Measurement and Research Department Reports 2000-3

IRT MODELS FOR SUBJECTIVE WEIGHTS OF OPTIONS OF MULTIPLE CHOICE QUESTIONS

H.H.F.M. Verstralen
N.D. Verhelst

Cito

Measurement and Research Department Reports 2000-3

IRT MODELS FOR SUBJECTIVE WEIGHTS OF OPTIONS OF MULTIPLE CHOICE QUESTIONS

H.H.F.M. Verstralen
N.D. Verhelst

Cito
Arnhem, 2000

Abstract

From earlier investigations it was found that the information from Multiple Choice (MC) questions could be increased about four fold by having the subject indicate the subset of options that he is unable to expose as false. In the present models this approach is generalized by having the subject distribute a number of 'taws' over the options, or draw a line after the options, such that the number of taws given to an option, or the line length reflects its subjective degree of correctness. It appears that even with values of the relevant parameters that seem modest, the information relative to binary scoring still is in excess of two. This means that with less than half the test length the same accuracy or reliability can be obtained as with binary scoring. If a few main fallacies can be reflected in the distractors of the items, the model can be applied to identify subjects with one of these fallacies.

Introduction

Several statistical models exist to describe the choice behavior of a student in the selection of an alternative of a *multiple choice* (MC) item. Among them Thissen and Steinberg (1984, 1997), Verstralen (1997), Verhelst and Verstralen, (in press), Verstralen and Verhelst (in press), Jansen and De Boeck (1998). The present approach can be viewed as a continuation of Verstralen (1997) and Verstralen and Verhelst (in press). Their approach is based on the assumption that a subject first selects a subset with 'possibly correct' alternatives, and subsequently randomly selects one alternative from this not observed subset. An important extra assumption in this model is that the subset always contains the correct alternative, and further, to convey information about the latent variable ϑ, it is assumed that the expected size of the subset shrinks with ϑ. As a side result from this approach it was found that the amount of information about ϑ would be substantially larger (in the order of four to five times larger) if the subset would have been observed. However, contrary to latent subsets, observed subsets are expected to violate the assumption that they always contain the correct option. Therefore, to tap the information about ϑ from the subject's evaluation of all the alternatives of an MC item, another data format with an accompanying model is needed. An obvious generalization of the subset of options, is the fuzzy set of options. The fuzzy set approach implies that the subject gives weights to the options that reflects the subjective correctness of the options. This type of data was already studied by Dirkzwager (1975,1996) in a more classical context. Here an IRT approach is developed for these data, where the weights are restricted to be integer. In the first model, the *Dirichlet Multinomial* (DM) model, it is assumed that a fixed number of taws is distributed over the options of an item. The fixed total number of taws causes the number of taws per option to be conditionally interdependent between options, given the ability of the subject. In case this may be difficult to accomplish by the subject, in the *Beta Binomial* (BB) model the maximum number of taws is fixed per option, not the total number of taws per item. This makes the weights over the options conditionally independent. In the next two sections the DM and BB models are treated, and the gain of information for both models is discussed in section 4. From section 5 on further details of the models are provided, and an estimation procedure is developed, along with a procedure to evaluate the fit of the model. Thereafter we discuss a diagnostic application of the models which allows to identify subjects with a common fallacy that can be represented in the distractors of the items in a test. Finally, the theory is applied to a generated data set among other things to check whether specific model violations are detected.

The Dirichlet-Multinomial model

A subject v is asked to distribute N taws over the alternatives of an item such that the distribution of taws reflects his subjective probability of correctness of the alternatives. Consequently, the data consist of vectors $\underline{n}_{vi} = (n_{vi0}, ..., n_{vi(J-1)})$, with n_{vij} the number of taws put by subject v on alternative j of item i. We first focus on the response of a single subject to one item. Therefore, the subscripts v and i are dropped.

Let $\underline{\pi} = \underline{\pi}(\vartheta) = (\pi_0(\vartheta), ..., \pi_{J-1}(\vartheta))$ be a model that describes the probability to put a taw on the options of a MC question with J alternatives as a function of the latent variable ϑ, where the subscript 0 indicates the correct option. Assume that

$$\pi_0' = \frac{\partial}{\partial\vartheta}\pi_0(\vartheta) > 0 \tag{1}$$

for all ϑ. Except for this assumption the function $\underline{\pi}$ will not be specified further in this section, and the next.

One could, of course, model the distribution of $\underline{n}$ at ϑ as the multinomial

$$p(\underline{n}; \vartheta) = \binom{N}{\underline{n}} \prod \pi_j^{n_j}, \tag{2}$$

with $\sum n_j = N$. However, this model is flawed, because it implies that information about ϑ increases without bound with the number of taws N. Indeed, for $\lim_{N\to\infty}$ the probabilities $\underline{\pi}(\vartheta)$, and, therefore ϑ, can be observed without error. To obtain a more plausible model we assume that the subject is not an acute perceiver of his own $\underline{\pi}$, and that his uncertainty about $\underline{\pi}$ is independent of N.

Let μ be a positive acuity parameter, and assume that the observed distribution of taws is realized in two stages. First a vector of probabilities $\underline{q} = \underline{q}(\vartheta)$ is drawn from a Dirichlet distribution

$$\underline{q}(\vartheta) \sim D(\underline{\pi}) = A(\mu\underline{\pi}) \prod q_j^{\mu\pi_j - 1}, \tag{3}$$

with the proportionality constant

$$A(\underline{x}) = \frac{\Gamma\left(\sum x_j\right)}{\prod \Gamma(x_j)}. \tag{4}$$

With this distribution, indexed by ϑ, $\mathcal{E}_\vartheta q_j = \pi_j$, and $\mathcal{V}ar_\vartheta(q_j) = \pi_j(1 - \pi_j)/(\mu + 1)$. From the mean and variance of q_j it follows that the higher μ, the closer q_j is drawn around π_j. Second, conditional on $\underline{q}$, $\underline{n}$ is drawn from the multinomial distribution

$$p(\underline{n}|\underline{q}) = \binom{N}{\underline{n}} \prod q_j^{n_j}. \tag{5}$$

The marginal distribution of $\underline{n}$ is then given by the *Dirichlet-Multinomial* (DM) distribution (Johnson & Kotz, 1969, sec. 11.8)

$$\begin{aligned} p(\underline{n};\vartheta) &= \binom{N}{\underline{n}} A(\mu\underline{\pi}) \int \prod q_j^{\mu\pi_j + n_j - 1} \underline{dq} \\ &= \binom{N}{\underline{n}} \frac{A(\mu\underline{\pi})}{A(\mu\underline{\pi} + \underline{n})}. \end{aligned} \tag{6}$$

The mean and (co)variance of DM distributed $\underline{n}$ are given by

$$\begin{aligned} \mathcal{E}n_j &= N\pi_j, \\ \mathcal{V}ar(n_j; N, \mu, \underline{\pi}) &= \frac{(\mu + N)}{(\mu + 1)} N\pi_j(1 - \pi_j). \\ \mathcal{C}ov(n_j, n_k; N, \mu, \underline{\pi}) &= -\frac{(\mu + N)}{(\mu + 1)} N\pi_j\pi_k. \end{aligned} \tag{7}$$

For $N = 1$ the variance is independent of μ. Using Formulas (4) and the second part of (6) it readily is found that for $N = 1$ the DM distribution does not depend on μ.

For $\mu \to 0$ the limiting variance is $N^2\pi_j(1 - \pi_j) = N^2\mathcal{V}ar(n_j; 1, ., \underline{\pi})$. This implies that for $\mu \to 0$, the subject first behaves as if he has just one taw to distribute as the multinomial $\underline{\pi}$, and puts the rest $N-1$ taws on the same option where he has put his first taw. Alternatively, one could say that for $\mu \longrightarrow 0$ the subject selects as his vector $\underline{q}$ the vector I_j with probability π_j, where I_j is row $j+1$ from the identity matrix I_J. And then puts all his N taws on option j with probability 1, according to the multinomial with parameters (N, I_j).

Theoretically, in this model for constant $\mu > 0$ the information about ϑ increases with N, but, for $0 < \pi_j < 1$, not without bound. The information is bounded by the uncertainty generated by $D(\underline{\pi})$. Because the Cramèr-Rao bound of the standard error of q_j equals $\sqrt{\frac{q_j(1-q_j)}{N}}$, $\underline{q}$ can be observed without error for $\lim_{N\to\infty}$. As stated before, the uncertainty generated by $D(\underline{\pi})$ depends inversely on μ.

The loglikelihood of the DM model is

$$\begin{aligned} \ln p(\underline{n};\vartheta) &= \ln \binom{N}{\underline{n}} + \ln A(\mu\underline{\pi}) - \ln A(\mu\underline{\pi} + \underline{n}) \qquad (8) \\ &= C + \ln\Gamma(\mu) - \ln\Gamma(\mu+N) - \sum_j \left[\ln\Gamma(\mu\pi_j) - \ln\Gamma(\mu\pi_j + n_j)\right]. \end{aligned}$$

Define $\sum_{i=0}^{-1} x_i = 0$, and using $\ln\Gamma(z+1) = \ln z + \ln\Gamma(z)$, then the result can be expressed as

$$\ln p(\underline{n};\vartheta) = C - \sum_{i=0}^{N-1} \ln(\mu + i) + \sum_j \sum_{i=0}^{n_j-1} \ln(\mu\pi_j + i), \qquad (9)$$

where only the last part depends on ϑ.

The Beta-Binomial Model

In the DM model subjects are to distribute a fixed amount of taws over the options of a multiple choice question. When considering the weight of a certain option a subject has to simultaneously consider the weights of the other options. And, perhaps move a taw from one option to another. It may be easier for subjects to assign up to N taws to one option at a time, without considering the other options. It will be the burden of this section to develop a model for data generated in this manner.

In the Beta-Binomial model it is hypothesized that a subject proceeds in two steps for every option of an item. For an option j he first he draws a random probability q from the Beta distribution with parameters $(\mu\pi_j, \mu(1-\pi_j))$.

$$p(q_j) \propto q_j^{\mu\pi_j}(1-q_j)^{\mu(1-\pi_j)}. \tag{10}$$

Next he draws a number n_j of taws from the Binomial distribution (N, q_j).

$$p(n_j) \propto q_j^{n}(1-q_j)^{N-n_j}. \tag{11}$$

The marginal probability (q_j integrated out) to observe n_j taws for an option in the Beta-Binomial (BB) is then given by (Jonson, and Kotz, 1969, sec. 3.11)

$$p(n_j;\vartheta) = \binom{N}{n_j}\frac{\Gamma(\mu)\Gamma(\mu\pi_j+n_j)\Gamma(\mu(1-\pi_j)+N-n_j)}{\Gamma(\mu+N)\Gamma(\mu\pi_j)\Gamma(\mu(1-\pi_j))}. \tag{12}$$

Mean and variance of n_j are the same as in the DM model, because the BB model is the marginal distribution of n_j in the DM model. However, the covariance vanishes in the BB model because the options are independent. After the same algebra as in the previous section on the DM model, we obtain for the BB model

$$\begin{aligned} \ln p(\underline{n};\vartheta) &= C - J\sum_{i=0}^{N-1}\ln(\mu+i) + \qquad (13) \\ &\sum_j\left(\sum_{i=0}^{n_j-1}\ln(\mu\pi_j+i) + \sum_{i=0}^{N-n_j-1}\ln(\mu(1-\pi_j)+i)\right). \end{aligned}$$

The limit for $\mu \to 0$ makes the BB model, like the DM model independent of N. For $\mu \to 0$ in the BB model all N taws are put on option j with probability π_j or no taws at all, for each option j independently. For $\mu > 0$ the information about ϑ increases with N, however, bounded by the uncertainty introduced by the Beta model. Because each option independently receives up to N taws, in the BB model there is no equivalent for binary scoring. For $N = 1$ each option may receive one taw. But, like the DM model, the BB model is also independent of μ for $N = 1$.

The information of the DM and BB models compared to binary scoring

For the DM model we will use

$$d(n,\pi) = \sum_{i=0}^{n-1} \left(\frac{1}{\mu\pi + i} \right)^2, \tag{14}$$

and for the BB model

$$b(n,\pi) = d(n,\pi) + d(N-n, 1-\pi). \tag{15}$$

First we will derive the Fisher-information about ϑ in the DM model. It follows from Formula (9) that the information function in the DM model for ϑ can be written as

$$\begin{aligned} I_{DM}(\vartheta; \mu, N) \;\; &= \;\; \mu^2 \mathcal{E}_\vartheta \left(\sum_j \pi_j'^2 d(n_j, \pi_j) \right) - \\ &\quad\;\; \mu \mathcal{E}_\vartheta \left(\sum_j \pi_j'' \sum_{i=0}^{n_j - 1} \frac{1}{\mu\pi_j + i} \right), \end{aligned} \tag{16}$$

where the DM model is indexed with ϑ, and where

$$\pi_j' = \frac{\partial}{\partial\vartheta} \pi_j. \tag{17}$$

We will show that the second line of Formula (16) vanishes. For all common densities or discrete probability functions $p(.)$ the order of differentiation and integration can be changed, so that

$$\begin{aligned} \mathcal{E}_{p(x)} (\ln p(x))' \;\; &= \;\; \int_x p(x) (\ln p(x))' dx \\ &- \;\; \int_x p'(x) dx = \left(\int_x p(x) dx \right)' = 1' = 0. \end{aligned} \tag{18}$$

Therefore,

$$\mathcal{E}_\vartheta \left(\ln p(\underline{n};\vartheta)\right)' = \mu \left(\sum_j \pi_j' \mathcal{E}_\vartheta \sum_{i=0}^{n_j-1} \frac{1}{\mu\pi_j + i} \right) = 0, \tag{19}$$

with n_j beta-binomially distributed. The value of $\underline{\pi}'$ is obviously arbitrary (because no model for $\underline{\pi}$ is specified as yet), except for the restriction

$$\sum_j \pi_j = 1 \to \sum_j \pi_j' = 0 \to \sum_j \pi_j'' = 0. \tag{20}$$

The arbitrariness of $\underline{\pi}'$ implies that

$$\mathcal{E}_\vartheta \sum_{i=0}^{n_j-1} \frac{1}{\mu\pi_j + i} \tag{21}$$

is independent of j, and, therefore, independent of the value of π_j, and depends only on μ and N (take in Formula (19), $\pi_0' = 1$, $\pi_j' = -1$ for some j, and zero otherwise). Because $\underline{\pi}''$ also obeys restriction (20), it immediately follows that the second line of Formula (16) vanishes. Therefore, we have that

$$I_{DM}(\vartheta;\mu,N) = \sum_j \left(\mu\pi_j'\right)^2 \mathcal{E}_\vartheta d(n_j,\pi_j). \tag{22}$$

Because n_j is beta-binomially distributed, I_{DM} can be numerically evaluated for all values of μ and N that are useful in practice. For the BB model $b(.)$ is substituted for $d(.)$.

Comparing the information to binary scoring, assuming equal trace lines for the distractors will yield a particularly attractive expression for the relative information. It is easily checked that for $N = 1$

$$I_{DM}(\vartheta;\mu,1) = \sum_j \left(\pi_j'\right)^2 \frac{1}{\pi_j}, \text{ and} \tag{23}$$

$$I_{BB}(\vartheta;\mu,1) = \sum_j \left(\pi_j'\right)^2 \left(\frac{1}{\pi_j} + \frac{1}{1-\pi_j}\right), \tag{24}$$

which are independent of μ. In the case of equal trace lines for the distractors, that is for $j > 0$ all π_j are equal, this simplifies to

$$I^*_{DM}(\vartheta;\mu,1) = \frac{(\pi_0')^2}{\pi_0(1-\pi_0)}, \text{ and} \tag{25}$$

$$I^*_{BB}(\vartheta;\mu,1) = (\pi_0')^2\left(\frac{1}{\pi_0(1-\pi_0)}+\frac{1}{1-\pi_0}+\frac{1}{J-2+\pi_0}\right). \tag{26}$$

$I^*_{DM}(\vartheta;\mu,1)$ equals the information function for binary scoring:

$$\begin{aligned} I_B(\vartheta) &= \pi_0\left((\ln\pi_0)'\right)^2+(1-\pi_0)\left((\ln(1-\pi_0))'\right)^2 \\ &= \frac{(\pi_0')^2}{\pi_0(1-\pi_0)} = I^*_{DM}(\vartheta;\mu,1), \end{aligned} \tag{27}$$

whereas $I^*_{BB}(\vartheta;\mu,1)$ equals $I_B(\vartheta)$ for $\pi_0 \to 0$, and equals twice this value for $\pi_0 \to 1$ (in Appendix C it is shown that this holds for all values of N). The relative information $RI_{DM} = I_{DM}/I_B$ is, therefore, given by

$$RI_{DM}(\vartheta;\mu,N) = \mu^2\pi_0(1-\pi_0)\left(\mathcal{E}_\vartheta d(n_0,\pi_0)+\sum_{j=1}\left(\frac{\pi_j'}{\pi_0'}\right)^2\mathcal{E}_\vartheta d(n_j,\pi_j)\right). \tag{28}$$

Using that $\sum \pi_j' = 0$, and under the assumption of identical trace lines for the distractors, it is found that

$$RI^*_{DM}(\vartheta;\mu,N) = \mu^2\pi_0(1-\pi_0)\left(\mathcal{E}_\vartheta d(n_0,\pi_0)+\frac{1}{J-1}\mathcal{E}_\vartheta d(n_1,\pi_1)\right), \tag{29}$$

where $\pi_1 = (1-\pi_0)/(J-1)$, and we have that $RI^*_{DM}(\pi_0(\vartheta);\mu,N)$ depends on ϑ only through π_0. Because for all items π_0 has the same range, $RI^*(\pi_0)$ represents an approximation to the relative information for all items with J options. The corresponding formula for the BB model again is found by substituting $b(.)$ for $d(.)$.

It is interesting to compare $RI^*_{DM}(\vartheta;\mu,N)$ for a certain value of N with its limit for $N \longrightarrow \infty$. As argued before this limit is given by the Dirichlet model, because for $N = \infty$, $\underline{q}$ can be observed without error. So instead of taking the limit for $N \longrightarrow \infty$ of I_{DM} in Formula (22), it suffices to take $-\mathcal{E}_{\mu\underline{\pi}}\left(\ln p(\underline{q})\right)''$, with p the Dirichlet distribution. The loglikelihood to observe $\underline{q}$, is given by

$$\ln p(\underline{q}) = \ln\Gamma(\mu) + \sum_j (\mu\pi_j - 1)\ln q_j - \ln\Gamma(\mu\pi_j). \tag{30}$$

The first two derivatives w.r.t. ϑ are

$$\begin{aligned} \ln p(\underline{q})' &= \mu\sum_j \pi_j' \left(\ln q_j - \psi(\mu\pi_j)\right), \text{ and} \\ \ln p(\underline{q})'' &= \mu\sum_j \pi_j'' \left(\ln q_j - \psi(\mu\pi_j)\right) - \mu\pi_j'^2\psi'(\mu\pi_j), \end{aligned} \tag{31}$$

with $\psi(z) = \frac{\partial}{\partial z}\ln\Gamma(z)$ the digamma function, and ψ' the derivative w.r.t. z. For the same reason as above after Formula (16), the part of the expectation of the summand with π_j'' in Formula (31) vanishes, and the expected information function is

$$-\mathcal{E}_{\mu\underline{\pi}}\left(\ln p(\underline{q})\right)'' = \mu^2\sum_j \pi_j'^2\psi'(\mu\pi_j). \tag{32}$$

So that the maximum of RI^*_{DM} is given by

$$\lim_{N\to\infty} RI^*_{DM}(\vartheta;\mu,N) = \mu^2\pi_0(1-\pi_0)\left(\psi'(\mu\pi_0) + \frac{1}{J-1}\psi'(\mu\frac{1-\pi_0}{J-1})\right). \tag{33}$$

In the same way the Beta model is the limit for the BB model. In Appendix B the information function for the Beta model is derived. Let

$$\psi_B'(\mu\pi) = \psi'(\mu\pi) + \psi'(\mu(1-\pi)), \tag{34}$$

and the corresponding formulas for the Beta model are found by substitution of $\psi_B'(.)$ for $\psi'(.)$.

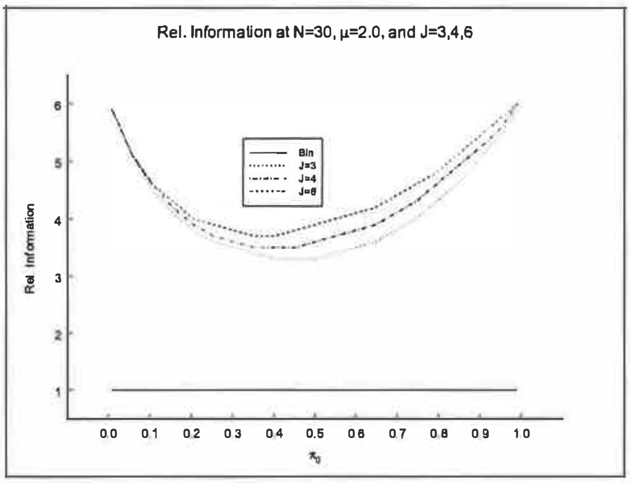

Figure 1: Relative information of the DM model at $N = 30$ taws, $\mu = 2.0$ and $J = 3, 4, 6$ compared to binary scoring.

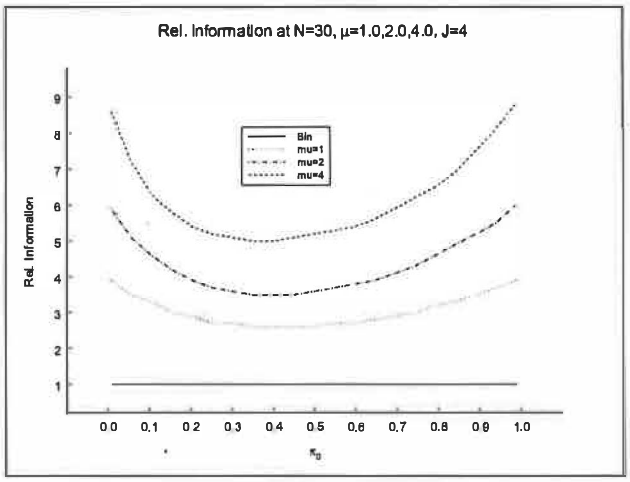

Figure 2: Relative information of the DM model at $N = 30$ taws, $\mu = 1.0, 2.0, 4.0$ and $J = 4$ compared to binary scoring.

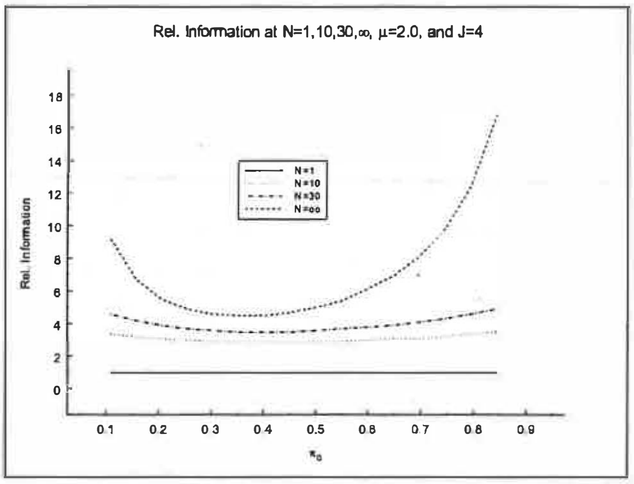

Figure 3: Relative information of the DM model at $N = 1, 10, 30, \infty$ taws, $\mu = 2.0$ and $J = 4$ compared to binary scoring.

Figures 1, 2, and 3 give an impression of the dependence of RI^*_{DM} on the values of respectively J, μ, and N. For μ the value 2.0 is chosen as a representative value. At $\mu = 2.0$, the Beta distribution for $\pi_0 = 0.5$, which is the marginal Dirichlet for $\pi_0 = 0.5$, is a constant. So at $\mu = 2.0$, $\pi_0 = 0.5$ q_0 is drawn from the uniform distribution on [0,1], and the standard deviation of q_0 equals 0.29. This seems quite large. Therefore, $\mu = 2.0$, perhaps, is an underestimate of the true value for this parameter. Figure 2 shows that the relative information increases steeply with μ. Consequently the value of μ determines to a large extent the gain of information that is to be expected from the DM-model. But even if $\mu = 1.0$, and with only 10 taws the smallest relative information still exceeds 2. Which means that comparable precision is obtained with less than half the test length compared to binary scoring. However, only parameter estimation with real data can clarify this issue. Because, the dependence of the relative expected information on μ, and N is the most conspicuous, we repeat Figures 2, and 3 for the BB-model in Figures 4, and 5. The relative information for the BB model starts at $\pi_0 \to 0$ at the same value as for the DM model. For $\pi_0 \to 1$ it becomes twice as large. At $\pi_0 = 0.5$ it is 1.6 times as large.

In practice, the information about ϑ most likely will not keep increasing with N, because from a certain value of N a lower estimate for μ will result.

Probably a subject v will turn out to have an optimum value for N, denoted by N_v. If given a lower number of taws than N_v to distribute, part of his potential to inform about his ability is left unused. If given a larger number of taws than N_v he is supposed to give more information than he is able to. This optimal value N_v is called his *resolution.* If the actual number of taws N does not match his resolution N_v one could assume that v first distributes N_v taws and next multiplies the resulting vector $\underline{n}$ by $c = N/N_v$, and then rounding such that the sum of taws equals N. Because in the present models μ is an item parameter it is assumed that the resolution of all subjects has about the same value.

Now, if a subject v has a resolution of N_v taws, but has to distribute $N = cN_v$ taws over the options at $\mu = \mu'$, it follows from Formula (7) that the estimated μ^* is given by

$$\begin{aligned} \mathcal{V}ar(cn_j; N_v, \mu', \underline{\pi}) &= c^2 \frac{(\mu' + N_v)}{(\mu' + 1)} N_v \pi_j (1 - \pi_j) \\ &= \frac{(\mu^* + cN_v)}{(\mu^* + 1)} cN_v \pi_j (1 - \pi_j), \end{aligned} \tag{35}$$

and so that

$$c\frac{(\mu' + N_v)}{(\mu' + 1)} = \frac{(\mu^* + cN_v)}{(\mu^* + 1)}. \tag{36}$$

For instance, at $\mu' = 2$, $N_v = 10$, and $c = 2$, we would expect an estimate $\mu^* = 1.71$[1].

It may seem an attractive idea to estimate $e = 1/c > 0$ as a resolution parameter. Unfortunately, substitution of eN for N and $e\underline{n}$ for $\underline{n}$ in Formula (2) reveals that $p(e\underline{n}; \vartheta) = 1$ for $e = 0$, and, therefore, that maximum likelihood is approached for $e \longrightarrow 0$, or $c \longrightarrow \infty$.

In the next section a model for $\underline{\pi}$ is introduced, whereafter parameter estimation and model tests will be treated.

[1]However, this moment estimate turns out to deviate from the maximum likelihood estimate, which, as appears from simulations, in this example turns out to be around 1.46. And this maximum likelihood estimate does not appear to be influenced by the number of items (5, 10, 20, and 40) nor by the number of subjects (200 and 1000, at 5 items).

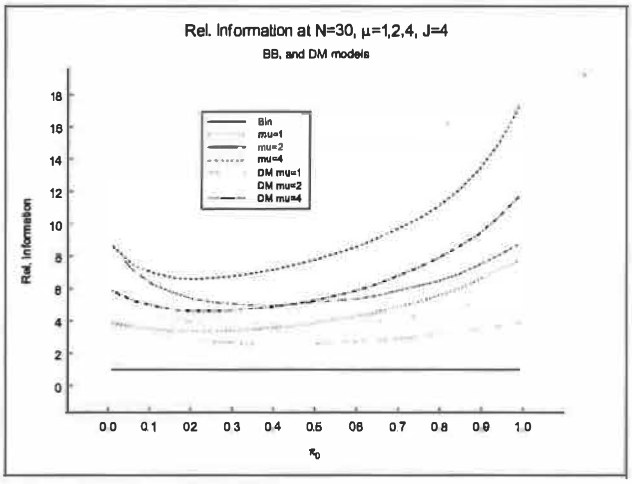

Figure 4: Relative information of the BB model at N=30 taws, $\mu_{.} = 1.0, 2.0, 4.0$, and J=4 compared to binary scoring. For comparison the same curves for the DM model are inserted. Corresponding curves coincide at their leftmost points.

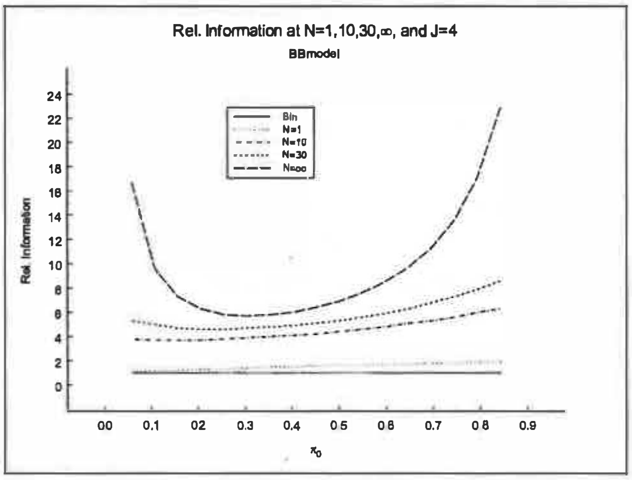

Figure 5: Relative information of the BB model at $N = 1, 10, 30, \infty$ taws, $\mu = 2.0$ and $J = 4$ compared to binary scoring.

A model for $\underline{\pi}$

To further develop the DM and BB models a model for $\underline{\pi}$ is to be chosen. Let

$$\zeta_j = \exp(\alpha\vartheta + \beta_j), \tag{37}$$

with α a discrimination parameter and β_j $(j > 0)$ a location parameter. Further, let $\tau_0 = 1$ and let $\gamma_j > 0$ $(j > 0)$ denote an attraction parameter, and let

$$\tau_j = \frac{\gamma_j}{1+\zeta_j},\ j > 0, \tag{38}$$

then π_j can be defined as

$$\pi_j = \frac{\tau_j}{\sum_k \tau_k}. \tag{39}$$

So a common discrimination for all options of an item is assumed, and only differences in discrimination between items are allowed. The main reason for a common steepness for options is that often the data on false options will be too sparse for accurate estimation of differences in discrimination between options. The parameter β determines the location of the option and γ its attractiveness.

The parameters β and γ can be simultaneously estimated reliably only with very large data sets. Their estimates are highly positively correlated, because increasing β, can be largely compensated for by increasing γ as well. There has to be a substantial amount of observations also at the lower values of ϑ, to reliably distinguish between the effects of β and γ. To see this τ_j can be rewritten as

$$\tau_j = \frac{1}{\gamma_j^{-1} + \exp(\alpha\vartheta + \beta_j - \ln\gamma_j)} = \frac{\gamma_j \exp(-\beta_j)}{\exp(-\beta_j) + \exp(\alpha\vartheta)}, \tag{40}$$

where the additive entangling of β_j and $\ln\gamma_j$, or the multiplicative entangling of $\exp(\beta_j)$ and γ_j, are clearly visible.

To accommodate not very large data sets one may choose between two restricted models:

1. $\gamma = 1$ for all items, (or impute a value per item)
2. $\beta = 0$ for all items, (or impute a value per item)

To distinguish the models the unrestricted model is called π_U, and the restricted models respectively π_β, where β is free to be estimated, and π_γ, where γ is free to be estimated. The restricted model π_β, with $\underline{\gamma} = \underline{1}$, has the property that

$$\lim_{\vartheta \to -\infty} \pi_j = \frac{1}{J} \text{ for } j = 0, \ldots J-1. \tag{41}$$

So all options are assumed equally attractive for low values of ϑ.

In the restricted model π_γ the relative attractiveness of distractors remains the same over the entire range of ϑ. It can easily be shown that the π_γ model is equivalent to the model in Verhelst and Verstralen (in press). These authors present the model in the parametrization

$$\begin{aligned} P(X &= 0|\xi) = \frac{\nu_0 \xi}{1 + \nu_0 \xi} \\ P(X &= j|\xi) = \frac{\nu_j}{1 + \nu_0 \xi}, \ (j > 0), \end{aligned} \tag{42}$$

with $\xi > 1$, and $\sum_{j=1} \nu_j = 1$. Without changing the model one could also choose $\nu_0 = 1$ as a normalization for $\underline{\nu}$, with $\sum_{j=1} \nu_j$ free to be estimated. Moreover, reparameterize $\xi = 1 + \exp(\alpha\vartheta)$, which is more general because $\alpha = 1$ in the original model, and substitute the symbol γ for ν, with $\gamma_0 = 1$, then the model can be written as

$$\begin{aligned} P(X &= 0|\vartheta) = \frac{\gamma_0 (1 + \exp(\alpha\vartheta))}{\sum_{j=1} \gamma_j + \gamma_0 (1 + \exp(\alpha\vartheta))} \\ P(X &= j|\vartheta) = \frac{\gamma_j}{\sum_{j=1} \gamma_j + \gamma_0 (1 + \exp(\alpha\vartheta))}, \ (j > 0), \end{aligned} \tag{43}$$

which is easily shown to be equal to Formula (39) by noting that ζ_j does not depend on j in the π_γ model. The parametrization given by Formula (42) is significant because it shows that a model, formally equivalent to the Rasch-model, allows for guessing by introducing a new interpretation of the parameters.

Parameter estimation

Parameter estimation can be accomplished by the EM-method (Dempster, a.o., 1977). Let ϑ be distributed as g. If we have a test with k MC items, the vector of observations of subject v is denoted as $\underline{\underline{n}}_v = (\underline{n}_{v1}, ..., \underline{n}_{vk})$. Then the complete data loglikelihood of observing $(\underline{\underline{n}}_v, \vartheta_v)$ is given by

$$\ell_v(\underline{\lambda}; \underline{\underline{n}}_v, \vartheta_v) = \sum_i \ln p_i(\underline{n}_{vi}; \vartheta_v) + \ln g(\vartheta_v), \tag{44}$$

with $\underline{\lambda}$ the parameter vector of the model. The marginal or expected loglikelihood is

$$M\ell_v(\underline{\lambda}; \underline{\underline{n}}_v) = \int \left(\sum_i \ln p_i(\underline{n}_{vi}; \vartheta) + \ln g(\vartheta) \right) g(\vartheta) d\vartheta. \tag{45}$$

The posterior of ϑ, given the observation of $\underline{\underline{n}}$ is given by

$$h(\vartheta; \underline{\lambda} | \underline{\underline{n}}) \propto \prod_i p_i(\underline{n}_i; \vartheta) g(\vartheta). \tag{46}$$

Because the first order derivatives of the marginal loglikelihood $M\ell = \sum_v M\ell_v$ are equal to the first order derivatives of the posterior expected loglikelihood, we have

$$M\ell + C = \sum_v \int \ell(\underline{\lambda}; \underline{\underline{n}}_v, \vartheta) h(\vartheta; \underline{\lambda} | \underline{\underline{n}}_v) d\vartheta, \tag{47}$$

with C a constant. An iterative EM-algorithm is obtained by distinguishing $\underline{\lambda}^*$ as the vector of parameters known from the previous iteration, and $\underline{\lambda}$ as the vector of parameters for which the following function has to be maximized in the current iteration

$$Q(\underline{\lambda}, \underline{\lambda}^*) = \sum_v \int \ell_v(\underline{\lambda}; \underline{\underline{n}}_v, \vartheta) h(\vartheta; \underline{\lambda}^* | \underline{\underline{n}}_v) d\vartheta. \tag{48}$$

Given $\underline{\lambda}^{t-1}$ from iteration $t-1$ in the E-step the first and second derivatives of $Q(\underline{\lambda}, \underline{\lambda}^*)$ w.r.t. $\underline{\lambda}$, evaluated at $\underline{\lambda}^{t-1}$, are obtained, and in the M-step values of $\underline{\lambda}^t$ are obtained by performing one Newton Raphson step

$$\underline{\lambda}^t = \underline{\lambda}^{t-1} + (-Q^{(2)})^{-1} Q^{(1)}, \tag{49}$$

where $Q^{(n)}$ denotes the m^n matrix of nth derivatives of Q w.r.t. $\underline{\lambda}$, and m denotes the number of elements of $\underline{\lambda}$. The formulas for the derivatives of Q can be found in Appendix A. Standard errors of estimation are found by the method developed in Louis (1982).

If we let $g = N(\overline{\vartheta}, 1)$, the normal distribution, the integral in Formula (48) can be approximated by the Gauss-Hermite procedure:

$$\int e^{-x^2} f(x)dx = \sum_{i=1}^{N} w_i f(x_i), \tag{50}$$

where $\underline{w}$ is a vector of weights, and $\underline{x}$ a vector of values of the argument of x. Values of $\underline{w}$ and $\underline{x}$ can be found for values of N from 2 through 190 by an algorithm in Press, a.o. (1992).

Whereas in the π_U and the π_β models the zero point of the scale can be fixed at $\overline{\vartheta} = 0$, this is not the case in the π_γ model. In the π_γ model $\overline{\vartheta}$ has to be estimated, because the origin of the scale is already fixed by choosing $\underline{\beta} = \underline{0}$. However, the convergence of the EM-algorithm while also estimating $\overline{\overline{\vartheta}}$ is excruciatingly slow. The additive entangling of γ_{ij} and β_{ij} as shown in Formula (40) holds, of course, also for γ_{ij} and $\overline{\beta}$, and, therefore, for $\overline{\vartheta}$ as well.

In Appendix A only one parameter μ is mentioned without a subscript or $\mu(.)$ as a function of something else. However, it is conceivable that μ depends on the subject, then we have μ_v, or on the item, μ_i, or even on the combination of subject and item, μ_{vi}. Except for μ_i these interpretations lead to too many parameters to estimate accurately. Because μ may be subject dependent, there seem to be four options to pursue:

1. Consider μ, like ϑ, as a nuisance parameter, and take the marginal likelihood w.r.t. the joint distribution of μ and ϑ

2. Consider μ as a function of ϑ, and estimate e.g.,

(a) μ_q for every Gauss-Hermite point, or for groups of adjacent Gauss-Hermite points

(b) $\mu = f(\vartheta)$, with f e.g., cubic, and estimate the four cubic parameters

3. Consider μ as an item parameter.

4. Consider μ as a constant.

If there were prior knowledge about μ option 4 would certainly be the most preferred. However, if there is not, option 3 results in the most simple EM-algorithm, because the matrix of second derivatives is blockdiagonal, with one small lower triangular matrix pèr item, and, therefore, is the best to investigate first. If there were an indication that μ is related to ability, option 2.a. would be preferable, because it enables to fit a linear, quadratic, cubic etc. to these estimates, and to judge whether the simplification to the direct estimation of a certain f is justified. Option 1 has the disadvantage that a double integral has to be calculated, which slows down the estimation tremendously. In the sequel option 3 is pursued.

Initial estimates

The iterative EM-procedure must be started with an initial guess about the value of the parameters. Initial estimates are generally obtained by adopting some simplifying assumptions. As mentioned in section 2 $\mathcal{E}n_j = N\pi_j(\vartheta)$. If, as a simplifying assumption we set $\vartheta = 0$, then $\pi_j(0) \approx n_j/(NV)$. If we apply the π_U model, and so have to estimate β and γ, the latter can be given an initial value of 1, as in the π_β-model. In both models we then have that

$$\beta_j = \ln\left(\frac{\pi_j(\vartheta)}{\pi_0(\vartheta)} - 1\right) \approx \ln\left(\frac{n_j + 0.5}{n_0 + 0.5} - 1\right). \tag{51}$$

With these initial estimates for β, the estimation procedure is very robust for general initial values for μ, and α, like 1 or 2.

If the π_γ-model is applied, and β is not a parameter, initial values for γ are found as

$$\gamma_j = (1+\exp(\alpha\overline{\vartheta}))\frac{n_j+0.5}{n_0+0.5}, \tag{52}$$

with α an initial estimate of α_i, and $\overline{\vartheta}$ of the mean of ϑ. $1+\exp(\alpha\overline{\vartheta}) = 2$ for $\overline{\vartheta} = 0$.

Testing the model

Model tests can be constructed using the framework of the Lagrange Multiplier (LM) test-statistic. An introduction to the LM-test within a larger context can be found in Buse (1982). The idea for the LM-test originates with Rao (1948) , there called the 'score test' , and with Aitchinson and Silvey (1958). An application within the context of IRT models can be found in Glas and Verhelst (1995), and Glas (1997, 1999). In general, to compute the LM-statistic restrictions on parameters are relaxed. For instance, one may release the restriction that α_i is equal for boys and girls, thereby replacing α_i with α_{ib} for the boys, and α_{ig} for the girls. Denote the U new parameters by $\underline{\xi}$. The likelihood function is then evaluated at the maximum likelihood estimates of the original parameters, and the U new parameters $\underline{\xi}$ at their latest values. In the example $\alpha_{ib} = \alpha_{ig} = \alpha_i$. The LM-test statistic can then be expressed as

$$LM = \ell^{(1)T}\left(-\ell^{(2)}\right)^{-1}\ell^{(1)}, \tag{53}$$

where the superscripts within parentheses denote order of differentiation w.r.t. the parameters of the relaxed model, and superscript T denotes transposition. LM is χ^2-distributed with degrees of freedom equal to the number of relaxed restrictions.

To obtain an especially simple procedure for the calculation of the LM statistic one focuses on implicit (0 or 1) or explicit constants in the model, and changes their status from constant to a variable parameter in the likelihood function. By this procedure all original parameters remain in the relaxed model. In general this is not the case. For instance, above α_i was replaced with α_{ib} and α_{ig}. Because the likelihood is evaluated at the maximum likelihood estimates of the original parameters, the elements of the first derivative corresponding to the original parameters are all equal to zero. Because all original parameters remain in the relaxed model this simplifies the computation. Denote the complete vector of original and new parameters as $\underline{\pi} = (\underline{\lambda}, \underline{\xi})$, and select with $F(\underline{\lambda})$ the vector of elements of the first derivatives of $M\ell$ with respect to the elements of $\underline{\lambda}$. Likewise $I(\underline{\lambda}, \underline{\xi})$ selects the part of the observed information matrix I with the rows for $\underline{\lambda}$, and the columns for $\underline{\xi}$. Then

$$LM = F(\underline{\xi})^T W^{-1} F(\underline{\xi}), \tag{54}$$

with

$$W = I(\underline{\xi}, \underline{\xi}) - I(\underline{\xi}, \underline{\lambda}) I(\underline{\lambda}, \underline{\lambda})^{-1} I(\underline{\lambda}, \underline{\xi}), \tag{55}$$

where $I(\underline{\lambda}, \underline{\lambda})^{-1}$ is already computed to obtain standard errors of the parameter estimates. If original parameters are replaced by new parameters by relaxation of restrictions, this simplification is not obtained.

In case the LM-test shows that the model is violated for a certain item, one may evaluate the size of the misfit with the first Newton-Raphson step of the new parameters, were the estimation continued after releasing the restrictions. These first steps are given by $F(\underline{\xi})^T W^{-1}$.

To calculate an LM-statistic for some or all parameters of an item, the respondents must be grouped independently of the response on the that particular item. This can be accomplished by using an independent background variable, like gender or parental educational level, but also by grouping the respondents into groups of homogeneous ability. The division into groups of homogenous ability must also be accomplished independently of the item. For that purpose the procedure developed in Verstralen (2000) can be used. There a computationally cheap procedure is developed to obtain an EAP-estimate of the person parameter that is independent of the item.

As noted in Verstralen (2000) the augmented model with group parameters is undetermined. Therefore, we introduce the restriction

$$\sum \xi_g = 0. \tag{56}$$

Diagnostic distractors

Suppose that in a certain subject matter a few main fallacies can be identified. And, moreover, suppose that a test with multiple choice questions can be constructed, where each distractor in an item relates to one of these fallacies. Then the DM-model allows, in principle, to identify the persons with a certain fallacy.

If a person is a victim of a certain fallacy, and does not know the correct response, he has a relatively large probability to put his taws on the option associated with that fallacy, and a low probability to put his taws on the other options. More technically, suppose first, without loss of generality, that for all items the index of the option indicates the type of fallacy. Now, let m be the option-index that represents the fallacy of a certain subject. Then we may assume that for all items it holds that when this subject does not know the correct response he is inclined to put his taws on m, at the cost of the other distractors. Now, if ϑ is estimated only on the basis of the number of taws put on distractors m, a lower estimate for ϑ is expected than on the basis of his complete vector $\underline{\underline{n}}$. Because m is a distractor its trace lines tend to be decreasing in ϑ with the implication that the lower ϑ the higher the expected number of taws put on m. For the same reason the number of taws put on distractors other than m will be lower, and a higher estimate for ϑ is to be expected if it is estimated only on the basis of the other distractors $j \neq m$.

Because the marginal distribution of n_j for the DM and the BB model as well are given by Formula (12), we have that

$$\begin{aligned}
\ln p(n_j;\vartheta) &= C + \ln\Gamma(\mu) - \ln\Gamma(\mu+N) - \\
&\quad \left(\begin{array}{c} \ln\Gamma(\mu\pi_j) - \ln\Gamma(\mu\pi_j + n_j) + \\ \ln\Gamma(\mu(1-\pi_j)) - \ln\Gamma(\mu(1-\pi_j) + N - n_j) \end{array} \right).
\end{aligned} \tag{57}$$

This results in a posterior distribution for ϑ

$$h(\vartheta;\underline{\lambda}|\underline{n}_j) \propto g(\vartheta)\prod_i p_i(n_{ij};\vartheta), \tag{58}$$

which enables to calculate an EAP-estimate for ϑ, and its posterior variance, denoted respectively by ϑ_j and s_j^2, on the basis of the numbers of taws put on options j. Denote the EAP-estimate and its posterior variance for ϑ on the basis of the complete vector $\underline{\underline{n}}$ by ϑ_c and s_c^2. The standardized difference

$$d_j = \frac{\vartheta_c - \vartheta_j}{\sqrt{s_c^2 + s_j^2}} \tag{59}$$

informs about the extra attraction of distractor j above that expected by the model. The larger d_j, the more extra attraction option j has for the subject. Because the data for ϑ_c contain the data for ϑ_j their estimates are conditionally positively correlated given ϑ. Therefore, the variance of d_j under the model is smaller than 1. Consequently, a significance test based on the assumption that d_j is standard normal distributed will turn out to be conservative. In the last part of the section below we will put this idea to a test.

A simulated example

As an example a 200 record data set was generated for the DM-π_β model ($\gamma = 1$), with responses on a test with 20 four-choice items ($J = 4$). Responses were generated for 20 taws as well as for 1 taw (converted to binary scores). The twenty items all have the same parameters, $\mu = 2.0$, $\alpha = 1.0$, $\underline{\beta} = -0.5, 0.0, 0.5$. The distribution of ϑ is the standard normal. The 20 taw response data set was analyzed using the DM-π_β model, and the binary data set using OPLM (Verhelst, a.o., 1995). With the DM model EAP-estimates of ϑ were obtained, and with OPLM WML-estimates of ϑ. Both are plotted against ϑ, and shown in Figures 6, and 7. The difference in accuracy is so large that it clearly shows in the scatter plots. The correlation EAP-ϑ equals 0.89, and the correlation WML-ϑ equals 0.76. The corresponding reliability coefficients are 0.80 and 0.58.

To test whether the Lagrange multiplier test is sensitive to a particular model violation we also generated a 200 record data set with μ of item 1 dependent on ϑ

$$\mu_1(\vartheta < 0) = 1, \text{ and } \mu_1(\vartheta \geq 0) = 3. \tag{60}$$

The other parameters are equal to the above simulation. The results for item 1 are shown in Table 1.

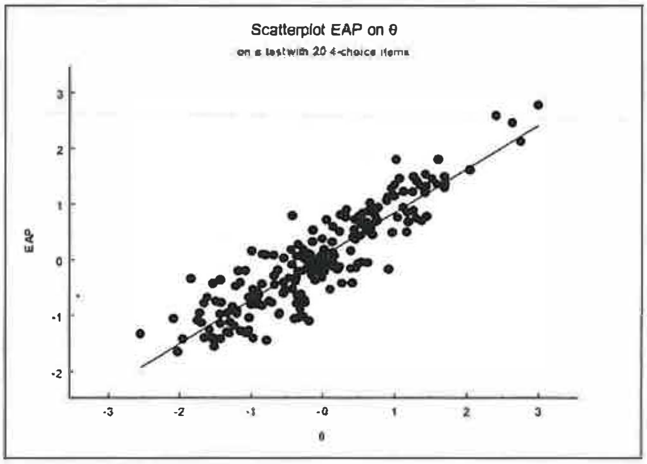

Figure 6: Scatterplot of DM-EAP estimates of ϑ against their true values, responses: 20 taws, $\mu = 2.0$.

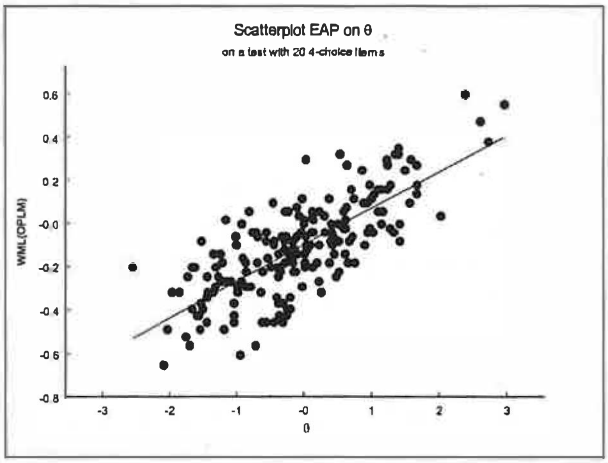

Figure 7: Scatterplot of OPLM-WML estimates of ϑ against their true values, binary responses.

Table 1
Calibration and Test results for item 1

Item	Grp	N	Par	Estimate	StErr	NR-Step	LM-Test	(df)	$P(\chi^2>LM)$
8	1	114	μ	1.483	0.108	-0.500	24.077	5	0.000
			α	0.705	0.147	-0.299			
			β_1	-0.703	0.350	0.166			
			β_2	-0.290	0.263	-1.177			
			β_3	0.293	0.231	0.013			
	2	86	μ			0.500			
			α			0.299			
			β_1			-0.166			
			β_2			0.177			
			β_3			-0.013			

As Table 1 shows, the model violation on μ_1 is clearly detected. The NR-Step for the lower group brings the estimate right on target, but for the higher group the estimate is too low (about 2 instead of 3). Moreover, the violation of μ_1 has its main effect not only on the estimate of μ_1, but also on the estimate of α_1. For the lower group (with a lower μ_1), also an appreciably lower value for α_1 is suggested, than for the higher group. The estimate of the vector $\underline{\beta}$ seems not to be affected by the model violation. Of the other 19 items just one item had a LM-test statistic with a 5%-significant p-value (p=0.032). The sum of the LM-statistics has a p-value of 0.055. Under the assumption that the LM-statistics of the items are independent χ^2-distributed, their sum is also χ^2-distributed, with the sum of the degrees of freedom. This assumption is, not exactly true, but it is approximately so. Although one of the 20 items was detected to clearly violate the model, this overall statistic with 200 records just raises some suspicion.

To test the idea about the diagnostic value of distractors that represent a certain fallacy, the same data set was generated, as used above, except for the last ten subjects 191 through 200. They were assumed to be a victim of a certain fallacy. Subject 191 suffered of the fallacy represented by distractors 1, subject 192 by distractors 2, 193 by 3, 194 again by 1, and so on. Their data were generated by changing their values for π_j, $j > 0$, as follows

$$
\begin{aligned}
\pi_m &= (1-\pi_0)\,(1.0-\varepsilon) \\
\pi_{j\neq m} &= (1-\pi_0)\,\frac{\varepsilon}{J-2},
\end{aligned}
\tag{61}
$$

with $\varepsilon = 0.05$, and m the index of the fallacy. So the probability to obtain a correct response was left unchanged, only the relative emphasis on the distractors was altered in favor of the fallacy. These data were analyzed with the DM-model and for all subjects the value of d_j was tested at the one-tailed 5% significance level. If $d_j > 1.65$ this was indicated by a $-j$ and if $d_j < -1.65$ this was indicated by a j. Just 6 of the normal 190 subjects showed one significant option. This shows that the test is rather conservative. Of the ten last subjects with one fallacy, all were marked, six of them for all three distractors, three of them for two, and one subject just for one distractor. Eight of them with the $-j$ for the fallacy distractor. For instance subject 192 was marked by 1 -2 3. Indicating that ϑ_2 for this subject was much lower than ϑ_c, and ϑ_j for the other two distractors much higher than ϑ_c. We may conclude that the process works for clear-cut fallacies.

Discussion

It is an open question in what way one can obtain valid data for the model. What points should be stressed in an instruction? What is an optimal choice for N the number of taws? If implemented in a CBT environment, what are the essential ingredients in a comfortable interface?

One could for instance be tempted to implement the attachment of weights to options by having the subject draw lines behind the options, where the length of the line reflects the subjectively perceived correctness. Moreover, it would be tempting to consider these data as continuous, and therefore, consider the data as an instance of $N = \infty$. This is especially tempting because the model predicts the highest information gain for $N = \infty$. In that case the limit for $N \to \infty$ of the DM and BB models would apply, which, as discussed in Section 2 is just respectively the Dirichlet or the Beta model. However, from Formulas (3) and (10) it appears that the probability for $q_j = 0$ vanishes. This means that if just one subject draws a line of length 0 behind just one option, the model cannot be estimated, because the whole data set has probability zero under the model. And one is bound to observe lots of lines of length zero, because for the better subjects some options simply are completely wrong.

This observation does not mean that one should not have subjects draw lines behind options. However, one should be aware that although subjective lines may seem continuous, they have not infinite subjective resolution. For instance, subjects will not perceive a difference between a pile of 30 taws and a pile of 31 taws, or a line of 30 millimeters and a line of 31 millimeters, when not shown side by side of course. Therefore, the lines given by subjects should be transformed to finitely accurate measures, as one does with length measurement, for instance rounded to the nearest mm, depending on the measuring instrument. To do this properly, research is needed as to the subjective resolution. Then the 'continuous ' line lengths are to be converted to an appropriate number of taws.

This research could, perhaps, be done within the framework of the DM, and BB models itself. One could, for instance, analyze a 'continuous' data set with a range of resolutions, such as one taw per 2 mm, per mm, per half a mm, etc. When the resolution of the analysis transgresses the resolution of the subjects, one would be trying to extract more information from the data than it contains, which should be balanced by a decrease in the estimates for μ. As an aside remark, this phenomenon will, in general, show up when the number of taws to distribute does not fit the resolution of the subject.

Another, be it less elegant, solution to the 'zero'-probability problem could be to exclude zero line lengths, by giving each line a minimum length, for instance 0.025 of the maximum length. With a four option MC question under the DM-framework this consumes 10% of the total weight to be distributed. What a reasonable minimum line length should be and what its influence, is to be investigated.

Appendix A : Derivatives of $Q(\underline{\lambda},\underline{\lambda}^*)$

Denote with

$$h_{vq} = h(\vartheta_q; \underline{\lambda}^* | \underline{\underline{n}}_v). \tag{62}$$

Let $\underline{\phi}$ be a vector of parameters then the nth order derivatives of Q have the following form

$$\frac{\partial^n}{\partial\phi_k ... \partial\phi_l} Q = \sum_{vq} h_{vq} \sum_i \frac{\partial^n}{\partial\phi_k ... \partial\phi_l} \ln p_{mi}(\underline{n}_{vi}; \vartheta_q), \tag{63}$$

with $m = D$ for the DM model and $m = B$ for the BB model. As already given in Formula (8) we have that

$$\ln p_{Di}(\underline{n}_{vi}; \vartheta_q) = C_{DM} + \ln\Gamma(\mu) - \ln\Gamma(\mu + N) - \sum_j \left[\ln\Gamma(\mu\pi_j) - \ln\Gamma(\mu\pi_j + n_j)\right], \tag{64}$$

for the DM model , and

$$\begin{aligned} \ln p_{Bi}(\underline{n}_{vi}; \vartheta_q) \quad = \quad & C_{BB} + J\left(\ln\Gamma(\mu) - \ln\Gamma(\mu + N)\right) - \\ & \sum_j \left(\begin{array}{c} \ln\Gamma(\mu\pi_j) - \ln\Gamma(\mu\pi_j + n_j) + \\ \ln\Gamma(\mu\,(1 - \pi_j)) - \ln\Gamma(\mu\,(1 - \pi_j) + N - n_j) \end{array} \right), \end{aligned} \tag{65}$$

for the BB model, where C_{DM} and C_{BB} are constants independent of the parameters.

Now let λ_k be a parameter of $\underline{\pi}$, then for Q we have the parameter vector $(\mu,\underline{\lambda})$. Furthermore, using subscript D for the DM model and B for the BB model let

$$
\begin{aligned}
\pi_j^{(k)} &= \frac{\partial}{\partial \lambda_k}\pi_j \qquad (66)\\
\psi'(x) &= \frac{\partial}{\partial x}\psi(x)\\
\psi_{Dj} &= \psi(\mu\pi_j) - \psi(\mu\pi_j + n_j) = -\sum_{k=0}^{n_j-1}\frac{1}{\mu\pi_j + k} \text{ (A\&S 6.3.5)}\\
\psi_{Bj} &= \psi(\mu(1-\pi_j)) - \psi(\mu(1-\pi_j) + N - n_j)\\
\psi'_{Dj} &= \psi'(\mu\pi_j) - \psi'(\mu\pi_j + n_j) = \sum_{k=0}^{n_j-1}\frac{1}{(\mu\pi_j + k)^2} \text{ (A\&S 6.4.6)}\\
\psi'_{Bj} &= \psi'(\mu(1-\pi_j)) - \psi'(\mu(1-\pi_j) + N - n_j)\\
\psi_D &= \psi(\mu) - \psi(\mu + N)\left(= -\sum_{k=0}^{N-1}\frac{1}{\mu + k}\right),\\
\psi_j &= \psi_{Dj} - \psi_{Bj}\\
\psi'_j &= \psi'_{Dj} + \psi'_{Bj}
\end{aligned}
$$

then the first and second derivatives of $\ln p_{Di}$ are

$$
\begin{aligned}
\frac{\partial}{\partial \mu}\ln p_{Di} &= \psi_D - \sum_j \pi_j \psi_{Dj} \qquad (67)\\
\frac{\partial}{\partial \lambda_k}\ln p_{Di} &= -\mu \sum_j \pi_j^{(k)} \psi_{Dj}\\
\left(\frac{\partial}{\partial \mu}\right)^2 \ln p_{Di} &= \psi'_D - \sum_j \pi_j^2 \psi'_{Dj}\\
\frac{\partial^2}{\partial \mu \partial \lambda_k}\ln p_{Di} &= -\sum_j \pi_j^{(k)}\left(\psi_{Dj} + \mu\pi_j\psi'_{Dj}\right) = \frac{\frac{\partial}{\partial \lambda_k}\ln p_{Di}}{\mu} - \mu\sum_j \pi_j \pi_j^{(k)}\psi'_{Dj}\\
\frac{\partial^2}{\partial \lambda_k \partial \lambda_l}\ln p_{Di} &= -\mu\sum_j \pi_j^{(k,l)}\psi_{Dj} - \mu^2 \sum_j \pi_j^{(k)}\pi_j^{(l)}\psi'_{Dj}.
\end{aligned}
$$

The derivatives for the BB model can be presented very similar in appearance, with an additive correction for the derivatives w.r.t. μ

$$
\begin{aligned}
\frac{\partial}{\partial \mu} \ln p_{Bi} &= J\psi_D - \sum_j \pi_j \psi_j + \psi_{Bj} \\
\frac{\partial}{\partial \lambda_k} \ln p_{Bi} &= -\mu \sum_j \pi_j^{(k)} \psi_j \\
\left(\frac{\partial}{\partial \mu}\right)^2 \ln p_{Bi} &= J\psi'_D - \sum_j \pi_j^2 \psi'_j + (1 - 2\pi_j)\psi'_{Bj} \\
\frac{\partial^2}{\partial \mu \partial \lambda_k} \ln p_{Bi} &= -\sum_j \pi_j^{(k)} \left(\psi_j + \mu \left(\pi_j \psi'_j - \psi'_{Bj}\right)\right) = \\
& \quad \frac{\frac{\partial}{\partial \lambda_k} \ln p_{Bi}}{\mu} - \mu \sum_j \pi_j^{(k)} \left(\pi_j \psi'_j - \psi'_{Bj}\right) \\
\frac{\partial^2}{\partial \lambda_k \partial \lambda_l} \ln p_{Bi} &= -\mu \sum_j \pi_j^{(k,l)} \psi_j - \mu^2 \sum_j \pi_j^{(k)} \pi_j^{(l)} \psi'_j .
\end{aligned}
\tag{68}
$$

Derivatives of $\underline{\pi}$:
In this section not mentioned entries are equal to 0. Let

$$
\begin{aligned}
f^{(\varphi)} &\doteq \frac{\partial}{\partial\varphi} f, \text{ and } f^{(\varphi,\psi)} \doteq \frac{\partial}{\partial\varphi\partial\psi} f \\
\delta_{jk} &= 1 \text{ for } j = k, \text{ and } \delta_{jk} = 0 \text{ for } j \neq k \\
\zeta_0 &\doteq 0 \\
\zeta_j &\doteq \exp(\alpha\vartheta + \beta_j), \ (j > 0) \\
\tau_j &\doteq \frac{\gamma_j}{1+\zeta_j} \\
\tau_j^{\beta} &\doteq \tau_j^{(\beta_j)}, \ \tau_j^{\beta\beta} \doteq \tau_j^{(\beta_j,\beta_j)} \\
\tau_j^{\gamma} &\doteq \tau_j^{(\ln\gamma_j)}, \ \tau_j^{\gamma\gamma} \doteq \tau_j^{(\ln\gamma_j, \ln\gamma_j)} \\
\tau_j^{\alpha\beta} &\doteq \tau_j^{(\alpha,\beta_j)}, \ \tau_j^{\alpha\gamma} \doteq \tau_j^{(\alpha,\ln\gamma_j)}, \ \tau_j^{\gamma\beta} \doteq \tau_j^{(\ln\gamma_j,\beta_j)} \\
\Xi &\doteq \Xi_\tau \doteq \sum_{k=0} \tau_k \\
\Xi_{\partial\tau} &\doteq \sum_{k=1} \tau_k^k \\
\Xi_{\beta\beta} &\doteq \sum_{k=1} \tau_k^{\beta\beta} \\
\pi_j &\doteq \frac{\tau_j}{\Xi}.
\end{aligned}
\tag{69}
$$

For instance τ_j^{jk} with $k \neq l$ equals 0. Then

$$
\begin{aligned}
\zeta_j^{(\alpha)} &= \vartheta\zeta_j && (70)\\
\zeta_j^{(\beta_j)} &= \zeta_j\\
\tau_j^{(\alpha)} &= \vartheta\tau_j^{\beta},\ \tau_j^{\beta} = \tau_j\left(\frac{\tau_j}{\gamma_j}-1\right),\ \tau_j^{\gamma} = \tau_j\\
\tau_j^{\alpha\alpha} &= \vartheta^2\tau_j^{\beta\beta}\\
\tau_j^{\alpha\beta} &= \vartheta\tau_j^{\beta\beta},\ \tau_j^{\beta\beta} = \tau_j^{\beta}\left(2\frac{\tau_j}{\gamma_j}-1\right)\\
\tau_j^{\alpha\gamma} &= \vartheta\tau_j^{\beta},\ \tau_j^{\beta\gamma} = \tau_j^{\beta},\ \tau_j^{\gamma\gamma} = \tau_j\\
\Xi^{\alpha} &= \vartheta\sum\tau_j^{\beta} = \vartheta\Xi_{\beta}\\
\Xi^{\beta_j} &= \tau_j^{\beta},\ \Xi^{\gamma_j} = \tau_j^{\gamma} = \tau_j\\
\Xi^{\alpha\alpha} &= \vartheta^2\sum\tau_j^{\beta\beta} = \vartheta^2\Xi_{\beta\beta}\\
\Xi^{\alpha\beta_j} &= \vartheta\tau_j^{\beta\beta},\ \Xi^{\beta_j\beta_j} = \tau_j^{\beta\beta}\\
\Xi^{\alpha\gamma_j} &= \tau_j^{\alpha\gamma} = \vartheta\tau_j^{\beta},\ \Xi^{\beta_j\gamma_j} = \tau_j^{\beta\gamma} = \tau_j^{\beta},\ \Xi^{\gamma_j\gamma_j} = \tau_j^{\gamma\gamma} = \tau_j\\
\pi_0^{(\lambda)} &= -\frac{\Xi^{(\lambda)}}{\Xi^2}\\
\pi_j^{(\lambda)} &= \frac{\tau_j^{(\lambda)}\Xi - \tau_j\Xi^{(\lambda)}}{\Xi^2}\\
\pi_0^{(\lambda,\kappa)} &= -\frac{\Xi^{(\lambda,\kappa)}\Xi - 2\Xi^{(\lambda)}\Xi^{(\kappa)}}{\Xi^3}\\
\pi_j^{(\lambda,\kappa)} &= \frac{\left(\begin{array}{c}\left(\tau_j^{(\lambda,\kappa)}\Xi - \tau_j^{(\lambda)}\Xi^{(\kappa)} - \tau_j^{(\kappa)}\Xi^{(\lambda)} - \tau_j\Xi^{(\lambda,\kappa)}\right)\Xi + \\ 2\tau_j\Xi^{(\lambda)}\Xi^{(\kappa)}\end{array}\right)}{\Xi^3}.
\end{aligned}
$$

Applying the Formulas for the first derivatives we have

$$\begin{aligned}
\pi_0^{(\alpha)} &= -\vartheta\frac{\Xi_\beta}{\Xi^2} \\
\pi_0^{(\beta_k)} &= -\frac{\tau_k^\beta}{\Xi^2} \\
\pi_0^{(\ln\gamma_k)} &= -\frac{\tau_k}{\Xi^2} \\
\pi_j^{(\alpha)} &= \frac{\vartheta\left(\tau_j^\beta\Xi - \tau_j\Xi_\beta\right)}{\Xi^2} = \frac{\nu_j^\alpha}{\Xi^2} \\
\pi_j^{(\beta_k)} &= \frac{\delta_{jk}\tau_j^\beta\Xi - \tau_j\tau_k^\beta}{\Xi^2} = \frac{\nu_j^{\beta_k}}{\Xi^2} \\
\pi_j^{(\ln\gamma_k)} &= \frac{\delta_{jk}\tau_j\Xi - \tau_j\tau_k}{\Xi^2} = \frac{\nu_j^{\ln\gamma_k}}{\Xi^2}.
\end{aligned} \tag{71}$$

And the second derivatives of $\underline{\pi}$ are given by

$$\begin{aligned}
\pi_0^{(\alpha,\alpha)} &= -\vartheta^2\frac{\Xi_{\beta\beta}\Xi - 2\Xi_\beta^2}{\Xi^3} \\
\pi_0^{(\alpha,\beta_l)} &= -\vartheta\frac{\tau_l^{\beta\beta}\Xi - 2\tau_l^\beta\Xi_\beta}{\Xi^3} \\
\pi_0^{(\alpha,\ln\gamma_l)} &= -\vartheta\frac{\tau_l^\beta\Xi - 2\tau_l\Xi_\beta}{\Xi^3} \\
\pi_0^{(\beta_k,\beta_l)} &= -\frac{\delta_{kl}\tau_k^{\beta\beta}\Xi - 2\tau_k^\beta\tau_l^\beta}{\Xi^3} \\
\pi_0^{(\beta_k,\ln\gamma_l)} &= -\frac{\delta_{kl}\tau_k^\beta\Xi - 2\tau_k^\beta\tau_l}{\Xi^3} \\
\pi_0^{(\ln\gamma_k,\ln\gamma_l)} &= -\frac{\delta_{kl}\tau_k\Xi - 2\tau_k\tau_l}{\Xi^3}
\end{aligned} \tag{72}$$

$$
\begin{aligned}
\pi_j^{(\alpha,\alpha)} &= \frac{\vartheta^2\left(\tau_j^{\beta\beta}\Xi - \tau_j \Xi_{\beta\beta}\right)\Xi - 2\vartheta\nu_j^{\alpha}\Xi_\beta}{\Xi^3} \\
\pi_j^{(\alpha,\beta_l)} &= \frac{\vartheta\left(\tau_j^{\beta}\tau_l^{\beta} - \tau_j\tau_l^{\beta\beta} + \delta_{jl}\left(\tau_j^{\beta\beta}\Xi - \tau_j^{\beta}\Xi_\beta\right)\right)\Xi - 2\nu_j^{\alpha}\tau_l^{\beta}}{\Xi^3} \\
\pi_j^{(\alpha,\ln\gamma_l)} &= \frac{\vartheta\left(\tau_j^{\beta}\tau_l - \tau_j\tau_l^{\beta} + \delta_{jl}\left(\tau_j^{\beta}\Xi - \tau_j\Xi_\beta\right)\right)\Xi - 2\nu_j^{\alpha}\tau_l}{\Xi^3} \\
\pi_j^{(\beta_k,\beta_l)} &= \frac{\left(\delta_{jkl}\tau_j^{\beta\beta}\Xi + \left(\delta_{jk} - \delta_{jl}\right)\tau_k^{\beta}\tau_l^{\beta} - \delta_{kl}\tau_j\tau_k^{\beta\beta}\right)\Xi - 2\nu_j^{\beta_k}\tau_l^{\beta}}{\Xi^3} \\
\pi_j^{(\beta_k,\ln\gamma_l)} &= \frac{\left(\delta_{jkl}\tau_j^{\beta}\Xi + \left(\delta_{jk} - \delta_{jl}\right)\tau_k^{\beta}\tau_l - \delta_{kl}\tau_j\tau_k^{\beta}\right)\Xi - 2\nu_j^{\beta_k}\tau_l}{\Xi^3} \\
\pi_j^{(\ln\gamma_k,\ln\gamma_l)} &= \frac{\left(\delta_{jkl}\tau_j\Xi + \left(\delta_{jk} - \delta_{jl}\right)\tau_k\tau_l - \delta_{kl}\tau_j\tau_k\right)\Xi - 2\nu_j^{\ln\gamma_k}\tau_l}{\Xi^3}.
\end{aligned}
\tag{73}
$$

Appendix B : The expected value of $\ln(q), \ln(\frac{q}{1-q})$, and $\ln(\frac{q}{1-q})^2$ with q Beta distributed

Below we give a proof of results part of which was already proved in a different way by Verhelst (1998).

Denote $\frac{\partial}{\partial \vartheta} f(\vartheta)$ as $f'(\vartheta)$, $\frac{\partial}{\partial \vartheta} \ln p(q;\vartheta)$ as $(\ln p)'(q)$, and let $\pi = \pi(\vartheta)$ be a function of ϑ, and let q be $Beta_\vartheta \doteq Beta(\mu\pi, \mu(1-\pi))$ distributed. (Note that for $Beta(\alpha,\beta)$, $\mu = (\alpha+\beta)$, $\pi = \frac{\alpha}{\alpha+\beta}$). Then we have that

$$\ln p(q) = (\mu\pi - 1)\ln q + (\mu(1-\pi) - 1)\ln(q-1) - c(\vartheta), \tag{74}$$

with

$$c(\vartheta) = \ln\Gamma(\mu\pi) + \ln\Gamma(\mu(1-\pi)) - \ln\Gamma(\mu). \tag{75}$$

Now

$$(\ln p)'(q) = \mu\pi' \ln\frac{q}{1-q} - c'(\vartheta), \tag{76}$$

with

$$c'(\vartheta) = \mu\pi' \left[\psi(\mu\pi) - \psi(\mu(1-\pi))\right]. \tag{77}$$

Because

$$\mathcal{E}_\vartheta \left[(\ln p)'(q)\right] = 0, \tag{78}$$

where $\mathcal{E}_\vartheta$ denotes expectation over $Beta_\vartheta = p(.;\vartheta)$, we then have that

$$\mathcal{E}_\vartheta \ln\frac{q}{1-q} = \psi(\mu\pi) - \psi(\mu(1-\pi)) = \frac{c'(\vartheta)}{\mu\pi'}. \tag{79}$$

Therefore,

$$\mathcal{E}_\vartheta \ln q = \psi(\mu\pi) - f(\mu), \tag{80}$$

where $f(\mu)$ at most depends on μ. Using Formula (79) and taking the limit for $\pi \to 0$, and therefore also for $q \to 0$, we find that $f(\mu) = \psi(\mu)$.

Further, we have that the expected information about ϑ equals

$$\begin{aligned}
\mathcal{E}_\vartheta \left[\left((\ln p)'(q) \right)^2 \right] &= \mathcal{E}_\vartheta \left[\left(\mu\pi' \ln \frac{q}{1-q} - c'(\vartheta) \right)^2 \right] \\
&= (\mu\pi')^2 \, Var_\vartheta \left[\ln \frac{q}{1-q} \right],
\end{aligned} \tag{81}$$

which also equals

$$-\mathcal{E}_\vartheta \left[(\ln p)''(q) \right] = -\mathcal{E}_\vartheta \left[\mu\pi'' \ln \frac{q}{1-q} - c''(\vartheta) \right], \tag{82}$$

where

$$c''(\vartheta) = \mu\pi'' \left(\psi(\mu\pi) - \psi(\mu(1-\pi)) \right) + (\mu\pi')^2 \left(\psi'(\mu\pi) + \psi'(\mu(1-\pi)) \right), \tag{83}$$

so that, using Formula (79), we have that the expected information about ϑ is given by

$$-\mathcal{E}_\vartheta \left[(\ln p)''(q) \right] = (\mu\pi')^2 \left(\psi'(\mu\pi) + \psi'(\mu(1-\pi)) \right), \tag{84}$$

and that

$$Var_\vartheta \left[\ln \frac{q}{1-q} \right] = \psi'(\mu\pi) + \psi'(\mu(1-\pi)). \tag{85}$$

Appendix C: Limits of $\mathbf{I}^*_{BB}/\mathbf{I}^*_{DM}$ for $\pi_0 \to 0$, and $\pi_0 \to 1$

From Formula (22) we have that

$$\frac{I_{BB}(\vartheta;\mu,N)}{I_{DM}(\vartheta;\mu,N)} = \frac{\sum_j \left(\mu\pi'_j\right)^2 \mathcal{E}_\vartheta b(n_j,\pi_j)}{\sum_j \left(\mu\pi'_j\right)^2 \mathcal{E}_\vartheta d(n_j,\pi_j)}. \tag{86}$$

Assuming from now equal trace lines for the distractors, and using $\sum \pi_j = 1$, and $\sum \pi'_j = 0$ this gives

$$\frac{I^*_{BB}(\vartheta;\mu,N)}{I^*_{DM}(\vartheta;\mu,N)} = \frac{(\mu\pi'_0)^2 \left(\mathcal{E}_\vartheta b(n_0,\pi_0) + \frac{1}{J-1}\mathcal{E}_\vartheta b(n_1,\frac{1-\pi_0}{J-1})\right)}{(\mu\pi'_0)^2 \left(\mathcal{E}_\vartheta d(n_0,\pi_0) + \frac{1}{J-1}\mathcal{E}_\vartheta d(n_1,\frac{1-\pi_0}{J-1})\right)}. \tag{87}$$

Denote by $p_j(n)$ the Beta-Binomial$(\vartheta;\mu,N)$ probability that n taws are put on option j. Using Formulas (14) and (15) and realizing that for $\pi_0 \to 0$ in $\mathcal{E}_\vartheta b(n_0,\pi_0)$ and $\mathcal{E}_\vartheta d(n_0,\pi_0)$ as well the summand $p_0(1)\frac{1}{(\mu\pi_0)^2}$ dominates it readily follows that $\lim_{\pi_0\to 0} \frac{I^*_{BB}(\vartheta;\mu,N)}{I^*_{DM}(\vartheta;\mu,N)} = 1$.

In the same vein for $\pi_0 \to 1$. First note that

$$p_0(n_0) = (J-1)p_1(N-n_0), \tag{88}$$

and therefore, that

$$\frac{1}{J-1}\mathcal{E}_\vartheta d(n_1,\frac{1-\pi_0}{J-1}) = \sum_{n_0=0}^{N} p_0(n_0) \sum_{i=0}^{N-n_0-1} \left(\frac{1}{\mu(1-\pi_0)+i(J-1)}\right)^2. \tag{89}$$

For $\pi_0 \to 1$, we have that for I^*_{BB} the sum is dominated by $2p_0(N-1)\frac{1}{(\mu(1-\pi_0))^2}$, one from $\mathcal{E}_\vartheta b(n_0,\pi_0)$ and the same summand from $\mathcal{E}_\vartheta b(n_1,\frac{1-\pi_0}{J-1})$, whereas I^*_{DM} is dominated only by $p_0(N-1)\frac{1}{(\mu(1-\pi_0))^2}$ from $\mathcal{E}_\vartheta d(n_1,\frac{1-\pi_0}{J-1})$, and it follows that $\lim_{\pi_0\to 1} \frac{I^*_{BB}(\vartheta;\mu,N)}{I^*_{DM}(\vartheta;\mu,N)} = 2$.

References

Abramowitz, M., & Stegun, I.A. (1972). *Handbook of mathematical functions*. New York: Dover.

Aitchinson, W., & Silvey, S.D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics, 28*, 813-828.

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

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note. *The American Statistician, 36*, 3, part 1, 153-157.

Dirkzwager, A. (1975). Computer-based testing with automatic scoring based on subjective probabilities. In: O. Lecarne & R. Lewis, (Eds.). Computers in education. Amsterdam: North Holland.

Dirkzwager, A. (1996). Testing with personal probabilities; eleven year olds can correctly estimate their personal probabilities. *Educational & Psychological Measurement, 56*, 957-971.

Glas, C.A.W. (1997). Testing the generalized partial credit model. In: M. Wilson, G. Engelhard, Jr., & K. Draney (Eds.). *Objective measurement, theory into practice, Vol. 4*. New Jersey: Ablex Publishing Corporation.

Glas, C.A.W. (1999). Modification indices for the 2-PL and the nominal response model. *Psychometrika, 64*, 273-294.

Glas, C.A.W., & Verhelst, N.D. (1995). Tests of fit for polytomous Rasch models. In: G.H. Fischer, & I.W. Molenaar (Eds.). *Rasch models, their foundation, recent developments and applications*. New York: Springer.

Jansen, R., & De Boeck, P. (1998). *Modelling partial knowledge in multiple-choice tests using elimination scoring.* Research Report KU Leuven, Report nr. 98-1.

Johnson, N.L., & Kotz, S. (1969). *Discrete distributions*. New York: Wiley.

Louis, Th., A. (1982). Finding the information matrix when using the EM- algorithm. *Journal of the Royal Statistical Society, Series B, 44*, 226-233.

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. (1992). *Numerical recipes in Pascal*. Cambridge: Cambridge University Press.

Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings* *of the Cambridge Philosophical Society, 44*, 50-57.

Thissen, D., & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika, 49*, 501-519.

Verhelst, N.D. (1998). *The expectation of ln(x) when x is beta distributed.* Note June 5, Arnhem: Cito.

Verhelst, N.D., Glas, C.A.W., & Verstralen, H.H.F.M. (1995). *OPLM: One-parameter logistic model. computer program and manual.* Arnhem: Cito.

Verhelst, N.D., & Verstralen, H.H.F.M. (in press). The Rasch model and multiple choice items. *Journal of Educational & Behavioral Statistics.* (Under editorial review).

Verstralen, H.H.F.M. (1997). *A logistic latent class model for multiple choice items.* Measurement & Research Department Reports, 97-1. Arnhem: Cito.

Verstralen, H.H.F.M., Verhelst, N.D., & Bechger, T.M. (2000). *A double hazard model for mental speed.* Measurement & Research Department Reports, 2000-1. Arnhem: Cito.

Verstralen, H.H.F.M., & Verhelst, N.D. (in press). *A latent response model for multiple choice questions.*

Recent Measurement and Research Department Reports:

2000-1 H.H.F.M. Verstralen, N.D. Verhelst, & T.M. Bechger. A Double Hazard Model for Mental Speed.

2000-2 T.M. Bechger, N.D. Verhelst, & H.H.F.M. Verstralen. Identifiability of Non-Linear Logistic Test Models.