Journal of Educational and Behavioral Statistics
Fall 2003, Vol. 28, No. 3, pp. 195–230

Using Data Augmentation and Markov Chain Monte Carlo for the Estimation of Unfolding Response Models

Matthew S. Johnson
Baruch College, City University of New York

Brian W. Junker
Carnegie Mellon University

Unfolding response models, a class of item response theory (IRT) models that assume a unimodal item response function (IRF), are often used for the measurement of attitudes. Verhelst and Verstralen (1993) and Andrich and Luo (1993) independently developed unfolding response models by relating the observed responses to a more common monotone IRT model using a latent response model (LRM; Maris, 1995). This article generalizes their approach, and suggests a data augmentation scheme for the estimation of any unfolding response model. The article introduces two Markov chain Monte Carlo (MCMC) estimation procedures for the Bayesian estimation of unfolding model parameters; one is a direct implementation of MCMC, and the second utilizes the data augmentation method. We use the estimation procedure to analyze three data sets, one simulated, and two from real attitudinal surveys.

Keywords: *data augmentation, latent response model, Markov chain Monte Carlo, unfolding response models*

1.

A number of methods have been developed in the past eight decades for the measurement of attitudes in surveys and other self-report instruments. Such attitudes range from students' attitudes towards topics of instruction or teaching styles, to changes in smokers' attitudes as they approach successive change (e.g., Noël, 1999), to citizens' attitudes towards major policy issues (e.g., Formann, 1988) or towards staying informed about politics generally (Muhlberger, 1999). One of the methods suggested in the literature to study such attitudes is the direct response method. In the direct response method respondents are given a set of J items (statements or other stimuli) and asked to examine each one and determine whether or not to endorse it.

Item response theory (IRT) can be used to model direct response unfolding data, if the central assumption of monotonicity of the item response functions (IRF's) is suitably modified. Standard, unidimensional IRT assumes that there is a single, real-valued latent variable θ being measured; in unfolding models the sign of θ represents the valence of the respondent's attitude (political liberalism versus conservatism, for example) and the magnitude $|\theta|$ represents the intensity of the respondent's attitude. When the items are statements that the respondent can endorse (denoted $X_j = 1$) or not (denoted $X_j = 0$), the usual item response function (IRF),

$$P_j(\theta) \equiv Pr\{\text{Subject endorses item } j \text{ given the subject is located at } \theta\},$$
$$= P[X_j = 1|\theta]$$

still plays a central role in modeling.

One of the original direct response methods, suggested by Thurstone (1927, 1928), prescribes that, in order to measure the unidimensional latent attitudes of subjects, θ, we should first premeasure the survey items. This premeasurement is achieved by giving the items to a number of judges who are asked to rank the items from one extreme to the other on the latent scale. These ranks are used to estimate the locations, β_j, of each item on the real line. The items are then given to survey respondents, and the location of these respondents on the real line is estimated as the average location of those items, which are endorsed.

Thurstone's procedure has the disadvantage that it requires two stages of estimation. Coombs (1964) suggests a method, which he calls *unfolding* that allows for the joint estimation of item and respondent locations with a single data collection step. His method, based on the assumption that subjects agree with those items, and only those items which are located near their location on the latent scale, implies that the IRF $P_j(\theta)$ for endorsing the item is deterministic, that is, $P_j(\theta)$ takes only the values 0 and 1. Both Thurstone's and Coombs' methods imply that the item response function is a unimodal function of θ. The assumption of unimodality of the IRF is in contrast to the usual assumption in applications of IRT to testing in which the probability that an examinee correctly responds to a test item is a nondecreasing function of θ. Item response models that assume a unimodal item response function are called *unfolding response models.*

In any unimodal item response model, we may define the *location of item j* as the point at which the IRF reaches a maximum (or the midpoint of the interval of all such points), that is,

$$\beta_j = \text{argmax}_\theta P_j(\theta).$$

We may further define the *latitude of acceptance* for each item, as the (minimum) radius of the set of θ for which the probability of endorsing the item is high,

$$\delta_j = \max\{\delta : \min[P_j(\beta_j - \delta), P_j(\beta_j + \delta)] \geq P_j(\beta_j) - \varepsilon\}$$

for some predetermined $\varepsilon > 0$.

Although Coombs's method allows for measurement of the survey stimuli using the survey responses, its assumption of a deterministic response function, specifically

$$P_j(\theta) = \begin{cases} 1, & \text{if } |\theta - \beta_j| < \delta_j; \\ 0, & \text{otherwise,} \end{cases}$$

is too restrictive. For example, if there are four ordered stimuli A, B, C, D, then at most 11-item response patterns are valid because no pattern can contain the triplet 1, 0, 1. For example a person who agreed with A and C, but not B, violates the deterministic model.

The inflexibility of the deterministic model prompted the development of probabilistic unfolding models (Davison, 1977; Andrich, 1988; Hoijtink, 1990; Luo, 1998). The probabilistic unfolding models assume conditional independence (CI), unidimensionality (U), and unimodality of the response function [i.e., the response function is single peaked] (SP). The probabilistic parameterizations of the IRF $P_j(\theta)$ allow for statistical inferences to be made on both the subject's location θ, and the item locations β_j. The squared logistic model and the hyperbolic cosine model are two examples of probabilistic unfolding models.

The squared logistic model

The simplest example of a probabilistic unfolding IRF is the squared logistic model (SLM; Andrich, 1988). The SLM assumes that the logit of the item response function is quadratic in the subject's location. The IRF is parameterized

$$P_j(\theta) = \frac{1}{1 + \exp\{(\theta - \beta_j)^2 - \gamma_j\}}. \tag{1}$$

The SLM IRF is symmetric around β_j and reaches a maximum of $(1 + e^{-\gamma_j})^{-1}$ when $\theta = \beta_j$.

The hyperbolic cosine model

Andrich and Luo (1993) and Verhelst and Verstralen (1993) independently developed an unfolding response model with the aid of a *latent response model* (LRM; Maris, 1995). The formulation assumes that each respondent has a one-dimensional trichotomous latent response that indicates one of "Disagree Below," "Agree," or "Disagree Above." The observed responses are then simply a mapping of these responses. A latent response of "Agree" maps to the observed response "Agree" and the two disagree latent responses map to the one observable "Disagree" category. If the latent responses are observed, rather than unobserved, the researcher has a large number of models from standard item response theory with which to measure the latent attitude θ. Verhelst and Verstralen (1993) and Andrich and Luo (1993) developed the hyperbolic cosine model [HCM] by assuming the partial credit model (PCM; Masters, 1982) for the latent responses. The resulting model parameterizes the item response model with the equation

$$P_j(\theta) = \frac{e^{\gamma_j}}{e^{\gamma_j} + 2\cosh(\theta - \beta_j)}. \tag{2}$$

The IRF is symmetric around the location β_j, and the maximal endorsement probability $P_j(\beta_j) = \frac{1}{1+2\exp\{-\gamma_j\}}$ is a function of what Andrich and Luo (1993) call the item-unit parameter γ_j. The item-unit parameter γ_j in Equation 2 is a parameter associated with the item that measures of how likely a subject located at β_j is to endorse item *j*. Figure 1 displays three HCM item response functions located at $\beta_1 = -2$, $\beta_2 = 0$, $\beta_3 = 3$ with item-unit parameters $\gamma_1 = \gamma_3 = 2$ and $\gamma_2 = 3$.

Section 2 introduces three examples that will be discussed throughout the article. The first is a survey from Formann (1988) that studies the opinions of individuals on the use of nuclear power plants. The second is a survey developed to study the psychological reasons people follow politics (Mulhberger, 1999). The third is an artificial data set simulated from the HCM to resemble Formann's nuclear power plant data, used to illustrate parameter recovery with the estimation procedures described below.

The LRM formulation for unfolding models is reviewed in Section 3. This formulation suggests a straightforward data augmentation method useful for estimation of such models. Section 4 explores the data augmentation method and suggests a generalization of the latent response formulation of the unfolding models sug-

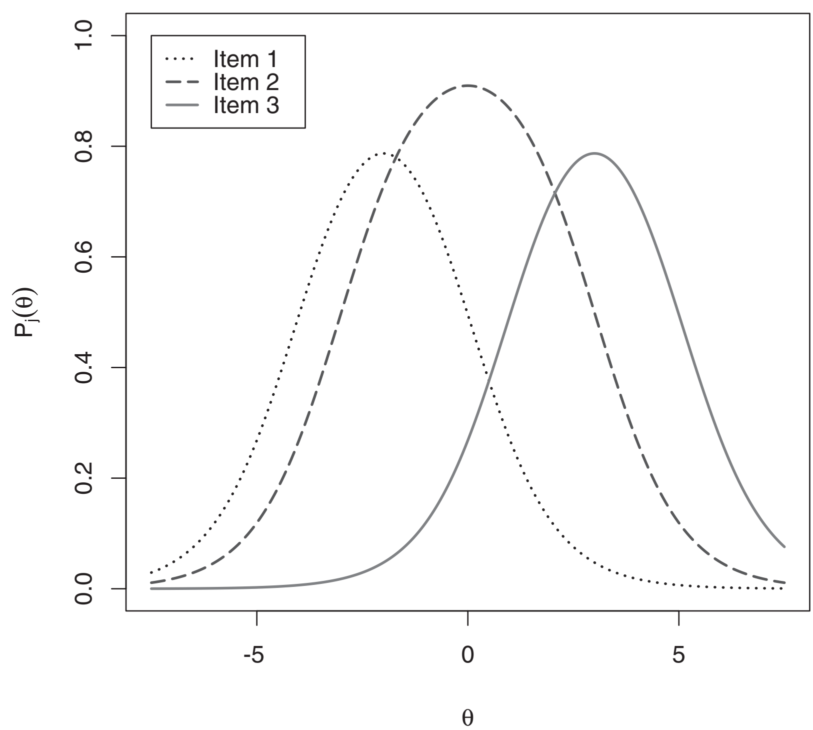

FIGURE 1. *The item response functions of the hyperbolic cosine model located at* $\beta_1 = -2$, $\beta_2 = 0$, $\beta_3 = 3$, *with item units* $\gamma_1 = \gamma_3 = 2$ *and* $\gamma_2 = 3$.

gested by Andrich and Luo (1993) and Verhelst and Verstralen (1993); the generalization is useful for unfolding response models such as the squared logistic model that do not naturally have a "nice" latent response formulation.

Estimating unfolding IRT models has been a serious problem in the literature (Hoijtink, 1990; Verhelst & Verstralen 1993) until quite recently (Roberts, Donoghue, & Laughlin, 2000; Luo, 2000; Maris & Maris, 2002). In Section 5.1 we review some methods of estimation that have been tried and some weaknesses of these methods, and in the remainder of Section 5 we introduce a hierarchical Bayesian framework for unfolding models and two MCMC approaches to estimating the model in this framework. In Section 6 we illustrate these methods using the three data sets introduced in Section 2, and illustrate the flexibility of the data-augmentation based estimation methods described in sections 4 and 5.

2. Examples

This article explores different approaches to the estimation of unfolding response model parameters using three examples. The first two examples come from actual surveys. The third data set is a set of data simulated from the hyperbolic cosine model (HCM; Verhelst & Verstralen, 1993; Andrich and Luo, 1993) in such a way to resemble real data.

2.1 Attitudes on Nuclear Power Plants

Formann (1988) uses five opinions on nuclear power to measure the attitudes of 600 college students on the subject. Each subject is asked to either agree with the statement [coded as 1], or disagree with the statement [coded as 0]. The five items, translated from German (Formann, 1988) are:

1. In the near future, alternate sources of energy will not be able to substitute nuclear energy.
2. It is difficult to decide between the different types of power stations if one carefully considers all the pros and cons.
3. Nuclear power stations should not be put into operation before the problems of radioactive waste have been solved.
4. Nuclear power stations should not be put into operation before it is proven that the radioactive waste caused by them is harmless.
5. The foreign [sic] power stations now in operation should be closed down.

2.2 Survey of Political Motivation

Muhlberger (1999) presents a data set containing 113 subjects' responses to 16 items studying the psychological reasons individuals have for following politics. Muhlberger suggested that the survey items could be split into two groups. The first group, containing seven items, are assumed to measure *political agency.* The second group, containing nine survey items are believed to measure *politically integrated motivation,* with one of these nine items possibly removed from this group. With this item removed from the list, Muhlberger hypothesizes that these items represent, in order, more “developmentally adequate responses.” These eight items are:

1. I follow politics because that’s what I’m expected to do (not true / true).
2. I follow politics so people won’t be upset with me (true / not true).
3. I follow politics because I will feel bad about myself if I don’t (true / not true).
4. I follow politics because it bothers me when I don’t (not true / true).
5. I follow politics because I want to learn new things (true / not true).
6. I follow politics because I think it’s important (true / not true).
7. I follow politics because it’s fun (not true / true).
8. I follow politics because I enjoy it (true / not true).

The subjects were asked to respond to the items on a 48-point scale. In some cases this scale indicates the level of agreement of the subject with the item (e.g., Item 1) where a score close to 48 indicates that the subject believes the statement is true. In other cases the scale is reversed (e.g., Item 2) where a response near 48 indicates that the subject believes the statement is not true about themselves. The differences in direction are indicated by the (not true / true) and (true / not true) parenthetical remarks at the end of each item.

In order to facilitate the discussion of the binary unfolding response models the items have been dichotomized. Responses from 1 to 24 have been coded as 0, and responses from 25 to 48 have been coded as 1. Histograms of the original responses appear in Figure 2.

Figure 2 contains barplots for each of the eight political motivation items. Each bar represents three response categories, with the number of respondents given by the height of the bar. For example, examining the barplot for the second item, we find that 60 respondents selected one of the first three response categories. The eight barplots seem to indicate that respondents tend to endorse low categories (between zero and ten) much more than high categories, and the middle categories (around 24) seem to be much more popular than surrounding categories (15–20 and 30–35). Although some information is lost because of the dichotomization, the effect is expected to be relatively small because of the response tendencies of the respondents.

2.3 Simulated Data

Using the HCM described in Equation 2, the responses of 1,000 individuals are simulated for each of five items. The model parameters used to simulate the data were selected to resemble the estimates reported for the Formann’s nuclear power plant data in Section 2. The latent attitudes, θ, were drawn from a normal distribution with mean zero and standard deviation $\sigma = 2.0$. The agreement, or disagreement with each of the five items was determined using location parameters $\beta = (-2.5, -1.0, -0.9, 1.5, 4.0)$, and item unit parameters $\gamma = (1.62, 1.72, 3.85, 4.05, 3.97)$. The five item response functions for the simulated data appear in Figure 3.

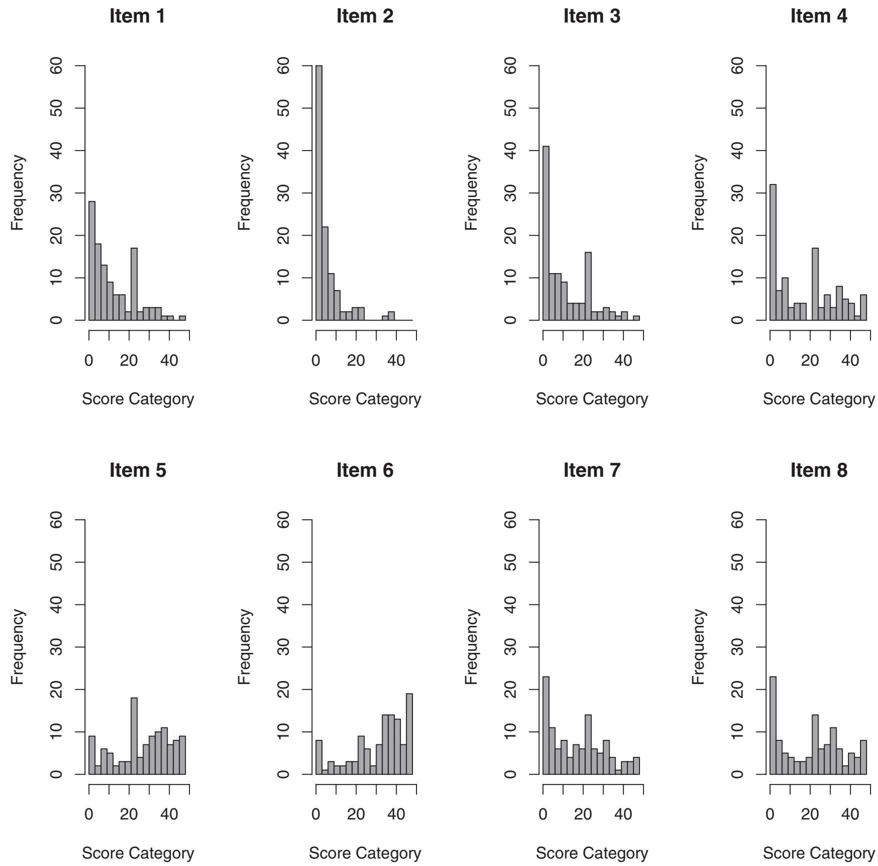

FIGURE 2. *Barplots of the number of observed responses in each of the forty-eight categories for the eight political motivation items. The total number of respondents on the survey was 113.*

3. The Unfolding Model as a Latent Response Model

Unfolding models developed by Andrich and Luo (1993), Verhelst and Verstralen (1993), and Roberts, Donoghue, and Laughlin (2000) assume the existence of an underlying latent response model (LRM; Maris, 1995) that defines the response behavior of the respondents. The LRM formulation assumes that the observed responses are simply standard responses from a monotone item response model that are recoded; this recoding parallels the ambiguity associated with the information lost using unfolding items.

Definition 3.1. Suppose there exists a set of M latent responses $\xi_{ij} = (\xi_{ij1}, \ldots, \xi_{ijM})^t$ *for each observed item response* X_{ij} *and a* $M \rightarrow 1$ *mapping of the latent responses called the condensation rule by Maris (1995), such that this mapping, denoted by* $C(\cdot)$, *maps the latent responses onto the observed responses,*

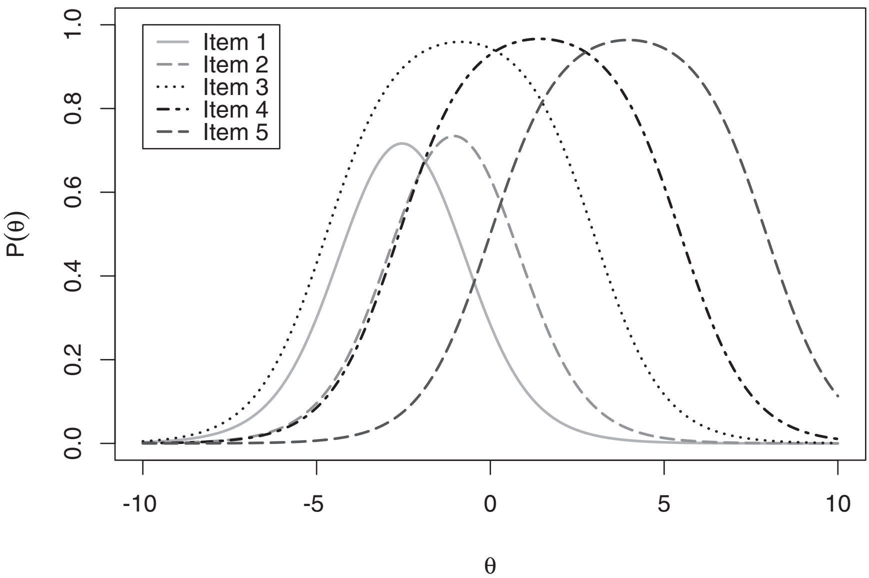

FIGURE 3. *The five hyperbolic cosine item response functions from which the responses of the 1,000 simulated respondents were simulated.*

$$X_{ij} = C(\xi_{ij}).$$

Then define the latent response model by the equation

$$Pr\{X_{ij} = x|\theta_i\} = E\{\delta[C(\xi_{ij}) - x|\theta]\}, \tag{3}$$

where $\delta(\cdot)$ is the Kronecker delta function:

$$\delta(y) = \begin{cases} 1 & y = 0 \\ 0 & otherwise. \end{cases}$$

Note: Although the focus here is on dichotomous responses, Maris' treatment allows the responses to be discrete or continuous, scalar or vector-valued.

Verhelst and Verstralen (1993) and Andrich and Luo (1993) assume that the latent response model is a monotone item response model with trichotomous responses. The basic assumptions are that each subject $i = 1, \ldots, N$ has an ordered, trichotomous latent response ξ_{ij} to each survey item $j = 1, \ldots, J$, indicating subject i's attitude or opinion regarding item j. For example, if a political survey is being conducted, a latent response of $\xi_{ij} = 0$ may indicate that subject i views item j as too far right politically for his liking. Similarly, $\xi_{ij} = 1$ indicates agreement of subject i with survey item j, and $\xi_{ij} = 2$ denotes disagreement of subject i with item j because the opinion in item j is too far left on the political scale.

Denote the observed response of subject i to item j by X_{ij}, where $X_{ij} = 0$ indicates disagreement of subject i to item j, and $X_{ij} = 1$ indicates agreement of the subject to the item. Note that the $X_j = 0$ is a catch-all-disagree category for item j. In the political survey example, the observed response is coded as disagree ($X_{ij} = 0$) whether a subject disagrees with an item because they believe the item is too far left or too far right for their liking. Mathematically this is written

$$X = C(\xi) = 1 - |1 - \xi|, \tag{4}$$

which Maris (1995) calls the *collapsing condensation rule.* Given the responses of N subjects to J unfolding-type survey items define the $N \times J$ response matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1J} \\ x_{21} & x_{22} & \dots & x_{2J} \\ \vdots & & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NJ} \end{bmatrix}$$

where element x_{ij} is the response of subject i to item j.

If the latent responses are observed, rather than unobserved, the researcher has a large number of models from item response theory to measure the latent attitude θ. Let

$$R_{jk}(\theta) = Pr\{\xi_j = k \mid \theta\} \tag{5}$$

denote such a model. Typically R_{jk} is defined by the following conditions.

Definition 3.2. An item response model with dichotomous or polytomous response variable ξ_j is called a monotone response model if the following three conditions hold:

- *(CI)* conditional independence: *given the subject's latent trait* θ, *the responses* $\xi = (\xi_1, \dots, \xi_J)$ *are conditionally independent. That is*

$$Pr\{\xi_j = x_1, \dots, \xi_J = x_j \mid \theta\} = \prod_{j=1}^{J} R_{jx_j}(\theta);$$

- *(U)* unidimensionality*: the latent trait is unidimensional,* $\theta \in IR$;
- *(M)* monotonicity *of the item response functions:* $Pr(\xi_j > t \mid \theta)$ *is non-decreasing in* θ, *for all j and t.*

See Holland and Rosenbaum (1986) for a similar development.

Combining Equations 3 and 5 leads to the general form of the latent response unfolding model:

$$P_j(\theta) = Pr\{X_j = 1 \mid \theta\} = R_{j1}(\theta), \tag{6}$$

$$1 - P_j(\theta) = Pr\{X_j = 0|\theta\} = R_{j0}(\theta) + R_{j2}(\theta). \tag{7}$$

A number of unfolding response models can be derived for the observed responses by choosing different forms of the latent response model $\mathbf{R}_j(\theta)$. One such model, the hyperbolic cosine model (Andrich & Luo, 1993; Verhelst & Verstralen, 1993) is summarized. For other examples of unfolding response models as latent response models see Roberts et al. (2000) and Maris and Maris (2002).

The Hyperbolic cosine model

Andrich and Luo (1993) and Verhelst and Verstralen (1993) model unfolding responses as a latent response model where the latent responses are assumed to follow a partial credit model (PCM; Masters, 1982). The partial credit model assumes that the adjacent category logits are linear in θ with unit slope. The probabilities of disagree below $\xi = 0$, agree $\xi = 1$, and disagree above $\xi = 2$ for the partial credit model can therefore be parameterized:

$$\begin{aligned}
R_{j0}(\theta) &= Pr\{\xi = 0|\theta\} = \frac{1}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}} \\
R_{j1}(\theta) &= Pr\{\xi = 1|\theta\} = \frac{\exp\{\theta - \beta_j + \gamma_j\}}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}} \\
R_{j2}(\theta) &= Pr\{\xi = 2|\theta\} = \frac{\exp\{2(\theta - \beta_j)\}}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}}
\end{aligned} \tag{8}$$

for an item located at β_j, and an item-unit of γ_j. $R_{j0}(\theta)$ and $R_{j2}(\theta)$ are monotone decreasing, and monotone increasing respectively. This setup induces the hyperbolic cosine model (HCM) for unfolding responses defined in Equation 2.

4. Augmentation Methods for Unfolding Models

The latent response formulation of the unfolding model suggests a simple data augmentation method for the analysis of unfolding responses. We generalize the latent response formulation of the unfolding model in order to define a data augmentation method for all dichotomous unfolding models, whether or not the model has a "nice" latent response formulation. This new data augmentation method requires that the IRF is *dominated* by the category response function for the middle response in a monotone IRT model (i.e., $\xi = 1$).

Definition 4.1. A response function $P_j(\theta)$ is dominated by the response function $Q(\theta)$ if and only if

$$P_j(\theta) \leq Q(\theta)\ \forall\theta$$

Proposition 4.1. Let $P_j(\theta)$ denote the unfolding IRF describing the probability of agreement for item j, and $\mathbf{R}_j(\theta) = [R_{j1}(\theta), R_{j2}(\theta), R_{j3}(\theta)]^t$ denote the vector of *category response functions for a monotone item response model for trichotomous responses. If $P_j(\theta)$ is dominated by the middle category response function $R_{j1}(\theta)$, and the conditional distribution of X given the latent response ξ is:*

$$Pr\{X_j = x \mid \theta, \xi = 0\} = \begin{cases} 1 & x = 0 \\ 0 & otherwise \end{cases}, \tag{9}$$

$$Pr\{X_j = x \mid \theta, \xi = 1\} = \begin{cases} 1 - \dfrac{P_j(\theta)}{R_{j1}(\theta)} & x = 0 \\ \dfrac{P_j(\theta)}{R_{j1}(\theta)} & otherwise \end{cases}, \tag{10}$$

$$Pr\{X_j = x \mid \theta, \xi = 2\} = \begin{cases} 1 & x = 0 \\ 0 & otherwise \end{cases}, \tag{11}$$

then the marginal (unconditional) distribution of X_j is defined by the unfolding response model $P_j(\theta)$.

Note: Setting $R_{j1}(\theta) = P_j(\theta)$ ($P_j(\theta)$ does dominate $P_j(\theta)$ according to Definition 1) leads to the original latent response model formulation of the unfolding model suggested by Verhelst and Verstralen (1993) and Andrich and Luo (1993).

Proposition 4.1 (whose proof is sketched in the Appendix) shows that a LRM with ordered, trichotomous latent responses can be constructed for any dichotomous-response unfolding model, as long as the middle category of the LRM dominates the IRF for endorsement for each dichotomous response variable. The proposition shows how to generalize the collapsing condensation rule in Equation 4 by re-weighting the probability of response in the LRM, so that the desired dichotomous unfolding IRF results, even when the middle category of the LRM does not match exactly the unfolding IRF. In this sense Proposition 4.1 is a sort of importance sampling result (Ripley, 1987) for constructing LRM representations of unfolding models. It is also important to note that Proposition 4.1 implies that LRM representations for unfolding models are not necessarily unique. The choice of representation cannot be made on psychometric grounds but rather as a matter of mathematical or computational convenience.

Theorem 4.1 derives a converse of Proposition 4.1, and provides the complete conditional probabilities needed to use an LRM representation in a data augmentation (c.f. Tanner, 1996, pp 90–136) strategy (for example, within an EM or MCMC algorithm) for estimating dichotomous unfolding models. A sketch of the proof of Theorem 4.1 can be found in the Appendix.

Theorem 4.1. *Suppose that $P_j(\theta)$ is the item response function describing the probability of agreement for item j, and that $\mathbf{R}_j(\theta) = [R_{j0}(\theta), R_{j1}(\theta), R_{j2}(\theta)]$ is a monotone item response model for trichotomous responses, such that $P_j(\theta)$ is dominated* *by the category response function* $R_{j1}(\theta)$*, for* $j = 1, \ldots, J$*. If the conditional distribution of the latent response* ξ *is defined by*

$$Pr\{\xi_j = k | \theta, X_j = 0\} = \begin{cases} \dfrac{R_{j0}(\theta)}{1 - P_j(\theta)} & k = 0 \\ \dfrac{R_{j1}(\theta) - P_j(\theta)}{1 - P_j(\theta)} & k = 1, \\ \dfrac{R_{j2}(\theta)}{1 - P_j(\theta)} & k = 2 \end{cases} \tag{12}$$

$$Pr\{\xi_j = k | \theta, X_j = 1\} = \begin{cases} 1 & k = 0 \\ 0 & otherwise \end{cases}, \tag{13}$$

then the marginal distribution of the latent response ξ *is the monotone item response model* $\mathbf{R}_j(\theta)$.

Note that under the deterministic LRM formulation of Equations 6 and 7, $Pr\{\xi = 1 | \theta, X = 0\}$ must be identically 0, as can be seen from Equation 10 since $P_j(\theta) = R_{j1}(\theta)$ for all θ. On the other hand, when $P_j(\theta) < R_{j1}(\theta)$ for some θ s, $Pr\{\xi = 1 | \theta, X = 0\}$ can be quite arbitrary as a function of θ, as illustrated by Example 4.1 and Figure 4.

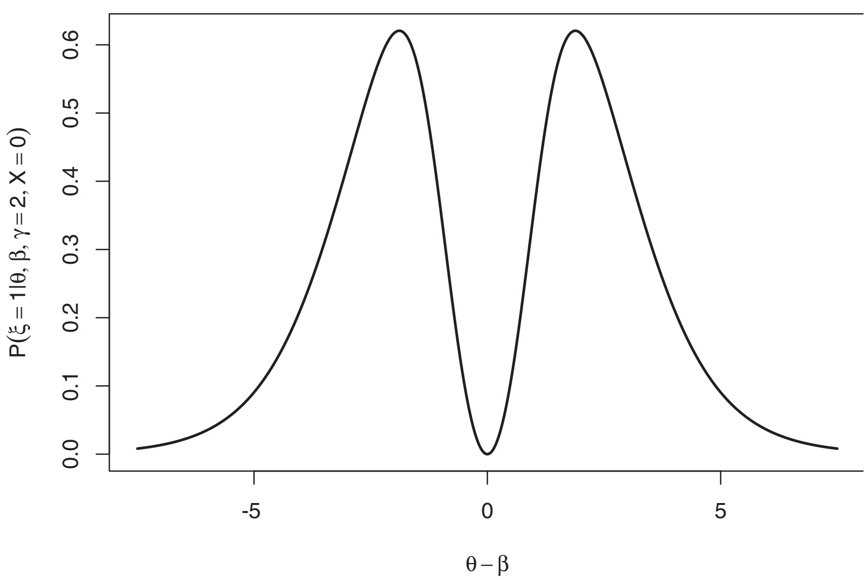

FIGURE 4. *Probability that a subject's latent response is* $\xi = 1$ *given that their observed response is* $X = 0$ *for the SLM/PCM data augmentation example in Example 4.1.*

Contrary to the LRM formulation of the unfolding model, the data augmentation method derived in Theorem 4.1 does not assume the latent agreement response ($\xi = 1$) implies the observed response is also agree ($X = 1$); rather it states the observed response is agree ($X = 1$) with probability equal to $\frac{P_j(\theta)}{R_{j1}(\theta)}$ when the latent response is agree ($\xi = 1$).

The beauty of Theorem 4.1 lies in its application to unfolding response models where the IRF is not the middle category response function of a well-known polytomous model. In such situations the result allows us to connect the model with any polytomous model with a dominating middle category response function.

One such unfolding response model is the SLM defined by assuming the logit of the endorsement probability is quadratic in θ. The following example uses Proposition 4.1 and Theorem 4.1 to derive a representation and a data augmentation procedure that connects the SLM to the well-known partial credit model.

Example 4.1 (Augmenting the squared logistic model with partial credit model latent responses). *Recall the parameterization of the squared logistic model (SLM) in* Equation 1 *that parameterizes the probability of endorsement for an individual with latent attitude* θ *by the equation*

$$P_j(\theta) = \left\{1 + e^{\gamma_j} \exp\left[(\theta - \beta_j)^2\right]\right\}^{-1}.$$

We develop a simple estimation algorithm for the SLM using a pre-existing program that estimates the parameters of the partial credit model [PCM] and noting that a PCM with a middle category response function of

$$\begin{aligned} R_{j1}(\theta) &= \frac{\exp\{\theta - \beta_j + \gamma_j\}}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}} \\ &= \frac{1}{\exp\{-(\theta - \beta_j) - \gamma_j\} + 1 + \exp\{(\theta - \beta_j) - \gamma_j\}} \\ &= \frac{1}{1 + 2e^{-\gamma_j}\cosh\{\theta - \beta_j\}} \end{aligned}$$

*dominates the SLM IRF P(*θ*).*

Applying Theorem 4.1 we estimate the model parameters of the SLM by iteratively simulating latent responses conditional on the current estimates of the model parameters according to the following probabilities

$$P(\xi_j = 1 \mid \theta, X = 0) = \frac{\exp\left\{(\theta - \beta_j)^2\right\} - \cosh(\theta - \beta_j)}{\left[1 + e^{\gamma_j}\cosh(\theta - \beta_j)\right]\exp\left\{(\theta - \beta_j)^2\right\}} \qquad (14)$$

$$P(\xi_j = 1 \mid \theta, X = 1) = 1$$

and then re-estimating the PCM parameters with these latent responses. Figure 4 *contains the plot of the function in Equation 14.*

5. Estimation

5.1 Maximum Likelihood

As Holland (1990) states, "one of the curiosities of estimating IRT models is the number of different procedures that all claim to result in 'maximum likelihood estimates'." One of these procedures known as the joint maximum likelihood [JML] procedure considers both respondent parameters [θ] and item parameters [(β, γ)] as fixed effects and maximizes the joint likelihood with respect to all parameters simultaneously. The joint likelihood of the model parameters given the response matrix **X** is

$$L(\beta, \theta; \mathbf{X}) = \prod_{i=1}^{N} \prod_{j=1}^{J} [P_j(\theta_i)]^{x_{ij}} [1 - P_j(\theta_i)]^{1-x_{ij}}, \tag{15}$$

where $P_j(\cdot)$ is a function of β_j, $\beta = (\beta_1, \ldots, \beta_J)$ and $\theta = (\theta_1, \ldots, \theta_N)$. The estimation procedure sets out to maximize this joint likelihood in Equation 15 with respect to the subject parameters [θ] and item parameters [β] simultaneously. DeSarbo and Hoffman (1986), Andrich (1988), Andrich and Luo (1993), Andrich (1996), and Roberts et al. (2000) all use this method for estimating the parameters of their respective models.

Andrich and Luo (1993) note that JML estimates for β when there is a fixed number of item parameters, and an increasing number of survey respondents are not consistent. This is also true for monotone IRT models such as the Rasch model (Andersen, 1970). For the HCM, Andrich and Luo (1993) notice problems for fifteen to twenty items answered by 500 subjects; and report that the noticeable inconsistency of estimates starts to disappear as the number of items reaches 100.

A second maximum likelihood procedure popular in random effects models is to condition on a simple (possibly multidimensional) sufficient statistic for the random effects or nuisance parameters, in this case θ. The resulting *conditional likelihood* is independent of the nuisance parameter θ. The parameters that maximize this conditional likelihood are called the conditional maximum likelihood (CML) estimates. Although CML estimation is not typically possible for unfolding response models, Johnson (2001) suggests a method for the CML estimation of the squared logistic model. The technique is similar to that introduced by Verhelst and Glas (1995) for the two-parameter logistic model for monotone item responses.

The final maximum likelihood procedure removes the latent variable θ from the likelihood in a different fashion. Maximum marginal likelihood (MML) estimation assumes a population distribution $G(\theta)$ for the nuisance parameter θ and integrates the parameter out of the likelihood. The procedure then maximizes the resulting marginal likelihood,

$$L_M(\beta, \gamma; \mathbf{x}) = \prod_{i=1}^{N} \int_{\Theta} \prod_{j=1}^{J} [P_j(\theta)]^{x_{ij}} [1 - P_j(\theta)]^{1-x_{ij}} \, dG(\theta), \tag{16}$$

with respect to the item parameters.

The marginal likelihood L_M depends on the choice of the population distribution $G(\cdot)$. An extreme example would be to assume the attitude distribution $G(\theta)$ is a point mass [e.g., $G(\theta) = \delta(\theta)$]. In this extreme case the marginal likelihood corresponds to the independence model likelihood. Maximizing the marginal likelihood in Equation 16 with respect to the location vector β and item-unit vector γ yields the MML estimates.

If the integration in Equation 16 is simple then MML estimation is straightforward. When no simple analytical solution exists, a numerical integration method is used. Marginal maximum likelihood estimation in item response models is often carried out using the EM-algorithm (Dempster et al., 1977) where the latent attitude variables, θ_i, are considered missing data.

Hoijtink (1990), Verhelst and Verstralen (1993), and Roberts et al. (2000) have employed MML estimation for a number of different unfolding response models. Section 6 provides MML estimates of the hyperbolic cosine model parameters and their approximate standard errors for the simulated and nuclear power plant data sets assuming G is the normal distribution with mean zero and variance σ^2. We carry out the MML estimation by approximating the normal distribution with a discrete distribution with mass $\frac{1}{50}$ on each of the normal quantiles $\sigma\Phi^{-1}\left(\frac{k}{51}\right)$, $k = 1, \ldots, 50$. The maximization was carried out using the function optim in the statistical package R (available from www.r-project.org), and standard errors were approximated using the inverse of the Hessian matrix.

Limitations of maximum likelihood estimation

Maximum likelihood estimation for unfolding response models tends to have unattractive features. Suppose the latent attitudes of five respondents are fixed and known at $\theta = (-3, -1, 0, 2, 3)$, and that Subject 1 and Subject 5 agree with a given item, and the other three subjects disagree with the item. The likelihood function for the item location, displayed in Figure 5, is bi-modal.

Any maximum likelihood procedure using numerical methods to optimize the likelihood might converge to the wrong mode, or even the anti-mode.

It should be noted that the multimodality of the likelihood has meaning in the unfolding response model. If the two most extreme subjects agree with an item and all other $N - 2$ subjects disagree with the item, then intuitively the item should be located near one of the two subjects that agreed with the item, and not in the middle of the latent dimension.

In situations where the likelihood for a parameter is multimodal, standard methods for the approximation of standard errors may not be valid, and standard summaries based on these standard errors do not express the multi-modal nature of the likelihood. The following section introduces a hierarchical Bayesian model for the analysis of unfolding responses. Because the Bayesian analysis summarizes the whole posterior distribution, the multimodal nature of the distribution need not be lost.

5.2 Hierarchical Bayes Formulation of the Hyperbolic Cosine Model

Bayesian estimation has a number of advantages over the maximum likelihood estimation of unfolding models. A Bayesian analysis of the unfolding response

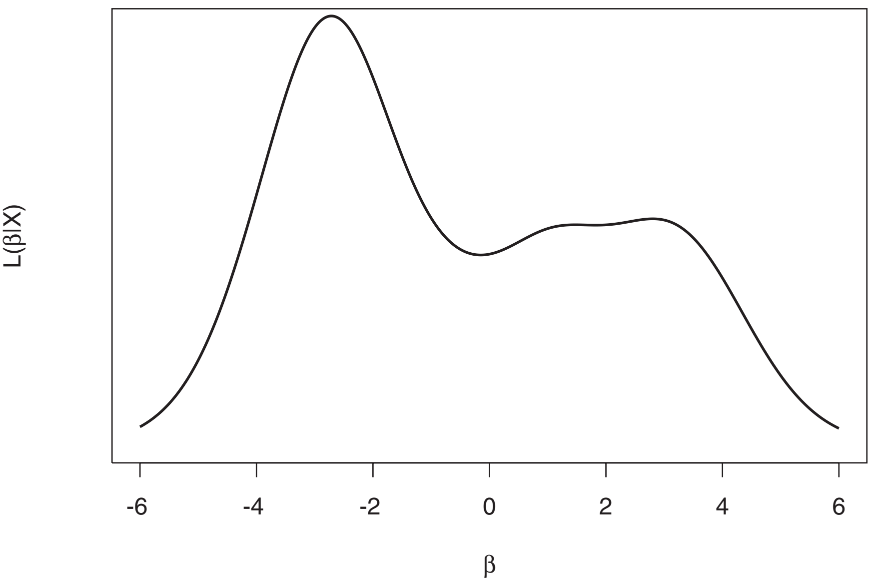

FIGURE 5. *The conditional likelihood of the location parameter* β *in the HCM given* $\gamma = 2$, $\beta = (-3, -1, 0, 2, 3)^t$, *and* $\mathbf{X} = (1, 0, 0, 0, 1)^t$.

model, like joint maximum likelihood, allows all parameters to be estimated. By specifying prior probability distributions for each of the unknown model parameters, the hierarchical Bayesian model defines the joint posterior distribution of the model parameters (conditional on the observed data). From the posterior distribution probabilistic summaries of the model parameters are calculated (e.g., credible intervals and quantiles) without relying on asymptotic results.

Because analytic solutions to the integrals needed to derive the posterior distribution are intractable, we use Markov chain Monte Carlo (MCMC; Gelman, Carlin, Rubin, & Stern, 1995) techniques to simulate parameter values from a Markov chain whose stationary distribution is the posterior distribution of the model parameters. With a large enough sample from this approximate posterior distribution, it is possible to calculate summary statistics (e.g., means, standard deviations) to any desired degree of accuracy.

For ease of exposition we focus on a Bayesian approach to the HCM in the remainder of Section 5 and most of Section 6. However the same ideas apply to other unfolding models, whether estimated directly or with an LRM-based data augmentation scheme. Indeed at the end of Section 6 we show how the same methodology can be applied to the SLM of Example 4.1. For specific unfolding response models there may be more efficient MCMC algorithms for the approximation of the posterior distribution of the model parameters. Maris and Maris (2002) for example develop a Gibbs sampler for the estimation of their model.

The Bayesian formulation of the hyperbolic cosine model is similar to that used in marginal maximum likelihood estimation, in that it assumes respondents' latent attitudes (θ) are drawn from some population, denoted $F_\Theta(\theta \mid \lambda)$, and that the probability of endorsement is defined by the hyperbolic cosine model in Equation 2:

$$\left.\begin{array}{lcc} \theta_i & \overset{iid}{\sim} & F_\Theta(\theta_i \mid \lambda) \\ P_j(\theta_i) & = & \dfrac{e^{\gamma_j}}{e^{\gamma_j} + 2\cosh(\theta_i - \beta_j)} \\ X_{ij} & \overset{indep}{\sim} & \text{Bernouli}(P_j(\theta_i)) \end{array}\right\}, \tag{17}$$

for $i = 1, \ldots, N,$ and $j = 1, \ldots, J,$ where λ is a vector of parameters describing the attitude distribution (e.g., the mean and variance).

In addition to assuming the Bernoulli model and the HCM probabilities, a Bayesian analysis must define prior distributions for all model parameters to reflect the uncertainty, or prior information, we have about these parameters. Let Π_λ, G_B, and H_Γ (with densities π, g, and h) denote the prior distributions for the model parameters λ, β and γ respectively. One type of prior information that is incorporated into the model through the prior distributions is constraining information that is necessary to ensure identifiability of the model. Two types of constraints are necessary in unfolding models:

- *Origin constraints:* This type of constraint is necessary because a constant c added to each attitude θ_i and to each item location β_j does not affect the likelihood. One such constraint requires the mean of the latent attitude distribution to be fixed. Typically the mean is fixed at zero. A second, equivalent constraint forces the sum of item locations to be fixed at some constant (typically zero).
- *Directional constraint:* A directional constraint is necessary because each attitude, and each item location can be multiplied by -1 without affecting the likelihood. Constraining the sign of a single item's location is one method for ensuring identifiability (e.g., $\beta_1 < 0$).

5.3 General Markov Chain Monte Carlo Method

The goal of MCMC methods is to simulate a random walk process in the parameter space $\phi = (\beta, \gamma, \theta, \ldots)$ that converges to the joint posterior distribution $f(\phi \mid X)$. The two Markov chains used here are the Gibbs sampler and the Metropolis-Hastings algorithm, both defined below.

Gibbs Sampler (Geman & Geman, 1984; Gelman et al., 1995): The parameter vector ϕ is split into d components, or subvectors, $\phi = (\phi_1, \ldots, \phi_d)$. For each iteration of the simulation, the d subvectors of ϕ are simulated from the posterior distribution of the subvector conditioned on the values of the other $d - 1$ subvectors.

For $t = 1, 2, \ldots, I$

For $s = 1, \ldots, d$

Sample $\phi_{-s}^{(t)} \sim f(\phi_s \mid \phi_{-s}^{(t-1)}, \mathbf{X})$.

where $\phi_{-s}^{(t-1)}$ represents all sub-vectors of the parameter vector, except for subvector *s,* at their current value.

$$\phi_{-s}^{(t)} = \left(\phi_1^{(t)}, \ldots, \phi_{s-1}^{(t)}, \phi_{s+1}^{(t)}, \ldots, \phi_d^{(t-1)}\right)$$

Because the distribution of $\phi^{(t+1)}$ given the current state of the parameter vector $\phi^{(t)}$ does not depend further on the history of the sequence $\{\phi^{(1)}, \ldots, \phi^{(t-1)}\}$, the sequence $\{\phi^{(1)}, \phi^{(2)}, \ldots\}$ is a Markov chain. As the length of the chain increases, the distribution of parameter vectors $\phi^{(t)}$ converges to the posterior distribution of ϕ given the observed data **X** (Tanner & Wong, 1987).

The Gibbs sampler requires a method to draw the parameter vector ϕ_s from the conditional posterior $f(\phi_s \mid \phi_{-s}^{(t-1)}, \mathbf{X})$. For instances where simulation from the conditional posterior is not straightforward, an alternative method is used. One alternative, used here, is the Metropolis-Hastings algorithm.

Metropolis-Hasting Algorithm (Metropolis & Ulam, 1949; Metropolis et al., 1953; Hastings, 1970; Chib & Greenberg, 1995). Given a target distribution $g(\psi \mid X)$ (e.g., the conditional posterior in the Gibbs sampler), which can be computed up to a constant, a sequence of random values can be simulated whose distribution converges to the target distribution using the following algorithm.

For $t = 1, \ldots, I$

1. Simulate a candidate ψ^* from a candidate distribution that depends on the previous value of the parameter.

$$\psi^* \sim J_t\left(\psi^* \mid \psi^{(t-1)}\right).$$

2. Calculate the acceptance probability

$$\alpha = \frac{g(\psi^* \mid \mathbf{X}) / J_t\left(\psi^* \mid \psi^{(t-1)}\right)}{g\left(\psi^{(t-1)} \mid \mathbf{X}\right) / J_t\left(\psi^{(t-1)} \mid \psi^*\right)}.$$

3. Set

$$\psi^{(t)} = \begin{cases} \psi^* & \textit{with probability } \min(1, \alpha) \\ \psi^{(t-1)} & \textit{otherwise.} \end{cases}$$

The following sections introduce the MCMC procedure for two hierarchical formulations of the hyperbolic cosine model of Andrich and Luo (1993). The first assumes the dichotomous responses follow the HCM, ignoring everything about the underlying latent responses. The second method takes advantage of data augmentation and existing methods for estimation of the partial credit model (PCM; Masters, 1982) to develop an MCMC estimation procedure.

5.4 Direct MCMC Implementation for the HCM

This section introduces the Markov chain Monte Carlo algorithm for the Bayesian estimation of the hyperbolic cosine model parameters. The algorithm draws parameter values from a Markov chain whose stationary distribution is the joint posterior distribution of all model parameters, including both item and subject parameters.

Algorithm 1. *Iterate the following steps for* $n = 1, \ldots, I$.

- *For each respondent* $i = 1, \ldots, N$ *sample the latent attitude* $\theta^{(n)}$ *by:*
  1. *Draw candidate* $\theta_i^* \sim J_t(\theta_i \mid \theta^{(n-1)})$ *independently for each* $i = 1, \ldots, N$.
  2. *Calculate the acceptance probability for each candidate*

$$\alpha_i = \frac{\left\{\prod_j \frac{\left[e^{-\gamma_j^{(n-1)}} \cosh\left(\theta_i^* - \beta_j^{(n-1)}\right)\right]^{1-x_{ij}}}{1/2 + e^{-\gamma_j^{(n-1)}} \cosh\left(\theta^* - \beta_j^{(n-1)}\right)}\right\} f\left(\theta_i^* \middle| \lambda^{(n-1)}\right) \Big/ J_n\left(\theta_i^* \middle| \theta_i^{(n-1)}\right)}{\left\{\prod_j \frac{\left[e^{-\gamma_j^{(n-1)}} \cosh\left(\theta_i^{(n-1)} - \beta_j^{(n-1)}\right)\right]^{1-x_{ij}}}{1/2 + e^{-\gamma_j^{(n-1)}} \cosh\left(\theta^{(n-1)} - \beta_j^{(n-1)}\right)}\right\} f\left(\theta_i^{(n-1)} \middle| \lambda^{(n-1)}\right) \Big/ J_n\left(\theta_i^{(n-1)} \middle| \theta_i^{(*)}\right)}.$$

  3. *Set*

$$\theta_i^{(n)} = \begin{cases} \theta_i^* & \textit{with probability } \min\{1, \alpha_i\} \\ \theta_i^{(n-1)} & \textit{otherwise} \end{cases}$$

- *Update each of the item locations* $\beta_j^{(n)}$ *for* $j = 1, \ldots, J$ *by:*
  1. *Choose a candidate location* $\beta_j^* \sim J_n(\beta_J \mid \beta^{(n-1)})$.
  2. *Calculate the acceptance probabilities for the candidate locations*

$$\alpha_j = \frac{\prod_i \left\{\frac{\left[e^{-\gamma_j^{(n-1)}} \cosh\left(\theta_i^{(n)} - \beta_j^*\right)\right]^{1-x_{ij}}}{1/2 + e^{\gamma_j^{(n-1)}} \cosh\left(\theta_i^{(n)} - \beta_j^*\right)}\right\} g\left(\beta_j^*\right) \Big/ J_n\left(\beta_j^* \middle| \beta_j^{(n-1)}\right)}{\prod_i \left\{\frac{\left[e^{-\gamma_j^{(n-1)}} \cosh\left(\theta_i^{(n)} - \beta_j^{(n-1)}\right)\right]^{1-x_{ij}}}{1/2 + e^{-\gamma_j^{(n-1)}} \cosh\left(\theta_i^{(n)} - \beta_j^{(n-1)}\right)}\right\} g\left(\beta_j^{(n-1)}\right) \Big/ J_n\left(\beta_j^{(n-1)} \middle| \beta_j^*\right)}.$$

  3. *Set*

$$\beta_j^{(n)} = \begin{cases} \beta_j^* & \textit{with probability } \min\{1, \alpha_j\} \\ \beta_j^{(n-1)} & \textit{otherwise} \end{cases}$$

- *Update each of the item unit parameters* $\gamma_j^{(n)}$ *for* $j = 1, \ldots, J$ *by:*
  1. *Choose a candidate item unit* $\gamma_j^* \sim J_n(\gamma_j \mid \gamma_j^{(n-1)})$.
  2. *Calculate the acceptance probabilities for the candidate locations*

$$\alpha_j = \frac{\prod_i \left\{\frac{\left[e^{-\gamma_j^*} \cosh\left(\theta_i^{(n)} - \beta_j^{(n)}\right)\right]^{1-x_{ij}}}{1/2 + e^{\gamma_j^*} \cosh\left(\theta_i^{(n)} - \beta_j^{(n)}\right)}\right\} h\left(\gamma_j^*\right) \Big/ J_n\left(\gamma_j^* \middle| \gamma_j^{(n-1)}\right)}{\prod_i \left\{\frac{\left[e^{-\gamma_j^{(n-1)}} \cosh\left(\theta_i^{(n)} - \beta_j^{(n)}\right)\right]^{1-x_{ij}}}{1/2 + e^{-\gamma_j^{(n-1)}} \cosh\left(\theta_i^{(n)} - \beta_j^{(n)}\right)}\right\} h\left(\gamma_j^{(n-1)}\right) \Big/ J_n\left(\gamma_j^{(n-1)} \middle| \gamma_j^*\right)}.$$

3. *Set*

$$\gamma_j^{(n)} = \begin{cases} \gamma_j^* & \textit{with probability } \min\{1, \alpha_j\} \\ \gamma_j^{(n-1)} & \textit{otherwise} \end{cases}.$$

- *Update the hyper-parameters, λ, for the distribution of latent attitudes by:*

1. *Draw candidate hyper-parameters $\lambda_k^* \sim J_n(\lambda \,|\, \lambda^{(n-1)})$ for $k = 1, \ldots, K$.*
2. *Calculate acceptance probabilities for the candidate hyper-parameters*

$$\alpha_k = \frac{\left[\prod_i f\left(\theta_i^{(n)} | \lambda_k^*; \lambda_{-k}^{(n-1)}\right)\right] \pi\left(\lambda_k^*\right) \Big/ J_j\left(\lambda_k^* | \lambda_k^{(n-1)}\right)}{\left[\prod_i f\left(\theta_i^{(n)} | \lambda_k^{(n-1)}; \lambda_{-k}^{(n-1)}\right)\right] \pi\left(\lambda_k^{(n-1)}\right) \Big/ J_j\left(\lambda_k^{(n-1)} | \lambda_k^*\right)}.$$

where $\lambda_{-k}^{(n-1)} = \left(\lambda_1^{(n)}, \ldots, \lambda_{k-1}^{(n)}, \lambda_{k+1}^{(n-1)}, \ldots, \lambda_k^{(n-1)}\right)$.

3. *Set*

$$\lambda_k^{(n)} = \begin{cases} \lambda_k^* & \textit{with probability } \min\{1, \alpha_k\} \\ \lambda_k^{(n-1)} & \textit{otherwise} \end{cases}.$$

5.5 Data Augmentation MCMC Method for Estimation

A second MCMC method for estimation of the HCM takes advantage of a pre-existing MCMC procedure for the estimation of the partial credit model. Recall from Section 3 the latent response model formulation of the hyperbolic cosine model. The following hierarchical model defines the latent response formulation of the hyperbolic cosine model:

$$\left.\begin{array}{ccc} \theta_i & \overset{iid}{\sim} & F_\Theta(\theta_i | \lambda)(24) \\ R_j(\theta_i) & = & \mathrm{PCM}(\theta_i, \beta_j)(25) \\ \xi_{ij} & \overset{indep}{\sim} & \mathrm{Multinomial}_3(1, R_j(\theta_i))(26) \\ X_{ij} & = & 1 - |1 - \xi_{ij}|(27) \end{array}\right\}, \tag{18}$$

where $\mathrm{PCM}(\theta_i, \beta_j)$ is the partial credit model for trichotomous choices defined in Equation 9.

The following algorithm describes a method for the estimation of the HCM parameters by taking advantage of a pre-existing MCMC algorithm for the PCM (see for example: Johnson et al., 1999; Patz & Junker, 1999; Patz et al., 2002). The MCMC algorithm augments the data by simulating the latent responses and then uses the existing MCMC procedure for the estimation of the model parameters given the latent responses. The first step of the algorithm applies Theorem 4.1 in order to simulate the latent responses ξ.

Algorithm 5.2. *Iterate the following steps for* $n = 1, \ldots, I$.

1. *Given the current values of the model parameters* $\beta^{(n-1)}$, $\gamma^{(n-1)}$, and $\theta^{(n-1)}$, *simulate the latent response vectors* ξ *from the conditional posterior distribution*

$$Pr\{\xi_{ij} = k \mid \beta^{(n-1)}, \gamma^{(n-1)}, X_{ij}\}$$

defined by substituting the probabilities from the partial credit model in Equation 9 into Equations 12 and 13. The conditional posterior distribution for ξ_{ij} *is defined by*

$$\xi_{ij} = \begin{cases} \begin{cases} 0 & \text{with probability } \dfrac{1}{1 + \exp\{2(\theta_i^{(n-1)} - \beta_j^{(n-1)})\}} \\ 2 & \text{with probability } \dfrac{\exp\{2(\theta_i^{(n-1)} - \beta_j^{(n-1)})\}}{1 + \exp\{2(\theta_i^{(n-1)} - \beta_j^{(n-1)})\}} \end{cases} & \text{if } X_{ij} = 0 \\ 1 & \text{if } X_{ij} = 1 \end{cases} . \qquad (19)$$

2. *Update the model parameters according to one step of the pre-existing MCMC procedure for the partial credit model. Let* $\beta^{(n)}$, $\gamma^{(n)}$, $\lambda^{(n)}$, and $\theta^{(n)}$ *denote these updated parameters estimates.*

5.6 Reparameterization

Preliminary analyses found that item unit parameters $\gamma_1, \ldots, \gamma_J$ estimated separately exhibit an identifiability problem for the items located away from the center of the attitude distribution. This is attributable to the fact that $2\cosh(t) \approx e(t)$ for $t >> 0$, and $2\cosh(t) \approx e(-t)$ for $t << 0$ which implies the HCM item response function satisfies:

$$\frac{1}{1 + 2e^{-\gamma_j}\cosh(\theta - \beta_j)} \approx \frac{1}{1 + e^{-\gamma_j}e^{\beta_j - \theta}} = \frac{1}{1 + e^{-\gamma_j + \varepsilon}\exp(\beta_j - \theta - \varepsilon)}$$

for $\beta_j >> 0$. In fact, if $\gamma_j = \beta_j + b_j$, then the HCM approaches the Rasch model with difficulty b_j as β_j increases. This near-indeterminacy caused the estimates of the unit parameter γ and the item location parameter β to be strongly correlated for extreme items in the MCMC procedure. In one instance this correlation was as strong as –0.99.

The problem of a near-indeterminacy between item units γ and item locations β has also been noted by Verhelst and Verstralen (1993) and Luo (2000). Luo (2000) suggests first estimating the model with item-units constant across all items. That is $\gamma_j = \gamma$ for all $j = 1, \ldots, J$. Let $\hat{\gamma}$ denote this estimated grand-item-unit parameter. With $\hat{\gamma}$ estimated, Luo (2000) re-estimates the unfolding model by relaxing the constraint on the item units so the average item-unit is equal to the estimated grand-item unit. That is, the item units are constrained by

$$\frac{1}{J}\sum_{j=1}^{J}\gamma_j = \hat{\gamma}.$$

This procedure works well for the JML method Luo (2000) uses. A hierarchical Bayesian version of this idea was implemented but performed no better than the original parameterization.

High correlation between item parameters γ_j and β_j suggest the MCMC procedures introduced here may perform more efficiently under another parameterization. The parameterization adopted here estimates the difference between the item unit and item location, and the sum of the two parameters. That is, the reparameterized model is

$$P_j(\theta) = \frac{1}{1 + \exp\{\theta - \eta_j\} + \exp\{\delta_j - \theta\}},$$

where

$$\eta_j = \beta_j + \gamma_j \text{ and } \delta_j = \beta_j + \gamma_j.$$

Using this reparameterization on the same set of data that produced an observed correlation of –0.99 between β and γ, produced a correlation of 0.04 between δ and η, thus allowing the MCMC algorithm to move through the parameter space more efficiently.

5.7 Prior Distributions

For the analyses in Section 6, prior distributions are defined for all HCM parameters. The two sets of item parameters, η and δ, are independent of each other, and all other model parameters, and are both normally distributed with mean zero and diagonal variance-covariance matrix with diagonal element τ. The hyperparameter τ can be chosen to reflect any prior information a researcher may have about the parameters. To reflect the fact that no prior information is available the analyses in the upcoming section use $\tau = 100$.

Conditional on the hyper-parameter vector $\lambda = (\mu, \sigma^2)$, the attitude distribution $F_\Theta(\theta|\lambda)$ is the normal distribution with mean μ and variance σ^2. The prior distribution of the attitude distribution variance is $\sigma^2 \sim \Gamma^{-1}$ (0.5, 0.5). To ensure identifiability of the model, the mean of the attitude distribution is fixed at zero ($\mu \equiv 0$).

5.8 Jumping Distributions

The MCMC algorithm proposed requires that jumping kernels be specified for each of the model parameters. We have found that normal distributions centered around the previous draw for the parameter work well for item parameters η_j and δ_j and the latent attitudes θ_i. Our algorithm draws

$$\left.\begin{array}{lll}\theta_i^{\mathring{a}} & \sim & N\left(\theta_i^{(n-1)}, \tau_\theta^2\right)\\ \eta_j^{\mathring{a}} & \sim & N\left(\eta_j^{(n-1)}, \tau_\eta^2\right)\\ \delta_j^{\mathring{a}} & \sim & N\left(\delta_j^{(n-1)}, \tau_\delta^2\right)\end{array}\right\},$$

independently for all $j = 1, \ldots, J$ and $i = 1, \ldots, N$. The variances of the jumping distributions are adjusted for each data set so that approximately 30% of the proposed draws are accepted by the Metropolis-Hastings algorithm (Gelman et al., 1995, pp 334–335). The conditional posterior distribution of σ^2 [assuming the inverse-gamma prior distribution $\sigma^2 \sim \Gamma^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)$] is the inverse-gamma distribution $\sigma^2 \sim \Gamma^{-1}\left(\frac{N+1}{2}, \frac{1}{2}\left\{1 + \sum \theta_i^{(n)2}\right\}\right)$; therefore the Metropolis-Hastings algorithm is not necessary for drawing σ^2.

6. Bayesian Analyses of the Three Example Data Sets

In the following sections we use the data augmentation Markov chain Monte Carlo algorithm to perform Bayesian estimation of the hyperbolic cosine model parameters for the three data sets introduced in Section 2. In each example we ran the MCMC algorithm for 100,000 iterations, removed the first 10,000 iterations as "burn-in," and thinned the remaining 90,000 iterations by selecting every tenth iteration. The results reported are thus based on 9,000 draws from the Markov chain.

Results from the direct Markov chain Monte Carlo implementation are indistinguishable from the data augmentation MCMC method up to Monte Carlo error. However, the data augmentation does take more computational time for each iteration and does require a larger number of iterations for burn-in than does the direct method. So, the data augmentation method's real value is in its ease of implementation.

For comparison we also compute marginal maximum likelihood estimates and their approximated standard errors, as described in Section 5.1, for the first two examples. The C++ source code for the MCMC estimation procedures and the R code for MML estimation are available by email from Matthew_Johnson@baruch.cuny.edu.

6.1 The Simulated Data

The data augmentation MCMC procedure is applied to the data simulated for five items from Section 2.3. Table 1 presents the estimated 95% equal-tailed credible intervals, and posterior medians found using the data augmentation MCMC procedure, and the marginal maximum likelihood estimates and standard errors for the five item locations [β], and item units [γ].

The 95% equal-tailed credible intervals for all item parameters contain the simulating value, and the posterior median is very close to the MLE for all item parameters except for the location and unit parameters for Item 5. The posterior distributions for Item 5's parameters are both positively skewed, and an approximation of the posterior mode more closely resembles the MLE.

The widths of the credible intervals and the standard errors of the MLEs tend to decrease for both the location and the item-unit parameters as the magnitude of the

TABLE 1
MCMC Estimated Posterior Medians, 95% Equal-Tailed Credible Intervals, Marginal Maximum Likelihood Estimates (MLE) and the Standard Error of the MLE for the Ten Sets of Item Parameters

	Bayesian		MML		
Parameter	*Mdn*	95% Credible interval	MLE	*SE*	True
Item 1					
Location β_1	−2.39	(−2.87, −1.97)	−2.39	0.22	−2.50
Unit γ_1	1.77	(1.47, 2.12)	1.76	0.15	1.62
Item 2					
Location β_2	−0.96	(−1.26, −0.65)	−0.98	0.16	−1.00
Unit γ_2	1.76	(1.57, 1.98)	1.85	0.09	1.72
Item 3					
Location β_3	−0.67	(−1.02, −0.37)	−0.67	0.17	−0.90
Unit γ_3	4.00	(3.71, 4.32)	4.11	0.12	3.85
Item 4					
Location β_4	1.25	(0.89, 1.64)	1.15	0.17	1.50
Unit γ_4	3.74	(3.42, 4.10)	3.79	0.13	4.05
Item 5					
Location β_5	5.69	(3.18, 12.16)	3.37	0.59	4.00
Unit γ_5	5.43	(3.01, 11.97)	3.18	0.53	3.97

Note. Simulating parameter values are provided in the right-most column.

item location decreases, indicating that the data provides more information about items near the middle of the attitude distribution than it does for those in the tails of the distribution.

The posterior median for the attitude standard deviation σ is 2.11 and the 95% equal-tailed credible interval for the parameter is (1.89, 2.34) which contains the true simulating $\sigma = 2.00$. The MLE for the attitude standard deviation is $\hat{\sigma} = 2.33$ and has a standard error of 0.12.

Table 2 contains the simulating value, posterior median, and 95% credible interval for each of ten simulees from the simulated data set. The first five simulees have somewhat unusual response patterns. The first simulee endorsed none of the items the second endorsed all items, the third endorsed statements that that are far from each other, the fourth endorsed every second item, and the fifth endorsed all but the central item. The final five simulees are examples of "good" response patterns (only adjacent items were selected). For all simulees, the 95% equal-tailed credible intervals contain the simulating value θ.

In this simulation study the width of the credible intervals tend to shrink as the number of endorsed items increases, and with "nice" response patterns (agree with adjacent items only). The widths of the intervals are approximately 12.3 when no items are endorsed, 5.6 when only the last item is endorsed, 6.2 when only extreme items are endorsed, 4.7 when the last two items are endorsed, and 4.4 when the

TABLE 2
MCMC Estimated Posterior Median and 95% Equal-Tailed Credible Intervals for Ten of the 1,000 Simulees' Latent Attitudes θ for the Simulated Data Set

Parameter	*Mdn*	95% Credible Interval	True	Response Pattern
Subject 1	–4.60	(–7.26, 5.00)	–4.20	00000
Subject 2	–0.25	(–2.11, 1.67)	–0.39	11111
Subject 3	–1.39	(–3.59, 2.53)	0.30	11001
Subject 4	–1.10	(–3.51, 2.70)	–0.41	10101
Subject 5	0.24	(–2.33, 2.56)	–0.21	11011
Subject 6	4.28	(1.31, 6.91)	3.19	00001
Subject 7	2.95	(0.64, 5.29)	2.47	00011
Subject 8	1.71	(–0.61, 3.94)	1.64	00111
Subject 9	0.68	(–1.31, 2.72)	0.48	01111
Subject 10	–0.34	(–2.64, 1.65)	–0.28	01110

Note. Actual parameter values, and the response patterns for the subjects are given in the two right-most columns.

last three, or middle three items are endorsed, and 3.9 when four or five items are endorsed.

6.2 Measuring Attitudes on Nuclear Power Plants

The analysis of the nuclear power plant data follows the same procedure as the analysis of the simulated data. Table 3 contains the MCMC (data augmentation method) approximated posterior medians, 95% equal-tailed credible intervals, MLEs and standard errors for the item locations and item-units for the five nuclear power items.

The HCM analysis of the nuclear energy data suggests the items are ordered from one to five along the latent scale, suggesting that the latent dimension being measured may run from pro-nuclear energy to anti-nuclear energy attitudes. If a subject has a highly negative latent attitude we believe that person is for nuclear energy; conversely a high positive value indicates the subject is against nuclear power. Like the simulated data set, the widths of the credible intervals for the item parameters, β and γ, decrease as the magnitude of item locations decrease, with a slight exception for the second and third items.

Items three through five, all anti-nuclear energy items have high item unit estimates, ranging from 3.85 to 4.05. This suggests that anti-nuclear energy respondents are much more likely to endorse these items than a respondent who is pro-nuclear energy is to endorse items one or two, the pro-nuclear energy items. The five response functions, calculated using the median of the posterior distributions for the items parameters, appear in Figure 3.

Table 4 contains the approximated posterior medians and 95% credible intervals for nine respondents selected for either their unusual or "nice" response patterns.

The posterior distributions of latent attitudes behave similarly to those from the simulated data set, with relatively wide credible intervals for respondents endorsing non-adjacent items, and narrower credible intervals for those individuals endorsing

TABLE 3
The MCMC Approximated Posterior Medians, 95% Credible Intervals and the Marginal Maximum Likelihood Estimates and Standard Errors for the Item Locations and Units for the Five Nuclear Energy Stimuli

	Bayesian		MML	
Parameter	*Mdn*	95% Credible interval	MLE	*SE*
Item 1	–2.54	(–3.33, –1.93)	–2.47	0.28
Location β_1	1.62	(1.22, 2.21)	1.57	0.19
Unit γ_1				
Item 2	–1.05	(–1.44, –0.64)	–1.05	0.20
Location β_2	1.71	(1.44, 1.99)	1.79	0.11
Unit γ_2				
Item 3	–0.87	(–1.29, –0.46)	–0.83	0.19
Location β_3	3.85	(3.46, 4.25)	3.96	0.15
Unit γ_3				
Item 4	1.42	(0.99, 1.91)	1.39	0.21
Location β_4	4.05	(3.60, 4.53)	4.16	0.18
Unit γ_4				
Item 5	3.97	(2.94, 5.53)	5.85	15.15
Location β_5	3.97	(3.03, 5.52)	5.85	15.14
Unit γ_5				
Attitude Distribution Variance σ^2	4.16	(3.08, 5.44)	5.28	0.75

TABLE 4
MCMC Estimated Posterior Median and 95% Equal-Tailed Credible Intervals for the Latent Attitudes of Nine of the 600 Survey Respondents

Parameter	*Mdn*	95% Credible Interval	Response Pattern
Subject 1	–4.42	(–7.08, 6.15)	00000
Subject 2	–0.44	(–2.20, 1.51)	11111
Subject 3	–1.31	(–3.65, 2.13)	10101
Subject 4	0.15	(–2.40, 2.37)	11011
Subject 5	4.05	(–0.71, 6.76)	00001
Subject 6	2.82	(0.53, 5.01)	00011
Subject 7	1.51	(–0.94, 3.81)	00111
Subject 8	0.55	(–1.41, 2.56)	01111
Subject 9	–0.47	(–2.82, 1.49)	01110

Note. The response patterns for the subjects are given in the last column.

more items. Figure 6 displays the posterior distributions for nine subjects' latent attitudes.

The posterior distributions for individuals endorsing none of the items (response pattern 00000) in this example is unimodal but severely positively skewed. The skewness is due to the fact that the unit parameter estimates for the three right-most items are much larger than the unit parameters for the two left-most items. This fact also leads to the fat tails for the posterior distributions for individuals endorsing only the last item, or the last two items (i.e., patterns 00001 and 00011).

6.3 Analysis of Muhlberger's Politically Integrated Motivation Survey Items

6.3.1 Analysis with the HCM

This section applies the Bayesian formulation of the hyperbolic cosine model (HCM; Andrich and Luo, 1993; Verhelst and Verstralen, 1993, see Chapter 2) to the political motivation data. Table 5 contains the parameter estimates for the HCM item parameters.

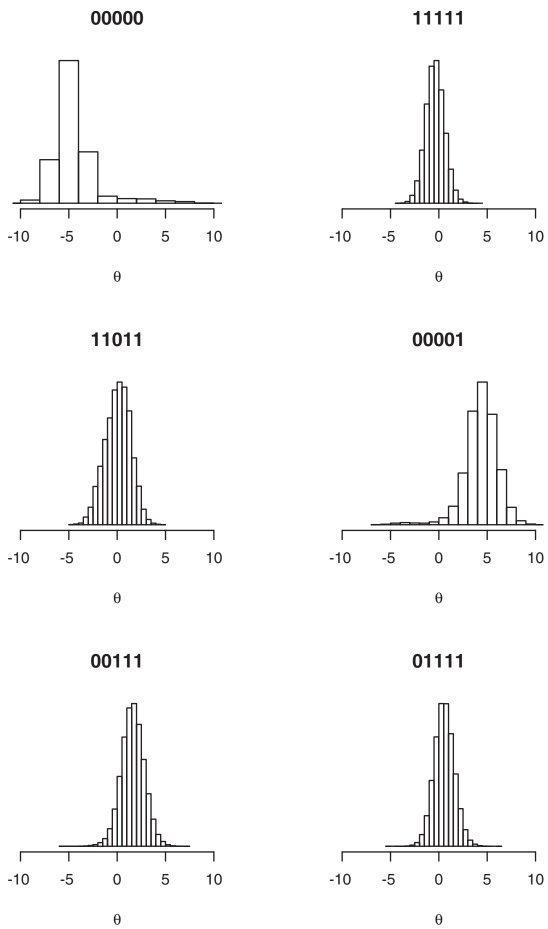

FIGURE 6. *Posterior distributions of latent attitudes for nine of the 600 subjects responding to the five questions about nuclear energy.*

The second and third columns of numbers in the table correspond to the median of the posterior distribution, and the 95% equal-tailed posterior credible interval for the items locations β_j, and item unit parameters γ_j found using the data augmentation MCMC procedure discussed earlier in this chapter. Muhlberger (1999) hypothesized that the items represent varying levels of political maturity. The low numbered items represent politically immature reasons for following politics, and the high numbered items represent mature reasons. The HCM parameter estimates suggest the items are ordered (1, 2, 3, 4, 5, 8, 7, 6). The HCM analysis provides more than just an ordering of the items, it also gives estimates for the locations of the eight items. For example, Item 3 is estimated to be closer than Item 5 to Item 4 on the latent dimension. The median of the posterior distribution of σ is 4.12; all items fall within one standard deviation of the average respondent.

Examining the item-unit parameters (γ_j) we find that item 2 has the lowest item-unit $\gamma_2 = -1.24$; this item was endorsed (responded above 24) by only three

TABLE 5
The MCMC Approximated Posterior Medians and 95% Equal-Tailed Credible Intervals for Item Locations and Item Units for Muhlberger's Political Motivation Data

Parameter	*Mdn*	95% Credible Interval
1. Expected		
Location β_1	−3.39	(−5.12, −1.69)
Unit γ_1	0.67	(−0.19, 1.62)
2. Upset		
Location β_2	−2.99	(−5.49, −0.58)
Unit γ_2	−1.24	(−2.63, −0.06)
3. Feel bad		
Location β_3	−1.89	(−3.60, −0.18)
Unit γ_3	−0.14	(−0.86, 0.52)
4. Bother		
Location β_4	−0.87	(−2.22, 0.61)
Unit γ_4	1.86	(1.21, 2.56)
5. Learn		
Location β_5	0.54	(−0.89, 1.91)
Unit γ_5	3.63	(2.88, 4.43)
6. Important		
Location β_6	1.50	(−0.16, 3.30)
Unit γ_6	5.17	(4.09, 6.42)
7. Fun		
Location β_7	1.25	(−0.27, 3.13)
Unit	1.69	(1.06, 2.40)
8. Enjoy Politics		
Location β_8	0.87	(−0.54, 2.67)
Unit γ_8	2.60	(1.93, 3.36)
Attitude Distribution Variance σ^2	17.02	(10.47, 26.61)

of the 113 respondents. The item with the largest item-unit parameter estimate is Item 6 ($\gamma_6 = 5.16$). Eighty-two of the 113 respondents responded positively to Item 6. Figure 7 illustrates how the probability of responding positively on a given item (the *y*-axis) is affected by the respondent's location on the political motivation scale (the *x*-axis).

Table 6 and Figure 8 study the posterior distribution of four of the 113 respondents' latent attitudes.

The posterior distributions of latent attitudes again show the distribution for an individual endorsing no items is bimodal, which was not the case in the previous example. Not only is the posterior for this individual bimodal, but also there is almost zero probability that the individual's latent attitude θ falls in the range (0, 3). A 95% equal-tailed credible interval ignores important information about the distribution. In this case it makes more sense to examine the 95% highest posterior density region [HPD]. For a subject endorsing none of the items the 95% HPD is $(-12.3, -2.2) \cup (4.5, 11.6)$.

6.3.2 Analysis with the squared logistic model

This section reanalyzes the political motivation data with the squared logistic model (SLM) defined in Equation 1. To estimate the SLM parameters we apply Theorem 4.1 to generate the latent responses ξ_{ij} so that they follow a simple monotone

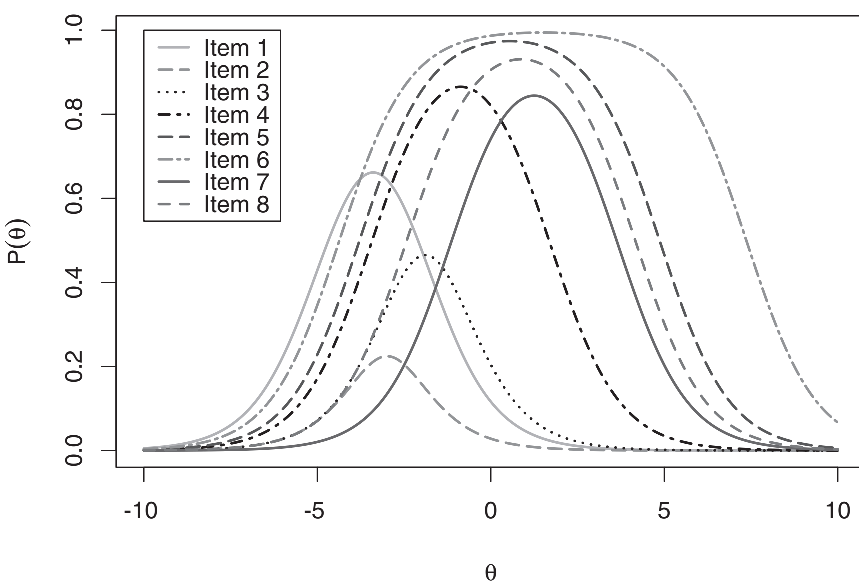

FIGURE 7. *The item response functions for the eight political agency items estimated using the hyperbolic cosine unfolding model.*

TABLE 6
The MCMC Approximated Posterior Medians and 95% Equal-Tailed Credible Intervals for the Latent Attitudes of Four of the 113 Respondents in Muhlberger's Political Motivation Data

Parameter	*Mdn*	95% Credible Interval	Response Pattern
Subject 1	−5.48	(−11.39, 10.41)	00000000
Subject 2	−0.88	(−2.75, −1.19)	01111111
Subject 3	−3.56	(−5.71, −1.52)	11000001
Subject 4	−0.94	(−3.70, 3.93)	00011100

item response model, and then use the augmented data to estimate the item parameters of that model. Table 7 contains the parameter estimates for the SLM model parameters and Figure 9 contains the estimated posterior densities of four respondents.

Although the estimates for the item location from the SLM are not on the same scale as the estimates of the HCM item locations, the two methods order the items exactly the same (1, 2, 3, 4, 5, 8, 7, 6) and the shapes of the posterior densities for the four respondents look almost identical. The difference in the scale is obvious when we compare the two estimates of the attitude distribution variances. The HCM estimates the standard deviation of the attitude distribution as $\sigma_h = 4.12$ whereas the SLM

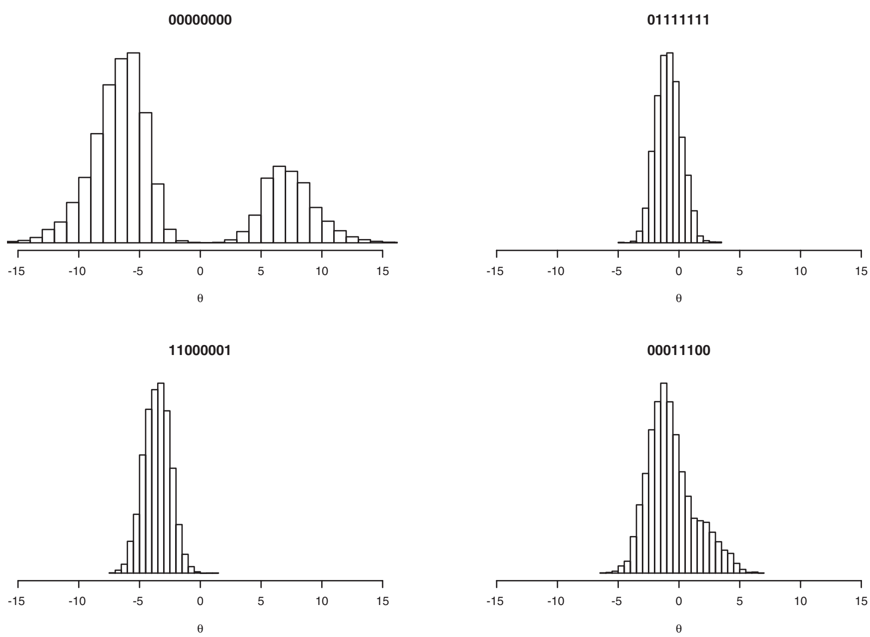

FIGURE 8. *Posterior distributions of latent attitudes for four of the 113 students responding to Muhlberger's survey of political motivation under the hyperbolic cosine model.*

TABLE 7
Posterior Medians and 95% Credible Intervals for the Squared Logistic Model Parameters for the Muhlberger (1999) Political Motivation Data Set

Parameter	*Mdn*	95% Credible Interval
Item 1		
Location	–1.42	(–1.93, –0.92)
Unit	0.11	(–0.77, 1.17)
Item 2		
Location	–1.35	(–2.27, –0.47)
Unit	–1.97	(–3.55, –0.56)
Item 3		
Location	–0.68	(–1.19, –0.21)
Unit	–0.48	(–1.24, 0.31)
Item 4		
Location	–0.05	(–0.41, 0.33)
Unit	1.08	(0.43, 1.82)
Item 5		
Location	0.71	(0.39, 1.04)
Unit	2.96	(2.12, 3.96)
Item 6		
Location	1.09	(0.75, 1.48)
Unit	4.95	(3.65, 6.86)
Item 7		
Location	0.93	(0.53, 1.29)
Unit	0.76	(0.17, 1.44)
Item 8		
Location	0.78	(0.42, 1.11)
Unit	1.73	(1.07, 2.52)
Attitude Distribution Variance	3.62	(2.41, 5.53)
Subject 1	3.39	(–4.28, 6.12)
Subject 2	–0.14	(–0.75, 0.53)
Subject 3	–1.28	(–1.86,–0.66)
Subject 4	–0.01	(–1.13, 1.95)

estimates the standard deviation as $\sigma_s = 1.91$. This difference in scale is accountable to the fact that the SLM item response functions approach zero as θ moves away from the item location at a much faster rate than the HCM item response functions.

7. Discussion

In the last 15 years several authors have developed probabilistic unfolding models as a way to study the responses to attitudinal questionnaires. In this article we describe two methods for the Bayesian estimation of unfolding responses models. One method is based on the direct implementation of Markov chain Monte Carlo methods; the second method takes advantage of a generalization of the latent response formulation of the unfolding response models.

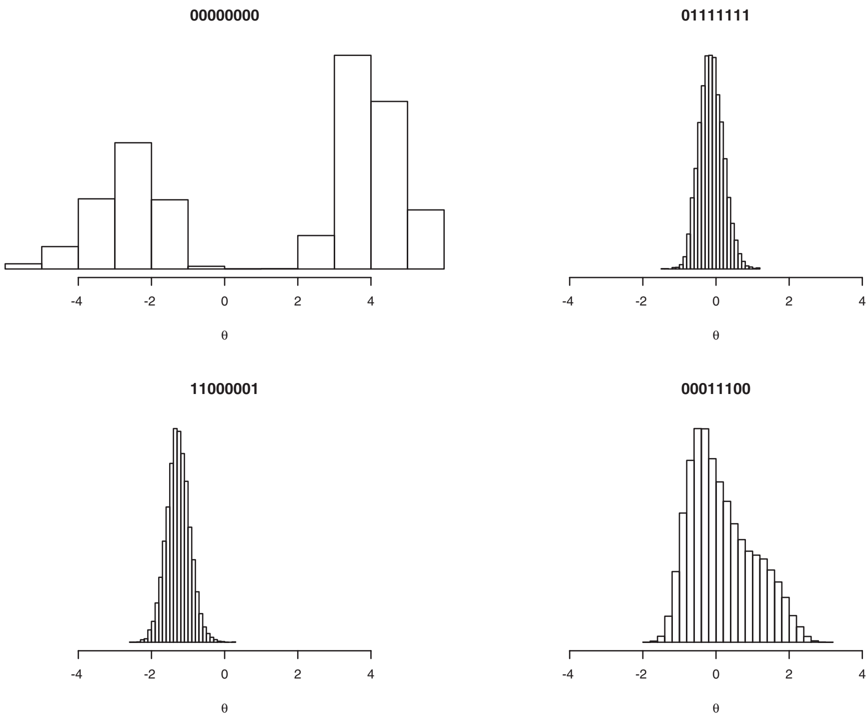

FIGURE 9. *Posterior distributions of latent attitudes for four of the 113 students responding to Muhlberger's survey of political motivation under squared logistic model.*

Bayesian methods have clear advantages over maximum likelihood estimation procedures for unfolding models. Even when the item locations are known, the likelihood for the respondents' latent attitudes are often multimodal, making MLEs difficult to find with numerical methods, and standard errors even more difficult to approximate. A Bayesian analysis on the other hand can construct any summaries of the model parameters without relying on asymptotic results. For example, we can construct the 95% highest posterior density (HPD) credible regions for each of the model parameters from the converged MCMC procedure. As we demonstrate in the examples, the HPD regions may or may not be connected sets.

The Bayesian analyses of the three example data sets demonstrate the advantages of the Bayesian analysis in unfolding methods. Joint maximum likelihood methods remove any individuals who endorse no items from the analysis claiming no information about these individuals can be extracted from the data, and that these individuals provide no information about item parameters. The Bayesian analysis, like MML estimation, does retrieve useful information from the data by assuming a latent attitude distribution for θ. However a Bayesian analysis is able to derive the posterior distributions of item locations and respondent locations simultaneously, whereas

MML estimation first estimates the item parameters and then uses empirical Bayes methods to estimate the respondent's locations.

One of the difficulties estimating unfolding models is the near-indeterminacy of the item-unit parameters for items that are far from the center of the attitude distribution. The re-parameterization we suggest is promising; however, because unfolding models approach monotone response models as the item location increases to infinity or decreases to negative infinity, it may be advantageous to develop an estimation algorithm that allows the extreme items (those located very high or low on the scale) to be parameterized as monotone (either increasing, or decreasing) items.

Appendix

A.1 Proof of Proposition 4.1

Proof. To find the conditional distribution of X given the latent variable θ marginalize the probabilities in Equations 9, 10 and 11 over the conditional distribution of the latent response ξ given θ. That is,

$$\begin{aligned}
Pr\{X_j = 1|\theta\} &= \sum_{k=0}^{2} Pr\{X_j = 1|\theta, \xi_j = k\}Pr\{\xi_j = k|\theta\} \\
&= 0 + \frac{P_j(\theta)}{R_{j1}(\theta)} R_{j1}(\theta) + 0 \qquad , \\
&= P_j(\theta) \\
Pr\{X_j = 0|\theta\} &= \sum_{k=0}^{2} Pr\{X_j = 0|\theta, \xi_j = k\}Pr\{\xi_j = k|\theta\} \\
&= R_{j0}(\theta) + \left(1 - \frac{P_j(\theta)}{R_{j1}(\theta)}\right)R_{j1}(\theta) + R_{j2}(\theta) \\
&= \sum_{k=0}^{2} R_{jk}(\theta) - P_j(\theta) \\
&= 1 - P_j(\theta)
\end{aligned}$$

so the required identities hold. □

A.2 Proof of Theorem 4.1

Proof. Note from Equations 12 and 13 that all probabilities are less than or equal to 1 for all θ.

$$\begin{aligned}
P_j(\theta) &\leq R_{j1}(\theta)\ \forall\theta && \text{[By Hypothesis]} \\
\Rightarrow 1 - P_j(\theta) &\geq 1 - R_{j1}(\theta) = 1 - [1 - R_{j0}(\theta) - R_{j2}(\theta)] && \left[\sum_k R_{jk}(\theta) = 1\right] \\
\Rightarrow 1 - P_j(\theta) &\geq R_{j0}(\theta) + R_{j2}(\theta) \geq R_{j0}(\theta) \\
\Rightarrow \frac{R_{j0}(\theta)}{1 - P_j(\theta)} &\leq 1
\end{aligned}$$

Similarly, $\frac{R_{j2}(\theta)}{1-P_j(\theta)} \leq 1$. For $k = 1$

$$1 - P_j(\theta) \geq 1 - R_{j1}(\theta) \geq P_j(\theta) - R_{j1}(\theta) \Rightarrow \frac{P_j(\theta) - R_{j1}(\theta)}{1 - P_j(\theta)} \leq 1.$$

So the functions in Equations 12 and 13 are between zero and one.

Secondly, the conditional distributions in Equations 12 and 13 must imply the correct marginal probabilities for the latent responses (conditional on the model parameters). For $\xi = 0$ the category response function is defined by

$$\begin{aligned} Pr\{\xi_j = 0|\theta\} &= \sum_k Pr\{\xi_j = 0|\theta, X = k\}Pr\{X = k|\theta\} \\ &= Pr\{\xi_j = 0|\theta, X = 0\}Pr\{X = 0|\theta\} \\ &\quad [Pr\{\xi = 0|\theta, X = 1\} = 0] \\ &= R_{j0}(\theta). \end{aligned}$$

Similarly $Pr\{\xi_j = 2\,|\theta\} = R_{j2}(\theta)$ for the conditional distributions defined in Equations 12 and 13. The category response function for $\xi = 1$, $R_{j1}(\theta)$, is

$$\begin{aligned} Pr\{\xi_j = 1|\theta\} &= \sum_k Pr\{\xi_j = 1|\theta, X = k\}Pr\{X = k|\theta\} \\ &= \left[\frac{R_{j1}(\theta) - 1}{1 - P_j(\theta)} + 1\right](1 - P_j(\theta)) + P_j(\theta) \\ &= R_{j1}(\theta) - 1 + 1 - P_j(\theta) + P_j(\theta) \\ &= R_{j1}(\theta). \end{aligned}$$

Hence the conditional distributions imply the correct marginal distributions. □

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B, 32,* 283–301.

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

Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology, 49,* 347–365.

Andrich, D., & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses." *Applied Psychological Measurement, 17,* 253–276.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician, 49,* 327–335.

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

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

Dempster, A. P., Laird, N., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–38.

DeSarbo, W. S., & Hoffman, D. L. (1986). Simple and weighted unfolding threshold models for the spatial representation of binary choice data. *Applied Psychological Measurement, 10,* 247–264.

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

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London, UK: Chapman & Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721–741.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57,* 97–109.

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

Holland, P. W. (1990). On the sampling theory foundations of item response models. *Psychometrika, 55,* 4, 577–601.

Holland, P. W. & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics, 14,* 1523–1543.

Johnson, M. S. (2001). *Parametric and non-parametric extensions to unfolding response models.* Doctoral dissertation, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA.

Johnson, M. S., Cohen, W. M., & Junker, B. W. (1999). *Measuring appropriability in research and development with item response models.* Technical report, Pittsburgh, PA: Department of Statistics, Carnegie Mellon University.

Luo, G. (1998). A general formulation for unidimensional latent trait unfolding models: Making explicit the latitude of acceptance. *Journal of Mathematical Psychology, 42,* 400–417.

Luo, G. (2000). A joint maximum likelihood estimation procedure for the Hyperbolic Cosine Model for single-stimulus responses. *Applied Psychological Measurement, 24*(1), 33–49.

Maris, E. (1995). Psychometric latent response models. *Psychometrika, 60,* 523–547.

Maris, G, & Maris, E. (2002). *Are attitude items monotone or single-peaked? An analysis using Bayesian Methods.* Citogroep Measurement and Research Department Report No. 2002-02. Arnhem, NE.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics, 21,* 1087–1092.

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association, 44,* 335–341.

Muhlberger, P. (1999). *A general unfolding, non-folding scaling model and algorithm.* Presented at the 1999 American Political Science Association Annual Meeting: Atlanta, GA.

Noël, Y. (1999). Recovering unimodal latent patterns of change by unfolding analysis: Application to Smoking Cessation. *Psychological Methods, 4,*(2), 173–191.

Patz, R., & Junker, B. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics, 24,* 342–366.

Patz, R., Junker, B. W., Johnson, M. S., & Mariano, L. (2002). The hierarchical rater model for rated test items and its application to large-scale educational test data. *Journal of Educational and Behavioral Statistics, 27,* 341–384.

Ripley, B. D. (1987). *Stochastic simulation.* New York: Wiley.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24,* 3–32.

Tanner, M., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82,* 528–550.

Tanner, M. A. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions.* Springer Series in Statistics, 3rd ed. New York: Springer-Verlag.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 278–286.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33,* 529–554.

Verhelst, N. D., & Glas, C. A. W. (1995). *Rasch models: Foundations, recent developments, and applications.* New York: Springer-Verlag.

Verhelst, N. D., & Verstralen, H. H. F. M. (1993). A stochastic unfolding model derived from the partial credit model. *Kwantitative Methoden, 42,* 73–92.

Authors

MATTHEW S. JOHNSON is Assistant Professor, Department of Statistics and Computer Information Systems, Baruch College, City University of New York, New York, NY 10010; Matthew_Johnson@baruch.cuny.edu. He specializes in the development and estimation of hierarchical Bayesian models for educational and behavioral data.

BRIAN W. JUNKER is Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh PA 15213; brian@stat.cmu.edu. He specializes in the statistics of latent variable models for measurement, and adaptations of parametric and nonparametric item response models to novel repeated measures settings.

Manuscript Received March 2003
Revision Received June 2003
Accepted June 2003