of attraction is greater than the probability of rejection.

The cumulative attraction is operationalized as $P^*_{ig}(\theta)$, where P^*_{ig} is defined as the probability of responding positively at a category boundary given all previous categories, conditional on θ. As we saw earlier, the probability of responding in a category can then simply be calculated as the probability of responding positively at a category boundary minus the probability of responding positively at the next category boundary:

$$P_{i_g} = P^*_{i_g} - P^*_{i_{g+1}}. \qquad (1.2)$$

Thus, the probability of responding in a particular category is modeled, logically, as the

difference between the cumulative probabilities of serial attraction to two adjacent categories. This obviates the need for estimating the probability of rejecting a given category, U_{ig}, or for estimating the component Mi_g.

We will now look at specific models in terms of the two broad theoretical classes of models that Samejima introduced, that is, the homogeneous case and the heterogeneous case.

The Homogeneous Case: Graded Response Models

Models that fall within the classification of a homogeneous case of a polytomous model have two notable elements. The first is that they are the only models actually operationalized in terms of Equation 1.2. The second notable feature is that homogeneous models have $P*_{ig}$, which are all the same shape within an item (Samejima, 1995, 1997b). This latter feature is the distinguishing theoretical feature of the homogeneous case and gives rise to its title.

Before we turn to the mathematical representation of these models, however, a note about nomenclature is in order. In keeping with what amounts to almost universal usage in the literature, the term *graded response model* (GRM), used in isolation, will refer here only to the logistic example of Samejima's (1969) original two models. The normal ogive example of the original two models will be referred to as the normal ogive version of the GRM. Any references to other specific models will use the specific names that are commonly connected with them. In keeping with standard usage, the terms homogeneous or heterogeneous “class” or “case” will imply a theoretical level of discussion rather than an operationalized model.

The Mathematical Model

As a practical application within Samejima's general framework, the GRM is the archetypal model in the framework, in part because it is one of the original models. It is also the polytomous IRT manifestation of Thurstone's method of successive intervals (Burros, 1955; Edwards & Thurstone, 1952; Rimoldi & Hormaeche, 1955) in terms of its structural attributes (e.g., see Andrich, 1995; Masters, 1982; Reiser, 1981; Tutz, 1990), where those attributes are described in the top half of Figure 1.4.

As is often noted in the applied literature (e.g., Cooke & Michie, 1997; Flannery, Reise, & Widaman, 1995; Fraley, Waller, & Brennan, 2000; Samejima, 1997b), the GRM is built on the two-parameter logistic (2PL) model because this dichotomous model is used as the function to obtain the cumulative boundary functions denoted by $P*_{ig}$. Theoretically, however, the $P*_{ig}$ could be represented by any appropriate mathematical function (Samejima, 1972, 1996). In practice, they have only ever been modeled by two types of function, the ubiquitous 2PL, as in

the GRM, and the 2-parameter normal ogive dichotomous model, in the generally ignored normal ogive version of the GRM. As mentioned above, both of these versions of the GRM were outlined in Samejima's (1969) seminal work.

The usual equation for a GRM CBRF is

$$P^*_{i_g} = \frac{e^{a_i(\theta - b_{i_g})}}{1 + e^{a_i(\theta - b_{i_g})}}, \qquad (4.1)$$

where a_i is the item discrimination parameter and b_{ig} is the boundary location parameter.

Given this definition of the CBRFs, the GRM arises directly from Equation 1.2. This gives the probability of responding in category g as the difference between the probabilities of responding in or above category g (i.e., $P*_{ig}$) and responding in or above category $g + 1$ (i.e., $P*_{ig+1}$). These $P*_{ig}$ were described earlier as Thurstone/Samejima CBRFs, and that is the role they play here.

For Equation 1.2 to completely define an item's ICRFs in terms of the available boundary functions, two additional definitions are required. The first is that the probability of responding in or above the lowest possible category for any item is defined as 1.0, across the entire range of θ. This is operationalized by postulating a boundary below the lowest category, where that boundary is designated $P^*_0(\theta)$ and is located at $-\infty$ Therefore, algebraically,

$$P^*_0 = 1, \qquad (4.2)$$

and the ICRF for the first category ($P*_1$) of any item becomes a monotonically decreasing function defined by Equation 1.2 as $P*_0 - P*_1$ which is equal to $1 - P*_1$.

The probability of responding in a category higher than the highest category available, designated $P^*_{m+1}(\theta)$, is defined to equal zero throughout the trait range. Algebraically,

$$P^*_{m+1} = 0, \qquad (4.3)$$

which means that the probability of responding in the highest category (P^*_{im}) equals the probability of responding positively at the highest category boundary (P^*_{im}), because $P_{im} = P^*_{im} - 0$ by Equation 1.2.

This general approach to parameterizing ordered polytomous items has several relevant features. One is that, unlike the polytomous Rasch models, there is no general model for GRM ICRFs. Each ICRF is modeled separately by Equation 1.2. Furthermore, boundary locations are defined at the inflection point of the CBRFs, that is, the point on the trait continuum where p =

.5. This also contrasts with polytomous Rasch models where boundary locations are defined at the crossing points of adjacent ICRFs. With the GRM, these ICRF crossing points need not be where the CBRF p = .5. A related result is that boundary reversals do not occur in the GRM by virtue of the definition of the boundaries as cumulative probability functions. This set of features accounts for the fact that it is common to see GRM CBRFs plotted explicitly, whereas it is not common to see polytomous Rasch model CBRFs plotted.

By definition, the cumulative CBRFs within items in models that belong to the homogeneous case of Samejima's (1972) framework are all the same shape. This is true whether they are modeled by the 2PL model, as in the GRM, or by the 2-parameter normal ogive, or are any other shaped functions (Samejima, 1996). In the case of the GRM, this effectively translates to mean that, within an item, the CBRFs have the same discrimination. Unlike polytomous Rasch models, category boundary discriminations cannot vary in the GRM, even if they are specified a priori and not estimated. Also unlike the Rasch models, but in a manner analogous to the GPCM, the GRM can model items with discrimination that varies across items.

Information

Beginning at the broadest level, test information is defined as the negative expectation of the second derivative of the log of the likelihood function of the person parameter θ (Hambleton & Swaminathan, 1985). Algebraically,

$$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right], \qquad (4.4)$$

where $I(\theta)$ denotes test information, conditional on θ, and L represents the likelihood function of a test. Of course, this always results in a positive value because the second derivative of the log of the likelihood function itself is always negative.

Test information, however, is also defined in terms of the sum of item information for the items in a test, assessed across the range of θ. We know that item information equals the squared IRF slope divided by conditional variance, from earlier discussions of information, where the first derivative of the IRF is indicative of slope. Furthermore, in the dichotomous case, the probability of a positive response (P_i) multiplied by the probability of a negative response (Q_i) is the measure of conditional variance. In algebraic terms,

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) = \sum_{i=1}^{n} \frac{P_i'(\theta)^2}{P_i(\theta)Q_i(\theta)}, \qquad (4.5)$$

where $I(\theta)$ denotes test information and $I_i(\theta)$ denotes item information, with both conditional on θ.

Information for Polytomous Models

Conceptualizing information in terms of polytomous IRT models involves somewhat more complex considerations than for the dichotomous case. Nevertheless, the basic principles are the same, and the foregoing description highlights the fact that there are different ways to represent information. Specifically, it can be represented in terms of derivatives with respect to relevant functions (Equation 4.4), in terms of conditional expectations (as described in the context of the RSM), or in terms of sums of component elements (Equation 4.5).

Information in the context of polytomous IRT models also can be represented in these three ways. The additional complexities come partly from the fact that item categories provide an extra, more detailed level to the set of component elements. Furthermore, the probability of responding in a category (P_{ig}) is no longer described by a single function but itself is based on the relationship among constituent boundary functions. Thus, information can be represented in terms of Pi_g, or equivalently, in terms of P^*_{ig}, where the boundary functions do not independently define ICRFs but do so in conjunction with each other. One consequence of these additional considerations is that components can no longer simply be summed to provide higher levels of information. Instead, aggregating components of information in the polytomous context often requires weighted sums rather than simple sums.

We start with item categories as the most basic level of aggregation. Samejima (1977, 1988, 1996, 1998) notes that item category information, denoted here as $I_{ig}(\theta)$, can be defined as the negative of the second derivative of the log of an ICRF:

$$I_{i_g}(\theta) = -\frac{\partial^2}{\partial\theta^2} \log P_{i_g}(\theta), \qquad (4.6)$$

which Samejima (1998) notes is equivalent to

$$I_{i_g}(\theta) = \left\{ \frac{P'_{i_g}(\theta)}{P_{i_g}(\theta)} \right\}^2 - \frac{P''_{i_g}(\theta)}{P_{i_g}(\theta)}. \qquad (4.7)$$

Information is evaluated conditionally with respect to θ (information varies as a function of θ). The quantity described here as category information has the usual interpretation for information, being the direct indicant of measurement precision provided by a given category (Baker, 1992).

Because an item's categories are not independent of one another, it is not possible to simply define item information as a sum of category information, even though this might seem logical at first glance. Instead, an intermediary term must be constructed that represents the share of information that each category provides to item information. Category share information is

obtained as a weighted component of category information, where I_{ig} is weighted by the response probability for a category across θ. Category share information is therefore designated

$$I_{i_g}(\theta)P_{i_g}(\theta). \qquad (4.8)$$

Item information is then a simple sum of category share information across all the categories of an item.

$$I_i = \sum_{g=0}^{m} I_{i_g}P_{i_g}. \qquad (4.9)$$

Samejima (1969) noted that, in conditional expectation terms, this is equivalent to describing item information as the expected value of category information. That is

$$I_i = E[I_{i_g}|\theta]. \qquad (4.10)$$

Furthermore, Samejima shows that category share information can also be obtained by subtracting the second derivative of an ICRF from the squared first derivative of that function divided by the probability value of that function at each point across θ. In mathematical terms,

$$I_{i_g}P_{i_g} = \frac{(P'_{i_g})^2}{P_{i_g}} - P''_{i_g}. \qquad (4.11)$$

Note that Equation 4.11 results from multiplying Equation 4.7 by Pi_g as required by the definition of category share information in Equation 4.8.

Given the status of ICRFs in the GRM as being comprised of differences between adjacent CBRFs, Equation 4.11 also can be represented in category boundary terms (Samejima, 1969). With the usual notation of P^*_{ig} to represent a CBRF, Equation 4.11 becomes

$$I_{i_g}P_{i_g} = \frac{(P^{*\prime}_{i_g} - P^{*\prime}_{i_{g+1}})^2}{(P^*_{i_g} - P^*_{i_{g+1}})} - (P^{*\prime\prime}_{i_g} - P^{*\prime\prime}_{i_{g+1}}). \qquad (4.12)$$

Then, from Equation 4.9, we know that we can obtain item information by simply summing either Equation 4.11 or Equation 4.12 over the categories of an item. However, Samejima (1969) also showed that the sum of the second derivatives of the ICRFs for an item equal zero. Thus, the elements to the right of the subtraction symbol in Equations 4.11 and 4.12 equal zero when summed across the categories of an item. Item information, when presented in terms of these two equations, therefore can be simplified to

$$I_i = \sum_{g=0}^{m} \frac{(P'_{i_g})^2}{P_{i_g}}, \qquad (4.13)$$

and in CBRF terms,

$$I_i = \sum_{g=0}^{m} \frac{(P^{*\prime}_{i_g} - P^{*\prime}_{i_{g+1}})^2}{(P^{*}_{i_g} - P^{*}_{i_{g+1}})}. \qquad (4.14)$$

Figure 4.1 Category Information (I_{ig}) for a Five-Category Item

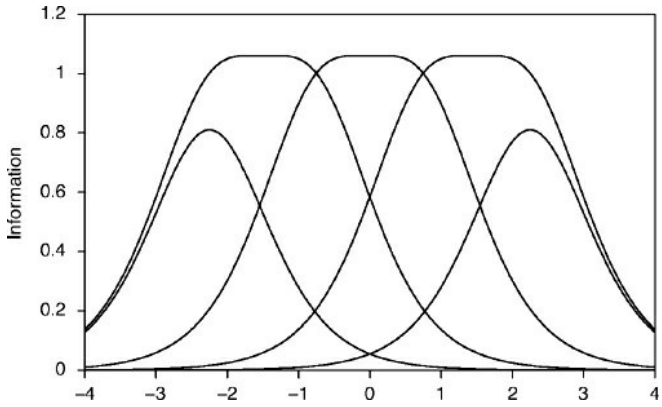
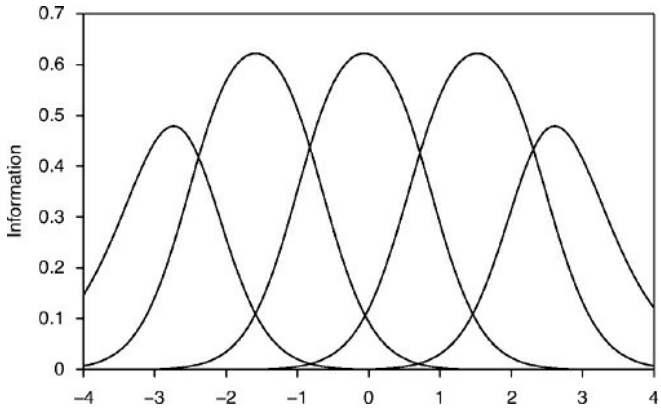


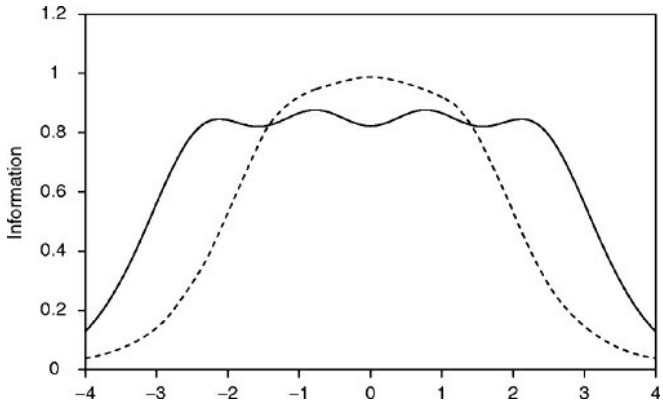
Figure 4.2 Category Share Information (I_{ig} P_{ig}) for a Five-Category Item



Figures 4.1 to 4.3 show examples of the category information, category share information, and item information, respectively, for the five-category item shown earlier in Figure 1.2. Note that this item is modeled with the more usual logistic version of the GRM. Interestingly, the category information functions for the normal ogive form of the GRM are dramatically different from those for the logistic GRM (see Samejima, 1983, for examples). However, subsequent information share and item information functions for the two manifestations of this model are very similar (Samejima, 1983). Figure 4.1 clearly shows that individual categories, in isolation, provide more information than do the same categories taken together (Figure 4.2).

Figure 4.3 Item Information for Two Five-Category Items

Note: Dashed lines are for an item with closer category boundaries.



Baker (1992), following Samejima (1975), notes that the distinction between response category information, I_ig, and category share information, $I_{ig}P_{ig}$, exists only in polytomous IRT models. This provides a useful reminder of the differences between dichotomous and polytomous IRT models. A difference highlighted here is that polytomous item categories overlap in their contribution to item information (and therefore test information), hence the need for weighted rather than simple sums of category information in the aggregation process. This is another manifestation of the situation described earlier concerning the fact that category discrimination is complexly defined as a combination of the amount a category discriminates and the distance between category boundaries.

The issue of category width also arises in cases where widely spaced category boundaries produce multimodal information functions for the nonextreme categories. This occurred in the demonstration item in Figure 4.3. An example of an item information function for more closely spaced categories is shown with a dashed line in Figure 4.3. Here, the category boundary locations are $b\mathbf{i_g}$= −1.25, −0.25, 0.25, and 1.25, for g = 1,…, 4, respectively, with a_i again equal to 1.8. This dashed function is unimodal and also illustrates the fact that closer boundaries provide more information over a smaller range of the trait scale than do more widely spaced boundaries.

It is useful to recall that the amount of information provided by a polytomous item increases as the number of categories for an item increases (Samejima, 1969), resulting in smaller standard errors for respondents' trait estimates as categories are added (Baker, 1992). Thus, an advantage that arises from accepting the added complexity of polytomous discrimination and information is that polytomous items provide greater amounts of information than do dichotomous items.

Relationship to Other IRT Models

The GRM occupies an interesting place among polytomous IRT models. As a difference model in Thissen and Steinberg's (1986) terminology, it constitutes the primary distinct type of polytomous IRT model to the Rasch and other divide-by-total models. However, as a member of Samejima's (1972) framework, it is related to the other models in the framework. In terms of models for ordered categorical data, this makes the GRM related to all of the divide-by-total models through their membership in the heterogeneous class of models (see Table 4.1).

The Heterogeneous Case

Two features distinguish the heterogeneous case from the homogeneous case in Samejima's (1972) framework. Unlike the homogeneous case, which requires ordered data, the heterogeneous class of models in Samejima's framework can accommodate nominal data. Bock's (1972) NRM represents this element of the heterogeneous class of models.

At the theoretical level, however, the fundamental difference between Samejima's two classes of models is in the shape of the cumulative boundary functions, P^*_{ig}. Whereas for models in the homogeneous case, these functions are the same shape within a given item, this need not be the case in the heterogeneous class of models, where the functions may have different shapes within an item (Samejima, 1988, 1996, 1997b).

Consider again the role, in the homogeneous case, of the P^*_{ig} functions, which is to define cumulative category boundaries that are used to obtain the probability of responding in any given category. Given this role, it would be a problem to have boundary functions with different slopes. The problem is that such functions would cross at some point on the trait continuum, leading to negative values for the differences that are meant to represent category response probabilities (P_{ig}). It makes no sense, however, to have probabilities less than zero.

This problem is avoided in the models that represent the heterogeneous class because the P^*_{ig} do not serve the practical role of defining category response probabilities in these models. In other words, the parameters describing the shape of the P^*_{ig} are not estimated in the heterogeneous case. The category probabilities represented by ICRFs are instead obtained by other means for models that are members of the heterogeneous class of models. This has already been described for the most prominent members of this class, the polytomous Rasch models, the GPCM, and their variants.

Indeed, it is generally the case that heterogeneous model ICRFs *must* be obtained through some means other than estimating the parameters of the P^*_{ig} for use in Equation 1.2. The

reason is that having a model with $P*_{ig}$ that can vary in shape within an item typically means that the functions are not the shape of logistic ogives (Thissen & Steinberg, 1986). This makes the task of estimating, or even defining, the parameters for such functions quite difficult, because the functions do not all have a simple common form.

It is possible to obtain the P^*_{ig} functions for models in the heterogeneous case by virtue of their definition as cumulative probability functions. However, in this case, the P^*_{ig} are obtained from the ICRFs rather than being used to produce the ICRFs, as they are in the homogeneous case. In the heterogeneous case, the P^*i_g are simply obtained by summing the probabilities of responding in the category or categories for the cumulative category boundary of interest and higher.

Representations for $P*i_g$ therefore can be produced for the Rasch and NRM ICRFs shown earlier. Figure 4.4 shows $P*_{ig}$ obtained by accumulating PCM ICRFs, whereas Figure 4.5 shows the $P*_{ig}$ obtained by accumulating the appropriate NRM ICRFs from Figure 2.1. Note that these functions are subtly different in shape. For both sets of functions, as one moves up the θ scale, we see functions whose asymptotes approach zero more quickly and simultaneously approach 1.0 more slowly. In other words, considering each ordered function in turn, we find that at any given value of θ, the associated probability value is always higher for earlier $P*i_g$ than for later functions. This ensures that the functions do not cross, that they differ in shape, and results in functions that are not logistic (or normal) ogives in shape. (Note that even though the functions have an ogive-like *s* shape, in mathematical terms, they do not meet the formal requirements for logistic, or normal, ogives.)

It may be obvious that these functions serve no practical, model-building purpose. Indeed, the only role they have is at the theoretical level, where they define the heterogeneous class of polytomous IRT models, enabling a level of comparison of the two distinct types of models within a common framework.

Figure 4.4 Heterogeneous Case $P*_{ig}$ Obtained by Accumulating PCM ICRFs

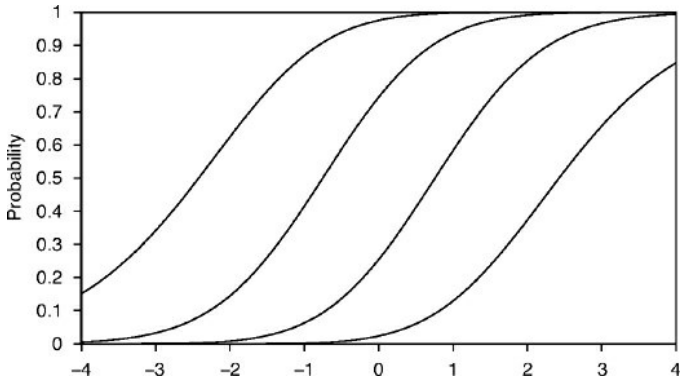
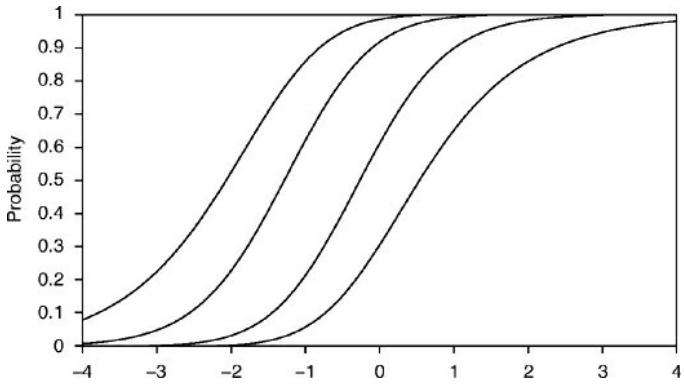


Figure 4.5 Heterogeneous Case $P*_{ig}$ Obtained by Accumulating NRM ICRFs



It is also of some interest that, even though Samejima (1972) was aware of the possibility of constructing models within the heterogeneous class in her theoretical framework, she concluded that they were unlikely to be useful or plausible (Samejima, 1979b, 1996, 1997a). As a result, she did not develop any specific, practical models within this theoretical class of models until quite recently, when she developed the acceleration model (Samejima, 1995). The acceleration model is an example of Samejima's ongoing interest in developing response models that represent a plausible psychological response process.

At a theoretical level, the acceleration model is itself, like the GRM, one of a family of potential acceleration-type models (Samejima, 1995, 1997b). As with the non-Samejima heterogeneous models, the acceleration model cannot be built on the cumulative category boundaries defined by $P*_{ig}$. Instead, in this case, it is constructed directly on the processing function $M*_{ig}$(Samejima, 1995, 1997b), which was noted earlier to be the theoretical basis for the $P*_{ig}$.

The acceleration model was developed specifically to provide a way to model, in detail, complex cognitive tasks (Samejima, 1995). In this context, a test item becomes a cognitive task that is made up of a number of problem-solving steps that represent the item categories. The steps are modeled by parameters that are, in turn, modified by another parameter that represents a

sum of the cognitive subprocesses that are required to complete each step of the task. In a sense, the acceleration model represents Samejima's exponential analog to the linear logistic models of Fischer described earlier.

From Homogeneous Class to Heterogeneous Class and Back

It has been noted that it is possible to obtain cumulative category boundary functions, P^*_{ig}, from item category functions, P_{ig}, for all heterogeneous class models, whether or not the P^*_{ig} have an algebraic definition. This can be done even when the ICRFs have not been obtained from the P^*_{ig}. This is despite the fact that the cumulative CBRFs (P^*_{ig}) were originally introduced as the means of obtaining GRM ICRFs. The possibility of obtaining P^*_{ig} from P_{ig} is a trivial process for the GRM because the model is identified by having specific category probabilities (P_{ig}) that result from differences between consecutive cumulative probabilities (P^*_{ig}). Summing category probabilities to obtain cumulative probabilities is therefore hardly informative in the GRM case.

In contrast, summing category probabilities to obtain cumulative boundaries is the practical mechanism that makes concrete the existence of the heterogeneous class as a component of Samejima's (1972) framework. Were it not possible to obtain P^*_{ig} from P_{ig}, it would not be possible to compare the NRM, Rasch models, and Tutz and Samejima's sequential models with the GRM in the context of a common framework.

Saying that cumulative probabilities can be obtained as the sum of specific category probabilities can be expressed algebraically (Thissen & Steinberg, 1986) as

$$P^*_{i_g} = \sum_{g=g}^{m} P_{i_g}. \qquad (4.15)$$

This, in turn, is equal to the ratio

$$P^*_{i_g} = \frac{\sum_{g=g}^{m} e^{Z_{i_g}}}{\sum_{g=1}^{m} e^{Z_{i_g}}}. \qquad (4.16)$$

The exponential term, Z_{ig}, varies from model to model in reasonably obvious ways, but several of the more relevant options will be laid out here. In the case of the NRM,

$$z_{i_g} = a_{i_g}\theta + c_{i_g}. \qquad (4.17)$$

In the case of the PCM,

$$z_{i_g} = \theta + b_{i_g}. \qquad (4.18)$$

In the case of the RSM,

$$z_{i_g} = \phi_i(\theta + b_i) + \kappa_g. \qquad (4.19)$$

In the case of the ELM, that is, the scoring function formulation of the PCM,

$$z_{i_g} = \phi_i(\theta + b_i) + \kappa_{i_g}. \qquad (4.20)$$

In the case of the GPCM,

$$z_{i_g} = a_i\left(\phi_i(\theta + b_i) + \sum \tau_{i_g}\right). \qquad (4.21)$$

When the appropriate expression for Z_{ig} from Equations 4.17 to 4.21 is placed in Equation 4.16 and summed over the relevant categories in the numerator and denominator, a cumulative boundary function, $P*_{ig}$, results. These are defined in the same way as GRM $P*_{ig}$ in that they provide the probability of responding in category *g* or higher. If one wants to think of $P*_{ig}$ purely as CBRFs, then this process describes how one can obtain globally defined CBRFs (the $P*_{ig}$) from locally defined, Rasch-type CBRFs. However, because these models are all examples of the heterogeneous class, the resulting $P*_{ig}$ need not be the same shape within an item. In fact, as already mentioned and shown in Figures 4.4 and 4.5, they are unlikely to even be logistic functions (Thissen & Steinberg, 1986).

Hemker (1996) addresses this issue by noting that divide-by-total models do not have a simple parametric form for cumulative category boundaries ($P*_{ig}$). Reciprocally, the difference models do not have a simple parametric form for locally defined, adjacent category boundaries.

However, in all other current examples of heterogeneous models, such as the adjacent category models, model parameters for functions other than $P*_{ig}$ are used to estimate ICRF probabilities. The ICRFs are then themselves used to obtain the shape of the $P*_{ig}$ if this is desired, through the process described above. The alternative ICRF model parameters include locally defined category boundary locations, or item and threshold locations. Typically, these are obtained in the context of divide-by-total (usually Rasch model) format.

A Common Misconception

We hope that the foregoing discussion of heterogeneous class models in Samejima's (1972) framework will have made reasonably clear what $P*_{ig}$ represents in the context of these models. Specifically, these cumulative functions are, in a sense, artificial constructions based on model parameters obtained in other contexts—often by first modeling items with a polytomous Rasch model. It is worthwhile, however, clarifying precisely what the $P*_{ig}$ do not represent in the heterogeneous case.

Samejima (1969) initially developed the GRM, and did so on the basis of 2PL-defined $P*_{ig}$

CBRFs, where the boundary discriminations can differ across items but not within items. This is the classic example of a specific model in the homogeneous case in her framework. Her subsequent expansion of the framework to include the heterogeneous case, where models are defined by P^*_{ig} that differ in shape *within* items, seems to have led occasionally to the impression that this means that "the" heterogeneous model is one where P^*_{ig} CBRFs are estimated using the 2PL model with item discriminations allowed to vary within items. This is not the case, and there is actually no such model. Estimating cumulative CBRFs in this way would result in crossing boundary functions, which would then produce negative category probabilities when P^*_{ig} differences were taken according to Equation 1.2.

Even the tactic of artificially constraining boundary differences to be positive does not solve the problem of crossing boundaries; it only avoids the problem of negative probabilities at the cost of having segments of the trait continuum where the sum of probabilities in all categories fails to add up to 1.0. This is a poor solution that tries to make the misrepresentation of heterogeneous case models into a workable model.

Ultimately, the only specific models in the heterogeneous class are models such as the NRM in the nominal case, together with the polytomous Rasch models and the acceleration model in the ordered categorical case. There is no such entity as a specific heterogeneous GRM.

Variations

Continuous Response Model

A variation of the GRM developed by Samejima (1973) is the continuous response model (CRM). This model is for data where responses can be made at any point on the response continuum, rather than being restricted to a finite set of discrete categories. The CRM can be thought of as a variation of the GRM because Samejima develops it as the limiting case of the ordered response situation. This is done in a manner similar to the approach described earlier for Müller's (1987) continuous Rasch model. Specifically, beginning with a finite segment of the response continuum, Samejima (1973) postulates an ordered categorical response situation with categories of equal size. These are successively halved in size until, in the extreme case of an infinite number of partitionings, the "categories" represent every point on the continuum, thereby degenerating to the continuous response level.

The CRM is also a variation of the GRM in that Samejima (1973) develops it within the homogeneous class of models in her framework. Müller's (1987) continuous Rasch model is an example of a continuous response model in Samejima's (1972) heterogeneous class of models.

The $P*$ functions can be modeled any number of ways in the CRM. Samejima (1973) outlines three possibilities for a continuous response type of model, the normal ogive and logistic functions, as well as a hybrid, double exponential function. Parameter estimation is a problem, and a solution is found only for the normal ogive-based model. Because this is the only practical implementation of the potential models, it is the model referred to above as the CRM.

Multiple-Choice GRM

Another interesting variant of the GRM is provided by Samejima (1979b). Previously, we introduced Samejima's modification of Bock's (1972) NRM, which allows for the possibility of modeling respondent guessing. If the response options of multiple-choice items have a known (or previously identified) order with respect to the latent trait, then these items could be modeled by the GRM instead of the NRM, which is required only when responses are unordered. However, the concern with guessing in the face of multiple-choice items would then remain. Samejima (1979b) shows that her modification for the NRM can be equally well applied to the usual logistic function version of the GRM as well as the less common normal ogive version, if either of these models is being applied to multiple-choice data.

Rating Scale GRM

Muraki (1990) develops a variant of the GRM that is also applied to both the logistic and normal ogive versions of the model. Muraki's contribution is to constrain the GRM into a form that is targeted to rating scales, that is, he develops a rating scale GRM (RS-GRM). He does this by reparameterizing the GRM CBRF location parameter, b_{ig}, into a single item location parameter, b_i, and a set of boundary location parameters, c_g, so that $P*_{ig}$ in the logistic version of the RS-GRM is now represented by

$$P_{i_g}^{*} = \frac{e^{a_i(\theta - b_i + c_g)}}{1 + e^{a_i(\theta - b_i + c_g)}}, \qquad (4.24)$$

instead of Equation 4.1.

This reparameterization is formally equivalent to the way the RSM is reparameterized relative to the PCM (Equations 3.1 and 3.5, respectively). The rationale for the reparameterization is also the same, that is, that adopting a rating scale for a set of items implies that the scale should be operating equivalently across items. This implies that specific categories should be the same size across items and not unique to each item. This is accomplished, in the RS-GRM, by estimating a single set of category boundary parameters, c_g, that is applied to all items that differ only in their scale location, which is indexed by the parameter b_i.

It is important to realize, however, that the RS-GRM category boundary parameters, c_g, are not the same parameters as the RSM threshold parameters, τ_g, although both serve the same function. The c_g are obtained in the global item context of GRM item location parameters, whereas the τ_g are obtained in the adjacent category context of Rasch model item location parameters.

The RS-GRM also differs from the RSM in that it can include an item discrimination parameter as well as a location parameter. Thus, although the category boundary locations are expected to be the same across items using the same rating scale, the items are permitted to vary in discrimination. This makes the RS-GRM closer to a rating scale version of the GPCM (or a generalized RSM) than the RSM itself, in terms of item parameters that may vary. Muraki (1990) does, however, also demonstrate an RS-GRM application with discrimination fixed across items. This form of the model is more directly comparable to the RSM in terms of parametric complexity.

It is interesting to note that, although the justification for the RS-GRM is the same as for the RSM, Muraki's motivation for developing the RS-GRM is actually to be able to separate the estimation of item parameters from that of the category boundary parameters. The demonstrated reason for this is so that he could compare changes in item location over time by looking at a single item parameter while keeping category boundaries fixed.

Summary of Samejima Models

Samejima (1972) developed a comprehensive framework for extended item formats. The framework arguably encompasses all polytomous IRT models. She has developed three models within this framework: the GRM (two versions), the CRM, and the acceleration model. In addition to her specific models, Samejima also developed a modification for the NRM and the GRM to accommodate potential guessing in multiple-choice items. Much of her more recent work has also focused on nonparametric IRT estimation procedures and other ways of obtaining cumulative boundary functions with more flexible shapes than that provided by a simple 2PL ogive.

Numerous other models have been described in Samejima's work in theoretical terms or as practical possibilities without ever being developed to a usable level. The logistic version of the GRM has been the model most commonly used by other researchers. This IRT extension of Thurstone's method of successive intervals has come to symbolize the global dichotomization approach to polytomous IRT model building that uses cumulative category boundaries as its basis. As such, the GRM is the only distinctive polytomous IRT modeling approach to the

divide-by-total approach (NRM, Rasch models, and their variants) that has been used in psychometric research. The sequential model approach of Tutz (1990) and Samejima (1995) is a third distinctive approach but has yet to generate a substantial research profile.

Making comparisons between the two major approaches is not straightforward. Unlike the dichotomous case, where the Rasch, 2PL, and 3PL models form a nested hierarchy of models, at least in mathematical terms, there is no such relationship among the two major approaches to polytomous IRT modeling. Therefore, comparing polytomous Rasch models and the GRM in terms of their respective number of parameters to be estimated can be unhelpful. For example, extending the PCM by allowing a variable item discrimination parameter produces the GPCM, not the GRM. These are two fundamentally different models, even though they are both “2-parameter” polytomous models.

An important consequence of the different definition of the boundary location parameters is that it allows Rasch models, and other models based on adjacent category comparisons, to decompose discrimination into an item and a category component. The cumulative boundary approach underlying the GRM allows only item discrimination to vary. This provides the Rasch models and their variants the capacity to represent the structure of a polytomous item in a more complex manner than is possible for the GRM. It suggests that these models may be more flexible than the GRM.

If polytomous Rasch and similar models are more flexible than the GRM, it would be because the adjacent category approach is inherently more flexible than the cumulative boundary approach. That is, modeling only adjacent categories is less restrictive in data modeling terms than modeling cumulative boundaries.

A comparison of structural flexibility is difficult because of the different definitions of the respective location parameters and their distinctive functions. However, Samejima's framework provides a basis of comparison, essentially by converting category probabilities (from ICRFs) based on adjacent category boundaries into cumulative boundaries. As was mentioned earlier, such transformed boundaries need not be the same shape (hence the definition of heterogeneous models). This leads Samejima (1997b) to conclude that heterogeneous models (such as the polytomous Rasch models) are likely to better fit data given the resulting greater variety of possible ICRF shapes.

Direct empirical comparisons are difficult to construct and therefore rare. Maydeu-Olivares et al. (1994) found that each type of model better fit data generated following their own structure. However, Dodd's (1984) dissertation, with data that were not generated to fit either model,

suggested that the PCM could hold its own against the GRM in terms of information provided, despite being a simpler model. This conclusion was further supported by the fact that a restricted GRM, with constrained discrimination, performed considerably more poorly than either the PCM or the GRM, even though it was included in the project as a fairer comparison for the simpler PCM.

Potential Weaknesses of the Cumulative Boundary Approach

In terms of sheer number of model variations, the adjacent category approach is clearly superior to the cumulative approach of the GRM. This flexibility was noted earlier as a strength of polytomous Rasch models. Kelderman's (1984) loglinear model framework is an example of this flexibility. Complementary work by Agresti (1997) describes general relationships between Rasch models and loglinear models for unordered and ordered categorical data. In this work, Agresti also describes relationships between these models and a cumulative boundaries model that is very closely related to Samejima's (1969) GRM. Despite this tenuous link to the more general statistical context of loglinear models, the cumulative boundaries approach underlying the GRM does not appear to embody the same level of flexibility as is found in polytomous Rasch models.

Possible Strengths of the Cumulative Boundary Approach

The GRM does allow item discriminations to vary from item to item, but relaxing the specific objectivity requirements of Rasch models allows them to also be extended in this way, as in the case of the GPCM. However, if the choice is made to pursue flexibility in the area of discrimination, the adjacent category approach is still more flexible because it allows the possibility of varying not only item discrimination but also category-level discrimination. This cannot be done in the GRM because of the resulting negative probabilities. However, being able to model category-level discrimination enhances model flexibility by enabling models with the capacity to model partly ordered and partly nominal data (e.g., the OPM), null categories, and completely ordinal data (NRM).

The area in which the GRM has a less equivocal advantage over adjacent category models is with respect to the issue of the meaning of the boundary location parameter. The interpretation of the boundary location parameter is unambiguous in the case of cumulatively defined boundaries and sequential models. However, it is not clear how one should interpret this parameter in the case of locally defined, adjacent category boundaries. It is generally agreed that it cannot legitimately be interpreted as modeling sequential response steps, as was initially proposed. What is not clear, however, is whether the ambiguous interpretation of this parameter has any important measurement consequences, over and above the conceptual difficulties that

it raises.

Samejima's research on polytomous item information has application across both major types of polytomous IRT models. Much of this application has been in the context of implementing polytomous item CAT procedures, although it is also useful for guiding test construction. The widespread applicability of her work on information complements Samejima's theoretical framework by providing a valuable practical tool that spans the spectrum of polytomous IRT models.

TABLE 4.2 Parameter Values and Response Probabilities for a Morality Item Using the Graded Response Model

			i		
Model		*1*	*2*	*3*	*4*
GRM	$a_i = 0.719$ b_{i_g}	−3.558	−0.878	1.146	4.114
			Item Category		
	1	*2*	*3*	*4*	*5*
$p\|\theta = -1.5$	0.185	0.425	0.260	0.112	0.017
$p\|\theta = 1.5$	0.026	0.128	0.284	0.431	0.132

A Practical Example

The same Likert-type response example item used to demonstrate the PCM, RSM, and GPCM is also used in the following practical example demonstrating the GRM. As was the case with the GPCM, five unique parameters must be estimated to model our example item with the GRM. The unique item discrimination parameter and four unique boundary location parameters are shown in Table 4.2. It must be remembered, however, that in this case, the b_{ig} represent the locations of cumulative boundary functions. Because these are a different type of function to the boundary functions of any other model for which practical examples have previously been provided, it is not meaningful to directly compare location parameters, even where the same item is being modeled.

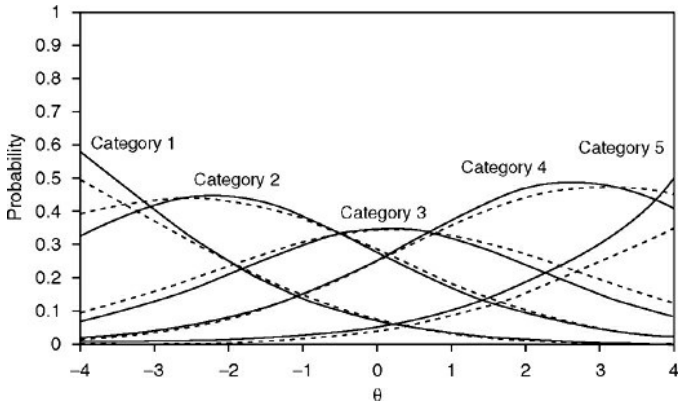
An example of the differences between the two types of boundary location parameters is that GRM-modeled CBRF locations are always, invariably sequential, unlike adjacent category boundary locations. However, it is still possible to obtain diagnostic information about item category functioning from GRM ICRFs, which will represent potentially problematic categories as “never most probable” in the same way that adjacent category model ICRFs do. Note that when modeled by the GRM, there is a segment of the trait continuum where a response is most probable for each of the five categories of our example item. That is, as was the case when this item was modeled by the adjacent category models, there is no category that is never the most probable category at any point on the trait continuum. This is demonstrated graphically in the

GRM ICRFs shown in Figure 4.6.

Although it is not meaningful to compare CBRFs across adjacent category and cumulative boundary models, the response probabilities modeled by the ICRFs are the same type of probability across both types of models. Therefore, response probabilities can be compared at specific trait levels, such as those shown in Table 3.4. Equivalent response probabilities for each item category, as modeled by the GRM, at θ = −1.5 and θ = 1.5 are shown in the bottom half of Table 4.2. These probabilities show that Category 2 is the most probable response at θ = −1.5, whereas at θ = 1.5, the most probable response category is Category 4. This is the same result as was obtained for the PCM and the GPCM. Perusal of specific response probability values, at both trait levels and across all five categories, shows that the probabilities obtained for the GRM are very similar to the GPCM results (see Table 3.4). A more comprehensive demonstration of this similarity is provided in Figure 4.6, where GPCM ICRFs are superimposed as dashed lines over the solid GRM ICRFs.

Figure 4.6 ICRFs for a Likert-Type Morality Item Modeled by the GRM

Key: Solid line = GRM; dashed line = GPCM.



The similarity of the two sets of results is likely to be due to the fact that both the GRM and GPCM specifically model a discrimination parameter for each item, rather than constraining discrimination to a fixed value as occurs for the PCM and RSM. It is particularly interesting that, ultimately, each of the five categories' response probabilities for the two models is so similar despite the model's respective ICRFs being built on very different types of CBRFs and, effectively, very different types of discrimination parameters. In effect, what has happened is that distinctively different parameter values have been estimated for each model's different types of a_i and b_{ig} parameters. This is most noticeable in the divergent a_i values (GRM a_i = 0.719; GPCM a_i = 0.395). Yet when these divergent values for the distinctive types of discrimination and boundary location parameters are combined within both types of model, they result in very similar category response probabilities across the trait continuum for our example

item.

http://dx.doi.org/10.4135/9781412985413.n4