Measurement and Research Department Reports 2004-5

An Introduction to the DA-T Gibbs Sampler for the Two-Parameter Logistic (2PL) Model and its Application

Gunter Maris
Timo M. Bechger

cito groep

Measurement and Research Department Reports 2004-5

An Introduction to the DA-T Gibbs Sampler for the Two-Parameter Logistic (2PL) Model and its Application

Gunter Maris

Timo M. Bechger

Citogroep
Arnhem, 2004

Abstract

The DA-T Gibbs sampler is proposed by Maris and Maris (2002) as a Bayesian estimation method for a wide variety of item response theory models. The present paper provides an expository account of the DA-T Gibbs sampler for the two-parameter logistic model. It further presents two applications that are of independent interest. The first concerns the estimation of classical test theory reliability. The second application is that the DA-T Gibbs sampler for the 2PL may be used to build Gibbs samplers for a wider class of item response theory models.

1. Introduction

Let $Y_{pi} = 1$ denote the event that person p gives the correct answer to item i, and θ_p his or her ability. Assume that there exists a latent response variable X_{pi} such that person p solves item i if X_{pi} is larger than a threshold δ_i. That is,

$$P(Y_{pi} = 1|\theta_p) = P(X_{pi} > \delta_i|\theta_p) \quad .$$

It is seen that the probability of a correct response depends on the threshold of the item as well as the ability of the respondent. The probability $P(Y_{pi} = 1|\theta_p)$ is called *the Item Response Function (IRF).*

Under the *two-parameter logistic (2PL) model* (Birnbaum, 1968), X_{pi} is assumed to follow a logistic distribution with mean $\alpha_i\theta_p$ and scale parameter $\beta = 1$ so that

$$\begin{aligned} P(X_{pi} > \delta_i|\theta_p, \alpha_i, \delta_i) &= \int_{-\infty}^{\infty} (x_{pi} > \delta_i) f(x_{pi}|\theta_p, \alpha_i) dx_{pi} \qquad (1) \\ &= \int_{-\infty}^{\infty} (x_{pi} > \delta_i) \frac{\exp(x_{pi} - \alpha_i\theta_p)}{[1 + \exp(x_{pi} - \alpha_i\theta_p)]^2} dx_{pi} \\ &= \frac{\exp(\alpha_i\theta_p - \delta_i)}{1 + \exp(\alpha_i\theta_p - \delta_i)} \quad , \end{aligned}$$

where $(x_{pi} > \delta_i)$ denotes an indicator variable that is one if $x_{pi} > \delta_i$, and zero otherwise. If the latent response variable X_{pi} is assumed to follow a normal distribution, we obtain what is known as *the two-parameter normal ogive (2NO) model.*

The *discrimination parameter* α_i determines how fast the probability of a correct answer changes as a function of ability. It is seen in Figure 1 that the IRFs of different items may cross if their discrimination parameters differ. If α_i is positive (negative), the probability of answering correctly is an increasing (decreasing) function of ability. Here, we allow both positive and negative values. It will be demonstrated below that it is easy to restrict the discrimination parameters to positive values. The *Rasch model* (Rasch, 1980) is a special case of the 2PL where all items have a discrimination parameter equal to one.

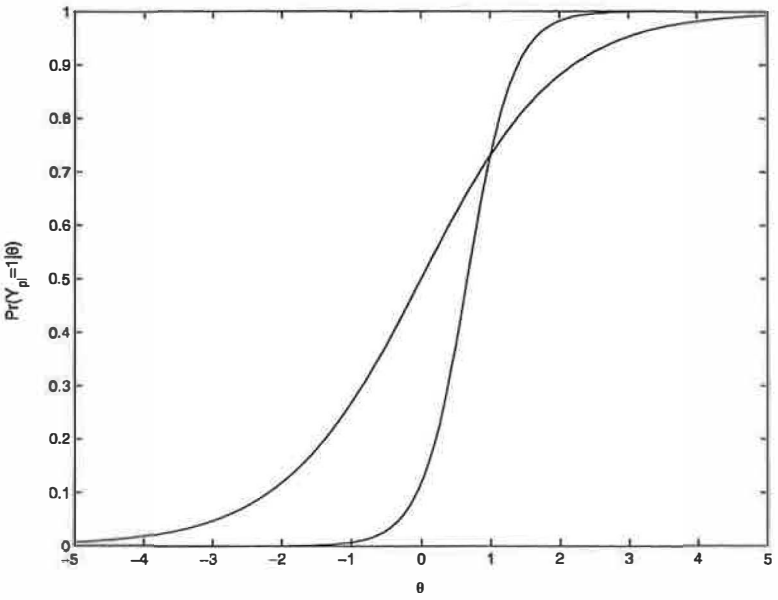

FIGURE 1.
IRFs for two 2PL items with different parameters.

As it stands, the 2PL is *unidentifiable*. Specifically,

$$P(Y_{pi} = 1|\theta_p, \alpha_i, \delta_i) = \frac{\exp(\alpha_i^*\theta_p^* - \delta_i^*)}{1 + \exp(\alpha_i^*\theta_p^* - \delta_i^*)}$$

where

$$\alpha_i^* = \alpha_i d, \quad \delta_i^* = \delta_i - \alpha_i c, \quad \theta_p^* = \frac{\theta_p - c}{d} \quad ,$$

and c and d are arbitrary constants. To deal with this indeterminacy we arbitrarily set $\alpha_1 = 1$, and $\delta_1 = 0$. This means that the item parameters must be interpreted relative to the first item.

The main purpose of this paper is to provide an expository account of Bayesian estimation of the 2PL focussing on the DA-T Gibbs sampler developed by Maris and Maris (2002). We provide references for further reading.

The outline of the paper is as follows. In the sections 2 to 4 we provide an expository account of the DA-T Gibbs sampler. Two new applications that go beyond the

estimation of the 2PL are discussed in Section 5. In paragraph 5.1, we demonstrate how the DA-T Gibbs sampler may be used to obtain an estimate of classical test theory reliability. In paragraphs 5.2 and 5.3, we demonstrate how the Gibbs sampler for the 2PL may be used to build Gibbs samplers for a wider class of models suited for polytomous items. The paper is concluded with a discussion in Section 6.

2. Gibbs Sampling

Let $\Lambda = (\Lambda_1, \ldots, \Lambda_m)$, $m \geq 2$, denote a vector of parameters. In Bayesian statistics, the unknown parameters are considered random variables. Bayes theorem states that *the posterior density* (the posterior, for short) of Λ given the observed data y is given by

$$f(\lambda|y) = \frac{f(y|\lambda)f(\lambda)}{f(y)} \quad ,$$

where $f(y|\lambda)$ denotes the likelihood function, and $f(y)$ the marginal likelihood function. *The prior density* $f(\lambda)$ (prior, for short) expresses substantive knowledge concerning the parameters prior to data collection. In Bayesian statistics, all inferences about the parameters are based upon the posterior.

The Gibbs sampler is an iterative procedure to generate parameter values $\lambda^{(0)}, \lambda^{(1)}, \ldots$ from the posterior. The first n generated values are discarded and the rest is considered to be a *dependent and identically distributed (did)* sample from the posterior. This means that

1. The distribution of $\Lambda^{(n+j)}$ given the data is the posterior for all $j > 0$.
2. Conditional upon the data, $\Lambda^{(n+j)}$ is *not* independent of $\Lambda^{(n+i)}$ for $(i \neq j)$.

We will now discuss *how* the Gibbs sampler works and *why*. Alternative explanations can be found, for instance, in Casella and George (1992), Tanner (1996), or Ross (2003). The reader is referred to Tierney (1994) for a more rigorous explanation.

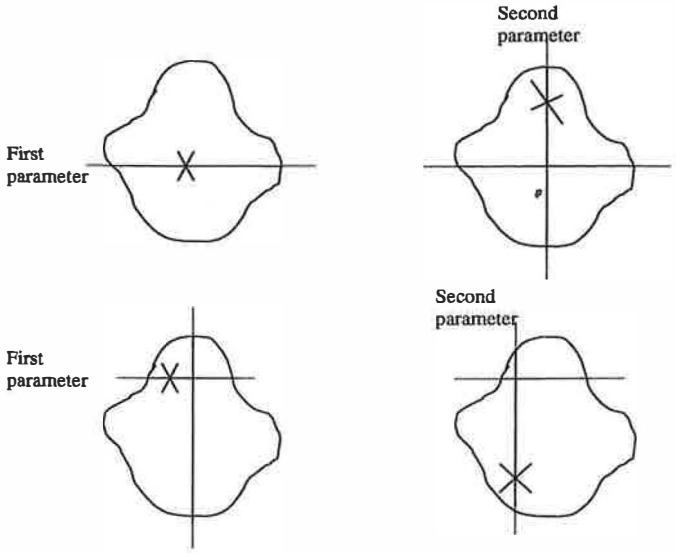

FIGURE 2.
Schematic representation of two iterations of the Gibbs sampler with two parameters

2.1. How

The procedure starts by choosing an initial value $\lambda^{(0)}$. Then, in each successive iteration, individual parameters are sampled independently from their so-called full conditional distribution. The order in which the parameters are sampled is arbitrary.

The full conditional distribution is the distribution given the observed data and the current value of all other parameters. Specifically, $\lambda_k^{(j+1)}$ (for $k = 1, \ldots, m$) is drawn from a density $f(\lambda_k | \lambda_{-k}^{(j)}, y)$, where

$$\lambda_{-k}^{(j)} \equiv (\lambda_1^{(j+1)}, \ldots, \lambda_{k-1}^{(j+1)}, \lambda_{k+1}^{(j)} \ldots, \lambda_m^{(j)})$$

To determine, up to a constant, the full conditional distribution of a parameter λ_k one may write down the density $f(\lambda_k, \lambda_{-k}^{(j)}, y)$ and remove all factors that are unrelated to λ_k.

An informal illustration is provided by Figure 2. The closed curve indicates *the support* of a two dimensional posterior; that is, the region containing possible values

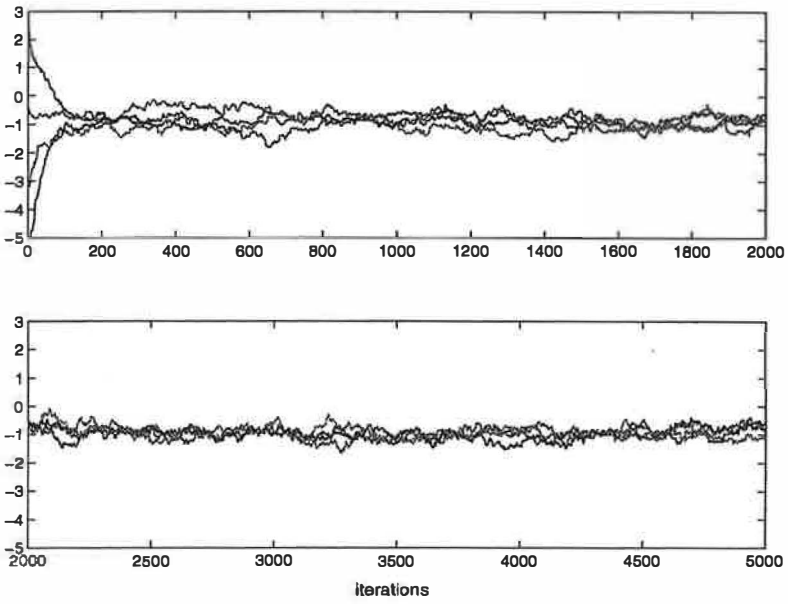

FIGURE 3.

Plot of sampled values against iterations.

of the parameters. The solid lines indicate the support of the full conditionals, and the crosses are simulated values. It is seen that the Gibbs sampler "walks" through the support of the posterior along horizontal and vertical lines. Note further that every region in the support can be reached.

With a *did* sample from the posterior we may use the *Monte Carlo* (MC) method to calculate an unbiased estimate of the posterior expectation of any function $g(\lambda, y)$:

$$\int g(\lambda, y) f(\lambda|y) d\lambda \approx \frac{1}{n_s} \sum_j g(\lambda^{(j)}, y) \quad ,$$

where n_s denotes the number of sampled values. That is, we approximate the expectation by the sample mean. The posterior probability that a parameter is smaller or equal to a constant t, for example, is estimated by

$$\frac{1}{n_s} \sum_j \left(\lambda_k^{(j)} \leq t\right) \quad .$$

The variance of the estimator of the posterior expectation can be estimated by the

variance over independent replications of the Gibbs sampler.

Unfortunately, there is no established way to determine an appropriate value for n. One option is to look at plots of $\lambda^{(1)}, \lambda^{(2)}, \ldots$ against iterations for a number of independent replications. An illustration with four independent replications is given in Figure 3. If, after n iterations, the values appear to fluctuate around a common stationary value, this may be taken as circumstantial evidence that n is large enough. In Figure 3, the plots appear to stabilize after about 1200 iterations. However, there is no way to be sure since we do not know what will happen after 5000 iterations. Other ways to assess the required number of iterations are discussed by Gelman and Rubin (1992).

2.2. Why

Let $\left\{\Lambda^{(n)}, n \geq 0\right\}$ denote a Markov chain. *A Markov chain* is a stochastic process such that each value depends only on its immediate predecessor; that is, for $n > 0$,

$$f(\lambda^{(n)}|\lambda^{(n-1)}, \ldots, \lambda^{(0)}) = f(\lambda^{(n)}|\lambda^{(n-1)}) \quad ,$$

The Gibbs sampler is a procedure to simulate a realization of a Markov chain. Figure 3 shows 5000 realizations of a Markov chain.

In general, to simulate one realization of a Markov chain we do as follows:

1. Choose $\lambda^{(0)}$.
2. For $n = 1, 2, \ldots$ draw $\lambda^{(n)}$ from $f(\lambda|\lambda^{(n-1)})$.

It is seen that the behaviour of the chain depends on the distribution of Λ given its previous value. This distribution is often called a *transition kernel.*

Suppose we intent to construct a Markov chain such that the marginal distribution of $\Lambda^{(n)}$ converges to the posterior distribution if n increases. This requires that:

1. The posterior is the invariant distribution.
2. The chain is irreducible.

Invariance means that if $\lambda^{(0)}$ is drawn from the posterior, then all subsequent values are also draws from the posterior. Suppose, for ease of presentation, that there are two parameters. The argument for the general case follows by mathematical induction. The posterior density is

$$f(\lambda_1, \lambda_2|y) = f(\lambda_1|\lambda_2, y)f(\lambda_2|y) \quad .$$

To sample from the posterior, we draw $\lambda_2^{(1)}$ from the marginal posterior distribution and then $\lambda_1^{(1)}$ from the distribution conditional upon $\lambda_2^{(1)}$. Note that the latter is a full conditional as defined in the previous paragraph.

We now set up a Markov chain to draw $\lambda_2^{(1)}$ from the marginal posterior distribution. Convergence is faster if the dependence between subsequent values is weaker. We set up a chain with a weak form of dependence known as *exchangeability*. Specifically, we ensure that $\Lambda_2^{(1)}$ and $\Lambda_2^{(0)}$ are independent conditional upon $\Lambda_1^{(0)}$. That is,

$$f(\lambda_2^{(0)}, \lambda_2^{(1)}|y) = \int f(\lambda_2^{(0)}|\lambda_1^{(0)}, y)f(\lambda_2^{(1)}|\lambda_1^{(0)}, y)f(\lambda_1^{(0)}|y)d\lambda_1^{(0)} \quad .$$

Furthermore, it one integrates $f(\lambda_2^{(0)}, \lambda_2^{(1)}|y)$ with respect to $\lambda_2^{(0)}$ (or $\lambda_2^{(1)}$), it is seen that $\Lambda_2^{(0)}$ and $\Lambda_2^{(1)}$ have the same marginal distribution; that is, the marginal posterior. It follows that the transition kernel of our chain equals

$$\begin{aligned} f(\lambda_2^{(1)}|\lambda_2^{(0)}, y) &= \frac{f(\lambda_2^{(0)}, \lambda_2^{(1)}|y)}{f(\lambda_2^{(0)}|y)} \\ &= \int f(\lambda_2^{(1)}|\lambda_1^{(0)}, y)\frac{f(\lambda_2^{(0)}|\lambda_1^{(0)}, y)f(\lambda_1^{(0)}|y)}{f(\lambda_2^{(0)}|y)}d\lambda_1^{(0)} \\ &= \int f(\lambda_2^{(1)}|\lambda_1^{(0)}, y)f(\lambda_1^{(0)}|\lambda_2^{(0)}, y)d\lambda_1^{(0)} \quad . \end{aligned}$$

Thus, to produce a value $\lambda_2^{(1)}$ from the posterior distribution we may use *the method of composition* (Tanner, 1996, section 3.3.2) as follows:

1. Draw $\lambda_2^{(0)}$ from the posterior.
2. Draw $\lambda_1^{(0)}$ from the full conditional $f(\lambda_1|\lambda_2^{(0)}, y)$.
3. Draw $\lambda_2^{(1)}$ from the full conditional $f(\lambda_2|\lambda_1^{(0)}, y)$.

It is seen that this procedure is a Gibbs sampler starting with a draw from the posterior.

With $\lambda_2^{(1)}$ drawn from the marginal posterior, we then draw $\lambda_1^{(0)}$ from the full conditional $f(\lambda_1|\lambda_2^{(1)}, y)$ and repeat the process with $\lambda_2^{(1)}$ replacing $\lambda_2^{(0)}$. Schematically, the sampling procedure may be depicted as in Figure 4 where the values generated by the Gibbs sampler are drawn inside a rectangle. It can be shown that these values are the realization of a Markov chain whose invariant distribution is, by construction, the posterior. There is no need to generate the values outside the rectangle.

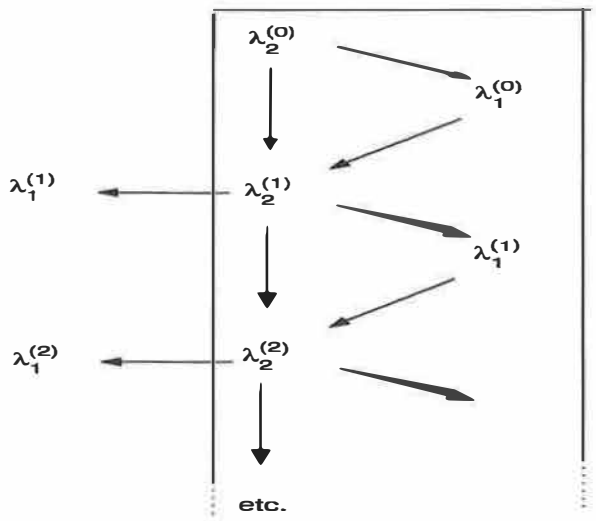

FIGURE 4.
Schematic picture of sampling procedure

Irreducibility refers to the fact that it must be possible to reach each region in the support of the posterior must be reached. That this requirement is met is

illustrated in Figure 2.

In the previous section we stated that for some value n, $\lambda^{(n+1)}, \lambda^{(n+2)}, \ldots$ can be considered a *did* sample from the posterior. We see that it is more correct to state that $\lambda^{(n+1)}, \lambda^{(n+2)}, \ldots$ are *approximately* a *did* sample from the posterior. The approximation improves if n increases.

3. A Brief History of the DA-T Gibbs Sampler

Consider the simple situation where there is only one person who took a test consisting of two Rasch items. We will use this unrealistically small example merely to illustrate the basic principles and explain how the DA-T Gibbs sampler emerged as a variant of the Gibbs sampler. It is arbitrarily assumed that the parameters are sampled in the order: θ, δ_1, δ_2.

Consider the full conditional of θ. The posterior is proportional to

$$\begin{aligned} f(\theta, \delta, y) &= f(y|\theta, \delta) f(\theta, \delta) \\ &= P(Y_1 = y_1|\theta, \delta_1) P(Y_2 = y_2|\theta, \delta_2) f(\theta) f(\delta_1) f(\delta_2) \end{aligned}$$

The last equality was established under two assumptions:

1. The parameters are *a priori* independent.
2. A person responds independently to different items. This assumption is called *Local Independence (LI).*

Ignoring all terms unrelated to θ it is found that:

$$\begin{aligned} f(\theta|\delta^{(j)}, y) &\propto P(Y_1 = y_1|\theta, \delta_1^{(j)})P(Y_2 = y_2|\theta, \delta_2^{(j)})f(\theta) \\ &= \frac{\exp\left[y_1\left(\theta - \delta_1^{(j)}\right)\right]}{1 + \exp\left(\theta - \delta_1^{(j)}\right)} \frac{\exp\left[y_2\left(\theta - \delta_2^{(j)}\right)\right]}{1 + \exp\left(\theta - \delta_2^{(j)}\right)} f(\theta) \\ &\propto \left(\frac{f(\theta)}{\prod_h \left[1 + \exp\left(\theta - \delta_h^{(j)}\right)\right]}\right) \exp\left[\theta(y_1 + y_2)\right] \quad . \end{aligned} \tag{2}$$

To our knowledge, $f(\theta|\delta^{(j)}, y)$ is not a familiar density but it is seen from (2) that it is a member of the exponential family. Although there are general algorithms to generate samples from any exponential family distribution, such algorithms require time and expertise to implement and may not be very efficient (Devroye, 1986). Other possibilities are discussed by Gilks and Wild (1992), or Chib and Greenberg (1995). We consider DA-Gibbs sampling.

The abbreviation DA stand for *Data Augmentation*. DA means that latent data is added as a parameter which, in some applications, results in simpler full conditional distributions (e.g., Tanner, 1996, chapter 5). In the present case, we include the latent response variables $X = (X_1, X_2)$ as a parameter and consider the following DA posterior

$$\begin{aligned} f(\theta, \delta, x|y) &\propto f(\theta, \delta, x, y) \\ &= f(y|x, \delta)f(x|\theta)f(\theta)f(\delta_1)f(\delta_2) \quad . \end{aligned}$$

To find the full conditional of θ we delete all factors that are unrelated to θ. This gives:

$$\begin{aligned} f(\theta|x, y) &\propto f(x|\theta)f(\theta) \\ &= f(x_1|\theta)f(x_2|\theta)f(\theta) \quad . \end{aligned} \tag{3}$$

Especially, when it is easy to sample from the prior distribution we would like the

product $f(x_1|\theta)f(x_2|\theta)f(\theta)$ in (3) to be of the same functional form. This occurs for the 2NO model (Albert, 1998). Specifically, if $f(x_i|\theta)$ $(i = 1, 2)$ and $f(\theta)$ are normal densities,

$$
\begin{aligned}
f(\theta|x,y) &\propto \exp\left(-\frac{1}{2}(x_1-\theta)^2\right)\exp\left(-\frac{1}{2}(x_2-\theta)^2\right)\exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right) \\
&= \exp\left(-\frac{1}{2}\left[(x_1-\theta)^2+(x_2-\theta)^2+\frac{1}{\sigma^2}(\theta-\mu)^2\right]\right) \\
&\propto \exp\left(-\frac{1}{2\left(\frac{1}{2+\sigma^{-2}}\right)}\left[\theta^2-2\frac{\sum_i x_i+\mu\sigma^{-2}}{2+\sigma^{-2}}\theta\right]\right) \text{ “complete the squares”} \\
&\propto \exp\left(-\frac{1}{2\left(\frac{1}{2+\sigma^{-2}}\right)}\left[\theta^2-2\frac{\sum_i x_i+\mu\sigma^{-2}}{2+\sigma^{-2}}\theta+\left(\frac{\sum_i x_i+\mu\sigma^{-2}}{2+\sigma^{-2}}\right)^2\right]\right) \\
&= \exp\left(-\frac{1}{2\left(\frac{1}{2+\sigma^{-2}}\right)}\left[\left(\theta-\frac{\sum_i x_i+\mu\sigma^{-2}}{2+\sigma^{-2}}\right)^2\right]\right) \quad .
\end{aligned}
$$

It follows that the full conditional of θ is a normal distribution. Under the Rasch model (or the 2PL), $f(x_i|\theta)$ is logistic. Unfortunately, it is a fact that the product $\prod_i f(x_i|\theta)f(\theta)$ is not a logistic function. Not even if the prior of θ is a logistic distribution. In fact, it is not a well-known density or even a member of the exponential family. Maris and Maris (2002) show that a transformation of the latent data may provide a solution.

The T in DA-T Gibbs stands for *Transformation*; the latent responses are transformed in such a way that all parameters are removed from their distribution. As an illustration, we will derive the full conditional distribution of θ in three steps: *First*, note that the probability density function $f(y|x,\delta)$ is an indicator function; $f(y|x,\delta) = 1$ if the observed responses match with the latent data, and $f(y|x,\delta) = 0$

otherwise. For example,

$$P(Y_1 = 1, Y_2 = 0|x, \delta) = (x_1 > \delta_1)(x_2 \leq \delta_2)$$
$$= \begin{cases} 1 \text{ if } x_1 > \delta_1 \text{ and } x_2 \leq \delta_2 \\ 0 \text{ otherwise} \end{cases} .$$

It is then readily seen that $f(\theta, \delta, x, y)$ is equal to

$$\left[(x_1 > \delta_1)^{y_1}(x_1 \leq \delta_1)^{1-y_1}(x_2 > \delta_2)^{y_2}(x_2 \leq \delta_2)^{1-y_2}\right] f(x|\theta)f(\theta)f(\delta) \quad .$$

Second, let $Z_i = X_i - \theta$ denote the standardized latent response variables. It follows that $x_i = z_i + \theta$ and it is readily found that $f(\theta, \delta, z, y)$ is

$$(z_1 + \theta > \delta_1)^{y_1}(z_1 + \theta \leq \delta_1)^{1-y_1}(z_2 + \theta > \delta_2)^{y_2}(z_2 + \theta \leq \delta_2)^{1-y_2} f(z)f(\theta)f(\delta) \quad ,$$

where

$$f(z) = f(z_1)f(z_2) = \frac{\exp(z_1)}{[1 + \exp(z_1)]^2}\frac{\exp(z_2)}{[1 + \exp(z_2)]^2} \quad .$$

It is seen that standardization of the latent response variables has removed the person parameter from the density of the latent data. *Third*, deleting all factors unrelated to θ we readily find that

$$f(\theta|\delta^{(j)}, z^{(j+1)}, y) \propto$$
$$\begin{cases} \left(-\infty < \theta \leq \min\left\{\delta_1^{(j)} - z_1^{(j+1)}, \delta_2^{(j)} - z_2^{(j+1)}\right\}\right) f(\theta) & \text{if } y_1 = 0, y_2 = 0 \\ \left(\delta_1^{(j)} - z_1^{(j+1)} < \theta \leq \delta_2^{(j)} - z_2^{(j+1)}\right) f(\theta) & \text{if } y_1 = 1, y_2 = 0 \\ \left(\delta_2^{(j)} - z_2^{(j+1)} < \theta \leq \delta_1^{(j)} - z_1^{(j+1)}\right) f(\theta) & \text{if } y_1 = 0, y_2 = 1 \\ \left(\max\left\{\delta_1^{(j)} - z_1^{(j+1)}, \delta_2^{(j)} - z_2^{(j+1)}\right\} < \theta < \infty\right) f(\theta) & \text{if } y_1 = 1, y_2 = 1 \end{cases}$$

It is seen that the full conditional of θ is the truncated prior distribution.[1] That is,

[1]It was implicitly assumed that the transformed latent responses are sampled first, followed by the ability parameters.

a distribution with a density of the form

$$\frac{(l < \theta \leq h) f(\theta)}{\int_l^h f(\theta) d\theta} \quad ,$$

where l and h are called *truncation constants*. The same is true for the other full conditional distributions but there is no need to derive them here as we proceed with the DA-T Gibbs sampler for the 2PL.

4. The DA-T Gibbs Sampler for the 2PL

4.1. The Prior

For our present purpose, it is convenient to assume that:

1. The parameters are *a priori* independent.

$$f(\theta, \delta, \alpha) = \prod_p f(\theta_p) \prod_i f(\delta_i) f(\alpha_i) \quad . \tag{4}$$

2. All prior distributions are logistic.

The choice of prior is essentially subjective and researchers are free to choose any prior distribution that they see fit. In any event, the effect of the prior on the posterior diminishes if more data are observed, provided the parameters are identifiable. If the value of a parameter is undetermined, no amount of data will diminish our uncertainty beyond the prior assumptions (see Dawid, 1979).

4.2. The Full Conditionals

The DA posterior of the 2PL is proportional to

$$f(\theta, \delta, \alpha, x, y) = f(y|x, \delta) f(x|\theta, \delta, \alpha) f(\theta, \delta, \alpha) \quad ,$$

where

$$f(y|x,\delta) = \prod_p \prod_i f(y_{pi}|x_{pi},\delta_i)$$
$$= \prod_p \prod_i (x_{pi} > \delta_i)^{y_{pi}} (x_{pi} \leq \delta_i)^{1-y_{pi}} \quad ,$$

and

$$f(x|\theta,\delta,\alpha) = \prod_p \prod_i f(x_{pi}|\theta_p,\delta_i,\alpha_i)$$
$$= \prod_p \prod_i \frac{\exp(x_{pi}-\alpha_i\theta_p)}{[1+\exp(x_{pi}-\alpha_i\theta_p)]^2} \quad .$$

It is seen that we assume persons to be independent of one another. Combining terms, we find that $f(\theta,\delta,\alpha,x,y)$ is equal to

$$\prod_p \prod_i (x_{pi} > \delta_i)^{y_{pi}} (x_{pi} \leq \delta_i)^{1-y_{pi}} \frac{\exp(x_{pi}-\alpha_i\theta_p)}{[1+\exp(x_{pi}-\alpha_i\theta_p)]^2} f(\theta,\delta,\alpha) \quad .$$

If we apply the transformation $z_{pi} = x_{pi} - \alpha_i\theta_p$ we find that the DA-T posterior $f(\theta,\delta,\alpha,z|y) \propto f(\theta,\delta,\alpha,z,y)$, and $f(\theta,\delta,\alpha,z,y)$ equals

$$\prod_p \prod_i (z_{pi}+\alpha_i\theta_p > \delta_i)^{y_{pi}} (z_{pi}+\alpha_i\theta_p \leq \delta_i)^{1-y_{pi}} \frac{\exp(z_{pi})}{[1+\exp(z_{pi})]^2} f(\theta,\delta,\alpha) \quad .$$

To determine the full conditionals we delete all terms unrelated to the parameter of interest. Thus, it is seen that each of the full conditionals is the truncated prior distribution.

1. The full conditional of z_{pi} is a logistic distribution with support

$$(z_{pi} > \delta_i - \alpha_i\theta_p)^{y_{pi}} (z_{pi} \leq \delta_i - \alpha_i\theta_p)^{1-y_{pi}}$$

2. The full conditional of δ_i $(i > 1)$ is a logistic distribution with support

$$\prod_p (\delta_i < z_{pi} + \alpha_i\theta_p)^{y_{pi}} (\delta_i \geq z_{pi} + \alpha_i\theta_p)^{1-y_{pi}}$$

3. The full conditional of α_i $(i > 1)$ is a logistic distribution with support

$$\prod_p (\alpha_i \theta_p > \delta_i - z_{pi})^{y_{pi}} (\alpha_i \theta_p \leq \delta_i - z_{pi})^{1-y_{pi}}$$

4. The full conditional of θ_p is a logistic distribution with support

$$\prod_i (\theta_p \alpha_i > \delta_i - z_{pi})^{y_{pi}} (\theta_p \alpha_i \leq \delta_i - z_{pi})^{1-y_{pi}}$$

Each of the support regions is given as the product of indicator functions. In the ensuing section we discuss how the deal with them. For notational convenience, we have deleted the superscripts. The superscripts indicated the order in which the parameters are sampled. It is not essential that they are given since the order in which the parameters are sampled is arbitrary.

4.3. Calculating the Truncation Constants

The support of each of the full conditionals is seen to be a product of indicator functions of the following form:

$$\prod_j (l_j < \lambda_i \leq h_j) = \left(\max_j \{l_j\} < \lambda_i \leq \min_j \{h_j\} \right) \quad , \tag{5}$$

where either $l_j = -\infty$ or $h_j = \infty$. Hence, each term $(l_j < \lambda_i < h_j)$ restricts the range of λ_i to a half open interval extending to either plus or minus infinity. Their product is the intersection of these intervals ranging from $\max_j \{l_j\}$ to $\min_j \{h_j\}$. Thus, $\max_j \{l_j\}$ and $\min_j \{h_j\}$ are the truncation constants for the full conditional.

The support for δ_i: The support for δ_i is a product of indicator functions over persons. We see that

$$l_p = \begin{cases} -\infty & \text{if } y_{pi} = 1 \\ z_{pi} + \alpha_i \theta_p & \text{if } y_{pi} = 0 \end{cases}$$

and

$$h_p = \begin{cases} z_{pi} + \alpha_i \theta_p & \text{if } y_{pi} = 1 \\ \infty & \text{if } y_{pi} = 0 \end{cases}$$

The support of α_i: Note that

$$\begin{aligned} &(\alpha_i \theta_p > \delta_i - z_{pi})^{y_{pi}} (\alpha_i \theta_p \leq \delta_i - z_{pi})^{1-y_{pi}} \qquad (6) \\ &= \begin{cases} (t_{pi} < \alpha_i < \infty)^{y_{pi}} (-\infty < \alpha_i \leq t_{pi})^{1-y_{pi}} & \text{if } \theta_p > 0 \\ (-\infty < \alpha_i < t_{pi})^{y_{pi}} (t_{pi} \leq \alpha_i < \infty)^{1-y_{pi}} & \text{if } \theta_p < 0 \end{cases} \end{aligned}$$

where

$$t_{pi} \equiv \frac{\delta_i - z_{pi}}{\theta_p} \quad . \qquad (7)$$

The indicator functions depend on the sign of θ_p because we divide by θ_p on both sides of the inequality sign in (6). The support for α_i is a product over persons. If $\theta_p > 0$,

$$l_p = \begin{cases} t_{pi} & \text{if } y_{pi} = 1 \\ -\infty & \text{if } y_{pi} = 0 \end{cases} \quad \text{and} \quad h_p = \begin{cases} \infty & \text{if } y_{pi} = 1 \\ t_{pi} & \text{if } y_{pi} = 0 \end{cases} .$$

If $\theta_p < 0$, then

$$l_p = \begin{cases} -\infty & \text{if } y_{pi} = 1 \\ t_{pi} & \text{if } y_{pi} = 0 \end{cases} \quad \text{and} \quad h_p = \begin{cases} t_{pi} & \text{if } y_{pi} = 1 \\ \infty & \text{if } y_{pi} = 0 \end{cases} .$$

The support of θ_p: Calculating the support for θ_p is very similar to calculating the support of α_i. First, note that

$$\begin{aligned} &(\alpha_i \theta_p > \delta_i - z_{pi})^{y_{pi}} (\alpha_i \theta_p \leq \delta_{i+1} - z_{pi})^{1-y_{pi}} \\ &= \begin{cases} \left(t^*_{pi} < \theta_p < \infty\right)^{y_{pi}} \left(-\infty < \theta_p \leq t^*_{pi}\right)^{1-y_{pi}} & \text{if } \alpha_i > 0 \\ \left(-\infty < \theta_p < t^*_{pi}\right)^{y_{pi}} \left(t^*_{pi} \leq \theta_p < \infty\right)^{1-y_{pi}} & \text{if } \alpha_i < 0 \end{cases} \end{aligned}$$

where

$$t^*_{pi} \equiv \frac{\delta_i - z_{pi}}{\alpha_i} \tag{8}$$

Thus, the indicator functions depend on the sign of α_i. Now, we have a product over items. If $\alpha_i > 0$, then

$$l_i = \begin{cases} t^*_{pi} & \text{if } y_{pi} = 1 \\ -\infty & \text{if } y_{pi} = 0 \end{cases} \quad \text{and} \quad h_i = \begin{cases} \infty & \text{if } y_{pi} = 1 \\ t^*_{pi} & \text{if } y_{pi} = 0 \end{cases}$$

If $\alpha_i < 0$,

$$l_i = \begin{cases} -\infty & \text{if } y_{pi} = 1 \\ t^*_{pi} & \text{if } y_{pi} = 0 \end{cases} \quad \text{and} \quad h_i = \begin{cases} t^*_{pi} & \text{if } y_{pi} = 1 \\ \infty & \text{if } y_{pi} = 0 \end{cases}$$

In practice, we consider each interval in (5) separately and increase (decrease) the lower bound (upper bound) of the intersection, each time we encounter an interval with a higher lower bound (lower upper bound). This is illustrated with the following psuedo-code to determine the truncation constants for the full conditional of θ_p:

$l = -\infty$

$h = \infty$

FOR $i = 1$ to the number of items

$t_{pi} = \frac{\delta_i - z_{pi}}{\alpha_i}$

IF $y_{pi} = 0$

IF $t_{pi} < h$ and $\alpha_i > 0$ then $h = t_{pi}$

IF $t_{pi} > l$ and $\alpha_i < 0$ then $l = t_{pi}$

IF $y_{pi} = 1$

IF $t_{pi} > l$ and $\alpha_i > 0$ then $l = t_{pi}$

IF $t_{pi} < h$ and $\alpha_i < 0$ then $h = t_{pi}$

END

It is important that none of the intersections is empty, otherwise we have a full conditional with empty support and the Gibbs sampler stops. For later references this is stated as a proposition.

Proposition 1. *In the 2PL, none of the full conditionals can have an empty support.*

Proof. For any parameter values at the jth iteration, we generate latent data such that

$$\prod_i \prod_p \left((z_{pi}^{(j+1)} > \delta_i^{(j)} - \alpha_i^{(j)}\theta_p^{(j)})^{y_{pi}} (z_{pi}^{(j+1)} \leq \delta_i^{(j)} - \alpha_i^{(j)}\theta_p^{(j)})^{1-y_{pi}} \right) = 1 \quad .$$

This means that at this point we are at a point inside the support of the posterior. Then, we draw, say, δ_i from

$$f(\delta_i|rest) \propto \left(\prod_p (z_{pi}^{(j+1)} + \alpha_i^{(j)}\theta_p^{(j)} > \delta_i)^{y_{pi}} (z_{pi}^{(j+1)} + \alpha_i^{(j)}\theta_p^{(j)} \leq \delta_i)^{1-y_{pi}} \right) f(\delta_i) \quad .$$

Since, the term within brackets is one for $\delta_i = \delta_i^{(j)}$, it follows that the support of the full conditional is not empty. The same is true for the other parameters. □

An informal illustration is given in Figure 2.

4.4. Sampling from a Truncated Logistic Distribution

As a starting point, we suppose that the reader is able to simulate from a uniform $(0, 1)$ distribution and we shall use the term *random numbers* to mean independent random variables from this distribution.

Let X denote a random variable with continuous distribution function F such that F^{-1} is computable. We can simulate X by simulating a random number u and then setting $x = F^{-1}(u)$. This is called *the inversion method.*

In the standard logistic distribution

$$F(x) = \text{Expit}(x) \equiv \frac{\exp(x)}{1 + \exp(x)} \quad ,$$

and

$$F^{-1}(u) = \text{Logit}(u) \equiv \log\left(\frac{u}{1-u}\right) \quad .$$

The variance of the standard logistic distribution is equal to $\frac{1}{3}\pi^2$ and its mean is equal to zero. If we desire a variance $\frac{1}{3}\pi^2\beta^2$, and a mean μ we calculate $\beta\,\text{Logit}(u)+\mu$.

It is straightforward to adapt the inversion method to sample from a truncated distribution. The distribution of a truncated random variable $l < X < h$ is given by

$$F_{tr}(t) = \frac{F(x) - F(l)}{F(h) - F(l)}$$

and

$$F_{tr}^{-1}(u) = F^{-1}\left\{u\left[F(h) - F(l)\right] + F(l)\right\} \quad .$$

It follows that

$$\beta\,\text{Logit}\left(u\left[\text{Expit}\left(\frac{h-\mu}{\beta}\right) - \text{Expit}\left(\frac{l-\mu}{\beta}\right)\right] + \text{Expit}\left(\frac{l-\mu}{\beta}\right)\right) + \mu$$

gives a simulated value from a truncated logistic distribution.

Figure 5 illustrates how the procedure works. Basically, we produce a random number u that is then transformed to u^* in the interval from $F(l)$ to $F(h)$. The corresponding value $F^{-1}(u^*)$ is a realization of the truncated variable.

4.5. Handling Incomplete Designs

In applications, the design of the study is often *incomplete.* This means that only a subset of the available items is administered to each person, and no responses are observed for items that were not administered. As an illustration we have drawn an incomplete design in Figure 6. The rectangles indicate which items where administered to which persons.

To adapt the Gibbs sampler to handle data collected in an incomplete design we need only ignore, for each person, the items that were not administered. To this

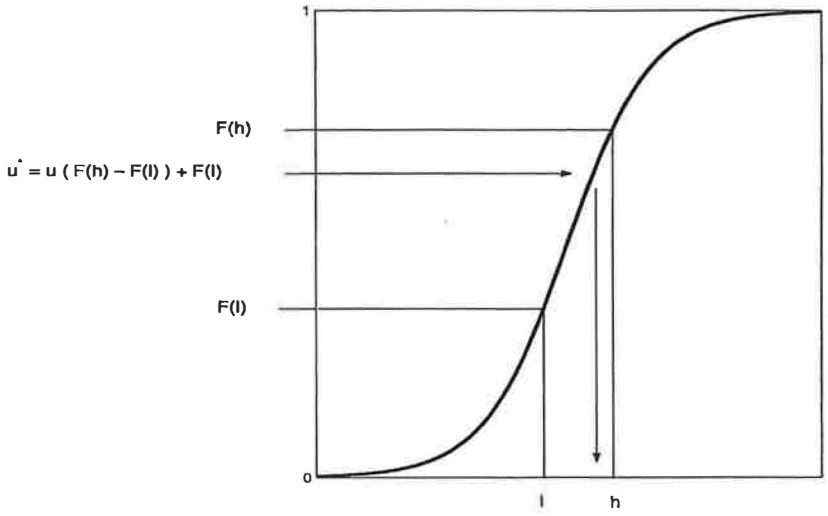

FIGURE 5.
Simulating from a truncated distribution

aim it is useful to construct a matrix d whose entries indicate which items where administered to each of the persons in the sample. Entry d_{pi} equals 1 if item i was administered to person p and zero otherwise (see Figure 6). The Gibbs sampler is unchanged except that nothing is done for person p and item i if $d_{pi} = 0$.

4.6. Sampling Under Restrictions

Researchers often hold prior ideas about the parameters that take the form of order restrictions on the parameters. They may, for instance, believe that item 1 is easier than item 2. In general, such restrictions are added to the range restrictions of the full conditionals.

Suppose, for example, that we add the restriction that $\alpha_2 > 0$. Then

$$f(y, \lambda) = f(y|\lambda)f(\lambda)(\alpha_2 > 0) \quad ,$$

and the full conditional of α_2 is:

$$f(\alpha_2|\lambda_{-\alpha_2}, y)(\alpha_2 > 0) \quad .$$

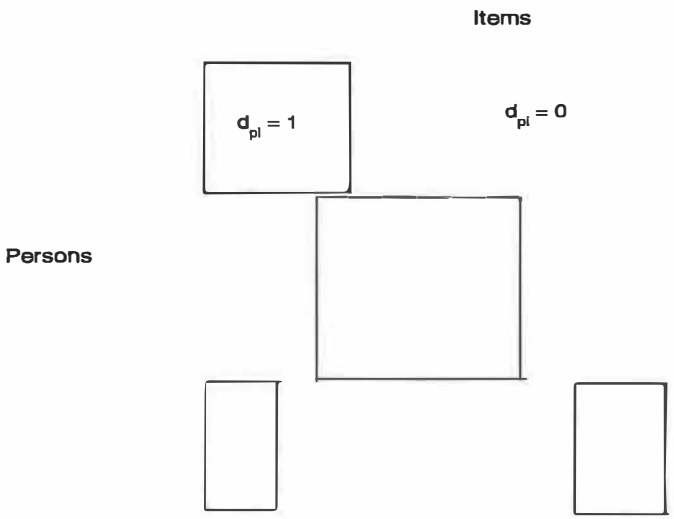

FIGURE 6.
Schematic picture of an incomplete design

Order restrictions among parameters are handled in the same way.

Restrictions may take many forms. Suppose, for example, that a researcher desire to estimate under the restriction that certain parameters take only integer values. In that case one simply chooses a discrete prior for these parameter.

4.7. A DA-T Gibbs Sampler for the Hierarchical 2PL

We now consider the hierarchical 2PL. The hierarchical model derives its name from the fact that a hierarchical structure is imposed upon the person parameters. That is, the person parameters are assumed to be a random sample from a particular distribution. This 2PL is formally equivalent to the marginal 2PL.

For our present purpose we assume that the person parameters are a random sample from a logistic distribution with mean μ_g and scale parameter $\beta_g > 0$. That is,

$$f(\theta_p|\mu_{g_p}, \beta_{g_p}) = \beta_g^{-1} \frac{\exp\left(\frac{\theta_p - \mu_{g_p}}{\beta_{g_p}}\right)}{\left[1 + \exp\left(\frac{\theta_p - \mu_{g_p}}{\beta_{g_p}}\right)\right]^2} \quad .$$

The subscript g_p denotes the population to which person p belongs, where subscript g denotes the population. For example, $g = 1$ may denote girls, and $g = 2$ boys.

In the hierarchical case, $f(\theta_p)$ in (4) is replaced by

$$f(\theta_p|\mu_{g_p}, \beta_{g_p}) f(\mu_{g_p}) f(\beta_{g_p}) \quad .$$

where $f(\mu_{g_p}) f(\beta_{g_p})$ denotes the prior density of the *hyperparameters*. In order to remove the hyperparameters from the distribution of η_p, and into the range restriction, we define the transformation

$$\eta_p \equiv \frac{\theta_p - \mu_{g_p}}{\beta_{g_p}} \Leftrightarrow \theta_p = \eta_p \beta_{g_p} + \mu_{g_p} \quad .$$

Applying this transformation gives the DA-T posterior

$$f(\delta, \alpha, \eta, \mu, \beta|y) \propto$$
$$\left(\prod_p \prod_i (z_{pi} + \alpha_i \left(\eta_p \beta_{g_p} + \mu_{g_p} \right) > \delta_i)^{y_{pi}} (z_{pi} + \alpha_i \left(\eta_p \beta_{g_p} + \mu_{g_p} \right) \leq \delta_i)^{1-y_{pi}} \right)$$
$$\left(\prod_p \prod_i f(z_{pi}) f(\eta_p) \right) f(\delta) f(\alpha) \left(\prod_g f(\mu_g) f(\beta_g) \right) \quad .$$

The resulting DA-T Gibbs sampler is only marginally different from the DA-T Gibbs sampler for the non-hierarchical 2PL.

5. Applications

5.1. Calculating Classical Test Theory Reliability

Consider a test with N_I items that was administered to N_p persons that are a *simple random* sample from some population. For our present purpose we will not explicitley distinguish between the items and the test and the random variable Y denotes an item or test score. It is assumed that the distribution of Y is defined by an IRT model; in this case the 2PL. Let λ_Y denote the parameter that characterizes the test or the item. The item or test parameters are random variables, each with

any suitable prior distribution. The population is assumed to have density function $f(\theta|\lambda_\theta)$ and the person parameters are assumed to be an *i.i.d.* sample from the postulated population distribution. Note that discrete and continuous variables are not explicitly distinguished.

The true score of a person with parameter θ is defined as the expectation $E[Y|\theta, \lambda_Y]$. The *reliability* of Y, ρ^2_Y, is defined as the proportion of true score variation in the population. That is,

$$\rho^2_Y(\lambda) = \frac{Var(E[Y|\theta, \lambda_Y])}{Var(Y)} \quad ,$$

where $\lambda = (\lambda_Y, \lambda_\theta)$. Under the assumption that the IRT model holds, the reliability is a function of λ.

Lord and Novick (1968) consider the following thought experiment, albeit in different wording: Draw a θ from the population and generate two independent responses y and y^* to the same item. The joint distribution of these responses is

$$f(y, y^*|\lambda) = \int f(y|\theta, \lambda_Y) f(y^*|\theta, \lambda_Y) f(\theta|\lambda_\theta) d\theta \quad , \tag{9}$$

where $f(y|\theta, \lambda_Y) = f(y^*|\theta, \lambda_Y)$. Equation 9 states that the response variables are exchangeable and henceforth they will be called *exchangeable replications*. The reliability of Y equals the correlation between exchangeable replications of Y. That is,

$$\rho^2_Y(\lambda) = Corr(Y, Y^*|\lambda)$$

(e.g., Bechger, Maris, Verstralen, & Béguin, 2003).

The variable Y need not be a continuous test score, and in practice it usually isn't. For example,

1. If Y is a discrete test score

$$\rho^2_Y(\lambda) = \frac{E[YY^*|\lambda] - (E[Y|\lambda])^2}{E[Y^2|\lambda] - (E[Y|\lambda])^2}$$

2. If $Y = 1$ if item i is answered correct and 0 if the answer was incorrect:

$$\rho^2_Y(\lambda) = \frac{P(Y = 1, Y^* = 1|\lambda) - [P(Y = 1|\lambda)]^2}{P(Y = 1|\lambda)\,[1 - P(Y = 1|\lambda)]}$$

   is the reliability of item i. It also equals the pairwise coefficient H of two exchangeable replications (Loevinger, 1948; Mokken, 1971; 1997).

3. If Y indicates whether a test score is above a certain threshold, $\rho^2_Y(\lambda)$ equals Cohen's κ (Cohen, 1960; Bechger, et al., 2003, proposition 2).

4. If Y equals an estimate of ability, $\rho^2_Y(\lambda)$ is the reliability of the estimated abilities.

We assume that "nature" has provided us with an identical and independently distributed (i.i.d.) sample $\mathbf{x}$ of size N_p from $P(X)$. The data matrix $\mathbf{x}$ contains all information that we have about the items and the persons. The objective is to use $\mathbf{x}$ to estimate the reliability of Y under the assumption that the 2PL holds. If we know the parameters, ρ^2_Y can be computed as accurately as desired (e.g., Bechger, et al., 2003). If there is uncertainty about the parameters the following sampling algorithm may be used to generate the empirical distribution of the reliability given the postulated IRT model:

1. Draw λ^* from the posterior, $f(\lambda|\mathbf{x})$, of λ given the observed data.
2. For $r = 1, \ldots, n$, draw *two* exchangeable replications:

   Draw θ^* from $f(\theta|\lambda^*_\theta)$

   Draw y_r and y^*_r, independently from $f(y|\theta^*, \lambda^*_Y)$.
3. Calculate the correlation between $(y_1, \ldots, y_n)$ and $(y^*_1, \ldots, y^*_n)$.

The value of n can be taken very large so that the correlation is determined exactly. Generating λ^*, is conveniently done using the DA-T Gibbs sampler. The number

of λ^*'s drawn determines the accuracy of the empirical distribution. The mean of the correlations may be taken as a point estimate of reliability and the standard deviation as its standard error. Note that the procedure assures that the point estimate is always between 0 and 1.

Besides calculating reliability, there are other useful things that one might do with a large number of exchangeable replications. Suppose, for instance, that c is the minimal required score to pass an examination. To estimate the probability of *inconsistent classification* one could look at the proportion of generated values where $y_r < c$ and $y_r^* > c$, or $y_r > c$ and $y_r^* < c$.

As an illustration we apply this procedure to a data set consisting of 300 responses to five dichotomous geometrical analogy items that was published by Rost (1996). We use the DA-T Gibbs sampler for the hierarchical Rasch model to generate (item and population) parameters from the posterior.[2] The estimated item reliabilities where:

$$0.390(0.057) \quad 0.397(0.053) \quad 0.40(0.050) \quad 0.392(0.048) \quad 0.380(0.049)$$

with standard errors given within parenthesis. The reliability of the summed item responses was estimated to be 0.756(0.025). A plot of the empirical distribution function of the summed responses, and a histogram with a superimposed normal density are shown in Figure 7. To illustrate additional possibilities of the proposed procedure it was calculated that 39%(1.64) of the respondents may be expected to get the same test score on two identical testing occasions. Note that, if we keep λ^* fixed, we obtain the same results that Bechger, et al. (2003) obtained using numerical integration. What these authors could not do was quantify the uncertainty involved. It is seen that this is very easy using MCMC methods.

[2]Earlier analyses with this data set revealed that the Rasch model fits satisfactorily (Rost, 1996, chapter 5; see also Bechger, Verstralen & Verhelst, 2002).

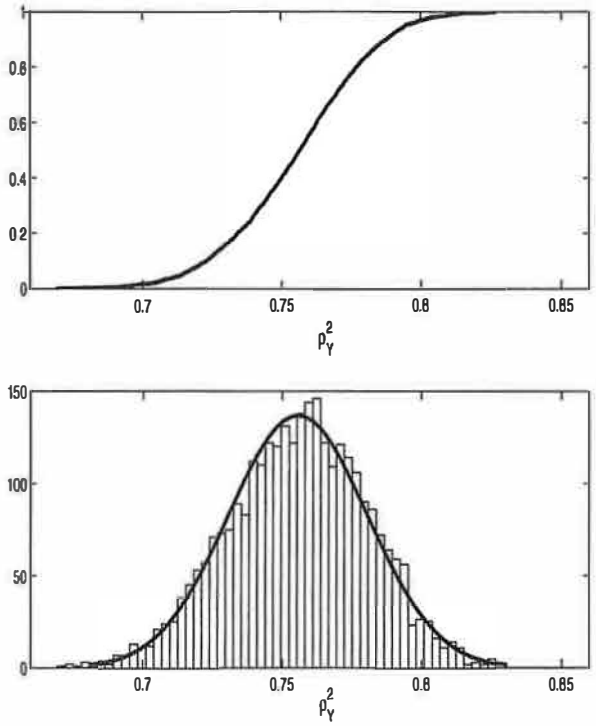

FIGURE 7.
Distribution of test reliability.

5.2. 2PL Mixture IRT Models

A 2PL Mixture Model (2PLMM) is the name that we have chosen for an IRT model that can be written as:

$$P(Y_{pi} = j|\theta, \lambda) = \sum_s P(Y_{pi} = j|S = s, \lambda_{y|s})P(S = s|\theta, \lambda_s) \quad ,$$

where $S = (S_1, \ldots, S_k)$ denotes discrete latent item responses. It is assumed that:

1. $Y_{pi}|S = s$, follows a multinomial distribution.
2. $P(S = s|\theta, \lambda_s)$ equals the likelihood of k locally independent 2PL items; λ_s contains the parameters of these items.
3. θ may be multi-dimensional.

2PLMMs are defined by restrictions on the distribution of Y_{pi} given $S = s$. Consider, for example, the *three-parameter logistic model* (3PL), which is often used in the American literature. In the 3PL, $k = 1$, and $S = 1$ if a person knows the correct answer and $S = 0$ if he doesn't know the answer. Consequently,

$$P(Y_{ip} = 1|S = 1, \lambda_{y|s}) = 1, \quad \text{and } P(Y_{ip} = 1|S = 0, \lambda_{y|s}) = \lambda_{y|s} \quad ,$$

where $\lambda_{y|s}$ is called a guessing parameter. In *latent response models* (Maris, 1995), θ is multi-dimensional but the probabilities $P(Y_{pi} = j|S = s)$ are known and equal to zero or one. An example is the conjunctive Rasch model (see Maris & Maris, 2002, section 2.3.2).

The DA-T Gibbs sampler for the 2PL can be used to build a Gibbs sampler for any 2PLMM. Specifically, at each iteration we draw a sample from the posterior

$$f(\theta, \lambda, s|y) \propto f(\theta, \lambda, s, y) \quad .$$

in three steps:

1. Generate latent discrete item responses from $f(s|\theta, \lambda, y)$. Due to LI, this step entails generating independent responses to each of the k items for each of the persons.
2. Generate θ and λ_s from $f(\theta, \lambda_s|s)$ using the DA-T Gibbs sampler.
3. Generate $\lambda_{y|s}$ from $f(\lambda_{y|s}|s, y)$.

Usually, step 3 is either very simple or unnecessary. It is especially simple if the prior of $\lambda_{y|s}$ is a truncated Dirichlet distribution, because then its full conditional is also a truncated Dirichlet distribution. In the 3PL, for instance, the full conditional of the quessing parameter is a then a truncated beta distribution. In latent response models, step 3 is unnecessary because $\lambda_{y|s}$ is known.

We will not discuss the 2PLMM in full generality but illustrate with two further examples how a DA-T Gibbs sampler can be constructed in practice. In both example, step 3 is uncessary but can be used to extend the models.

5.2.1. The Nedelsky Model for Ability Measurement

Consider a multiple-choice (MC) item i with $J_i + 1$ options arbitrarily indexed $0, 1, \ldots, J_i$. For convenience, 0 indexes the correct alternative. The other J_i answers are incorrect. *The Nedelsky Model* (NM) is based upon the idea that a person responds to a MC question by first eliminating the incorrect answers he recognizes as wrong and then guesses *at random* from the remaining answers.

The probability that wrong answer j is recognized as *wrong* by a respondent with ability θ is modelled as a 2PL. That is, for $j = 1, \ldots, J_i$,

$$P(S_{ij} = 1|\theta) = \frac{\exp(\alpha_i\theta - \delta_{ij})}{1 + \exp(\alpha_i\theta - \delta_{ij})} \quad ,$$

where S_{ij} denotes a random variable that indicates whether alternative j is recognized to be wrong. Thus, we may think of each distractor as a dichtomous 2PL item where a correct answer is produced if the distractor is seen to be wrong. The parameter δ_{ij} now represents the threshold that must be passed to recognize that option j of item i is wrong.

Define *a latent subset* $\mathbf{S}_i$ by the vector $(0, S_{i1}, \ldots, S_{iJ_i})$. Assuming independence among the options given θ, the probability that a subject with ability θ chooses any latent subset $\mathbf{s}_i$ is given by

$$\begin{aligned} P(\mathbf{S}_i = \mathbf{s}_i|\theta) &= \prod_{j=1}^{J_i} \frac{\exp(\alpha_i\theta - \delta_{ij})^{s_{ij}}}{1 + \exp(\alpha_i\theta - \delta_{ij})} \\ &= \frac{\exp\left(\alpha_i\theta s_i^+ - \sum_{j=1}^{J_i} s_{ij}\delta_{ij}\right)}{\prod_{j=1}^{J_i}\left[1 + \exp(\alpha_i\theta - \delta_{ij})\right]} \quad , \end{aligned}$$

where $s_i^+ \equiv \sum_{j=1}^{J_i} s_{ij}$ denotes the number of distractors that are recognized as wrong.

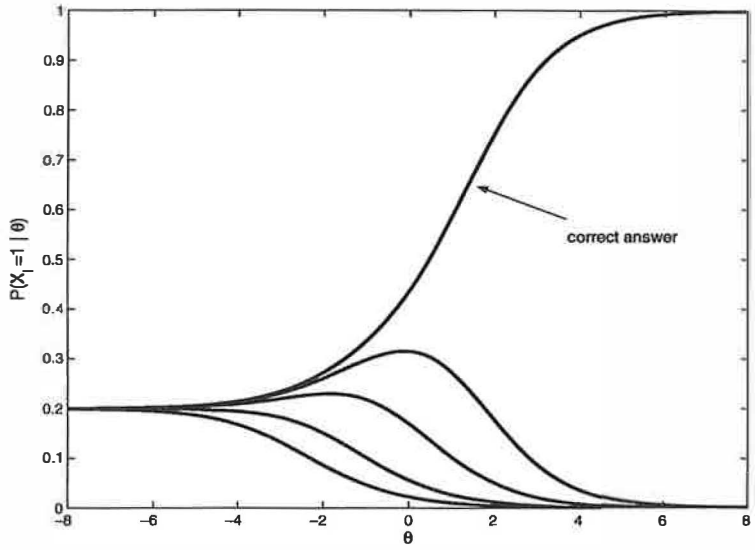

FIGURE 8.

$P(Y_i = j|\theta)$ against θ for a Nedelsky item with five categories.

It is seen that $P(\mathbf{S}_i = \mathbf{s}_i|\theta)$ is the likelihood of J_i independent 2PL items.

Once a latent subset is chosen, a respondent guesses *at random* from the remaining answers. Thus, the conditional probability of responding with option j to item i, given latent subset $\mathbf{s}_i$, is given by:

$$P(Y_i = j|\mathbf{S}_i = \mathbf{s}_i) = \frac{1 - s_{ij}}{\sum_{h=0}^{J_i}(1 - s_{ih})} \quad ,$$

where $\sum_{h=0}^{J_i}(1 - s_{ih}) = J_i + 1 - s_i^+$ denotes the number of alternatives to choose from.

Combining the two stages of the response process, we find that the conditional probability of choosing option j with item i is equal to

$$P(Y_i = j|\theta) = \sum_{\mathbf{s}_i} \frac{1 - s_{ij}}{\sum_{h=0}^{J_i}(1 - s_{ih})} P(\mathbf{S}_i = \mathbf{s}_i|\theta) \quad .$$

Figure 8 shows a plot of these probabilities for an item with five categories. Four properties of the NM are readily seen:

1. $\lim_{\theta \to -\infty} P(Y_i = j|\theta) = \frac{1}{J_i+1}$, for $j = 0, \ldots, J_i$.
2. $P(Y_i = 0|\theta)$ is an increasing function of θ and $\lim_{\theta \to \infty} P(Y_i = 0|\theta) = 1$.
3. The probability of a correct response is always larger than the probability of a distractor.

Note that if an item has only two answer categories; wrong and correct, the NM equals the 3PL with the guessing parameter in this model fixed at $\frac{1}{2}$. A more detailed discussion of the NM can be found in Bechger, Maris, Verstralen, and Verhelst (2004).

We will now derive a DA-T Gibbs sampler for the NM. First, we introduce some notation. Let y denote a data matrix with responses of N_p persons to N_I items; $y_{pi} = j$ if the option j was chosen by respondent p. Similarly, let $\mathbf{s}$ contains latent subsets $\mathbf{s}_{ip}$, where $\mathbf{s}_{ip}$ denotes the latent subset of respondent p with the ith item. The vector θ contains N_p abilities and the vector δ the parameters of N_I items. The vector $\delta_i = (a_i, \delta_{i1}, \ldots, \delta_{iJ_i})$ contains the parameters of the ith item.

The posterior of the NM model is

$$f(\theta, \delta|y) = \sum_{\mathbf{s}} f(\theta, \delta, \mathbf{s}|y) \quad ,$$

We proceed by drawing a sample from

$$f(\theta, \delta, \mathbf{s}|y) \propto f(\theta, \delta, \mathbf{s}, y)$$

and then ignore the latent subsets. It is seen that the latent subsets are the discrete latent item responses in this model. We consider two full conditionals; $f(\theta, \delta|y, \mathbf{s}) = f(\theta, \delta|\mathbf{s})$, and $f(\mathbf{s}|\theta, \delta, y)$. The Gibbs sampler proceeds by repeating the following two steps:

1. Draw latent subsets from $f(\mathbf{s}|\theta, \delta, y)$.
2. Draw θ and δ from $f(\theta, \delta|\mathbf{s})$.

Using LI and Bayes theorem it is seen that,

$$\begin{aligned} f(\mathbf{s}|\theta,\delta,y) &= \frac{f(y|\mathbf{s})f(\mathbf{s}|\theta,\delta)}{f(y|\theta,\delta)} \\ &= \frac{\prod_p \prod_i P(y_{pi}|\mathbf{s}_{ip})P(\mathbf{s}_{ip}|\theta_p,\delta_i)}{\prod_p \prod_i P(y_{pi}|\theta_p,\delta_i)} \\ &= \prod_p \prod_i \frac{P(y_{pi}|\mathbf{s}_{ip})P(\mathbf{s}_{ip}|\theta_p,\delta_i)}{\sum_{\mathbf{s}_i} P(y_{pi}|\mathbf{s}_i)P(\mathbf{s}_i|\theta_p,\delta_i)} \\ &= \prod_p \prod_i P(\mathbf{s}_{ip}|\theta_p,\delta_i,y_{pi}) \quad . \end{aligned}$$

Hence, sampling from $f(\mathbf{s}|\theta,\delta,y)$ entails independently drawing $N_p \times N_I$ latent subsets $\mathbf{s}_{ip}$ with probabilities:

$$P(\mathbf{s}_j|\theta_p,\delta_i,y_{pi}) = \frac{P(y_{pi}|\mathbf{s}_j)P(\mathbf{s}_j|\theta_p,\delta_i)}{\sum_{\mathbf{s}_i} P(y_{pi}|\mathbf{s}_i)P(\mathbf{s}_i|\theta_p,\delta_i)}$$

for each of the possible subset $\mathbf{s}_j$, $j = 1,\ldots,2^{J_i}$, where

$$P(y_{pi}|\mathbf{s}_j)P(\mathbf{s}_j|\theta_p,\delta_i) \propto \frac{1-s_{j(y_{pi})}}{\sum_{h=0}^{J_i}(1-s_{ih})} \exp\left(\alpha_i\theta s_j^+ - \sum_{k=1}^{J_i} s_{jk}\delta_{ik}\right) \quad .$$

To this aim, one makes a list of subsets, calculates the probabilities $P(\mathbf{s}_j|\theta_p,\delta_i,y_{pi})$, and chooses a random subset from the list (see Appendix).

The density $f(\theta,\delta|\mathbf{s})$ is the posterior density of the 2PL model when the subsets are considered data. This means that we may use the DA-T Gibbs sampler for the 2PL model in the second step considering the generated latent subsets as item responses.

Note that the NM has many parameters and hence a large number of persons is required to estimate the item parameters with reasonable precision. As an illustration we provide, in Figure 9, recovery plots of true values against estimated posterior means, for a (small) data set with 20 tri-chotomous items and 200 persons. It is seen that recovery is not particularly good. With 100 items and 3000 persons, the recovery of the item parameters was satisfactorily. As illustrated with Figure 10, the recovery of the person parameters still leaves to be desired.

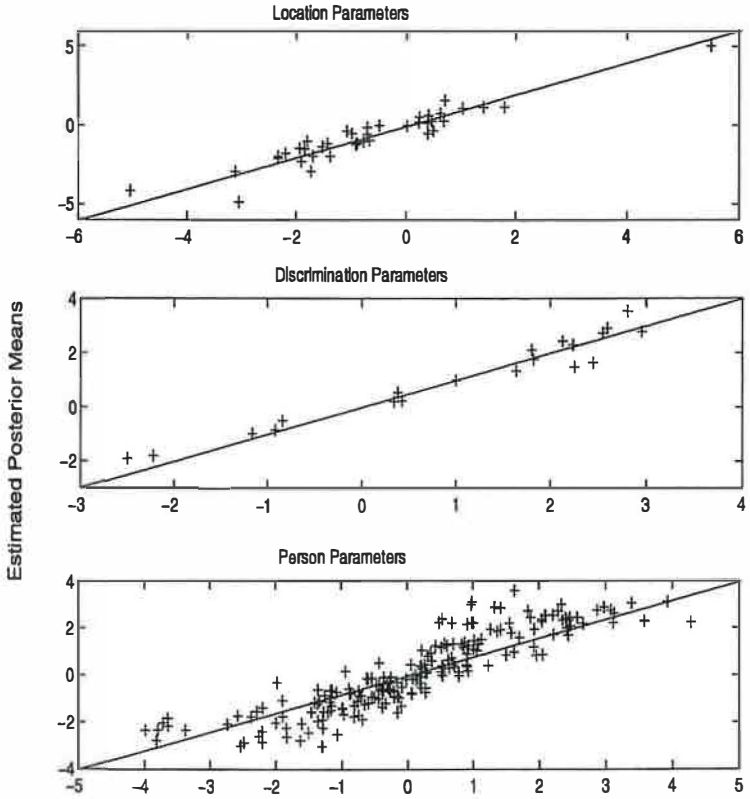

FIGURE 9.

A typical recovery plot for an analysis with 20 items and 200 persons. The parameter values that where used to generate the data are on the horizontal axes. The estimated posterior means are on the vertical axes.

5.2.2. A Nedelsky Model for Opinion Measurement

Consider a MC item i with J_i+1 options that are arbitrarily indexed $1, 2, \ldots, J_i+1$. Now, we assume that the options refer to statements that are presented to the respondent; options 1 to J_i express an opinion while the last option allows the respondents to express that they agree with none of the options. Respondents are allowed to choose more than one option. The $J_i + 1$ option may, for instance, state "no opinion". In practice, it is not necessary to present this option. If we assume that persons refuse to answer when they agree to none of the options, missing responses

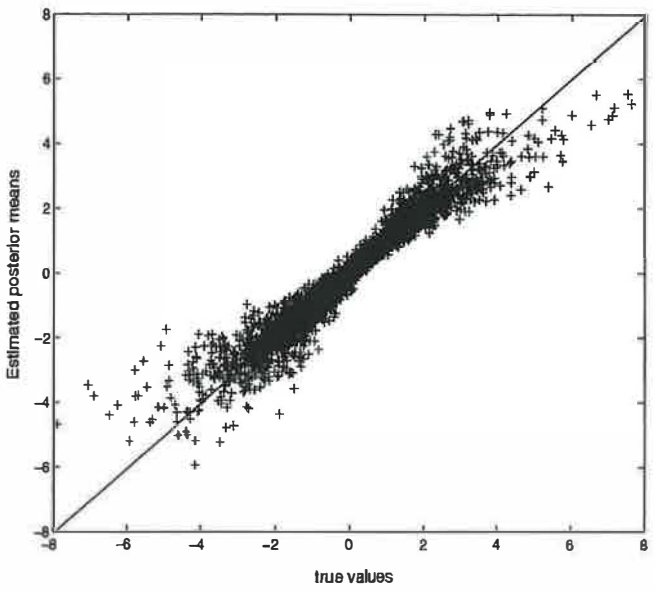

FIGURE 10.

Recovery plot of person parameters for an analysis with the NM with 100 items and 3000 persons.

may be coded as responses in the no opinion category. Thus, respondents are allowed to choose one or more options, or none of the options.

Similar to the NM, the solution process is assumed to consist of two stages. In the first stage, a respondent decides to which options he agrees. In the second stage, the respondent chooses his response randomly from the options he agrees to. We will now describe the two stages.

Let $\mathbf{E}_i = (E_{i1}, \ldots, E_{iJi})$ represent *a latent opinion*; $E_{ih} = 1$ if a respondent agrees to option h of item i, and zero otherwise. It is assumed that

$$P(E_{ij} = 1|\theta_p) = \frac{\exp(d_{ij}\theta_p - \delta_{ij})}{1 + \exp(d_{ij}\theta_p - \delta_{ij})} \quad .$$

It is seen that E_{ij} is modelled as a latent 2PL item. The person parameter θ_p should now be interpreted as the respondent's *attitude*, and the parameter δ_{ij} as the difficulty to agree with option j of item i. The symbol d_{ij} denotes a constant that is either $+1$ or -1. If $d_{ij} = 1$, the probability to agree with an option is increasing

in θ_p, and if $d_{ij} = -1$ this probability is decreasing.

Suppose, for example, that θ_p represents the respondent's political orientation ranging from left to right. Assume further that item i inquires which of a list of candidates a respondent would choose for president. It is reasonable to assume that the probability to choose a left-wing candidate decreases as θ_p moves to the right, while the probability to choose a right-wing candidate increases. Hence, we would set d_{ij} to one if j is a right-wing candidate, and to minus one if j denotes a left-wing candidate.

Assuming LI, the probability that a subject with attitude θ_p chooses any latent opinion $\mathbf{e}_i$ is given by the likelihood of J_i independent items; that is,

$$P(\mathbf{E}_i = \mathbf{e}_i|\theta_p) = \frac{\exp\left(d_i\theta_p e_i^+ - \sum_j e_{ij}\delta_{ij}\right)}{\prod_j \left[1 + \exp(d_{ij}\theta_p - \delta_{ij})\right]} \quad ,$$

where $e_i^+ = \sum_f e_{if}$ denotes the number of options endorsed, and $d_i = \sum_j e_{ij}d_{ij}$.

Let the random variable S_{ij} indicate whether option j is considered, and define $\mathbf{S}_i$ by the vector $(S_{i1}, \ldots, S_{iJ_i+1})$. As in the NM for ability measurement, we refer to $\mathbf{S}_i$ as a latent subset. The difference is that, here, $S_{ij} = 1$ if the jth option was in the subset, and there is no correct alternative. It is assumed that,

$$s_{ij} = \begin{cases} e_{ij}, & \text{if } j \leq J_i \\ \prod_{f=1}^{J_i}(1 - e_{if}), & \text{if } j = J_i + 1 \end{cases} . \tag{10}$$

For the first J_i options, $s_{ij} = e_{ij}$, which implies that alternative j is taken into consideration if $e_j = 1$ and the respondent agrees with the option. For the last option, $s_{iJ_i+1} = 1$ if and only if all e_{if} are zero and the respondent holds no opinion, otherwise s_{iJ_i+1} equal zero. It follows from (10) that

$$P(S_{ij} = 1|\theta) = \begin{cases} P(E_{ij} = 1|\theta) \text{ for } j \leq J_i \\ \prod_f P(E_{if} = 0|\theta) \text{ for } j = J_i + 1 \end{cases}$$

Note that there is a one-one correspondence between latent opinions and latent

subsets. Specifically, $\mathbf{s}_i = (\mathbf{e}_i, \gamma)$, where $\gamma = 1$ if $\mathbf{e}_i = \mathbf{0}$, and zero otherwise. When $J_i = 3$, for instance,

$$\mathbf{e}_i = \begin{matrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} \Leftrightarrow \mathbf{s}_i = \begin{matrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{matrix}$$

Once a latent subset is chosen, a respondent guesses at random from the options in the subset. Thus, the conditional probability of responding with option j to item i, given latent subset $\mathbf{s}_i$, is given by:

$$P(Y_i = j | \mathbf{S}_i = \mathbf{s}_i) = \frac{s_{ij}}{\sum_{h=1}^{J_i+1} s_{ih}},$$

where $Y_i = j$ denotes the event that the respondent chooses option j, and $\sum_{h=1}^{J_i+1} s_{ih}$ is the number of options in the subset. Note that,

$$P(Y_i = J_i + 1 | \mathbf{S}_i = \mathbf{s}_i) = \begin{cases} 1 & \text{if } \mathbf{s}_i = (0, 0, \ldots, 1) \\ 0 & \text{otherwise} \end{cases}$$

Combining the two stages of the answer process, we find that the conditional

probability of choosing option j with item i is equal to

$$
\begin{aligned}
P(Y_i = j|\theta) &= \sum_{\mathbf{s}_i} P(Y_i = j, \mathbf{S}_i = \mathbf{s}_i|\theta) \\
&= \sum_{\mathbf{s}_i} P(Y_i = j|\mathbf{S}_i = \mathbf{s}_i) P(\mathbf{S}_i = \mathbf{s}_i|\theta) \\
&= \sum_{\mathbf{s}_i} \frac{s_{ij}}{\sum_{h=1}^{J_i+1} s_{ih}} P(\mathbf{S}_i = \mathbf{s}_i|\theta) \\
&= \sum_{\mathbf{e}_i} \frac{e_{ij}}{\sum_{h=1}^{J_i+1} e_{ih}} P(\mathbf{E}_i = \mathbf{e}_i|\theta) \\
&= \begin{cases} \sum_{\mathbf{e}_i; \mathbf{e}_i \neq \mathbf{0}} \frac{e_{ij}}{\sum_{h=1}^{J_i} e_{ih}} P(\mathbf{E}_i = \mathbf{e}_i|\theta) \text{ for } j \leq J_i \\ \prod_f P(E_{if} = 0|\theta) \text{ for } j = J_i + 1 \end{cases}
\end{aligned}
$$

The penultimate equality can be established because there is a one-one correspondence between opinions and subsets. It is seen that the subsets were merely introduced to develop the model but are no longer needed.

The Gibbs sampler for the present model is only superficially different from that of the NM for ability measurement. Let the vector $\mathbf{e}$ contain latent opinions $\mathbf{e}_{ip}$, where $\mathbf{e}_{ip}$ denotes the latent opinion of respondent p with the ith item. To produce a sample from $f(\theta, \delta, \mathbf{e}|y)$ we repeate the following two steps:

1. Draw latent opinions from $f(\mathbf{e}|\theta, \delta, y)$.
2. Draw θ and δ from $f(\theta, \delta|\mathbf{e})$.

Similar to the NM for ability measurement, it is found that

$$
f(\mathbf{e}|\theta, \delta, y) = \prod_p \prod_i P(\mathbf{e}_{ip}|\theta_p, \delta_i, y_{pi}) \quad ,
$$

and sampling from $f(\mathbf{e}|\theta, \delta, y)$ entails independently drawing $N_p \times N_I$ latent opinions. Specifically, $\mathbf{e}_{ip}$ is drawn from the set of possible latent opinions such that each

possible opinion $\mathbf{e}_k$ is has probability $P(\mathbf{e}_k|\theta_p, \delta_i, y_{pi})$ to be drawn. If $y_{pi} < J_i + 1$:

$$\begin{aligned} P(\mathbf{e}_k|\theta_p, \delta_i, y_{pi}) &\propto P(y_{pi}|\mathbf{e}_k)P(\mathbf{e}_k|\theta_p, \delta_i) \\ &= \frac{e_{k(y_{pi})}}{\sum_{h=1}^{J_i} e_{kh}} P(\mathbf{e}_k|\theta_p, \delta_i) \end{aligned}$$

If $y_{pi} = J_i + 1$, $\mathbf{e}_j$ is $\mathbf{0}$ with probability 1. Thus, if $y_{pi} < J_i + 1$ we draw any of the opinions where $e_{k(y_{pi})} = 1$. If $y_{pi} = J_i + 1$, the latent opinion is the zero vector.

As before, $f(\theta, \delta|\mathbf{e})$ is seen to be the posterior density of the 2PL model when the latent opinions are considered data so that we may use the DA-T Gibbs samples for the 2PL in this step. Since there is no discrimination parameter in this case, there is no need to sample discrimination parameters; α_i is fixed at minus one or plus one.

6. Discussion

In this lengthy research report we have given an expository account of the DA-T Gibbs sampler for the 2PL. Our focus has been on Gibbs sampling. As a consequence, a number of important issues related to the theory and practice of Bayesian statistics where ignored or only mentioned in passing. As a courtesy to the reader, we mention a few of these issues and provide references where more information can be obtained.

1. Parameter estimation. Here, we mentioned posterior means as point estimates but there are other possibilities such as the posterior mode which is more easily determined with an EM-algorithm (Dempster, Laird, & Rubin, 1973). As an interval estimate, Baysians often report a *the highest posterior density region.* That is, the smallest region of the parameter space which contains a particular percentage of the mass of the posterior distribution. Calculation of highest posterior density regions is discussed, for example, by Tanner (1993, section 2.5), and Chen, Shao, and Ibrahim (2000).

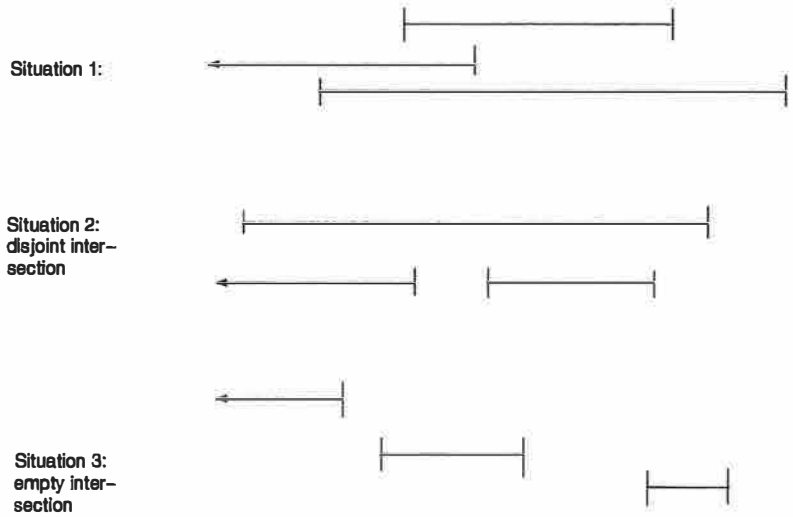

FIGURE 11.

Three possible situations involving range restrictions.

2. Bayesian evaluation of model fit. There are three common ways to determine model fit:

   Bayes factors: See Kass and Raftery (1995)

   Prior predictive checks: See Box (1983)

   Posterior predictive checks: See Rubin (1984)

3. The definition and importance of identifiability in Baysian statistics is discussed by Dawid (1979).

The most intricate aspect of the DA-T Gibbs sampler is the determination of the support of the full conditionals. In Figure 11 we have illustrated three situations that may occur:

1. The intersection is a single interval.
2. The intersection consists of a number of disjoint intervals.
3. The intersection is empty.

Here, we have implicitely made a number of decisions to ensure that we are in the first situation. It is not always that easy. Suppose, for instance, that we had parameterized the 2PL in the more common way as follows:

$$P(Y_{pi} = 1|\theta_p) = \frac{\exp(\alpha_i(\theta_p - \delta_i))}{1 + \exp(\alpha_i(\theta_p - \delta_i))} .$$

Then, we would sometimes encounter the second situation. For example, the support for the full conditional of α_i is now

$$\prod_p \left(\frac{z_{pi}}{\alpha_i} < \delta_i - \theta_p\right)^{y_{pi}} \left(\frac{z_{pi}}{\alpha_i} \geq \delta_i - \theta_p\right)^{1-y_{pi}} .$$

It is seen that the restriction depends on the sign of both α_i, and z_{pi}. Furthermore, as illustrated in Figure 12, the intervals may be disjoint for persons where $y_{pi} = 0$ and persons where $y_{pi} = 1$.

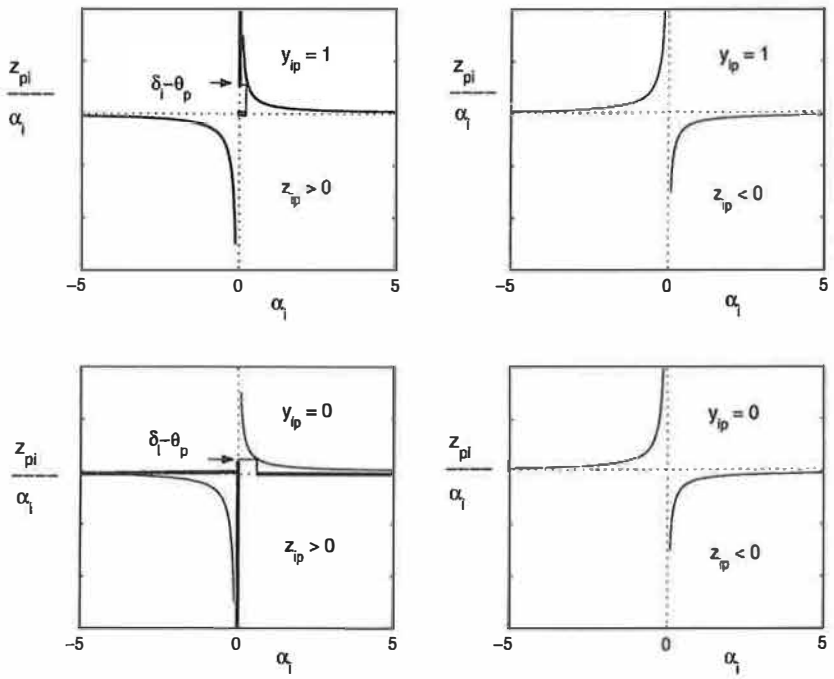

FIGURE 12.

Range restrictions for α_i with a different parameterization of the 2PL.

The DA-T Gibbs sampler has three advantages over alternative methods as the EM-algorithm. The main advantage is that it facilitates testing of restrictions involving multiple parameters. For example, to determine whether $\delta_1 < \delta_2 < \delta_3$ we simply calculate the proportion of the sampled threshold parameters where this is the case. If this proportion is high, it is likely that the hypothesis holds. Second, the method is easy to program. Third, once one has a firm understanding of the idea underpinning the DA-T Gibbs sampler, it is quite easy to develop DA-T Gibbs samplers for other IRT models provided these models can be written as models with continuous latent responses (see Maris & Maris, 2002 for further examples). For models with discrete latent responses, however, it is usually easy to develop a DA-Gibbs sampler.

A disadvantage of Gibbs sampling is that it can be very time-consuming, depending on the complexity and size of the analysis. If one needs to do large-scale analyses on a regular basis it is worthwhile to spend some time programming a faster estimation algorithm.

7. References

Albert, J. H. (1992). Bayesian estimation of normal ogive item response functions using Gibbs sampling. *Journal of Educational Statistics*, **17**, 251-69

Albert, J. H., & Chib, S. (1993). Baysian analysis of binary and polytomous response data. *Journal of the American Statistical Association*, *88*(422), 669-679.

Bechger, T. M. , Maris, G. & Verstralen, H. H. F. M. (2004). The Nedelsky model for multiple-choice items. In *New developments in categorical data analysis for the social and behavioral sciences* edited by Andries van der Ark, Marcel Croon and Klaas Sijtsma. New-York: Lauwrence Erlbaum.

Bechger, T. M., Maris, G., Verstralen, H. H. F. M. & Béguin, A. A. (2003). Using Classical Test Theory in combination with Item Response Theory. *Applied Psychological Measurement*, **27**, 319-334.

Bechger, T. M., Verstralen. H. H. F. M., & Verhelst, N. D. (2002). Equivalent LLTMs. *Psychometrika*, **67**, 120-134.

Birnbaum, A. (1968). Some Latent Trait Models and their Use in Inferring an Examinee's Ability. pp. 395-479 In Lord, F. M. and Novick, M. R. *Statistical Theories of Mental Test Scores*. Reading:Addison-Wesley.

Box, G. E. P. (1983). An apology for ecumenism in statistics. In *Scientific inference, data analysis:Theory and practice.* Cambridge, Massachusetts: M.I.T. Press.

Casella, G. & George, E. I.(1992). Explaining the Gibbs sampler. *The American Statistician*, **46**, 167-174.

Chen, M. H., Shao, Q. M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**(4), 327-335.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and*

Psychological Measurement,20, 37-46.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion) *Journal of the Royal Statistical Society*, **41**, 1-31.

Devroye, L. (1986). *Non-uniform random variate generation*. New-York: Springer-Verlag.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B, **39**, 1-138.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-472.

Gilks, W., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337-348.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Association*. **90**, 773-795.

Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin*, *45*, 507-530.

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

Maris, G., & Maris, E. (2002). A MCMC-method for models with continuous item responses *Psychometrika*, **67**, 335-350.

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*. Den Haag: Mouton.

Mokken, R. J. (1997). Nonparametric models for dichotomous responses. In W.J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory*(pp. 351-367). New-York: Springer.

Rasch, G. (1980). *Probabilitistic models for some intelligence and attainment*

testst. Chicago: The University of Chicago Press.

Ross, S. M. (2003). *Introduction to probability models*. 8th Edition. New-York: Academic Press.

Rost, J. (1996). *Testtheory, testkonstruction* [Test theory, test construction]. Bern, Germany: Verlag Hans Huber.

Rubin, D. B. (1984). Bayesian justifiable frequency calculations for the applied statistician. *Annals of Statistics*, **12**, 1151-1172.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22**, 1701-1762.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, **54**, 427-450.

8. Appendix

8.1. Draw latent Subsets or latent Opinions from a List

In practice, the number of answer alternatives is at most 8 and the set of possible opinions or subsets is small enough to be hard coded in the software. Thus, we have a list of objects and we must choose at random one of the objects. Each object on the list has a probability p_i to be chosen such that $\sum_i p_i = 1$. The probabilities were given in the text.

Let U denote a random number; that is, U is distributed uniformly on (0, 1) so that

$$P\left(\sum_{j=1}^{i-1} p_j < U < \sum_{j=1}^{i} p_j\right) = p_i \quad .$$

This means that, if we generate a random number u, and choose the set labelled i if

$$\sum_{j=1}^{i-1} p_j < u < \sum_{j=1}^{i} p_j \quad ,$$

the probability to choose i is p_i. In practice, this is done as follows:

1. Generate u. Set $sm = h = 0$.
2. Set $h = h + 1$, and $sm = \sum_{i=1}^{h} p_i$.
3. If $u < sm$ choose set i and stop. Otherwise go to 2.

8.2. Generating Data from a 2PL

Let p_{ic} denote the probability to answer correct to item i. Generating a response to item i proceeds as follows:

1. Generate a random number u
2. if $u \leq p_{ic}$ the answer is correct, otherwise it is incorrect.

To generate responses to N_I items we simply repeat this N_I times.