A Beta Item Response Model for Continuous Bounded Responses
Yvonnick Noel and Bruno Dauvier
Applied Psychological Measurement 2007 31: 47
DOI: 10.1177/0146621605287691

The online version of this article can be found at:
http://apm.sagepub.com/content/31/1/47

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Applied Psychological Measurement can be found at:

Email Alerts: http://apm.sagepub.com/cgi/alerts
Subscriptions: http://apm.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://apm.sagepub.com/content/31/1/47.refs.html

>> Version of Record - Dec 12, 2006
What is This?
A Beta Item Response Model for Continuous Bounded Responses

Yvonnick Noel and Bruno Dauvier, University of Rennes 2, France

An item response model is proposed for the analysis of continuous response formats in an item response theory (IRT) framework. With such formats, respondents are asked to report their response as a mark on a fixed-length graphical segment whose ends are labeled with extreme responses. An interpolation process is proposed as the response mechanism in this situation, from which the authors show that a beta distribution on the response naturally follows. The two natural parameters of the beta are expressed as monotonic functions of person and item distance on some latent continuum. It is shown that a logistic expected response function results from a simple choice for these monotonic functions. A joint maximum likelihood estimation procedure is presented that performs well in a simulation study. An application on mood items is also presented, with very good fit properties. Index terms: item response theory, beta distribution, continuous bounded responses, visual analogue scales, attitude measurement

Introduction

Categorical Versus Continuous Rating Scales

During the past 40 years, a great deal of work has been devoted to the probabilistic modeling of Likert responses. As Andrich (1996) pointed out, this response format, initially proposed by Likert (1932), has gained success mainly because it was easier to use than other measuring procedures available at that time (Thurstone, 1927, 1928). The question of how many responses categories should be retained in questionnaire design has been discussed at length in psychological literature (Cox, 1980). Although there is support for the view that more response categories tend to reduce extreme responses and may produce more discrimination, a common practice seems to favor 5- to 7-point formats. Known limitations of short-term memory capacity (Miller, 1956) are sometimes invoked to justify this choice. A Likert response format with a small number of response categories has the advantage of being easily understood and manipulated by a majority of persons but is likely to provide coarser measurements. On the other hand, introducing more response categories leads to an explosion in the number of parameters in categorical response models.

In this context, it is justified to wonder whether collecting continuous responses in attitude measurement is psychologically meaningful and practically feasible. If the answer is positive, then clearly, there would be many technical advantages to use such a response measurement procedure. In particular, probabilistic models for continuous data are likely to be more parsimonious in parameters.

In this article, the authors are interested in a continuous response format (CRF) in which respondents report their agreement with items by tracing a vertical mark somewhere on a fixed-length horizontal segment, with ends that are labeled “0% agreement” and “100% agreement,” respectively (other labeling would of course be possible, depending on the context). No other semantic anchor is
provided. This is sometimes referred to as visual analogue response scales. The measurement is then the distance (up to an arbitrary precision) of this mark to the left end of the segment. The measurement is, strictly speaking, continuous. In support with the use of such response formats and at odds with a common belief, it can be noted that Miller (1956) himself reports that absolute judgments of the position of a pointer in a linear interval are much more fine-grained than judgments on other kind of stimuli: Up to 16 different locations are easily distinguished.

Static CRF (a single response for each item) has already been used in personality assessment (Bejar, 1977; Ferrando, 2001) and is commonly used in pain intensity assessment (e.g., Morin & Bushnell, 1998). Alternatively, dynamic CRF data may be recorded online using ad hoc mechanical devices for the assessment of an ongoing emotional experience when listening to music (Brittin, 1996) or watching films (Goldin et al., 2005). Only the static CRF will be considered in this article.

Continuous responses may have several interesting properties in practice. Although several applied studies have reported high correlations between Likert and CRF responses and encourage using the simpler Likert format (e.g., Guyatt, Townsend, Berman, & Keller, 1987), some empirical evidence exists that subtle interactions may appear only with CRF data (Russel & Bobko, 1992) or that CRF responses show greater sensitivity in the measurement of change (Pfennings, Cohen, & van der Ploeg, 1995). Grant et al. (1999) also report greater sensitivity and reliability of CRF scores. Albaum, Best, and Hawkins (1981) argue that they allow for better interindividual discrimination, and McKelvie (1978) reports that respondents find continuous scales more pleasing to use and more consistent.

With the development of computerized testing, it may be that CRF will gain some popularity in the future (de Leeuw & Nicholls, 1996): Continuous responses are indeed easily collected through a computer interface where responses are entered by positioning a cursor along a slider.

Continuous Responses Models

By contrast with categorical responses, not many proposals have been made for the analysis of continuous bounded responses in an item response theory (IRT) framework (Ferrando, 2001; Mellenbergh, 1994; Müller, 1987; Samejima, 1973, 1974).

Mellenbergh (1994) proposed a simple linear response model that may be viewed as an IRT version of the factor-analytic model. Although it may in practice give insightful results, neither a linear response function nor a normal distribution for the response is strictly speaking appropriate to the modeling of bounded data.

Müller (1987) proposed a continuous rating scale model based on a response mechanism where a latent response variable, originally unbounded and following a normal distribution, is finally truncated during the response process to fit the response format constraint. The model is a direct extension of Andrich’s (1978) rating scale model for categorical responses, in which the number of response categories is arbitrarily increased, and category thresholds are assumed to follow a uniform distribution. Müller’s model belongs to the Rasch family and so has the interesting specific objectivity property.

Ferrando (2001) elaborated on a similar idea of a truncation mechanism, to extend the linear response model to a nonlinear congeneric model that takes the bounded nature of the data into account. Besides the fact that both models do not lead to a closed-form expected response function, it is not clear why respondents would psychologically truncate (rather than rescale, for instance) a latent response to fit the response format.

Samejima (1973, 1974) proposed a continuous response model (CRM) that is obtained as a limiting case of the graded response model (Samejima, 1969) when the number of response categories is increased to infinity. A software implementation for this model has been written by Wang and Zeng (1998). A nice property of the CRM, when the normal ogive is chosen as the intercategory regression function, is that the score information remains constant for all values of the latent
attitude. However, this model is complex in that it assumes a two-level response process: Assuming a normal distribution for some latent response, it takes the bounded nature of the observed data into account through a monotonic (e.g., logistic) transform of the latent response. The choice of the logistic (or any other sigmoid function) as a latent-to-manifest response transformation function is somewhat arbitrary and has no psychological interpretation. As a drawback, it also leads to a null density for the extreme responses, which is not very realistic.

By contrast, instead of transforming the data or truncating a real distribution, a new response mechanism is proposed in which an implicit evaluation of the scale boundaries is at the core of the response process. This will be shown to lead to a well-known distribution model on the response, which has compact support and from which person and item characteristics are derived. This approach has the advantage of defining a density on the whole (closed) response interval and will be shown to give an expected response function that can be written in closed form.

This article is organized as follows: The first section presents a hypothetical interpolation response mechanism, from which a beta distribution model for the manifest response is derived. Some properties of the beta are briefly summarized, and this distributional model is further specified in the next section by reexpressing the distribution parameters as a function of person and item parameters (or “ability” and “difficulty” parameters) to obtain a beta item response model. The third section extends this basic model by adding an extra dispersion parameter. Quality of parameter estimation is then tested in a simulation study, and the model is finally applied to real mood data.

An Interpolation Response Mechanism

By contrast with Likert response scales, CRF provides no semantic reference points except at the boundaries of the response segment. It is then assumed that, faced with CRF, persons construct their responses in reference to latent relevance values (or proximity judgments) they grant to both extreme responses, reflecting how close to each one they feel to be.

Psychological values are denoted by \( v_0 \) and \( v_1 \), granted to the left and right ends, respectively, and it is assumed that the observed responses are scaled to lie in \([0; 1]\). The response variable \( X \) is then assumed to be of the following form:

\[
X = \frac{\lambda_0 v_0 + \lambda_1 v_1}{v_0 + v_1} = \frac{v_1}{v_0 + v_1},
\]

(1)

where \( \lambda_0 = 0 \) and \( \lambda_1 = 1 \) are the arbitrary graphic locations of the segment boundaries. This may be viewed as an interpolation process: Respondents put a check at a distance of the left boundary that is proportional to the relative value they give to the extreme positive answer. This simple mechanism closely resembles commonly used models in choice theory (Luce, 1959) or reinforcement learning (Herrnstein, 1961).

The random part of the model is introduced by defining a distribution for \( v_0 \) and \( v_1 \). Because those implicit values are thought of as nonnegative quantities, it is assumed for convenience that

\[
\begin{align*}
v_0 &\sim \Gamma(m, s), \\
v_1 &\sim \Gamma(n, s).
\end{align*}
\]

(2)

Latent values are assumed to follow a gamma distribution with a common scale parameter. This is reasonable given that respondents have to take both reference points into account to construct a single response. This choice of the gamma for the value distribution is not the only one possible...
but has the attractive feature of leading to a known distribution for \( X \). From these assumptions, it is known that (Kotz & Johnson, 1982, p. 229)

\[
X = \frac{v_0}{v_0 + v_1} \sim \beta(m, n).
\]

The beta density, with two parameters \( m \) and \( n \), is defined as

\[
f(x; m, n) = \frac{\Gamma(m + n)}{\Gamma(m)\Gamma(n)} x^{m-1}(1-x)^{n-1} \quad \text{for} \quad x \in [0; 1], m, n > 0
\]  

with first moments:

\[
E(X; m, n) = \mu = \frac{m}{m + n},
\]

\[
V(X; m, n) = \frac{mn}{(m + n)^2(m + n + 1)} = \mu(1 - \mu) \left[ \frac{1}{m + n + 1} \right],
\]

for a random response \( X \). Note that, by contrast with the binomial, the variance does not depend simply on the mean. The term \( \eta = \frac{1}{m + n + 1} \) may thus be viewed as a pure dispersion parameter.
As the beta density is defined on \([0; 1]\), it is well suited to the modeling of bounded continuous responses, which may easily be rescaled to lie in \([0; 1]\). Depending on the values of \(m\) and \(n\), the beta may show very different shapes, as illustrated in Figure 1. Non-null densities for the extreme values are clearly possible. For \(m = n = 1\), the density is uniform. For some parameter values (e.g., \(m = 1/4\) and \(n = 1/4\)), it may also show a bimodal shape, but this is not expected to occur on the kind of data under interest below.

**A Beta Model of Response**

**The Basic Model**

The usual psychometric situation is now considered, where \(N\) persons give ratings on a set of \(p\) items. A beta model of response variable \(X_{ij}\) from subject \(i\) \((i = 1, \ldots, N)\) to item \(j\) \((j = 1, \ldots, p)\) may be formulated as

\[
f(x_{ij}; m_{ij}, n_{ij}) = \frac{\Gamma(m_{ij} + n_{ij})}{\Gamma(m_{ij})\Gamma(n_{ij})} x_{ij}^{m_{ij} - 1} (1 - x_{ij})^{n_{ij} - 1} \quad \text{with} \quad x_{ij} \in [0; 1],
\]

where \(x_{ij}\) is a realization of \(X_{ij}\). Because \(m_{ij}\) and \(n_{ij}\) are exponents on the \(x_{ij}\) and \((1 - x_{ij})\) factors, respectively, they may be interpreted as “acceptance” and “refusal” parameters. These distribution parameters are now reexpressed as a function of two sets of structural parameters—that is, person \((\theta_i)\) and item \((\delta_j)\) parameters—in the following way:

\[
m_{ij} = \exp\left(\frac{\theta_i - \delta_j}{2}\right),
\]

\[
n_{ij} = \exp\left(-\frac{\theta_i - \delta_j}{2}\right).
\]

Item acceptance is thus an increasing function of person-item distance, and item refusal a decreasing function of person-item distance on some latent continuum. The exponential ensures that the parameters take on strictly positive values. The one-half scaling factor is not strictly necessary but somewhat simplifies the expressions below. With this reparameterization, the response distribution in equation (6) is taken over the possible manifest responses given an attitude true score \(\theta_i\) and a fixed item parameter \(\delta_j\). No assumption is made for the distribution of the true scores.

Under this model, the conditional expected response is

\[
\mu_{ij} = E(X_{ij}; m_{ij}, n_{ij}) = \frac{m_{ij}}{m_{ij} + n_{ij}}
\]

and after reparameterization:

\[
E(X_{ij}; \theta_i, \delta_j) = \frac{\exp\left(\frac{\theta_i - \delta_j}{2}\right)}{\exp\left(\frac{\theta_i - \delta_j}{2}\right) + \exp\left(-\frac{\theta_i - \delta_j}{2}\right)}
\]

\[
= \frac{1}{1 + \exp\left(-\left(\theta_i - \delta_j\right)\right)}.
\]
A logistic expected response function thus naturally emerges from the model (equation (6)), with \( \theta_i \) and \( \delta_j \) parameters having the familiar interpretation of subject ability (or attitude) and item difficulty (or value), respectively. The variance function is

\[
V(X_{ij}; \theta_i, \delta_j) = \frac{\mu_{ij}(1 - \mu_{ij})}{1 + \exp\left(\frac{\theta_i - \delta_j}{2}\right) + \exp\left(-\frac{\theta_i - \delta_j}{2}\right)}
\]

\[
= \frac{\mu_{ij}(1 - \mu_{ij})}{1 + 2 \cosh\left(\frac{\theta_i - \delta_j}{2}\right)}.
\]

This is an even function of person-item distance, reaching its maximum for \( \theta_i = \delta_j \).

Under this form, it is called the one-parameter beta response model (BRM-1). It is both simple and easily interpretable.

In practice, however, it may appear too constraining to assume a common shape for the variance function across items. Items may greatly vary in the way they are understood or judged relevant by respondents, and this is likely to result in different conditional variance functions. Also note that for \( \theta_i = \delta_j \), the response density is uniform under this first model. Whether these assumptions are realistic may be tested by comparing this model to the extended form below.

**An Extended Model With Dispersion Parameters**

An extended model that will allow different items to have different response variabilities is obtained by setting

\[
m_{ij} = \exp\left\{\frac{\theta_i - \delta_j + \tau_j}{2}\right\},
\]

\[
n_{ij} = \exp\left\{-\left(\theta_i - \delta_j\right) + \tau_j\right\}.
\]

The expected response remains unchanged:

\[
\mu_{ij} = E(X_{ij}; \theta_i, \delta_j) = \frac{\exp\left\{\frac{\theta_i - \delta_j + \tau_j}{2}\right\}}{\exp\left\{\frac{\theta_i - \delta_j}{2}\right\} + \exp\left\{-\left(\theta_i - \delta_j\right) + \tau_j\right\}},
\]

\[
= \frac{1}{1 + \exp\left\{-\left(\theta_i - \delta_j\right)\right\}},
\]

whereas the variance function shape is now item specific:

\[
V(X_{ij}; \theta_i, \delta_j, \tau_j) = \frac{\mu_{ij}(1 - \mu_{ij})}{1 + \exp\left\{\frac{\theta_i - \delta_j + \tau_j}{2}\right\} + \exp\left\{-\left(\theta_i - \delta_j\right) + \tau_j\right\}}
\]

\[
= \frac{\mu_{ij}(1 - \mu_{ij})}{1 + 2 \phi_j \cosh\left(\frac{\theta_i - \delta_j}{2}\right)},
\]

with \( \phi_j = \exp\left(\frac{\tau_j}{2}\right) \), an (inverse) dispersion parameter. The higher \( \phi_j \), the lower the response variance given \( \theta_i \) and \( \delta_j \). The \( \phi_j \) parameter may thus be interpreted as a precision measure.
Under this form, it is called the two-parameter beta response model (BRM-2), and BRM-1 is just the special case in which \( \phi_j = 1 \); \( \phi_j \). Figure 2 displays the expected response function and the 95% confidence region for a sample item with \( \tau_j = 4 \). The conditional response densities for five different values of \( \theta_i - \delta_j \) are also plotted. The modal response function,

\[
Mo(X_{ij}; m_{ij}, n_{ij}) = \frac{m_{ij} - 1}{(m_{ij} - 1) + (n_{ij} - 1)},
\]

is added to make clear that in real data modeling, the estimated expected response function will in general appear at some distance of the data highest density zone.

For \( \theta_i = \delta_j \) (i.e., for the maximum of response variance), the (symmetric) response distribution is \( \beta(\phi_j, \phi_j) \), so a 95% response confidence interval at \( \theta_i = \delta_j \) could well be used as a measure of item imprecision. Because the sample space is bounded, some a priori chosen value (say, 0.80) for

---

**Figure 2**

Expected Response Function and Response Densities Under the Two-Parameter Beta Response Model (BRM-2)

*Note.* The continuous curve represents the expected response function and the dashed curve the modal (highest density) response function (\( \tau_j = 4 \)). The 2.5% and 97.5% response quantile functions are plotted as dotted curves. The two black dots represent the boundaries of the response 95% confidence interval for \( \theta_i = \delta_j \).
this central confidence interval may be used as an additional item selection criterion, besides standard item fit measures (see the real data example below).

From equations (11) and (12), it can be seen that for person scores at symmetric locations around some fixed item location—say, \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \), such that \( (\hat{\theta}_1 - \delta_j) = -(\hat{\theta}_2 - \delta_j) \)—conditional response distributions will show symmetric shapes with respect to the \( X = 0.5 \) axis. That is, if \( X_j | \hat{\theta}_1 \sim \beta(m_j, n_j) \), then \( X_j | \hat{\theta}_2 \sim \beta(n_j, m_j) \).

Although the interpretation of BRM parameters is familiar, by analogy with the well-known Rasch models, an important difference must be underlined: The logistic here has no distributional interpretation. It is a conditional expected response function of the form \( E(X | \theta, \delta, \tau) \), or a regression function of item response on the latent trait. Moreover, by comparison with Samejima’s (1973) CRM, it is not an ad hoc transformation function on the latent response but an expected rating function that mathematically emerges from the hypothetical interpolation response mechanism.

### Parameter Estimation

The model parameters are regarded as fixed (no distributional assumption is made), and a joint maximum likelihood (JML) procedure is used to estimate person and item parameters.

Assuming that responses are independent across persons and locally independent for each person given the model, the BRM-2 model likelihood, as a function of the whole set \( \Theta \) of \( \theta_i \), \( \delta_j \), and \( \tau_j \) parameters, given the whole data set \( X \), is

\[
L(\Theta; X) = \prod_{i=1}^{N} \prod_{j=1}^{p} f(x_{ij}|\Theta) = \prod_{i=1}^{N} \prod_{j=1}^{p} \frac{\Gamma(m_{ij} + n_{ij})}{\Gamma(m_{ij})\Gamma(n_{ij})} x_{ij}^{m_{ij}-1}(1-x_{ij})^{n_{ij}-1},
\]

and the log-likelihood function is

\[
\ln L(\Theta; X) = \sum_{i=1}^{N} \sum_{j=1}^{p} \left\{ \ln \Gamma(m_{ij} + n_{ij}) - \ln \Gamma(m_{ij}) - \ln \Gamma(n_{ij}) + (m_{ij} - 1) \ln x_{ij} + (n_{ij} - 1) \ln (1 - x_{ij}) \right\}.
\]

The partial derivatives of the log-likelihood function with respect to \( \theta_i \), \( \delta_j \), and \( \tau_j \) are (see the appendix)

\[
\frac{\partial \ln L(\Theta; X)}{\partial \theta_i} = \sum_{j=1}^{p} \left\{ \frac{m_{ij}}{2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] - \frac{n_{ij}}{2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] \right\},
\]

\[
\frac{\partial \ln L(\Theta; X)}{\partial \delta_j} = \sum_{i=1}^{N} \left\{ -\frac{m_{ij}}{2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{n_{ij}}{2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] \right\},
\]

\[
\frac{\partial \ln L(\Theta; X)}{\partial \tau_j} = \sum_{i=1}^{N} \left\{ \frac{m_{ij}}{2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{n_{ij}}{2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] \right\},
\]

where \( \Psi(x) = \partial \ln \Gamma(x) / \partial x \) is the digamma function.
Parameter estimates are obtained by setting these first derivatives to zero and solving for the parameters using an iterative numerical optimization routine (Newton-Raphson). Note that the model on $m_{ij}$ and $n_{ij}$ is overparameterized in equations (11) and (12), and one additional constraint is needed. Because there is no natural origin in the measurement of attitude, the following is arbitrarily set:

$$\sum_j \delta_j = 0.$$  

Because there are many parameters to estimate, this study does not solve over the whole set of parameters by inverting a full Hessian in the Newton-Raphson procedure. An alternating strategy iterates over the set of item parameters until some convergence criterion is reached, with the proper constraint, and then optimizes over the set of person parameters and finally over dispersion parameters. The second partial derivatives needed are as follows (see the appendix):

$$\frac{\partial^2 \ln L(\Theta; X)}{\partial \theta^2_i} = \sum_{i=1}^{N} \left\{ \frac{m_{ij}}{4} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{n_{ij}}{4} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] + \Psi'(m_{ij} + n_{ij}) \left( \frac{m_{ij} - n_{ij}}{2} \right)^2 - \Psi'(n_{ij}) \left( \frac{n_{ij}}{2} \right)^2 \right\},$$  

$$\frac{\partial^2 \ln L(\Theta; X)}{\partial \delta^2_j} = \sum_{i=1}^{N} \left\{ \frac{m_{ij}}{4} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{n_{ij}}{4} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] + \Psi'(m_{ij} + n_{ij}) \left( \frac{n_{ij} - m_{ij}}{2} \right)^2 - \Psi'(n_{ij}) \left( \frac{n_{ij}}{2} \right)^2 \right\},$$  

$$\frac{\partial^2 \ln L(\Theta; X)}{\partial \tau^2_j} = \sum_{i=1}^{N} \left\{ \frac{m_{ij}}{4} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{n_{ij}}{4} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] + \Psi'(m_{ij} + n_{ij}) \left( \frac{m_{ij} + n_{ij}}{2} \right)^2 - \Psi'(n_{ij}) \left( \frac{n_{ij}}{2} \right)^2 \right\},$$

where $\Psi'(x) = \partial \Psi(x) / \partial x$ is the trigamma function, and $m_{ij}$ and $n_{ij}$ are as defined in equations (11) and (12).
Approximate Standard Errors

From likelihood theory, it is known that asymptotic variances of the parameters are obtained as the diagonal elements of minus the inverse expected full Hessian matrix. A common simplification is adopted here, which is to estimate approximate standard errors by

\[
\hat{\sigma}_\pi = \left\{ -E \left[ \frac{\partial^2 \ln L(\pi; X)}{\partial \pi^2} \right] \right\}^{-\frac{1}{2}},
\]

where \( \pi \) is the parameter under interest. The expected second derivatives are given by the following (see the appendix):

\[
E \left[ \frac{\partial^2 \ln L(\Theta; X)}{\partial \theta_i^2} \right] = \sum_{j=1}^{p} \left\{ \Psi'(m_{ij} + n_{ij}) \left( \frac{m_{ij} - n_{ij}}{2} \right)^2 - \Psi'(m_{ij}) \left( \frac{m_{ij}}{2} \right)^2 \right\},
\]

(26)

\[
E \left[ \frac{\partial^2 \ln L(\Theta; X)}{\partial \delta_j^2} \right] = \sum_{i=1}^{N} \left\{ \Psi'(m_{ij} + n_{ij}) \left( \frac{n_{ij} - m_{ij}}{2} \right)^2 - \Psi'(m_{ij}) \left( \frac{m_{ij}}{2} \right)^2 \right\},
\]

(27)

\[
E \left[ \frac{\partial^2 \ln L(\Theta; X)}{\partial \tau_j^2} \right] = \sum_{i=1}^{N} \left\{ \Psi'(m_{ij} + n_{ij}) \left( \frac{m_{ij} + n_{ij}}{2} \right)^2 - \Psi'(m_{ij}) \left( \frac{m_{ij}}{2} \right)^2 \right\},
\]

(28)

The negative of these expectations represents the Fisher information functions of the BRM-2 parameters. The test information function, defined as

\[
I(\theta_i) = -E \left[ \frac{\partial^2 \ln L(\Theta; X)}{\partial \theta_i^2} \right]
= - \sum_{j=1}^{p} \left\{ \Psi'(m_{ij} + n_{ij}) \left( \frac{m_{ij} - n_{ij}}{2} \right)^2 - \Psi'(m_{ij}) \left( \frac{m_{ij}}{2} \right)^2 \right\}
= \sum_{i=1}^{N} \left\{ \Psi'(m_{ij} + n_{ij}) \left( \frac{n_{ij}}{2} \right)^2 \right\},
\]

is displayed in Figure 3 for several values of \( \tau_j \). As is the case with categorical cumulative models, the maximal item information is obtained for \( \theta_i - \delta_j = 0 \). Higher values of \( \tau_j \) are associated with higher information levels. So selecting items on the basis of their precision level, in addition to standard fit measures, as suggested above, amounts to selecting the most informative items.

Initial Estimates

A good selection of initial estimates is helpful in obtaining rapid convergence. The following estimates at Step 0 were found to be very good in practice:

\[
\hat{\theta}_i^{(0)} = \sigma \left( \frac{\sum_{j=1}^{p} x_{ij}}{p} \right),
\]
where $\sigma(.)$ is a centering-scaling function that transforms its argument into a zero mean and unit variance variable. At Step 0, all $\tau_j$ parameters are set to 0.

The expected second derivatives in equations (26), (27), and (28) will also be used in the parameter estimation procedure, in place of the empirical second derivatives, thus constituting an approximate Fisher scoring scheme. Each $\pi$ parameter is then iteratively improved at Step $t$ following

\[
\pi^{(t)} = \pi^{(t-1)} - \alpha E \left[ \frac{\partial^2 \ln L(\pi; X)}{\partial \pi^2} \right]^{-1} \left[ \frac{\partial \ln L(\pi; X)}{\partial \pi} \right],
\]

where $\alpha$ is a backtracking factor initially set to 1, which may be repeatedly reduced (by a factor of 2, for instance) in those cases when the parameter update does not strictly decrease minus the log-likelihood.
A Simulation Study

A simulation study is presented below to test the performance of this approximate method for recovering the model parameters. This study is also intended to provide some information on the behavior of the INFIT and OUTFIT statistics (Wright & Masters, 1982) for detecting ill-fitting items.

Fifty simulated data sets were generated for each of the 45 conditions resulting from factorially combining five levels of sample size (100, 200, 300, 500, and 1,000 observations) with nine levels of test length (10, 15, 20, 25, 30, 35, 40, 45, and 50 items). For each replication, subject parameters where drawn from a $N(0, 1)$ distribution, item difficulties were equally spaced between $-3$ and $+3$, and dispersion parameters were drawn at random from a uniform distribution in the $[0; 3]$ interval. Note that this gives a new set of person parameters and a new set of dispersion parameters in each replication, whereas the set of item difficulty parameters remains the same. This within-condition variability was introduced to allow for meaningful between-condition comparisons, as far as accuracy and fit statistics are concerned.

An R program was written to obtain JML estimates of the BRM parameters (R Development Core Team, 2003). In every condition, four statistics were computed:

1. correlations between true and estimated parameter values,
2. root mean squared error (RMSE): $\text{RMSE}_{\pi_i} = \left[ \frac{1}{N} \sum_{i=1}^{N} (\hat{\pi}_i - \pi_i)^2 \right]^{1/2}$
   and $\text{RMSE}_{\pi_j} = \left[ \frac{1}{p} \sum_{j=1}^{p} (\hat{\pi}_j - \pi_j)^2 \right]^{1/2}$,
3. mean absolute deviation (MAD): $\text{MAD}_{\pi_i} = \frac{1}{N} \sum_{i=1}^{N} |\hat{\pi}_i - \pi_i|$ 
   and $\text{MAD}_{\pi_j} = \frac{1}{p} \sum_{j=1}^{p} |\hat{\pi}_j - \pi_j|$,
4. the observed to true variance ratio (VR): $\text{VR} = \frac{\sigma^2_{\hat{\pi}}}{\sigma^2_{\pi}}$.

Results of this simulation experiment are plotted in Figures 4, 5, and 6 for $\theta_i$, $\delta_j$, and $\tau_j$ parameters, respectively. In all cases, the algorithm converged very quickly: The mean computation time with the largest data sets ($N = 1,000, p = 50$) was $2'46''$ (CPU 2 GHz on a Linux machine) in purely interpreted code.

To evaluate the influence of test length and sample size on quality of estimation, 12 two-way ANOVAs (one for each of the four statistics and the three parameter sets) with interaction were performed, taking test length and sample size as between-cases factors. Correlation ratios were computed and are reported in Table 1. $F$ values and significance levels are not of much interest here, given the nature of the “dependent variables” (which cannot reasonably be assumed to be normally distributed). The significance levels are reported with asterisks in Table 1, and only the correlation ratios are commented on here. They are denoted as $\eta^2_p$ (size effect of sample size), $\eta^2_I$ (size effect of test length), and $\eta^2_{I \times P}$ (size effect of sample size by test length interaction) in what follows.

Estimation of $\theta_i$ Parameters

Correlations between true and estimated values of person parameters were never lower than .95 over all conditions (see Figure 4), even with as few as 100 persons and 10 items. Correlation values, RMSE, and MAD were clearly improved with increasing test lengths (.915 < $\eta^2_I$ < .953). There seemed to be very little effect, if any, of sample size on these three measures of quality of estimation for person parameters (.000 < $\eta^2_p$ < .001). Sample size, however, seemed to affect the
variance ratio a little more, although the effect remained moderate ($\eta_p^2 = .003$). Although joint maximum likelihood estimates are notoriously biased, such that the variance ratio is not expected to fall to 1, it is striking that this ratio falls below 1.1 with as few as 15 items and 100 persons. On the whole, it can be observed that, provided the model is correct, estimating the $\theta_i$ parameters on 100 persons and 20 items results in an expected correlation between true and estimated parameters greater than .97, a mean absolute deviation lower than .2, and a variance ratio lower than 1.1.

Estimation of $\delta_j$ Parameters

The quality of estimation appears to be even better for difficulty parameters. This is of course expected, given that sample size was always greater than test length in the simulations. But the recovery of difficulty parameters appears to be impressively good also by comparison with what is usually observed in categorical item response models, at least for small data sets.

The average correlation between true and estimated difficulty parameters is very close to 1, whatever the number of items (see Figure 5). The quality of estimation was mainly affected by sample size, although less strongly than person parameters depended on test length ($\.741 < \eta_p^2 < .801$), as far as average correlations, RMSE, and MAD are concerned. Again, variance ratios display a more complex picture and seemed to be mainly affected by test length ($\eta_I^2 = .16$) and moderately by sample size ($\eta_p^2 = .031$) and their interaction ($\eta_{I \times p}^2 = .017$).

On the whole, it is noted that, provided the model is correct, estimating the parameters on as few as 200 persons and 10 items results in a correlation between true and estimated difficulty parameters very close to 1, a MAD lower than .1, and a variance ratio lower than 1.05.

### Table 1

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Sample Size</th>
<th>Test Length</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.001**</td>
<td>.915***</td>
<td>.002</td>
</tr>
<tr>
<td>RMSE</td>
<td>.000***</td>
<td>.953***</td>
<td>.001*</td>
</tr>
<tr>
<td>MAD</td>
<td>.001***</td>
<td>.949***</td>
<td>.001**</td>
</tr>
<tr>
<td>Variance ratio</td>
<td>.003**</td>
<td>.582***</td>
<td>.008</td>
</tr>
<tr>
<td>Difficulty parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.741***</td>
<td>.020***</td>
<td>.015***</td>
</tr>
<tr>
<td>RMSE</td>
<td>.800***</td>
<td>.003***</td>
<td>.005**</td>
</tr>
<tr>
<td>MAD</td>
<td>.801***</td>
<td>.006***</td>
<td>.006***</td>
</tr>
<tr>
<td>Variance ratio</td>
<td>.031***</td>
<td>.161***</td>
<td>.017*</td>
</tr>
<tr>
<td>Dispersion parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.615***</td>
<td>.010***</td>
<td>.012***</td>
</tr>
<tr>
<td>RMSE</td>
<td>.550***</td>
<td>.277***</td>
<td>.002</td>
</tr>
<tr>
<td>MAD</td>
<td>.517***</td>
<td>.309***</td>
<td>.003</td>
</tr>
<tr>
<td>Variance ratio</td>
<td>.073***</td>
<td>.196***</td>
<td>.008</td>
</tr>
</tbody>
</table>

*Note. RMSE = root mean squared error; MAD = mean absolute deviation. *
*p < .05; **p < .01; ***p < .001.
Estimation of $\tau_j$ Parameters

Estimation of dispersion parameters, as assessed by correlations, RMSE, and MAD, was mainly affected by sample size ($0.517 < \eta^2_P < 0.615$) but also by test length, to some extent ($0.01 < \eta^2_I < 0.309$). By contrast, the variance ratio was mainly determined by test length ($\eta^2_I = 0.196$) and, to a lesser extent, by sample size ($\eta^2_P = 0.073$). The 100-person condition seemed to perform badly on all four measures. On the whole, the condition involving 200 persons and 20 items gave a correlation between true and estimated dispersion parameters of about .98, an average MAD lower than .2, and a variance ratio lower than 1.2. This could be considered as the minimal acceptable size for a data set, as far as quality of estimation of dispersion parameters is concerned.

Figure 4
Estimation of Respondents’ Abilities

Note. Statistics are averaged over 50 replications.
Fit Statistics

In every condition, adapted versions of the OUTFIT and INFIT statistics (Wright & Masters, 1982) were computed for all items. Define

$$\hat{\mu}_{ij} = \frac{1}{1 + \exp(-\hat{\theta}_i + \delta_j)}$$
Figure 6
Estimation of Dispersion Parameters

Note. Statistics are averaged over 50 replications.

and

$$S_{ij}^2 = \frac{\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})}{1 + \hat{m}_{ij} + \hat{n}_{ij}}$$

$$= \frac{\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})}{1 + \exp(\hat{\theta}_i - \hat{\delta}_j + \hat{\tau}_j) + \exp(\hat{\delta}_j - \hat{\theta}_i + \hat{\tau}_j)},$$

where $\hat{\theta}_i$ and $\hat{\delta}_j$ are estimated values of the corresponding unknown parameters, and let

$$z_{ij} = \frac{x_{ij} - \hat{\mu}_{ij}}{S_{ij}}.$$
The item OUTFIT and INFIT $u_j$ and $v_j$ statistics are given by

$$u_j = \frac{1}{N} \sum_{i=1}^{N} z_{ij}^2$$

and

$$v_j = \frac{\sum_{i=1}^{N} S_{ij}^2 z_{ij}^2}{\sum_{i=1}^{N} S_{ij}^2} = \frac{\sum_{i=1}^{N} (x_{ij} - \hat{\mu}_{ij})^2}{\sum_{i=1}^{N} S_{ij}^2}.$$

The INFIT statistic may be viewed as an information-weighted measure of discrepancy and will thus be less affected by extreme respondents. Similarly, person OUTFIT and INFIT statistics may be computed by summing and averaging over items in the expressions above. Note that, by contrast with binomial or multinomial models, the response variance (equation (15)) under the BRM-2 does not simply depend on the expectation and heavily depends on the $\tau_j$ parameters. Both under- and overdispersion are thus important to detect, and both low and high values of INFIT and OUTFIT may indicate misfit.

In each condition and replication, these statistics were computed, and the maximal and minimal values observed over all items (or, respectively, persons) were taken as global indices of misfit for the whole set of items (or, respectively, persons). The empirical distributions of maximal and minimal INFIT and OUTFIT values, for both items and persons, are summarized in Table 2: Quantiles of order 0% (minimum), 5%, 10%, 90%, 95%, and 100% (maximum) of these extreme distributions are reported.

Over all samples and replications, the minimal item OUTFIT values were between .599 and .984, with an approximate 90% confidence interval being [.70; .93]. The maximal item OUTFIT values were between 1 and 2.66, with an approximate 90% confidence interval being [1.08; 1.54]. A practical method for assessing whether a given set of items forms a good BRM scale is thus to check that item OUTFIT values lie in the [.7; 1.5] interval. If one wants to base judgment on the INFIT information-weighted index, a similar reasoning leads to using the [.7; 1.3] interval instead.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Minimum</th>
<th>5th Percentile</th>
<th>10th Percentile</th>
<th>90th Percentile</th>
<th>95th Percentile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item fit statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimal OUTFIT</td>
<td>0.599</td>
<td>0.708</td>
<td>0.746</td>
<td>0.923</td>
<td>0.936</td>
<td>0.984</td>
</tr>
<tr>
<td>Maximal OUTFIT</td>
<td>1.009</td>
<td>1.076</td>
<td>1.093</td>
<td>1.425</td>
<td>1.549</td>
<td>2.656</td>
</tr>
<tr>
<td>Minimal INFIT</td>
<td>0.509</td>
<td>0.721</td>
<td>0.758</td>
<td>0.931</td>
<td>0.941</td>
<td>0.983</td>
</tr>
<tr>
<td>Maximal INFIT</td>
<td>0.988</td>
<td>1.045</td>
<td>1.057</td>
<td>1.251</td>
<td>1.292</td>
<td>1.888</td>
</tr>
<tr>
<td>Person fit statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimal OUTFIT</td>
<td>0.059</td>
<td>0.149</td>
<td>0.178</td>
<td>0.468</td>
<td>0.494</td>
<td>0.596</td>
</tr>
<tr>
<td>Maximal OUTFIT</td>
<td>1.551</td>
<td>2.077</td>
<td>2.293</td>
<td>6.483</td>
<td>7.821</td>
<td>25.603</td>
</tr>
<tr>
<td>Minimal INFIT</td>
<td>0.009</td>
<td>0.101</td>
<td>0.139</td>
<td>0.472</td>
<td>0.503</td>
<td>0.62</td>
</tr>
<tr>
<td>Maximal INFIT</td>
<td>1.472</td>
<td>1.706</td>
<td>1.776</td>
<td>3.274</td>
<td>3.67</td>
<td>6.031</td>
</tr>
</tbody>
</table>
Similarly, one may want to detect ill-fitting persons by using the criterion intervals [.15; 7.8] for person OUTFIT and [.1; 3.6] for person INFIT. This is just a crude indication, however, because these results depend on the parameter values that were used in the simulation and do not take into account possible sample size and test length effects.

**Application on Real Data**

An application of the BRM on mood data collected on 464 persons is illustrated. Subjects were undergraduate and graduate students in psychology. They were asked to report on 121-mm continuous response scales how much they agreed with 19 negative and positive mood items, from 0% to 100% agreement. Note that, within a reasonable range (75-125 mm), the choice for the response segment length appears to have a negligible effect, if any, on continuous ratings (Hubbard, Little, & Allen, 1989).

This item set was constructed by adding 16 mood adjectives to the 14 items of the Hospital Anxiety-Depression scale (Zigmond & Snaith, 1983). From this first pool, 11 items that specifically concerned a general activity level (aroused, stimulated, sluggish, etc.) were removed, to keep in the final set those items that clearly addressed positive versus negative mood states.

Twenty-two respondents had missing values and were discarded from the analysis. The responses were measured up to the half-millimeter as distances from the left end of the scale to the respondent’s mark. Then they were rescaled to [0; 1].

Note that, by contrast with simulated data, it may happen that respondents mark the very end of the visual scale, such that their response is recoded as exactly 0 or 1. Of course, the log-likelihood cannot be computed with such values, and 0 and 1 response values were arbitrarily changed to .0001 and .9999, respectively. It may appear necessary in practice to notify respondents that such extreme responses are psychologically unrealistic in most cases.

Respondents were instructed not to measure the response graphic line and to give their answers “as if they were tuning a radio receiver.” This instruction was intended to prevent respondents from “likertizing” their responses: Ferrando (2003) showed that this may happen with such response scales and that proper instructions are efficient in limiting this phenomenon.

If the hypothesis of a latent bipolar negative to positive mood continuum is correct, it is expected that the response functions will decrease for the negative items and increase for the positive ones. Note that the model definition in equation (14) constrains the expected response function to be monotonically increasing. The bipolar structure of the scale is dealt with by fixing to −1 the implicit slope for the negative mood items. Equivalently, response variables on the negative mood items may simply be replaced by their complement to 1 (Klinkenberg, 2001). The corresponding back-transformation is then performed on the response functions at the end of the estimation process.

Both BRM-1 and BRM-2 were fitted to the data, and the Akaike information criterion (AIC) was computed for both of them. Their respective AIC values were −3864.2 and −7086.6, so the BRM-2 is clearly better (even though the item orderings were almost perfectly the same).

Estimates of \( \delta_j \) and \( \tau_j \) BRM-2 parameters are reported in Table 3, with their approximate standard errors and the corresponding item INFIT and OUTFIT statistics.

The negative mood items clearly scale from lighter (“worrying thoughts”) to heavier (“frightened”) negative feelings, and positive mood items range from weaker (“I can laugh and see the funny side of things”) to stronger positive emotions (“euphoric”). The OUTFIT and INFIT statistics are very close to 1 and remain in the acceptable ranges discussed in the previous section. This item set is thus acceptable as an (illustrative) mood scale. A plot of INFIT and OUTFIT statistics for items
and persons is also displayed in Figure 7, which helps detect ill-fitting persons. Only 1 person out of 442 lies outside the [.1; 3.6] INFIT decision interval, and 2 of them lie outside the [.15; 7.8] OUTFIT decision interval, which represents acceptable outlier proportions of 0.2% and 0.5%.

Although all items look acceptable in terms of model fit, they are not all equally informative. If items are to be selected on the basis of precision, in the sense defined above, then one is free to fix any acceptable confidence interval width between 0 and 1 (say, 0.80) and discard items for which response variabilities are too large. From Table 3, it can be seen that this would lead to discarding six items, with four of them corresponding to extreme, probably infrequently experienced feelings (frightened, panic, elated, and euphoric).

The corresponding expected rating curves are plotted in Figures 8 and 9, along with the real data. Both subsets of negative and positive mood items appear on two separate graphs for clarity but of course have been scaled in the same BRM analysis.

Note that the black curves represent response conditional expectations and that, because the beta is not symmetric in general, these will appear at some distance of the distribution mode, for a given value of \( \theta_i \). This is why, at first glance, it seems that these curves do not go “through the middle” of the data, as would be the case with a Gaussian model or any other symmetric model.

Table 3
Two-Parameter Beta Response Model (BRM-2) Parameter Estimates for the Mood Data

<table>
<thead>
<tr>
<th>Item</th>
<th>Location ((\delta_j))</th>
<th>Approximate Standard Error</th>
<th>Dispersion ((\tau_j))</th>
<th>Approximate Standard Error</th>
<th>Imprecision(^a)</th>
<th>OUTFIT</th>
<th>INFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negative mood</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frightened</td>
<td>-0.942</td>
<td>0.055</td>
<td>0.382</td>
<td>0.125</td>
<td>0.858*</td>
<td>1.003</td>
<td>1.066</td>
</tr>
<tr>
<td>Panic</td>
<td>-0.649</td>
<td>0.055</td>
<td>0.166</td>
<td>0.115</td>
<td>0.883*</td>
<td>1.074</td>
<td>1.076</td>
</tr>
<tr>
<td>Unhappy</td>
<td>-0.603</td>
<td>0.042</td>
<td>1.799</td>
<td>0.121</td>
<td>0.674</td>
<td>0.856</td>
<td>0.824</td>
</tr>
<tr>
<td>Butterflies</td>
<td>-0.398</td>
<td>0.051</td>
<td>0.615</td>
<td>0.115</td>
<td>0.830*</td>
<td>0.829</td>
<td>0.885</td>
</tr>
<tr>
<td>Sad</td>
<td>-0.379</td>
<td>0.039</td>
<td>2.092</td>
<td>0.123</td>
<td>0.635</td>
<td>0.829</td>
<td>0.865</td>
</tr>
<tr>
<td>Tense—wound up</td>
<td>0.523</td>
<td>0.038</td>
<td>1.988</td>
<td>0.123</td>
<td>0.649</td>
<td>1.044</td>
<td>1.042</td>
</tr>
<tr>
<td>Annoyed</td>
<td>0.649</td>
<td>0.039</td>
<td>1.899</td>
<td>0.123</td>
<td>0.661</td>
<td>0.878</td>
<td>0.892</td>
</tr>
<tr>
<td>Nervous</td>
<td>0.781</td>
<td>0.043</td>
<td>1.439</td>
<td>0.120</td>
<td>0.722</td>
<td>1.010</td>
<td>0.986</td>
</tr>
<tr>
<td>Worrying thoughts</td>
<td>0.907</td>
<td>0.046</td>
<td>1.057</td>
<td>0.118</td>
<td>0.773</td>
<td>0.958</td>
<td>0.980</td>
</tr>
<tr>
<td><strong>Positive mood</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laugh—see funny side</td>
<td>-0.736</td>
<td>0.045</td>
<td>1.465</td>
<td>0.118</td>
<td>0.719</td>
<td>1.013</td>
<td>1.070</td>
</tr>
<tr>
<td>Enjoy things</td>
<td>-0.728</td>
<td>0.043</td>
<td>1.760</td>
<td>0.120</td>
<td>0.679</td>
<td>1.019</td>
<td>0.997</td>
</tr>
<tr>
<td>Enjoy a good book</td>
<td>-0.612</td>
<td>0.052</td>
<td>0.510</td>
<td>0.113</td>
<td>0.843*</td>
<td>1.022</td>
<td>1.015</td>
</tr>
<tr>
<td>Pleased</td>
<td>-0.159</td>
<td>0.026</td>
<td>3.865</td>
<td>0.129</td>
<td>0.429</td>
<td>0.902</td>
<td>0.894</td>
</tr>
<tr>
<td>Happy</td>
<td>-0.132</td>
<td>0.026</td>
<td>3.839</td>
<td>0.129</td>
<td>0.431</td>
<td>0.860</td>
<td>0.875</td>
</tr>
<tr>
<td>Cheerful</td>
<td>-0.032</td>
<td>0.029</td>
<td>3.363</td>
<td>0.128</td>
<td>0.481</td>
<td>0.972</td>
<td>0.927</td>
</tr>
<tr>
<td>Look for enjoyment</td>
<td>0.188</td>
<td>0.039</td>
<td>1.857</td>
<td>0.122</td>
<td>0.666</td>
<td>0.972</td>
<td>0.923</td>
</tr>
<tr>
<td>Content</td>
<td>0.319</td>
<td>0.034</td>
<td>2.596</td>
<td>0.126</td>
<td>0.571</td>
<td>0.940</td>
<td>0.928</td>
</tr>
<tr>
<td>Elated</td>
<td>0.876</td>
<td>0.048</td>
<td>0.823</td>
<td>0.117</td>
<td>0.804*</td>
<td>1.320</td>
<td>0.971</td>
</tr>
<tr>
<td>Euphoric</td>
<td>1.127</td>
<td>0.049</td>
<td>0.792</td>
<td>0.116</td>
<td>0.808*</td>
<td>1.423</td>
<td>1.054</td>
</tr>
</tbody>
</table>

a. For each item \( j \), the imprecision values are the widths of response confidence intervals at \( \theta_i = \delta_j \). Values above 0.80 are marked by asterisks, for illustration purposes (see text).
Modal response curves are plotted in gray to illustrate this point. This is a property of the model, as is clear from Figure 2 and by comparing equations (17) and (16).

This possibility of a visual inspection is a clear advantage of continuous over categorical data. It strongly helps the analyst to get a sense of what is going on in the data, particularly when there is misfit.

**Discussion**

In this article, a very simple item response model was proposed to deal with continuous bounded responses. A beta model for the response density is derived from the hypothesis of an interpolation response mechanism, which has some connections with well-known choice models.
models. In particular, the approach of deriving a response density from a latent assessment mechanism is similar to the one adopted in discrete choice models, which derive a logistic choice model from a hypothesis on the (Gumbel) distribution of latent values granted to several alternatives (McFadden, 1974).

Expressing the beta distribution parameters as simple monotonic functions of person and item parameters allowed the formulation of an expected response function that has a familiar logistic shape. The interpretation of item and person parameters is then straightforward: The item location is

Note. The black curve represents the expected rating function, and the gray one represents the modal response function. Modal values are not defined outside [0; 1]—hence, the floor and ceiling effects.
Figure 9
Expected Rating Curves for the Positive Mood Items

Note. The black curve represents the expected rating function, and the gray one represents the modal response function. Modal values are not defined outside [0; 1]—hence, the floor and ceiling effects.
that value of attitude for which a person is expected to give a .5 rating to that item on a continuous response scale.

The number of parameters to be estimated in the BRM is lower than the one commonly required with categorical models. Under the BRM-1, only $N + p$ parameters are needed, which is the same as in a Rasch model for binary data, but with all the advantages of a continuous and presumably richer measurement. The number of parameters to be estimated under the BRM-2 ($N + 2p$) still favorably compares with the one needed under a polytomous Rasch or partial credit model, which depends on the number of item response categories.

The quality of parameter estimation by joint maximum likelihood was shown to be excellent with as few as 200 respondents and 20 items, which also contrasts with the usual requirements of categorical cumulative models. Nevertheless, joint maximum likelihood estimates are notoriously biased, and more work is needed to see whether other numerical methods could perform even better. Unfortunately, as model parameters appear as exponents in the model, no sufficient statistic exists that would allow the maximization of a conditional log-likelihood. A marginal maximum likelihood approach is another possibility that still needs to be explored.

The model proposed in this article for continuous bounded responses is not thought to be better than previous proposals on all aspects. In particular, no sufficient statistics for item and person parameters are defined, and this may be considered a serious problem from a Rasch scaling perspective. However, the derivation from a hypothetical response mechanism and a familiar closed-form expression for the expected response functions are attractive features.

Appendix

Derivatives of the Log-Likelihood Function

First Derivatives of the Log-Likelihood Function

The log-likelihood function for the BRM-2 is

$$
\ln L(\Theta; X) = \sum_{j=1}^{N} \sum_{j=1}^{p} \left\{ \ln \Gamma(m_{ij} + n_{ij}) - \ln \Gamma(m_{ij}) - \ln \Gamma(n_{ij}) + (m_{ij} - 1) \ln x_{ij} + (n_{ij} - 1) \ln (1 - x_{ij}) \right\}.
$$

For any person-specific parameter $\pi_i$, the first derivative is given by

$$
\frac{\partial \ln L(\Theta; X)}{\partial \pi_i} = \sum_{j=1}^{p} \left\{ \Psi(m_{ij} + n_{ij}) \left[ \frac{\partial m_{ij}}{\partial \pi_i} + \frac{\partial n_{ij}}{\partial \pi_i} \right] - \Psi(m_{ij}) \frac{\partial m_{ij}}{\partial \pi_i} - \Psi(n_{ij}) \frac{\partial n_{ij}}{\partial \pi_i} \right. \\
+ \ln x_{ij} \frac{\partial m_{ij}}{\partial \pi_i} + \ln (1 - x_{ij}) \frac{\partial n_{ij}}{\partial \pi_i} \right\}
$$

$$
= \sum_{j=1}^{p} \left\{ \frac{\partial m_{ij}}{\partial \pi_i} [\Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij}] \\
+ \frac{\partial n_{ij}}{\partial \pi_i} [\Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln (1 - x_{ij})] \right\},
$$

where $\Psi(x) = \frac{\partial \ln \Gamma(x)}{\partial x}$ is the digamma function.
Analogously, for any item-specific parameter $\pi_j$, the first derivative is given by

$$\frac{\partial \ln L(\Theta; X)}{\partial \pi_j} = \sum_{i=1}^{N} \left\{ \frac{\partial m_{ij}}{\partial \pi_j} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{\partial n_{ij}}{\partial \pi_j} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] \right\}.$$  \hspace{1cm} (A3)

From equations (11) and (12),

$$\begin{align*}
\frac{\partial m_{ij}}{\partial \theta_i} &= \frac{m_{ij}}{2}, \\
\frac{\partial m_{ij}}{\partial \delta_j} &= -\frac{m_{ij}}{2}, \\
\frac{\partial m_{ij}}{\partial \tau_j} &= \frac{m_{ij}}{2}, \\
\frac{\partial n_{ij}}{\partial \theta_i} &= -\frac{n_{ij}}{2}, \\
\frac{\partial n_{ij}}{\partial \delta_j} &= \frac{n_{ij}}{2}, \\
\frac{\partial n_{ij}}{\partial \tau_j} &= \frac{n_{ij}}{2}.
\end{align*}$$  \hspace{1cm} (A4)

Replacing these expressions in equations (A2) and (A3) leads to equations (20), (21), and (22).

**Second Derivatives of the Log-Likelihood Function**

The second derivative with respect to any person-specific parameter $\pi_i$ is then given by

$$\frac{\partial^2 \ln L(\Theta; X)}{\partial \pi_i^2} = \sum_{j=1}^{p} \left\{ \frac{\partial^2 m_{ij}}{\partial \pi_i^2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{\partial^2 n_{ij}}{\partial \pi_i^2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] + \frac{\partial^2 m_{ij}}{\partial \pi_i \partial \pi_j} \left[ \frac{\partial m_{ij}}{\partial \pi_i} + \frac{\partial m_{ij}}{\partial \pi_j} \right] - \Psi'(m_{ij}) \frac{\partial m_{ij}}{\partial \pi_i} \frac{\partial m_{ij}}{\partial \pi_j} \right\}$$

$$= \sum_{j=1}^{p} \left\{ \frac{\partial^2 m_{ij}}{\partial \pi_i^2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln x_{ij} \right] + \frac{\partial^2 n_{ij}}{\partial \pi_i^2} \left[ \Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - x_{ij}) \right] + \Psi'(m_{ij} + n_{ij}) \left[ \frac{\partial m_{ij}}{\partial \pi_i} + \frac{\partial n_{ij}}{\partial \pi_i} \right]^2 - \Psi'(m_{ij}) \left( \frac{\partial m_{ij}}{\partial \pi_i} \right)^2 \right\}.$$  \hspace{1cm} (A5)

where $\Psi'(x) = \partial \Psi(x) / \partial x$ is the trigamma function.

Analogously, the second derivative of the log-likelihood function with respect to any item-specific parameter $\pi_j$ is given by
From equations (11) and (12),
\[
\frac{\partial^2 m_{ij}}{\partial \theta_i^2} = \frac{m_{ij}}{4}, \quad \frac{\partial^2 m_{ij}}{\partial \delta_j^2} = \frac{m_{ij}}{4}, \quad \frac{\partial^2 m_{ij}}{\partial \tau_j^2} = \frac{m_{ij}}{4},
\]
\[
\frac{\partial^2 n_{ij}}{\partial \theta_i^2} = \frac{n_{ij}}{4}, \quad \frac{\partial^2 n_{ij}}{\partial \delta_j^2} = \frac{n_{ij}}{4}, \quad \frac{\partial^2 n_{ij}}{\partial \tau_j^2} = \frac{n_{ij}}{4},
\]
(A7)

which leads to equations (23), (24), and (25).

**Expected Values of the Second Derivatives**

Expectations of the second derivatives are useful to get approximate standard errors and for estimating the parameters in a Fisher scoring approach. Considering the response variable \(X_{ij} \sim Be(m_{ij}, n_{ij})\) and denoting by \(L_{ij}\) the model likelihood function, it is known from likelihood theory that

\[
E \left[ \frac{\partial \ln L_{ij}}{\partial m_{ij}} \right] = 0,
\]
\[
E \left[ \frac{\partial \ln L_{ij}}{\partial n_{ij}} \right] = 0,
\]

from which the following is obtained:

\[
E[\Psi(m_{ij} + n_{ij}) - \Psi(m_{ij}) + \ln X_{ij}] = 0,
\]
\[
E[\Psi(m_{ij} + n_{ij}) - \Psi(n_{ij}) + \ln(1 - X_{ij})] = 0,
\]
or

\[
E[\ln X_{ij}] = \Psi(m_{ij}) - \Psi(m_{ij} + n_{ij}),
\]
\[
E[\ln(1 - X_{ij})] = \Psi(n_{ij}) - \Psi(m_{ij} + n_{ij}),
\]

where the expectation is taken over \(X_{ij}\). Finally, taking the expectation on both sides of equations (33) and (34), somewhat simpler expressions are obtained:

\[
E \left[ \frac{\partial^2 \ln L(\Theta; \mathbf{X})}{\partial \pi_j^2} \right] = \sum_{j=1}^{N} \left\{ \Psi'(m_{ij} + n_{ij}) \left[ \frac{\partial m_{ij}}{\partial \pi_i} + \frac{\partial n_{ij}}{\partial \tau_i} \right]^2 - \Psi'(m_{ij}) \left( \frac{\partial m_{ij}}{\partial \pi_i} \right)^2 
- \Psi'(n_{ij}) \left( \frac{\partial n_{ij}}{\partial \tau_i} \right)^2 \right\},
\]
Replacing the first and second derivatives of \( m_{ij} \) and \( n_{ij} \) with respect to \( \theta_i \), \( \delta_j \), and \( \tau_j \) parameters by their expressions in equations (A4) and (A7), equations (26), (27), and (28) are finally obtained.

References


Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity


**Acknowledgment**

The authors thank Catherine Trottier, University of Montpellier III, France, for helpful comments on this work.

**Author’s Address**

Address correspondence to Yvonnick Noel, Département de Psychologie, Université de Rennes 2, Place du Recteur Henri Le Moal, C S 24307, F-35043 Rennes Cedex, France; e-mail: yvonnick.noel@uhb.fr. An R program for estimating the parameters in the BRM is available upon request from the first author.