Journal of Statistical Software

May 2007, Volume 20, Issue 7. http://www.jstatsoft.org/

Modeling of Responses and Response Times with the Package cirt

Jean-Paul Fox
University of Twente

Rinke Klein Entink
University of Twente

Wim van der Linden
University of Twente

Abstract

In computerized testing, the test takers' responses as well as their response times on the items are recorded. The relationship between response times and response accuracies is complex and varies over levels of observation. For example, it takes the form of a trade-off between speed and accuracy at the level of a fixed person but may become a positive correlation for a population of test takers. In order to explore such relationships and test hypotheses about them, a conjoint model is proposed. Item responses are modeled by a two-parameter normal-ogive IRT model and response times by a lognormal model. The two models are combined using a hierarchical framework based on the fact that response times and responses are nested within individuals. All parameters can be estimated simultaneously using an MCMC estimation approach. A R-package for the MCMC algorithm is presented and explained.

Keywords: Hierarchical IRT model, MCMC, response times, FORTRAN.

1. Introduction

When computerized tests are administered, not only are the responses to the test items but also the times used to produce them are automatically recorded. The information in the response times may help to improve routine operations in testing, such as item calibration, adaptive item selection, latent ability estimation, as well as to explore and measure factors that influence the performances on the test.

The issue of how to model response times has been approached from three different angles. One approach is to model the response times with time parameters added to a regular item response theory (IRT) model (see, e.g., Roskam, 1997; Thissen, 1983; and Verhelst, Verstraalen, and Jansen, 1997). A second approach is characterized by modeling the response times separately from the responses (see, e.g., Maris, 1993, and Scheiblechner, 1979). Van der Linden (2006) discusses a selection of these models for response times on test items. In a third approach, introduced in van der Linden (2007), the response times and responses are modeled hierarchically. At the first level, both the distributions of the responses and response times are assumed to follow separate models, each with a different set of person and item parameters. The person parameters represent the speed and accuracy (or ability) of the test taker on the items. A test taker's choice of speed and accuracy is generally constrained by a tradeoff. But since the speed and accuracy is assumed to be stationary during the test, the tradeoff can be ignored. Hence, at this first level of modeling, the item responses and response times can be assumed to be conditionally independent given the speed and accuracy parameters. However, at the second level, these parameters are allowed to be dependent. This leads to a hierarchical modeling framework in which the relation between speed and accuracy is dependent on the level of modeling.

Since response times have a natural lower bound at zero, their logarithm is modeled. Their distribution is assumed to be normal. The choice of a lognormal distribution is a classic one in response-time research. For response times on test items, it was made earlier, for example, by Thissen (1983), Schnipke and Scrams (1997), and van der Linden, Scrams, and Schnipke (1999). Each of these studies showed a good fit of response times to a lognormal distribution. In the present paper, both the binomial distribution of the responses and the normal distribution of the log response times are given a traditional item-response theory (IRT) parameterization. The binomial parameter for the responses has the structure of the two-parameter normal-ogive model (Lord and Novick, 1968). The distribution of the response times has a parameterization close to that of an IRT model for continuous response data (see, e.g., Samejima, 1973; Shi and Lee, 1998). Since the responses and response times are conditionally independent, their joint distribution is the product of a binomial and a normal distribution. This product can be considered as a conjoint IRT model for the analysis of discrete and continuous data for measuring test takers's speed and ability on test items.

A novel approach to the necessity of introducing identifying restrictions for the conjoint model is followed. In this approach, the restrictions are incorporated in the prior structure such that both the model is identified and a Gibbs sampler for estimating the model parameters can be used. The approach facilitates the use of informative proper priors as well as a Bayes factor for testing statistical hypothesis. The Gibbs sampler was programmed in FORTRAN and can be used in R with a package of functions called **cirt**. The package enables users to model patterns of responses and response times as a conjoint IRT model and to estimate and check the model. In a simulation study, the beneficial effects of modeling response-time data jointly with response data were assessed by comparing the accuracies of the ability estimates in a stand-alone IRT and a conjoint IRT approach. This was done for different covariances between the speed and ability parameter, different sample sizes, and different numbers of items.

The model is described in Section 2. The implementation of the Gibbs sampler for estimating the model parameters is described in Section 3. In the next section, a brief overview of procedures for testing the fit of the model is given. The package **cirt** is described in Section 5; the description includes a full listing of the input and output variables. The simulation study is reported in Section 6. Finally, a few possible generalizations are formulated.

2. A conjoint IRT modeling approach

A hierarchical modeling procedure is followed. At the lowest level, separate models are defined for the responses and response times. At a second level, a distributional structure is defined for the model parameters. Subsequently, hyperprior distributions are specified for the parameters of these distributions.

2.1. Models at level 1

Item responses to a set of items indexed $k = 1, \ldots, K$ are taken to be stored in an $N \times K$ data matrix $\mathbf{y}$. The response patterns are exploited to characterize both the test takers and the items. A two-parameter IRT model is used to define a mathematical relationship between the probabilities of the responses and the person and item parameters (see, e.g., Lord and Novick, 1968). Let θ_i denote the ability of test taker i. Then, the probability of a correct response to item k is defined as:

$$P(y_{ik} = 1 \mid \theta_i, a_k, b_k) = \Phi(a_k\theta_i - b_k), \tag{1}$$

where a_k and b_k are generally known as the discrimination parameter and difficulty parameter of item k, respectively, and $\Phi(\cdot)$ denotes the normal cumulative distribution function.

Response-time distributions have a natural lower-bound at zero and, for that reason, are skewed to the right. A lognormal distribution is used to model the response times which are taken to be stored in an $N \times K$ matrix $\mathbf{t}$. It is assumed that each respondent chooses to complete the items at a speed that can be represented by a parameter denoted as ζ_i. The time needed to complete an item also depends on item characteristic parameters. They are denoted as ϕ_k and λ_k, and can be seen as a discrimination and time-intensity parameter, respectively. We introduce a random variable defined as

$$\begin{aligned} t_{ik} &= \exp\bigl(z_{ik} + \phi_k\bigl(\tilde{\lambda}_k - \zeta_i\bigr)\bigr) && (2)\\ &= \exp\bigl(z_{ik} + \lambda_k - \phi_k\zeta_i\bigr), && (3)\end{aligned}$$

where $z_{ik} \sim \mathcal{N}(0, \sigma_t^2)$. It follows that the response time is distributed lognormally with mean $-\phi_k\zeta_i + \lambda_k$ and variance one. Its distribution function can be given as

$$\begin{aligned} P\bigl(t_{ik} \leq t'_{ik})\bigr) &= P\bigl(\log t_{ik} \leq \log t'_{ik}\bigr) && (4)\\ &= P\bigl(z_{ik} + \lambda_k - \phi_k\zeta_i \leq \log t'_{ik}\bigr) && (5)\\ &= P\bigl(z_{ik} \leq \log t'_{ik} - \bigl(\lambda_k - \phi_k\zeta_i\bigr)\bigr) && (6)\\ &= \Phi\bigl(\log t'_{ik} - \bigl(\lambda_k - \phi_k\zeta_i\bigr)\bigr), t'_{ik} > 0. && (7)\end{aligned}$$

The mean of the random variable t_{ik} is the value t'_{ik} such that

$$\Phi\bigl(\log t'_{ik} - \bigl(\lambda_k - \phi_k\zeta_i\bigr)\bigr) = .5, \tag{8}$$

and hence the mean response time for individual i to item k equals $\exp(\lambda_k - \phi_k\zeta_i)$. Hence, increasing the time intensity λ_k leads to a positive shift of the location of the time distribution on the item. Likewise, an increase in the speed parameter ζ_i leads to a negative shift.

As already noted, modeling response times as a lognormal distribution is a classic choice. A lognormal model with a simpler decomposition of the mean parameter was proposed by

Schnipke and Scrams (1997). However, their model was not used to describe the distribution of the response time for a fixed person and item but as a convenient summary of the empirical distributions of the times on the items in a bank across its history of test takers. The parameterization in (2)–(3) corresponds closely to that of the two-parameter IRT model for continuous responses developed by Samejima (1973).

An implicit assumption of the model in (2)-(3) is that the speed parameter remains constant during the test. This means that, whatever the conditions under which the test is taken, the test takers are assumed to settle on a level of speed at the beginning of the test and then stick to it. Because of this feature, the model is able to deal with response times collected both when the test takers know that their response times are observed and when they do not know. However, the model is unable to deal with changes in speed, for example, due to fatigue or the adoption of a new strategy during the test.

2.2. Hierarchical structure at levels 2 and 3

A bivariate normal distribution is defined for the ability and speed parameters of the test takers,

$$(\theta, \zeta) \sim \mathcal{N}_2(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P)$$

where

$$\begin{aligned} \boldsymbol{\mu}_P &= (\mu_\theta, \mu_\zeta) \\ \boldsymbol{\Sigma}_P &= \begin{pmatrix} \sigma_\theta^2 & \rho \\ \rho & \sigma_\zeta^2 \end{pmatrix}. \end{aligned}$$

Parameter ρ denotes the covariance between the person parameters. The distribution is postulated to be empirical but the postulate can be interpreted in two different ways. First, it can be considered to represent a population of persons that take the test. The distribution can then be used as the sampling distribution of a random test taker from the population. Second, it can be considered as a direct approximation of the distribution of the person parameters in the data set. From a Bayesian perspective, if the test takers can be treated as exchangeable, the distribution can then be used as a common prior for the person parameters. Although both interpretations lead to formally identical procedures, throughout this paper we will use the terminology associated with the second interpretation.

As a hyperprior for the covariance matrix $\boldsymbol{\Sigma}_P$, an inverse-Wishart distribution with degrees of freedom ν_P and scale parameter V_P is chosen. The question of how to specify the vector of means $\boldsymbol{\mu}_P$ is addressed in the next section.

In the same way, a multivariate normal distribution is specified for the item parameters of the response and response-time models,

$$(\log a_k, b_k, \log \phi_k, \lambda_k) \sim \mathcal{N}(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I) \tag{9}$$

This assumption allows for the fact that the item parameters within each measurement model usually correlate. In addition, it allows the item parameters to correlation between the measurement models. This feature helps us to deal with the fact that more difficult items typically require more time to complete than relatively easy items. Thus, we use the full covariance matrix

$$\begin{aligned}
\boldsymbol{\Sigma}_I &= \begin{pmatrix} \boldsymbol{\Sigma}_{a,b} & \boldsymbol{\Sigma}_{(a,b),(\phi,\lambda)} \\ \boldsymbol{\Sigma}_{(\phi,\lambda),(a,b)} & \boldsymbol{\Sigma}_{\phi,\lambda} \end{pmatrix} && (10)\\
&= \begin{pmatrix} \sigma_a & \sigma_{a,b} & \sigma_{a,\phi} & \sigma_{a,\lambda} \\ \sigma_{b,a} & \sigma_b & \sigma_{b,\phi} & \sigma_{b,\lambda} \\ \sigma_{\phi,a} & \sigma_{\phi,b} & \sigma_\phi & \sigma_{\phi,\lambda} \\ \sigma_{\lambda,a} & \sigma_{\lambda,b} & \sigma_{\lambda,\phi} & \sigma_\lambda \end{pmatrix}. && (11)
\end{aligned}$$

As a hyperprior for $(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I)$, a normal-inverse-Wishart distribution is chosen. That is,

$$\begin{aligned}
\boldsymbol{\Sigma}_I &\sim Inv-Wishart_{\nu_I}\left(V_I^{-1}\right) && (12)\\
\boldsymbol{\mu}_I \mid \boldsymbol{\Sigma}_I &\sim \mathcal{N}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_I/\kappa\right), && (13)
\end{aligned}$$

where ν_I and V_I are the degrees of freedom and scale matrix of the inverse Wishart distribution, $\boldsymbol{\mu}_0$ is the prior mean and κ the number of prior measurements.

Observe that the discrimination parameters in both measurement models are defined on a logarithmic scale; therefore, they are restricted to be positive.

The hierarchical structure induces a shrinkage estimation method for all measurement-model parameters. In fact, the two covariance structures introduce a relationship between the observed response and response-time data. A simultaneous estimation procedure will be used that allows us to use collateral information about each of the parameters: The response times serve as collateral information that is used to estimate the parameters of the response model. Conversely, the responses are used as collateral information when estimating the parameters of the response-time model. As a result, an increase in bias will be obtained but estimation error will be reduced (van der Linden, Klein Entink, and Fox 2007).

2.3. Incorporating identifying restrictions in the priors

Two-parameter IRT models are usually identified by fixing the zero and unit of its scale. Typically, this is done by setting the mean and variance of θ equal to a fixed value, or by putting similar restrictions on the item parameters.

The conjoint IRT model can be identified in the same way; the restrictions are now imposed on the mean vector and a covariance matrix. For example, it is sufficient to set $\boldsymbol{\mu}_P = 0$ and $\sigma_\theta^2 = 1$, and $\prod_k \phi_k = 1$. The first restriction sets the mean of the speed and ability parameters equal to zero, which implies that the mean of the time-intensity parameters of the items is equated to the mean log response times and the mean ability is absorbed in the mean of the item difficulties, respectively.

When using a Markov chain Monte Carlo (MCMC) algorithm for parameter estimation, we now have to sample from a restricted covariance matrix for the person parameters. The restrictions are therefore incorporated directly into the prior distributions. Observe that the prior has to assign probability one to $\boldsymbol{\mu}_P = 0$ and $\sigma_\theta = 1$ and hence, $\theta_i \sim \mathcal{N}(0,1)$. As the bivariate distribution of θ and ζ is normal, the same holds for the conditional distribution of $\zeta_i \mid \theta_i$,

$$\zeta_i \mid \theta_i \sim \mathcal{N}\left(\rho\theta_i, \sigma_\zeta^2 - \rho^2\right). \qquad (14)$$

Let $\tilde{\sigma}_\zeta^2 = \sigma_\zeta^2 - \rho^2$. Then, the following prior distributions are specified for ρ and $\tilde{\sigma}_\zeta^2$,

$$\begin{aligned} \rho &\sim \mathcal{N}\left(\bar{\rho}, \sigma_\rho^2\right) && (15) \\ \tilde{\sigma}_\zeta^{-2} &\sim \mathcal{G}\left(g_1, g_2\right), && (16) \end{aligned}$$

where $\mathcal{G}$ denotes the gamma distribution. For the multivariate case, McCullogh, Polson, and Rossi (2000) showed that there is a one-to-one correspondence between $\boldsymbol{\Sigma}_P$ and $\left(\sigma_\theta^2, \rho, \tilde{\sigma}_\zeta^{-2}\right)$. This way a prior distribution has been specified that assigns probability one to a diagonal element of the covariance matrix being equal to one.

3. An MCMC algorithm

An implementation of the Gibbs sampling algorithm, introduced by (Geman and Geman, 1984; Tanner and Wong, 1987), for the conjoint model is described. If all conditional posterior distributions are specified, a Gibbs sampler can be used to simulate draws from them, which results in a sequence of random variables that converges in distribution to the joint posterior distribution of all free parameters.

For the response model in (1), a data augmentation step is introduced to make Gibbs sampling feasible (see Albert, 1992). The model defines a nonlinear relationship between the probability of a correct response and the ability parameter. Let $f(a_k\theta_i - b_k)$ be the equivalent normal deviate of $a_k\theta_i - b_k$. Thus,

$$f\left(a_k\theta_i - b_k\right) = \Phi^{-1}\left(\Phi\left(a_k\theta_i - b_k\right)\right), \quad (17)$$

and, therefore,

$$P\left(z_{ik} \leq f\left(a_k\theta_i - b_k\right)\right) = \Phi\left(a_k\theta_i - b_k\right) \quad (18)$$

where z_{ik} is a normal random variable with distribution function Φ. As a result, a linear relationship is established between the new variable z_{ik} and the ability parameter. The normal ogive model can therefore be stated as a linear regression structure,

$$z_{ik} = a_k\theta_i - b_k + \epsilon_{ik}, \quad (19)$$

with $\epsilon_{ik} \sim \mathcal{N}(0,1)$. In addition, response y_{ik} is an indicator of z_{ik} being positive.

The vector of augmented data $\mathbf{z}_i = (z_{i1}, \ldots, z_{iK})$ minus the vector of difficulty parameters, $\mathbf{b}^t$, and the similar vector of response times $\log \mathbf{t}_i = (\log t_{i1}, \ldots, \log t_{iK})$ minus the vector of time intensity parameters, $\boldsymbol{\lambda}^t$, are stacked in a vector $\mathbf{z}_i^*$. Then, both measurement models can be presented as a linear regression structure,

$$\begin{aligned} \mathbf{z}_i^* &= \left(\mathbf{a} \oplus -\boldsymbol{\phi}\right)(\theta_i, \zeta_i)^t + \mathbf{e}_i && (20) \\ &= \mathbf{x}_I \boldsymbol{\Omega}_i + \mathbf{e}_i && (21) \end{aligned}$$

where $\mathbf{e}_i \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{e_K}\right)$ where $\boldsymbol{\Sigma}_{e_K} = I_K \oplus \sigma_t^2 I_K$.

Similarly, let $\mathbf{z}_k = (z_{1k}, \ldots, z_{nk})^t$ and the vector of log response times, $\log \mathbf{t}_i = (\log t_{1k}, \ldots, \log t_{nk})^t$, to item k be stacked in a vector $\mathbf{z}_k^*$. Define covariate matrices $\mathbf{H}_\theta$ and $\mathbf{H}_\zeta$ as $\left(\boldsymbol{\theta}, -\mathbf{1}_n\right)$ and as $\left(-\boldsymbol{\zeta}, \mathbf{1}_n\right)$, respectively. A regression structure for the item parameters can be presented as

$$\begin{aligned} \mathbf{z}_k^* &= \left(\mathbf{H}_\theta \oplus \mathbf{H}_\zeta\right)(a_k, b_k, \phi_k, \lambda_k)^t + \mathbf{e}_k && (22) \\ &= \mathbf{x}_P \boldsymbol{\Lambda}_k + \mathbf{e}_k, && (23) \end{aligned}$$

where $\mathbf{e}_k \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{e_n}\right)$ where $\mathbf{\Sigma}_{e_n} = I_n \oplus \sigma_t^2 I_n$.

MCMC algorithm

Initial values for the parameters can be obtained by fitting both measurement models separately using, for example, **BILOG-MG** (Zimowski, Muraki, Mislevy, and Bock, 1996) for the response model and maximum-likelihood estimation for the response-time model.

Step 1. According to (19), sample augmented data given the values for the item and ability parameters.

Step 2. Sample values for the item parameter from $p(\mathbf{\Lambda}_k \mid \mathbf{z}_k^*, \mathbf{\Omega}, \boldsymbol{\mu}_I, \mathbf{\Sigma}_I)$ for $(k = 1, \ldots, K)$. From Lindley and Smith (1972), it follows that a product of a normal likelihood and a normal prior leads to a normal posterior distribution. So, from (23) and (9), it follows that

$$\begin{aligned} p\left(\mathbf{\Lambda}_k \mid \mathbf{z}_k^*, \mathbf{\Omega}, \boldsymbol{\mu}_I, \mathbf{\Sigma}_I, \sigma_t^2\right) &= p\left(\mathbf{z}_k^* \mid \mathbf{\Lambda}_k, \mathbf{\Omega}, \sigma_t^2\right) p\left(\mathbf{\Lambda}_k \mid \boldsymbol{\mu}_I, \mathbf{\Sigma}_I\right) / p\left(\mathbf{z}_k^* \mid \mathbf{\Omega}, \boldsymbol{\mu}_I, \mathbf{\Sigma}_I\right) && (24) \\ &= \psi\left(\mathbf{\Lambda}_k \mid \boldsymbol{\mu}_{\Lambda_k}, \mathbf{\Sigma}_\Lambda\right), && (25) \end{aligned}$$

where $\mathbf{\Sigma}_\Lambda^{-1} = \mathbf{x}_P^t \mathbf{\Sigma}_{e_n}^{-1} \mathbf{x}_P + \mathbf{\Sigma}_I^{-1}$ and $\boldsymbol{\mu}_{\Lambda_k} = \mathbf{\Sigma}_\Lambda \left(\mathbf{x}_P^t \mathbf{\Sigma}_{e_n}^{-1} \mathbf{z}_k^* + \mathbf{\Sigma}_I^{-1} \boldsymbol{\mu}_I\right)$ and $\psi(\cdot)$ is the normal density function.

Step 3. Sample values for the ability speed parameters from a multivariate normal distribution. Analogous to Step 2, the full conditional posterior distribution is constructed from a multivariate normal likelihood, (21) and a multivariate normal prior distribution as

$$\begin{aligned} p\left(\mathbf{\Omega}_i \mid \mathbf{z}_i^*, \mathbf{\Lambda}, \boldsymbol{\mu}_P, \mathbf{\Sigma}_P, \sigma_t^2\right) &= p\left(\mathbf{z}_i^* \mid \mathbf{\Omega}_i, \mathbf{\Lambda}\right) p\left(\mathbf{\Omega}_i \mid \boldsymbol{\mu}_P, \mathbf{\Sigma}_P\right) / p\left(\mathbf{z}_i^* \mid \mathbf{\Lambda}, \boldsymbol{\mu}_P, \mathbf{\Sigma}_P\right) && (26) \\ &= \psi\left(\mathbf{\Omega}_i \mid \boldsymbol{\mu}_{\Omega_i}, \mathbf{\Sigma}_{\Omega_i}\right), && (27) \end{aligned}$$

where $\mathbf{\Sigma}_{\Omega_i}^{-1} = \mathbf{x}_I^t \mathbf{\Sigma}_{e_K}^{-1} \mathbf{x}_I + \mathbf{\Sigma}_P^{-1}$ and $\boldsymbol{\mu}_{\Omega_i} = \mathbf{\Sigma}_{\Omega_i}\left(\mathbf{x}_I^t \mathbf{\Sigma}_{e_K}^{-1} \mathbf{z}_i^* + \mathbf{\Sigma}_P^{-1} \boldsymbol{\mu}_P\right)$. The prior with the identifying restrictions is used for $\mathbf{\Sigma}_P$, that is,

$$\mathbf{\Sigma}_P = \begin{pmatrix} 1 & \rho \\ \rho & \sigma_\zeta^2 \end{pmatrix}. \tag{28}$$

Step 4. The hyperprior parameters are related to a multivariate normal model for the person parameters, $\left(\rho, \tilde{\sigma}_\zeta^2 = \sigma_\zeta^2 - \rho^2\right)$, or a multivariate model for the item parameters, $\left(\boldsymbol{\mu}_I, \mathbf{\Sigma}_I\right)$.

- From (14), it follows that the hyperprior parameters ρ and $\tilde{\sigma}_\zeta^2$ are the parameters of a linear regression of $\boldsymbol{\zeta}$ on $\boldsymbol{\theta}$ with a conjugate prior. Therefore, the conditional distribution of ρ is given by

$$p\left(\rho \mid \boldsymbol{\theta}, \boldsymbol{\zeta}, \tilde{\sigma}_\zeta^2, \bar{\rho}, \sigma_\rho^2\right) = \psi\left(\mu_\rho, \Sigma_\rho\right) \tag{29}$$

where $\mu_\rho = \Sigma_\rho^{-1}\left(\boldsymbol{\theta}^t \boldsymbol{\zeta} + \sigma_\rho^{-2} \bar{\rho}\right)$ and $\Sigma_\rho^{-1} = \tilde{\sigma}_\zeta^{-2}\left(\boldsymbol{\theta}^t \boldsymbol{\theta}\right) + \sigma_\rho^{-2}$. The full conditional distribution of $\tilde{\sigma}_\zeta^2$ is thus inverse gamma with shape and scale parameters $g_1 + n/2$ and $g_2 + \left(\boldsymbol{\zeta} - \rho\boldsymbol{\theta}\right)^t\left(\boldsymbol{\zeta} - \rho\boldsymbol{\theta}\right)/2$, respectively.

- The full conditional posterior distribution of $(\boldsymbol{\mu}_I, \boldsymbol{\Sigma}_I)$ has a normal-inverse-Wishart distribution (e.g., Gelman, Carlin, Stern, and Rubin, 2004). It follows that

$$p\big(\boldsymbol{\mu}_I \mid \boldsymbol{\Sigma}_I, \boldsymbol{\mu}_0, \boldsymbol{\Lambda}, V_I\big) = \psi\big(\big(\kappa\boldsymbol{\mu}_0 + K\bar{\boldsymbol{\Lambda}}\big)/(\kappa + K), \boldsymbol{\Sigma}_I/(K+\kappa)\big), \tag{30}$$

where $\bar{\boldsymbol{\Lambda}} = \sum_k \boldsymbol{\Lambda}_k/K$. Subsequently, the full conditional of $\boldsymbol{\Sigma}_I$ is an inverse Wishart with parameter $K + \nu_I$ and scale parameter $V_I + \sum_k \big(\boldsymbol{\Lambda}_k - \bar{\boldsymbol{\Lambda}}\big)\big(\boldsymbol{\Lambda}_k - \bar{\boldsymbol{\Lambda}}\big)^t + \frac{\kappa K}{\kappa + K}\big(\bar{\boldsymbol{\Lambda}} - \boldsymbol{\mu}_0\big)\big(\bar{\boldsymbol{\Lambda}} - \boldsymbol{\mu}_0\big)^t$.

- The full conditional distribution of diagonal element σ_t^2 of $\boldsymbol{\Sigma}_e$ is the inverse gamma with parameter $g_1 + nK/2$ and scale parameter $g_2 + 1/2\sum_{i,k}\left(\log t_{ik} - (\lambda_k - \phi_k\zeta_i)\right)^2$ using a conjugated inverse gamma prior with parameters g_1 and g_2.

4. Goodness of fit

The fit of the measurement models to response and response-time data can be assessed through residual analysis. The actual observation $\log t_{ik}$ is then evaluated under the posterior predictive density. That is, the probability of observing a value smaller than $\log t_{ik}$ can be estimated by

$$P\big(\log t_{ik}^* < \log t_{ik} \mid \mathbf{y}, \mathbf{t}\big) \approx \sum_m \Phi\big(\log t_{ik} \mid \zeta_i^{(m)}, \phi_k^{(m)}, \lambda_k^{(m)}\big)/M, \tag{31}$$

given M iterations of the MCMC algorithm. Probabilities close to zero or one correspond to observations that are unlikely under the model.

Aggregated observed values over persons or items can be used to check the fit of the model to specific items or persons. The probability of observing a response y_{ik} under the model equals

$$p_{ik} = \Phi\big(y_{ik} \mid \mathbf{y}, \mathbf{t}\big) \approx \sum_m \Phi\big(y_{ik} \mid \theta_i^{(m)}, a_k^{(m)}, b_k^{(m)}\big)/M. \tag{32}$$

If the model holds, the random variables p_{ik} follow a uniform distribution according to theorem on probability integral transformations. (That is, the probability of occurrence is the same for all values of p_{ik}). This can be tested. for instance, by comparing the estimated moments with the true moments of the uniform distribution, checking if equally spaced intervals contain equal numbers of values p_{ik}, or by using the implicit smoothing in their empirical cumulative distributions and plotting these against the identity line.

Specific model restrictions can be tested via the Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, and van der Linde, 2002). The DIC is an integrated measure of model fit and complexity. It was developed for comparing complex hierarchical models where the number of parameters is not clearly defined. Let $\boldsymbol{\omega}$ denote the set of model parameters, then a deviance is defined as

$$D(\boldsymbol{\omega}) = -2\log p\big(\mathbf{y}, \mathbf{t} \mid \boldsymbol{\omega}\big) + 2\log p\big(\mathbf{y}, \mathbf{t}\big). \tag{33}$$

When comparing models, it can be assumed without loss of generality that $p\big(\mathbf{y}, \mathbf{t}\big) = 1$ for all models. The effective number of parameters in the models is defined by

$$p_D = \overline{D(\boldsymbol{\omega})} - D(\hat{\boldsymbol{\omega}}), \tag{34}$$

where $\overline{D(\boldsymbol{\omega})}$ is the posterior mean deviance and $\hat{\boldsymbol{\omega}}$ is the posterior mean of the parameters. The DIC can now be formulated as

$$\begin{aligned} DIC &= D(\hat{\boldsymbol{\omega}}) + 2p_D \qquad (35)\\ &= \sum_{i,k}\Big[y_{ik}\log\Phi\big(\hat{a}_k\hat{\theta}_i - \hat{b}_k\big) + (1-y_{ik})\log\big(1-\Phi\big(\hat{a}_k\hat{\theta}_i - \hat{b}_k\big)\big) \qquad (36)\\ &\quad + \log\Phi\big(\log t_{ik} - \big(\hat{\lambda}_k - \hat{\phi}_k\hat{\zeta}_i\big)\big)\Big] + 2p_D. \end{aligned}$$

Due to the fact that $D(\boldsymbol{\omega})$ is available in closed form, the MCMC algorithm can be used to compute $\overline{D(\boldsymbol{\omega})}$ by taking the sample mean of the simulated values of $D(\boldsymbol{\omega})$. The term $D(\hat{\boldsymbol{\omega}})$ is computed by plugging the mean of the simulated values of $\boldsymbol{\omega}$ into $D(\cdot)$. In general, models with larger numbers of effective parameters, p_D, are penalized by the DIC. Hence, the criterion prefers models of less complexity.

5. Package cirt

This package for R (R Development Core Team 2006) contains three user-callable functions: one for simulating data according to a conjoint IRT model (`generate`), one for estimating the model via the MCMC estimation procedure (`estimate`), and one for summarizing the results (`summarize`). The functions are stored in the package called **cirt**. Descriptions of the functions as well as an example can be found in the help files by typing `?cirt`.

For larger samples and required number of iterations, the use of an MCMC algorithm can be time consuming. Its current implementation was therefore programmed in Visual Pro FORTRAN (version 8) (Intel 2004), whereas the **IMSL** FORTRAN statistics library (version 5) (Visual Numerics 2004) was used for random number generation and sampling from the probability distributions. In this way, a dynamic link library application was created, *irtrt.dll,* for use as a subprogram in R for Microsoft Windows.

Function `generate` has arguments `N` and `K` for the number of persons and items, respectively. Two optional arguments are `rho` and `corbl`; the former specifies the correlation between θ and ζ, the latter the correlation between item difficulty b and item-time intensity λ. Output of the function are the generated model parameters and response patterns, stored in a list in the following order: a, b, ϕ, λ, θ, ζ, $\mathbf{y}$, and $\mathbf{t}$.

Function `estimate` has arguments `Y`, `Time`, `N`, `K` and `iter` where `Y` denotes the $N \times K$ matrix of the responses and `Time` the $N \times K$ matrix of the log response times. Matrix $\mathbf{y}$ should contain 1 for a correct response, 0 for an incorrect response, and 9 for a missing observation. Missing data are always assumed to be missing by design; the estimates are based on the observed data only, no imputation method is used. Object *iter* specifies the desired number of iterations. Starting values are generated automatically. The output consists of a list containing the sampled values from the marginal posterior distributions of all model parameters and some of the model-fit criteria discussed in the previous section. The header of the function shows which parameter is stored where in the list. The optional argument `PL=1` restricts the item response model to the one-parameter normal-ogive model (i.e., with $a_k = 1$ for all k). The default is `PL=2`, which leads to estimation of the two-parameter item response model. The optional argument `index=1` restricts the variance of the speed parameter to one. This option can be used when, in spite of the identifying restrictions above, the model is still poorly identified.

Function `summary` has arguments `out` and `burnin`, where object `out` contains the list of sampled values from the marginal posterior distributions of structural model parameters (produced by `estimate`) and `burnin` determines the number of burn-in iterations for the MCMC chain. `summary` generates a report which gives the EAP estimates and posterior standard deviations of the model parameters. Further, a DIC estimate is given that can be used for model comparison. These estimates are based on the drawn samples where the first number of samples, as specified by `burnin`, are discarded as the burn-in period.

MCMC chains are always available as output of the function `estimate`. Their convergence can be checked using the **boa** software, which is available in library format from The Comprehensive R Archive Network (CRAN), at `http://cran.r-project.org/`. **boa** is an R/S-PLUS program that enables the computation of convergence diagnostics and statistical and graphical analysis of Monte Carlo sampling output. The **boa** software produces posterior estimates, trace plots, density plots, and several convergence diagnostics.

6. Response-time based IRT parameter estimation

Results from a simulation study to examine the performances of the conjoint IRT model under different configurations are presented. The primary interest was in the influence of response times on the parameter estimates for the response model. Specifically, the gain of efficiency of using response times as collateral information when estimating the ability parameters was assessed.

The following quantities were varied: number of items: 5, 10 and 20; number of persons: 500 and 1000); and the correlation between the ability and speed parameters: 0.00, 0.20, 0.30, 0.40 and 0.50. So, in total there were 30 different conditions. The item parameters were sampled from a multivariate normal distribution with mean $\boldsymbol{\mu}_0 = (1, 0, 1, 0)$ and covariance matrix with diagonal elements equal to .5 and off-diagonal elements equal to zero. The item parameters were chosen not to correlate in the present study (although their estimates were allowed to do so). Every condition was replicated ten times and the estimates in Table 1 to 3 are based on averages over these replications.

For each condition, the same (proper) prior distributions were specified. Covariance matrix V_I was chosen to be a diagonal matrix with elements .01 to specify vague information about the item parameters. In addition, $\boldsymbol{\mu}_0 = (0, 0, 0, 0)$, $\kappa = 10$, and $\nu_I = 4$. Finally, a vague normal prior for the correlation coefficient was specified with parameters $\bar{\rho} = 0$ and $\sigma^2_\rho = 10$. For each condition a total of 10 data sets were simulated. For each of the 10 data sets the MCMC procedure was iterated $12,000$ times and the first $2,000$ iterations were discarded when the means, variances, and Bayesian confidence intervals of the model parameters were estimated. The final parameter estimates are averaged values over the 10 data-specific estimates. The accuracy of the parameter estimates was investigated by comparing them to the true generating values. Mean squared errors were computed to summarizes the differences.

In Table 1, the estimates of correlation parameter ρ and their standard deviations for the different conditions are given. It can be seen that the correlations were estimated very accurately. Also, the accuracy increases with the number of items and persons.

In Table 2, the MSEs of the estimated ability and speed parameters are presented. As

		Correlation Coefficient				
Persons	Items	.00	.20	.30	.40	.50
$n = 500$	5	0.02(0.06)	0.23(0.06)	0.33(0.06)	0.41(0.06)	0.49(0.05)
	10	0.00(0.05)	0.19(0.05)	0.30(0.05)	0.41(0.05)	0.51(0.04)
	20	0.00(0.05)	0.21(0.05)	0.31(0.05)	0.40(0.05)	0.50(0.04)
$n = 1000$	5	0.00(0.04)	0.19(0.04)	0.31(0.04)	0.41(0.04)	0.50(0.04)
	10	0.00(0.04)	0.20(0.04)	0.30(0.04)	0.40(0.04)	0.50(0.03)
	20	0.01(0.03)	0.20(0.03)	0.30(0.03)	0.40(0.03)	0.50(0.03)

Table 1: Posterior estimate of correlation between ability and speed.

			Correlation Coefficient				
Variable	Persons	Items	.00	.20	.30	.40	.50
Ability	500	5	0.67	0.65	0.70	0.66	0.61
		10	0.46	0.42	0.38	0.37	0.38
		20	0.23	0.24	0.24	0.22	0.23
	1000	5	0.67	0.64	0.62	0.61	0.64
		10	0.44	0.41	0.43	0.39	0.38
		20	0.24	0.23	0.24	0.23	0.22
Speed	500	5	0.31	0.33	0.28	0.33	0.31
		10	0.16	0.20	0.16	0.18	0.16
		20	0.10	0.09	0.09	0.09	0.09
	1000	5	0.30	0.32	0.30	0.34	0.32
		10	0.17	0.19	0.18	0.16	0.16
		20	0.09	0.08	0.09	0.09	0.08

Table 2: Mean squared error of ability estimate and speed estimate.

		Correlation Coefficient				
Persons	Items	.00	.20	.30	.40	.50
500	5	0.67(1.01)	0.65(1.02)	0.70(1.06)	0.66(1.06)	0.61(1.09)
	10	0.46(1.01)	0.42(1.00)	0.38(1.03)	0.37(1.05)	0.38(1.07)
	20	0.23(1.00)	0.24(1.01)	0.24(1.02)	0.22(1.02)	0.23(1.04)
1000	5	0.67(1.00)	0.64(1.01)	0.62(1.02)	0.61(1.05)	0.64(1.10)
	10	0.44(1.00)	0.41(1.01)	0.43(1.01)	0.39(1.04)	0.38(1.06)
	20	0.24(1.00)	0.23(0.99)	0.24(1.01)	0.23(1.03)	0.22(1.04)

Table 3: MSE of ability estimate and relative efficiency as a ratio of MSE of the null model ability estimate over the conjoined IRT model ability estimate.

expected, the accuracy of the estimates of the ability and speed parameters increased with the number of items and persons, respectively. But the improvement was independent of the correlation between the parameters. Apparently, in our setup, the likelihood dominated the chosen priors too much to yield an effect for the correlation.

Our next results suggest that the test length moderates the impact of the correlation on the estimates of the person parameters. A comparison made was between the usual two-parameter normal ogive model without any additional information from the response times (null model) and the conjoint model in this paper (alternative model). Following de la Torre and Patz (2005), the relative efficiency was computed as the ratio of the MSE of the ability estimates under the two models. A ratio greater than one indicated higher efficiency of the conjoint ability estimates due to information in the response times. Table 3 shows both the MSEs of ability estimates under the null model and the relative efficiencies of the conjoint model. It can be seen that the MSEs under the null model do not change when the correlation or the number of persons increased but that better ability estimates were obtained for longer tests lengths. However, a higher correlation led to higher efficiency but the gain of efficiency decreased with the length of the test.

7. Concluding remarks

In computerized testing, response times can be easily obtained. A conjoint IRT model was proposed to deal with these data in addition to the regular responses. The model was incorporated in a hierarchical framework that integrates these two sources of information and an MCMC estimation procedure was presented to enable the simultaneous estimation of all model parameters. Although the framework is more complex as to its structure and estimation, its use can be beneficial when response times are observed, for example, for ability estimation.

The conjoint IRT model can be generalized to account for guessing (through the choice of a three-parameter IRT model for the responses) or for items with a polytomous response format. Other generalizations that might be interesting are a multidimensional ability structure and mixtures as distributions of the ability and speed parameters to capture possible differences in item solving strategies between test takers.

References

Albert JH (1992). “Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling.” *Journal of Educational Statistics*, **17**, 251–269.

de la Torre J, Patz RJ (2005). “Making the Most of What We Have: A Practical Application of Multidimensional Item Response Theory in Test Scoring.” *Journal of Educational and Behavioral Statistics*, **30**, 295–311.

Gelman A, Carlin JB, Stern HS, Rubin DB (2004). *Bayesian Data Analysis.* Chapman & Hall/CRC, New York, 2nd edition.

Geman S, Geman D (1984). “Stochastic Relaxation, Gibbs Distribution, and the Bayesian

Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Intel (2004). *Visual FORTRAN Compiler for Windows (Version 8.0)*. Intel Corporation, Santa Clara, CA. URL http://www.intel.com/.

Lindley DV, Smith AFM (1972). "Bayes Estimates for the Linear Model." *Journal of the Royal Statistical Society B*, **34**, 1–41.

Lord FM, Novick MR (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA.

Maris E (1993). "Adaptive and Multiplicative Models for Gamma Distributed Variables, and Their Application as Psychometric Models for Response Times." *Psychometrika*, **58**, 445–469.

McCullogh RE, Polson NG, Rossi PE (2000). "A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters." *Journal of Econometrics*, **99**, 173–193.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org/.

Roskam EE (1997). "Models for Speed and Time-Limit Tests." In WJ van der Linden, RK Hambleton (eds.), "Handbook of Modern Item Response Theory," pp. 187–208. Springer, New York.

Samejima F (1973). "Homogeneous Case of the Continuous Response Level." *Psychometrika*, **38**, 203–219.

Scheiblechner H (1979). "Specific Objective Stochastic Latency Mechanisms." *Journal of Mathematical Psychology*, **19**, 18–38.

Schnipke DL, Scrams DJ (1997). "Representing Response Time Information in Item Banks." *LSAC Computarized Testing Report 97-09*, Law School Admission Council, Newton, PA.

Shi JQ, Lee SY (1998). "Bayesian Sampling-based Approach for Factor Analysis Models with Continuous and Polytomous Data." *British Journal of Mathematical and Statistical Psychology*, **51**, 233–252.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society B*, **64**, 583–639.

Tanner MA, Wong WH (1987). "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association*, **82**, 528–550.

Thissen D (1983). "Timed Testing: An Approach Using Item Response Theory." In DJ Weiss (ed.), "Latent Trait Test Theory and Computarized Adaptive Testing," pp. 179–203. Academic Press, New York.

van der Linden WJ (2006). "A Lognormal Model for Response Times on Test Items." *Journal of Educational and Behavioural Statistics*, **31**, 181–204.

van der Linden WJ (2007). "A Hierarchical Framework for Modeling Speed and Accuracy on Test Items." *Psychometrika*. In press.

van der Linden WJ, Klein Entink RH, Fox JP (2007). "IRT Parameter Estimation with Response Times as Collateral Information." *Applied Psychological Measurement*. Submitted.

van der Linden WJ, Scrams DJ, Schnipke DL (1999). "Using Response-time Constraints to Control for Speededness in Computarized Adaptive Testing." *Applied Psychological Measurement*, **23**, 195–210.

Verhelst ND, Verstraalen HHFM, Jansen MG (1997). "A Logistic Model for Time Limit Tests." In WJ van der Linden, RK Hambleton (eds.), "Handbook of Modern Item Response Theory," pp. 169–185. Springer, New York.

Visual Numerics (2004). **IMSL** *Numerical Libraries (Version 5.0)*. Visual Numerics, Houston, Texas. URL http://www.vni.com.

Zimowski MF, Muraki E, Mislevy RJ, Bock RD (1996). **Bilog MG** *3 - Multiple-group IRT Analysis and Test Maintenance for Binary Items*. Scientific Software International, Inc., Lincolnwood, IL. URL http://www.ssicentral.com/.

Affiliation:

Jean-Paul Fox
University of Twente
Department of Research Methodology, Measurement and Data Analysis
Enschede, The Netherlands
E-mail: Fox@edte.utwente.nl
URL: http://users.edte.utwente.nl/Fox/

Rinke Klein Entink
University of Twente
E-mail: R.H.KleinEntink@gw.utwente.nl
URL: http://www.kleinentink.eu/

Wim van der Linden
University of Twente
E-mail: W.J.vanderLinden@gw.utwente.nl
URL: http://www.gw.utwente.nl/omd/en/members/vanderlinden.doc

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/

Volume 20, Issue 7 *Submitted:* 2006-10-01
May 2007 *Accepted:* 2007-02-22