Characterizing the manifest probability distributions of three latent trait models for accuracy and response time.

Marsman, M.[1], Sigurdardóttir, H.[2], Bolsinova, M.[3], & Maris, G.[1, 3]
1 University of Amsterdam
2 Tilburg University
3 ACTNext

Correspondence concerning this article should be addressed to:
Maarten Marsman
University of Amsterdam, Psychological Methods
Nieuwe Achtergracht 129B
PO Box 15906, 1001 NK Amsterdam, The Netherlands
E-mail may be sent to m.marsman@uva.nl.

Abstract

In this paper we study the statistical relations between three latent trait models for accuracy and response times: the Hierarchical model of van der Linden (2007, *Psychometrika*, *72*(3), 629–650), the Signed Residual Time model proposed by Maris and van der Maas (2012, *Psychometrika*, *77*(4), 615–633), and the Drift Diffusion model as proposed by Tuerlinckx and De Boeck (2005, *Psychometrika*, *70*(4), 629–650). What sets these three models apart is that the HM and the DM either assume or imply that accuracy and response times are independent given the latent trait variables, while the SM does not. In this paper we investigate the impact of this conditional independence property —or a lack thereof— on the manifest probability distribution for accuracies and response times, and reveal how the accuracy and response time for an item *i* are related in the manifest distributions that are implied by the three latent trait models. Our approach to characterize the manifest probability distributions is related to the Dutch Identity (Holland, 1990, *Psychometrika*, *55*(6), 5–18), and we discuss the relation between these two approaches.

Keywords: Drift Diffusion model, Dutch Identity, graphical model, Hierarchical model, item response theory, response times, Signed Residual Time model, conditional independence

Introduction

In this paper we aim to clarify the statistical relations between three latent trait models for accuracy and response times: the Hierarchical model (HM) of van der Linden (2007), the Signed Residual Time model (SM) of Maris and van der Maas (2012), and the Drift Diffusion model (DM) proposed by Tuerlinckx and De Boeck (2005). These three models originate from different backgrounds and differ in many respects. Because of this, the way that these models relate to each other has been the topic of several publications. For instance, van Rijn and Ali (2017a) evaluated how the three models perform in practical situations, and Molenaar, Tuerlinckx, and van der Maas (2015a, 2015b) show that two of the models, the DM and the HM, can be formulated as a generalized linear model. Here we investigate the manifest probability distributions that are implied by the three latent trait models.

A major distinction between the three latent trait models is in the relation they stipulate between accuracy and response time after conditioning on the latent trait variables. Whereas responses are independent of response times after conditioning on the latent variable(s) in both the DM and the HM, this is not the case for the SM. This distinction inspired several statistical tests for the conditional independence between responses and response times (van der Linden and Clas, 2010; Bolsinova and Maris, 2016; Bolsinova and Tijmstra, 2016), and new statistical models to model residual dependencies (Bolsinova, De Boeck, and Tijmstra, 2017; Bolsinova, Tijmstra, and Molenaar, 2017). In this paper, we will study what the impact is of this conditional independence assumption —or a lack thereof— on the manifest distribution of accuracy and response times that is implied by the three latent trait models.

The remainder of this paper is structured as follows. In the next section, we introduce the three latent trait models. We will focus on versions of the models that either use or imply the two-parameter logistic model for the marginal distribution of response accuracies. We then introduce our approach to characterize the manifest probability distributions. After having introduced the three latent trait models and our approach to express their manifest distributions we set out to characterize the manifest expressions. Our paper ends with a discussion of the results.

Models

In this section we introduce the three latent trait models. We will assume that the item parameters are fixed constants, but that the accuracies, the response times, and the latent variables are random variables. Since these variables are assumed to be random everywhere in this paper, we do not distinguish between (vectors of) random variables and their realizations. We use $\mathbf{x}$ to denote a vector of p response accuracies —$x_i \in \{0, 1\}$— and use $\mathbf{t}$ to denote a vector of p response times —$t_i \in \mathbb{R}^+$ for the HM and DM and $t_i \in (0, \mathrm{d}_i)$ for the SM, see below.

The Hierarchical Model

The HM, as proposed by van der Linden (2007), is a general statistical framework for modeling accuracies and responses times that is based on the idea that there are two latent traits at work; ability θ governs the response accuracy distribution and speed η the response time distribution. Importantly, the response accuracy distribution is assumed to be independent of speed η given ability θ, the response time distribution is assumed to be independent of ability θ given speed η, and it is also assumed that accuracies and response times are independent given the full set of latent traits, i.e.,

$$p(\mathbf{x}, \mathbf{t} \mid \theta, \eta) = p(\mathbf{x} \mid \theta)\, p(\mathbf{t} \mid \eta).$$

This set-up provides a plug-and-play framework for modeling accuracy and response times: the measurement model for ability θ —marginal distribution of accuracies $p(\mathbf{x} \mid \theta)$— can be chosen independently of the measurement model for speed η —marginal distribution of response times $p(\mathbf{t} \mid \eta)$.[1] The HM is concluded with a model for the two latent traits.

Different measurement models for ability θ have been used in the literature. For example, van der Linden (2007) used normal ogive models, Bolsinova, De Boeck, and Tijmstra (2017) used their logistic counterparts, while Zhan, Jiao, and Liao (2018) used cognitive diagnosis models instead. Here, we use the two-parameter logistic model,

$$p(\mathbf{x} \mid \theta) = \prod_{i=1}^{p} p(x_i \mid \theta) = \prod_{i=1}^{p} \frac{\exp\left(x_i\, \alpha_i(\theta + \beta_i)\right)}{1 + \exp\left(\alpha_i(\theta + \beta_i)\right)}, \tag{1}$$

where α_i is an item discrimination parameter and β_i is an item easiness parameter. There are two reasons for using the two-parameter logistic model here. Firstly, the marginal distribution of accuracies $p(\mathbf{x} \mid \theta)$ that is implied by the versions of the SM and DM that are used here is also a two-parameter logistic model. Secondly, the two-parameter logistic model is a member of the exponential family of distributions, which will be convenient for expressing the manifest probability distribution $p(\mathbf{x}, \mathbf{t})$.

Different measurement models for speed η have also been used in the literature. For example, van der Linden (2006) used a log-normal distribution, Fox, Klein Entink, and van der Linden (2007) used a linear factor model for log-transformed response times, and Klein Entink, van der Linden, and Fox (2009) used models based on Box-Cox transformations of response times. In this paper we will use the log-normal distribution,

$$p(\mathbf{t} \mid \eta) = \prod_{i=1}^{p} p(t_i \mid \eta) = \prod_{i=1}^{p} \sqrt{\frac{\phi_i}{2\pi}}\, \frac{1}{t_i}\, \exp\left(-\frac{1}{2}\, \phi_i\, \left(\ln(t_i) + \eta - \xi_i\right)^2\right), \tag{2}$$

where ξ_i is an item time intensity parameter and ϕ_i an item precision parameter. The item precision ϕ_i is the response-time analogue of the item discrimination in the two-parameter logistic model; larger values of ϕ_i imply that speed explains a larger portion of the log-time variance. The log-normal distribution is a common choice for the measurement model for speed η in the HM framework, and is also a member of the exponential family of distributions.

To conclude the HM we specify a distribution for ability θ and speed η. Typically, a bivariate normal distribution is used in which ability and speed are correlated. To identify the model, however, the means of the bivariate normal need to be constrained to zero and the marginal variance of ability needs to be constrained to one.[2]

[1] We will use "marginal distribution of x" to refer to the distribution

$$p(x \mid \theta) = \int_{\mathbb{R}^+} p(x, t \mid \theta)\, \mathrm{d}t,$$

and we will use "manifest distribution of x" to refer to the distribution

$$p(x) = \int_{\mathbb{R}} p(x \mid \theta)\, p(\theta)\, \mathrm{d}\theta.$$

[2] Alternatively, one may choose an item i and constrain β_i and ξ_i to zero and constrain α_i to one for this item.

The Signed Residual Time Model

The SM has been proposed by Maris and van der Maas (2012) as a measurement model for ability θ in the context of tests with item-level time limits. The model was specifically designed for tests that use the following scoring rule,

$$\mathrm{s} = \sum_{i=1}^{p} \mathrm{s}_i = (2x_i - 1)(\mathrm{d}_i - t_i),$$

where d_i is the time limit for item i. This scoring rule encourages persons to work fast but punishes guessing: residual time $\mathrm{d}_i - t_i$ is gained when the response is correct, but is lost when the response is incorrect. Van Rijn and Ali (2017b) demonstrated that the SM is also appropriate for applications where these time limits are not specified a priori, but "estimated" from the observed response time distributions.

The SM specifies the following distribution for accuracy $\mathbf{x}$ and response times $\mathbf{t}$:

$$p(\mathbf{x}, \mathbf{t} \mid \theta) = \prod_{i=1}^{p} (\theta + \beta_i) \frac{\exp\left((2x_i - 1)(\mathrm{d}_i - t_i)(\theta + \beta_i)\right)}{\exp(\mathrm{d}_i(\theta + \beta_i)) - \exp(-\mathrm{d}_i(\theta + \beta_i))}, \tag{3}$$

where β_i is an item easiness parameter. Observe that the SM is an exponential family model, and that the scoring rule s is the sufficient statistic for ability θ. The SM has been generalized by van Rijn and Ali (2017b) allowing the items to differ in their discriminative power even when the time limits are the same across items. In this paper we will use the standard version of the SM.

Whereas the HM characterizes the joint distribution of accuracies and response times by specific choices of the marginals $p(\mathbf{x} \mid \theta)$ and $p(\mathbf{t} \mid \eta)$, the SM directly specifies a joint distribution for accuracies and response times $p(\mathbf{x}, \mathbf{t} \mid \theta)$. By integrating out the response times we obtain the marginal distribution for accuracies $p(\mathbf{x} \mid \theta)$. Maris and van der Maas (2012) show that this marginal distribution is the two-parameter logistic model in Eq. (1), where the item discrimination α_i is equal to the item time limit d_i. In a similar way, we obtain the marginal distribution of response times $p(\mathbf{t} \mid \theta)$ by summing out the accuracies,

$$p(\mathbf{t} \mid \theta) = \prod_{i=1}^{p} (\theta + \beta_i) \frac{\exp((\mathrm{d}_i - t_i)(\theta + \beta_i)) + \exp(-(\mathrm{d}_i - t_i)(\theta + \beta_i))}{\exp(\mathrm{d}_i(\theta + \beta_i)) - \exp(-\mathrm{d}_i(\theta + \beta_i))}.$$

An alternative specification of the SM is in terms of accuracies and that what Maris and van der Maas (2012, p. 624) refer to as *pseudo* response times $\mathbf{t}^*$. *Pseudo* response times are obtained from response times through the transformation

$$t_i^* = \begin{cases} t_i & \text{if } x_i = 1, \\ \mathrm{d}_i - t_i & \text{if } x_i = 0. \end{cases} \tag{4}$$

This transformation from response times to *pseudo* response times is one-to-one, so that no information is lost. One convenient feature of using *pseudo* response times instead of response times is that the *pseudo* response times and accuracies are (conditionally) independent in the SM, i.e.,

$$p(\mathbf{x}, \mathbf{t}^* \mid \theta) = p(\mathbf{x} \mid \theta) p(\mathbf{t}^* \mid \theta),$$

where the marginal distribution $p(\mathbf{t}^* \mid \theta)$ is equal to,

$$p(\mathbf{t}^* \mid \theta) = \prod_{i=1}^{p} (\theta + \beta_i) \frac{\exp\left(-t_i^* (\theta + \beta_i)\right)}{1 - \exp\left(-\mathrm{d}_i (\theta + \beta_i)\right)}. \tag{5}$$

The Drift Diffusion Model

The DM was introduced by Ratcliff (1978) as a model for two-choice experiments. In the DM, evidence for either choice accumulates over time until a decision boundary is reached. One way to characterize this evidence-accumulation process is in terms of a Wiener process with constant drift and volatility, and absorbing upper and lower boundaries (Cox & Miller, 1970). The drift μ of the diffusion process determines how fast information is accumulated, the volatility σ determines how noisy the accumulation process is, and the distance between the two boundaries α determines how much evidence needs to be accumulated before a choice is made. The process has two additional parameters; a bias parameter z that indicates the distance from the starting point to the lower boundary, and the non-decision time $T_{(er)}$. A commonly used simplification of the DM assumes that the process is unbiased $z = \frac{1}{2}\alpha$.

The DM has been extended to model differences between persons and tasks. For example, Tuerlinckx and De Boeck (2005) proposed to decompose the drift μ of the accumulation process into a person and an item part —i.e., $\mu = \theta + \beta_i$— and to treat the distance between the boundaries as an item characteristic —i.e., $\alpha = \alpha_i$. The person component θ in the drift specification carries the interpretation of an ability in item response theory models, as a higher value of θ implies an increased probability of choosing the correct alternative. To identify the DM, Tuerlinckx and De Boeck (2005) fixed the volatility σ to one. The joint distribution of decision times and the chosen alternatives —i.e., response accuracies if the upper and the lower boundaries correspond to the correct and incorrect responses— is then equal to:

$$p(\mathbf{x}, \mathbf{t} \mid \theta) = \prod_{i=1}^{p} \frac{\pi}{\alpha_i^2} \exp\left(\frac{1}{2}\alpha_i\,(2x_i - 1)\,(\theta + \beta_i) - \frac{1}{2}\,(t_i - T_{(er)})\,(\theta + \beta_i)^2\right) \tag{6}$$
$$\times \sum_{n=1}^{\infty} \sin\left(\frac{1}{2}\pi n\right) \exp\left(-\frac{1}{2\alpha_i^2}\pi^2 n^2\,(t_i - T_{(er)})\right).$$

Both the SM and the DM directly specify a joint distribution of accuracies and decision times (response times) that is based on one latent trait; ability θ. In contrast to the SM, however, accuracies and response times are independent given ability in the (unbiased) DM. The marginal $p(\mathbf{x} \mid \theta)$ is the two-parameter logistic model in Eq. (1), where the discrimination parameter equals the distance between the boundaries in the diffusion process, and the easiness parameter is an item effect on drift of the diffusion process. The marginal $p(\mathbf{t} \mid \theta)$ is equal to

$$p(\mathbf{t} \mid \theta) = \prod_{i=1}^{p} \frac{2\pi}{\alpha_i^2} \cosh\left(\frac{1}{2}\alpha_i(\theta + \beta_i)\right) \exp\left(-\frac{1}{2}\left(t_i - T_i^{(er)}\right)(\theta + \beta_i)^2\right)$$
$$\times \sum_{n=1}^{\infty} \sin\left(\frac{1}{2}\pi n\right) \exp\left(-\frac{1}{2\alpha_i^2}\pi^2 n^2 \left(t_i - T_i^{(er)}\right)\right).$$

Even though the marginal distribution $p(\mathbf{x} \mid \theta)$ is a member of the exponential family, neither the marginal distribution $p(\mathbf{t} \mid \theta)$ nor the joint distribution $p(\mathbf{x}, \mathbf{t} \mid \theta)$ is a member of the exponential family. The primary reason that the latter cannot be written in exponential family form is because it implies a statistic $\mathrm{s}_1 = \mathrm{s}_1(t) = -\frac{1}{2}\sum_i t_i$ that is sufficient for θ^2 and another statistic $\mathrm{s}_2 = \mathrm{s}_2(t) = \sum_i \alpha_i\, x_i - \sum_i \beta_i\, t_i$ that is sufficient for θ. If we express the former as $\eta = \eta(\theta) = \theta^2$, we end up with an exponential family model subject to constraints on the parameters θ and η: η is functionally related to ability θ. This is known as a curved exponential family model (Efron, 1975, 1978).

Characterizing Manifest Probabilities of Latent Trait Models

Consider the general case of an *item response theory (IRT)* model for accuracy and response times in exponential family form:

$$p(\mathbf{x}, \mathbf{t} \mid \boldsymbol{\zeta}) = \prod_{i=1}^{p} p(x_i, t_i \mid \boldsymbol{\zeta}) = \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})} \mathrm{b}_i \exp\left(\mathbf{s}_i^{\mathsf{T}} \boldsymbol{\zeta}\right), \tag{7}$$

where $\mathrm{Z}_i(\boldsymbol{\zeta})$ is a normalizing constant, $\mathrm{b}_i = \mathrm{b}_i(x_i, t_i)$ a base measure that does not depend on the value of the latent variable $\boldsymbol{\zeta}$, and where $\mathbf{s}_i = \mathbf{s}_i(x_i, t_i)$ is a (possibly vector valued) statistic. For the HM and DM $\boldsymbol{\zeta} = (\theta, \eta)^{\mathsf{T}}$, and for the SM $\boldsymbol{\zeta} = \theta$. In our approach to express the manifest probabilities for models of the same form as Eq. (7) we will make use of the following latent variable distribution.[3]

Definition 1. *For models $p(\mathbf{x}, \mathbf{t} \mid \boldsymbol{\zeta})$ that are of the form of Eq.* (7), *we may define the latent variable distribution*

$$g(\boldsymbol{\zeta}) = \frac{1}{\mathrm{Z}} \prod_{i=1}^{p} \mathrm{Z}_i(\boldsymbol{\zeta})\, k(\boldsymbol{\zeta}), \tag{8}$$

where $k(\boldsymbol{\zeta})$ is a kernel density and Z is the normalizing constant of $g(\boldsymbol{\zeta})$. For every kernel distribution $k(\boldsymbol{\zeta})$ for which the normalizing constant Z is finite, i.e.,

$$0 < \mathrm{Z} = \int_{\Omega_{\zeta}} \prod_{i=1}^{p} \mathrm{Z}_i(\boldsymbol{\zeta})\, k(\boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\zeta} < \infty,$$

where Ω_{ζ} is the support of $\boldsymbol{\zeta}$, $g(\boldsymbol{\zeta})$ is a proper distribution.

The distribution in Definition 1 was inspired by the latent trait distribution that has been introduced with the latent variable expression of a graphical model from physics known as the Ising (1925) model by Kac (1968, see also Marsman et al., 2018; Epskamp, Maris, Waldorp, & Borsboom, 2018), but a similar construction can also be found in, for instance, Cressie and Holland (1983, Eq. A9) and McCullagh (1994). We can now state our first result.

Theorem 1. *When $p(\mathbf{x}, \mathbf{t} \mid \boldsymbol{\zeta})$ is of the form of Eq.* (7) *and the latent variable distribution $g(\boldsymbol{\zeta})$ in Eq.* (8) *is a proper distribution, then the manifest distribution $p(\mathbf{x}, \mathbf{t})$ is given by*

$$p(\mathbf{x}, \mathbf{t}) = \frac{1}{\mathrm{Z}} \prod_{i=1}^{p} \mathrm{b}_i\, \mathbb{E}_{k(\zeta)} \left(\prod_{i=1}^{p} \exp\left(\mathbf{s}_i^{\mathsf{T}} \boldsymbol{\zeta}\right) \right),$$

where Z is a normalizing constant.

We omit the simple proof of Theorem 1, which requires one to fill in the definitions of the models in Eqs. (7) and (8), and then integrate out the latent variable $\boldsymbol{\zeta}$. Theorem 1 shows that for any latent trait model of the form of Eq. (7), combined with a latent variable distribution of the form of Eq. (8), the manifest distribution can be characterized in terms of the base measures b_i and the moment

[3] For the DM we face the issue that η is functionally related to θ. This forces us to consider a variant of our approach to characterize its manifest probability distribution $p(\mathbf{x}, \mathbf{t})$. We discuss this variant of our approach when we express the manifest distribution of the DM.

generating function of the kernel distribution $k(\boldsymbol{\zeta})$. This is similar to the Dutch Identity that was proposed by Holland (1990) and was further developed by Ip (2002) and Hessen (2012). A version of the Dutch Identity for models $p(\mathbf{x}, \mathbf{t} \mid \boldsymbol{\zeta})$ of the form in Eq. (7) is provided in Appendix A.

Observe that Theorem 1 characterizes the manifest distribution in terms of a moment generating function. For the practical application of Theorem 1, it is important to find a convenient form for the moment generating function. This is also the case for the Dutch Identity, where general analytic solutions have only be found for the (extended) Rasch model (Cressie & Holland, 1983; Holland, 1990; Tjur, 1982). For other latent trait models, an assumption is made about the posterior distribution of the latent variable to compute the manifest distribution $p(\mathbf{x}, \mathbf{t})$. In a similar way, a kernel $k(\boldsymbol{\zeta})$ has to be chosen for the application of Theorem 1. The relation between these two assumptions is discussed in Appendix A.

One compelling reason to choose a normal kernel distribution $k(\boldsymbol{\zeta})$ is that we can then conveniently use the Gaussian identity,

$$\exp\left(\frac{1}{2}a^2\right) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(a\zeta - \frac{1}{2}\zeta^2\right) \mathrm{d}\zeta,$$

to express the manifest probability distributions analytically.

Corollary 1. *If* $p(\mathbf{x}, \mathbf{t} \mid \zeta)$ *is of the form in Eq.* (7)*, and the latent variable distribution* $g(\boldsymbol{\zeta})$ *is of the form in Eq.* (8) *with a multivariate normal kernel* $k(\boldsymbol{\zeta})$ *having a mean vector* $\mathbf{m}$ *and covariance matrix* $\mathbf{V}$*, then (i)*

$$p(\mathbf{x}, \mathbf{t}) = \frac{1}{\mathrm{Z}} \exp\left(\sum_{i=1}^{p} \ln \mathrm{b}_i + \left[\sum_{i=1}^{p} \mathbf{s}_i\right]^{\mathsf{T}} \mathbf{m} + \frac{1}{2}\left[\sum_{i=1}^{p} \mathbf{s}_i\right]^{\mathsf{T}} \mathbf{V} \left[\sum_{i=1}^{p} \mathbf{s}_i\right]\right), \tag{9}$$

where Z is a normalizing constant, and (ii) the posterior distribution $g(\boldsymbol{\zeta} \mid \mathbf{s}_1, \ldots, \mathbf{s}_p)$ *is multivariate normal with mean vector* $\mathbf{m} + \mathbf{V}\left[\sum_{i=1}^{p} \mathbf{s}_i\right]$ *and covariance matrix* $\mathbf{V}$.

We omit the proof of Corollary 1, which requires one to insert in Theorem 1 the moment generating function of the multivariate normal distribution, i.e.,

$$\mathbb{E}\left(\exp\left(\mathbf{r}^{\mathsf{T}}\boldsymbol{\zeta}\right)\right) = \exp\left(\mathbf{r}^{\mathsf{T}}\mathbf{m} + \frac{1}{2}\mathbf{r}^{\mathsf{T}}\mathbf{V}\mathbf{r}\right),$$

using $\mathbf{r} = \sum_i \mathbf{s}_i$ as the interpolating parameter vector. Corollary 1 mirrors the results for assuming posterior normality of the latent trait in combination with the Dutch Identity as evidenced in Corollary 1 of Holland (1990), Corollary 1 of Ip (2002), and Theorem 2 of Hessen (2012), see also the log-multiplicative association models of Anderson and Vermunt (2000), and Anderson and Yu (2007), and the fused latent and graphical IRT model of Chen, Li, Liu, and Ying (2018). In fact, our Corollary 1 shows that when we use the latent variable model $g(\boldsymbol{\zeta})$ from Eq. (8) with a normal kernel $k(\boldsymbol{\zeta})$, the posterior distribution of the latent variable is also normal (see also Appendix A). Similarly, it can be shown that when we assume that the posterior distribution is multivariate normal with mean vector $\mathbf{m} + \mathbf{V}\left[\sum_{i=1}^{p} \mathbf{s}_i\right]$ and covariance matrix $\mathbf{V}$, then Corollary 1 of Holland (1990) assumes a latent variable distribution $g(\boldsymbol{\zeta})$ of the form of Eq. (8) with a multivariate normal kernel $k(\boldsymbol{\zeta})$ with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{V}$. We will use Corollary 1 to characterize the manifest probability distributions.

The Manifest Probabilities of the Three Latent Trait Models

In this section we characterize the manifest probability distributions $p(\mathbf{x}, \mathbf{t})$ for the three latent trait models. We consider each of the models in turn, and shortly discuss the results for each of them.

The Hierarchical Model

The version of the HM that is considered here is of the form

$$p\left(\mathbf{x}, \mathbf{t} \mid \boldsymbol{\zeta} = (\theta, \eta)^{\mathsf{T}}\right) g\left(\boldsymbol{\zeta} = (\theta, \eta)^{\mathsf{T}}\right) = p(\mathbf{x} \mid \theta)\, p(\mathbf{t} \mid \eta)\, g(\theta, \eta),$$

where the marginal $p(\mathbf{x} \mid \theta)$ is the two-parameter logistic model introduced in Eq. (1), the marginal $p(\mathbf{t} \mid \eta)$ is the log-normal distribution introduced in Eq. (2), and $g(\theta, \eta)$ is of the form of Eq. (8) using a bivariate normal kernel distribution $k(\theta, \eta)$ for ability θ and speed η, using a mean vector $\mathbf{m} = \mathbf{0}$ and covariance matrix

$$\mathbf{V} = \begin{pmatrix} 1 & \rho\, v_\eta \\ \rho\, v_\eta & v_\eta^2 \end{pmatrix}.$$

To come to an expression for the manifest probability distribution of this version of the HM we first rewrite the conditional distribution of accuracies and response times $p(\mathbf{x}, \mathbf{t} \mid \theta, \eta)$ to fit the form of Eq. (7). To this aim, we introduce the statistic

$$\mathbf{s}_i = \mathbf{s}_i(x_i, t_i) = \begin{pmatrix} x_i\alpha_i \\ -\ln(t_i)\,\frac{1}{2}\phi_i \end{pmatrix},$$

the base measures

$$\mathrm{b}_i = \mathrm{b}_i(x_i, t_i) = \exp\left(x_i\alpha_i\beta_i - \ln(t_i)^2\,\frac{1}{2}\phi_i - \ln(t_i)\,(1 - \phi_i\xi_i)\right),$$

and normalizing constants

$$\mathrm{Z}_i(\theta, \eta) = \frac{\sqrt{2\pi}}{\sqrt{\phi_i}}\,\left\{\exp\left(\frac{1}{2}\,\phi_i\,(\eta - \xi_i)^2\right) + \exp\left(\frac{1}{2}\,\phi_i\,(\eta - \xi_i)^2 + \alpha_i(\theta + \beta_i)\right)\right\}.$$

Having expressed the conditional distribution of accuracies and response times in the form of Eq. (7), we can now apply Corollary 1 to obtain the manifest distribution of the accuracies $\mathbf{x}$ and response times. It is slightly less cumbersome to characterize this manifest distribution for log-transformed response times $u_i = \ln(t_i)$ instead of response times directly. The manifest probability distribution of accuracies and log-transformed response times that results is equal to

$$p(\mathbf{x}, \mathbf{u}) = \frac{1}{\mathrm{Z}}\exp\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{u}\end{pmatrix}^{\mathsf{T}}\begin{pmatrix}\boldsymbol{\alpha}\odot\boldsymbol{\beta}\\ \boldsymbol{\phi}\odot\boldsymbol{\xi}\end{pmatrix} + \frac{1}{2}\begin{pmatrix}\mathbf{x}\\ \mathbf{u}\end{pmatrix}^{\mathsf{T}}\begin{pmatrix}\boldsymbol{\alpha}\,\boldsymbol{\alpha}^{\mathsf{T}} & -\rho\, v_\eta\,\boldsymbol{\alpha}\,\boldsymbol{\phi}^{\mathsf{T}}\\ -\rho\, v_\eta\,\boldsymbol{\phi}\,\boldsymbol{\alpha}^{\mathsf{T}} & v_\eta^2\,\boldsymbol{\phi}\,\boldsymbol{\phi}^{\mathsf{T}}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{u}\end{pmatrix}\right)\,\omega(\mathbf{u}),$$

where $\odot$ refers to Hadamard's product, and $\omega(\mathbf{u})$ is the base measure of a normal distribution

$$\omega(\mathbf{u}) = \prod_{i=1}^{p}\omega_i(u_i) = \prod_{i=1}^{p}\exp\left(-\frac{1}{2}\phi_i\, u_i^2\right),$$

with precisions ϕ_i. Observe that for this manifest distribution the association between accuracy and log-response time is of the opposite sign of the correlation ρ between speed η and ability θ:

faster responses correspond to correct answers when $\rho > 0$; slower responses correspond to correct answers when $\rho < 0$.

There are two important characteristics that can be observed from the manifest probability distribution of our version of the HM. Firstly, the base measure $\omega(\mathbf{u})$ that is used here stipulates an *a priori* restriction on the variance of the speed parameter ν_η^2 and the item specific precisions ϕ_i. To see this, observe that for this base measure the manifest distribution is a proper probability distribution—i.e., integrates to one—if and only if

$$\nu_\eta^2 < \min_i \left(\frac{1}{\phi_i}\right).$$

Since the variance of the $\log(t_i)$ is equal to,

$$\text{Var}(\log(t_i)) = \mathbb{E}(\text{Var}(\log(t_i) \mid \eta)) + \text{Var}(\mathbb{E}(\log(t_i) \mid \eta)) = \frac{1}{\phi_i} + \nu_\eta^2,$$

this restriction on ν_η^2 and ϕ_i implies that the speed variable can account for less than 50% of the total variance of the $\log(t_i)$.

A second important characteristic that can be observed from the manifest probability distribution of our version of the HM concerns the association between accuracies and log response times. First, note that when the association between an accuracy x_i and log-transformed response time u_i in the manifest distribution is equal to zero, these variables are independent conditional upon the remaining accuracies and log-transformed response times:

$$\rho\, \nu_\eta\, \alpha_i\, \phi_i = 0 \iff x_i \perp\!\!\!\perp u_i \;\Big|\; \left(\mathbf{x}^{(i)}, \mathbf{u}^{(i)}\right)^{\mathsf{T}}.$$

Thus, x_i and u_i are conditionally independent whenever either the correlation between speed and ability is zero; the discrimination of item i is zero; the precision of item i is zero and/or the *a priori* variance of the speed variable is zero. Observe, however, that when the discrimination of item i is zero the accuracy for item i is completely independent of ability, and the same holds for its precision and the dependency between the associated response time and speed. Thus, the only non-trivial way that accuracy and response times are independent in the manifest expression of our version of the HM is when ability and speed are *a priori* independent. But this entails the extreme case that all of the accuracies would be independent from all of the response times.

The Signed Residual Time Model

There are two versions of the SM that are considered here. The first version of the SM stipulates a distribution of accuracies and residual response times, and the second version of the SM stipulates a distribution of accuracies and *pseudo* response times. We will first characterize the manifest probability distribution of accuracies and residual times, and revert to *pseudo* response times after that.

The manifest distribution of accuracy and residual response time. The SM was introduced in Eq. (3) and can be expressed in the exponential family form of Eq. (7) with sufficient statistics

$$\mathrm{s}_i = \mathrm{s}_i(x_i, t_i) = (2x_i - 1)(\mathrm{d}_i - t_i),$$

base measures

$$\mathrm{b}_i = \mathrm{b}_i(x_i, t_i) = \exp\left((2x_i - 1)(\mathrm{d}_i - t_i)\beta_i\right),$$

and normalizing constants

$$\mathrm{Z}_i(\theta) = \frac{\exp\left(\mathrm{d}_i(\theta + \beta_i)\right) - \exp\left(-\mathrm{d}_i(\theta + \beta_i)\right)}{\theta + \beta_i}. \tag{10}$$

Having expressed the conditional distribution of accuracies and residual times of the SM in the form of Eq. (7) we can now use Corollary 1 to obtain their manifest distribution. Assuming a normal kernel with mean $m = 0$ and variance $V = 1$, Corollary 1 leads us to the following manifest distribution,

$$p(\mathbf{x}, \mathbf{t}) = \frac{1}{\mathrm{Z}} \exp\left(\left[(2\mathbf{x} - \mathbf{1}_p) \odot (\mathbf{d} - \mathbf{t})\right]^\mathsf{T} \boldsymbol{\beta} + \frac{1}{2}\left(2\mathbf{x} - \mathbf{1}_p\right)^\mathsf{T} (\mathbf{d} - \mathbf{t})\,(\mathbf{d} - \mathbf{t})^\mathsf{T} \left(2\mathbf{x} - \mathbf{1}_p\right)\right),$$

where $\mathbf{1}_p$ is the unit vector of length p. One way to write this distribution more succinctly is to express it in terms of the random variables $y_i = (2x_i - 1)$ and residual times $r_i = \mathrm{d}_i - t_i$, which gives

$$p(\mathbf{y}, \mathbf{r}) = \frac{1}{\mathrm{Z}} \exp\left(\mathbf{y}^\mathsf{T} (\mathbf{r} \odot \boldsymbol{\beta}) + \mathbf{y}^\mathsf{T} \left(\frac{1}{2} \mathbf{r}\, \mathbf{r}^\mathsf{T}\right) \mathbf{y}\right). \tag{11}$$

In this model, larger residual times (faster responses) are associated with an increased probability that a person responds accurately to easy items ($\beta_i > 0$) and a decreased probability to respond accurately to difficult items ($\beta_i < 0$). The association between the response accuracies of different items is positive and increases with increasing residual response times (faster responses).

There are two important characteristics that can be observed from the manifest probability distribution of the SM. Firstly, we observe that this manifest expression closely resembles a graphical model from physics that is known as the Ising model (Lenz, 1920; Ising, 1925). The Ising model is characterized by the following probability distribution over realizations of $\mathbf{y}$

$$p(\mathbf{y}) = \frac{1}{\mathrm{Z}} \exp\left(\mathbf{y}^\mathsf{T} \boldsymbol{\mu} + \mathbf{y}^\mathsf{T} \boldsymbol{\Sigma}\, \mathbf{y}\right),$$

where $\boldsymbol{\mu}$ denotes a vector of p intercepts μ_i, and $\boldsymbol{\Sigma}$ is a symmetric $p \times p$ matrix of pairwise interactions σ_{ij}. However, where the intercepts and interactions are fixed effects in the Ising network model, the intercepts μ_i and pairwise interactions σ_{ij} are random effects in the manifest expression $p(\mathbf{y}, \mathbf{r})$, with $\mu_i = r_i \times \beta_i$ and $\sigma_{ij} = \frac{1}{2} r_i \times r_j$. The SM provides a novel way to model the associations of the Ising model as random effects.

A second important characteristic that can be observed from the manifest probability distribution of the SM concerns the association between accuracies and residual response times. Whereas we have found that for the manifest distribution of the HM that accuracy can be independent of response time in a non-trivial manner with $\rho = 0$, accuracy cannot be independent of residual time in a non-trivial manner for the manifest expression of the SM. This can be observed, for example, from the conditional distribution of accuracy and residual response time for an item i given the responses and residual times of the remaining items

$$p\left(y_i, r_i \mid \mathbf{y}^{(i)}, \mathbf{t}^{(i)}\right) = \frac{\exp\left(y_i r_i \left[\beta_i + \sum_{j \neq i} y_j r_j\right]\right)}{\frac{\exp\left(\beta_i + \sum_{j \neq i} y_j r_j\right) - \exp\left(-\beta_i - \sum_{j \neq i} y_j r_j\right)}{\beta_i + \sum_{j \neq i} y_j r_j}},$$

from which it is clear that there are no values of β_i (or $\mathbf{y}^{(i)}$ and $\mathbf{r}^{(i)}$) that render y_i and r_i as (conditionally) independent.

The manifest distribution of accuracy and *pseudo* response time. An alternative formulation of the SM is in terms of accuracies and *pseudo* response times, which is of the form

$$p(\mathbf{x}, \mathbf{t}^* \mid \theta) = p(\mathbf{x} \mid \theta)\, p(\mathbf{t}^* \mid \theta),$$

where the marginal $p(\mathbf{x} \mid \theta)$ is the two-parameter logistic model in Eq. (1), and the marginal $p(\mathbf{t}^* \mid \theta)$ is given in Eq. (5).

To characterize the manifest distribution of accuracies and *pseudo* response times for this version of the SM we can take two approaches. Firstly, we may express the conditional distribution of accuracies and *pseudo* response times $p(\mathbf{x}, \mathbf{t}^* \mid \theta)$ in the exponential family form of Eq. (7) and then apply Corollary 1 using a normal kernel distribution with mean $m = 0$ and variance $V = 1$ to this conditional distribution. Alternatively, we may rewrite the sufficient statistic for residual response times in the manifest distribution in Eq. (11) through the relation

$$\mathrm{s}_i^* = x_i \mathrm{d}_i - t_i^* = \begin{cases} \mathrm{d}_i - t_i & \text{if } x_i = 1 \\ -(\mathrm{d}_i - t_i) & \text{if } x_i = 0 \end{cases} \Longleftrightarrow (2x_i - 1)(\mathrm{d}_i - t_i) = y_i r_i = \mathrm{s}_i.$$

Both approaches lead to the following manifest distribution of accuracies and *pseudo* response times

$$p(\mathbf{x}, \mathbf{t}^*) = \frac{1}{\mathrm{Z}} \exp\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{t}^*\end{pmatrix}^{\mathsf{T}} \begin{pmatrix}\mathbf{d} \odot \boldsymbol{\beta}\\ -\boldsymbol{\beta}\end{pmatrix} + \frac{1}{2} \begin{pmatrix}\mathbf{x}\\ \mathbf{t}^*\end{pmatrix}^{\mathsf{T}} \begin{pmatrix}\mathbf{d}\,\mathbf{d}^{\mathsf{T}} & -\mathbf{d}\,\mathbf{1}_p^{\mathsf{T}}\\ -\mathbf{1}_p\,\mathbf{d}^{\mathsf{T}} & \mathbf{1}_p\,\mathbf{1}_p^{\mathsf{T}}\end{pmatrix} \begin{pmatrix}\mathbf{x}\\ \mathbf{t}^*\end{pmatrix}\right). \tag{12}$$

Observe that for this model both the associations between accuracies and the associations between *pseudo* response times are positive, yet the associations between accuracies and *pseudo* response times are negative: correct responses are associated to smaller *pseudo* time values (faster response times); incorrect responses are associated to larger *pseudo* time values (slower response times).

The manifest distribution of accuracies and *pseudo* response times of the SM is of the same form as the manifest distribution of accuracies and the log transformed response times of the HM, and thus they share certain characteristics. For example, both models share the following Markov property: When the association between an accuracy and a response time for an item i is equal to zero, then this implies that these two variables are independent conditional upon the remaining accuracies and response times,

$$-\frac{1}{2}\mathrm{d}_i = 0 \Longleftrightarrow x_i \perp\!\!\!\perp t_i^* \;\Big|\; \left(\mathbf{x}^{(i)}, \mathbf{t}^{*(i)}\right)^{\mathsf{T}}.$$

However, it is immediately clear that this association is never zero in practice, since $d_i = 0$ implies a zero second time limit for item i.

The Drift Diffusion Model

The version of the DM that is considered here —introduced in Eq. (6)— cannot be expressed in the exponential family form of Eq. (7). This means that we cannot make use of Theorem 1 or Corollary 1 to express its manifest probability distribution. However, we can express the conditional distribution of accuracy and response time in the following form

$$p(\mathbf{x}, \mathbf{t} \mid \theta) = \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\theta)}\, \mathrm{b}_i \exp\left(-s_{1i}\theta^2 + s_{2i}\theta\right).$$

With the DM expressed in this form we can use a latent variable distribution of the form of Eq. (8) in combination with a normal kernel density $k(\theta)$ to come to an analytic expression for the manifest probability distribution.

To express the DM in this form we introduce the statistics

$$
\begin{aligned}
s_{1i} &= \frac{1}{2} t_i \\
s_{2i} &= \alpha_i\, x_i - \beta_i\, t_i,
\end{aligned}
$$

the base measures

$$
\mathrm{b}_i = \mathrm{b}_i(x_i, t_i) = \exp\left(\alpha_i \beta_i\, x_i - \frac{1}{2}\beta_i^2\, t_i\right) \sum_{n=1}^{\infty} \sin\left(\frac{1}{2}\pi n\right) \exp\left(-\frac{1}{2\alpha_i^2}\pi^2 n^2 \left(t_i - T_{(er)}\right)\right),
$$

and normalizing constants

$$
\mathrm{Z}_i(\theta) = \frac{\alpha_i^2}{\pi} \exp\left(\frac{1}{2}\alpha_i(\theta + \beta_i) - \frac{1}{2} T_{(er)}(\theta + \beta_i)^2\right).
$$

We can now define a latent trait distribution $g(\theta)$ of the form of Eq. (8).

When we use a latent trait distribution $g(\theta)$ of the form of Eq. (8) with a normal kernel density $k(\theta)$ with mean $m = 0$ and variance $V = 1$, the manifest distribution for our version of the DM can be expressed as follows

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{t}) &= \int_{\mathbb{R}} \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\theta)} \mathrm{b}_i \exp\left(-s_{1i}\theta^2 + s_{2i}\theta\right) \frac{1}{\mathrm{Z}} \prod_{i=1}^{p} \mathrm{Z}_i(\theta) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\theta^2\right) \mathrm{d}\theta \\
&= \frac{1}{\mathrm{Z}\sqrt{2\pi}} \prod_{i=1}^{p} \mathrm{b}_i \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left[2\sum_{i=1}^{p} s_{1i} + 1\right]\theta^2 + \frac{1}{2}\left[2\sum_{i=1}^{p} s_{2i}\right]\theta\right) \mathrm{d}\theta \\
&= \frac{1}{\mathrm{Z}\sqrt{2\pi}} \prod_{i=1}^{p} \mathrm{b}_i \int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left[2\sum_{i=1}^{p} s_{1i} + 1\right]\theta^2 + \left[\sum_{i=1}^{p} s_{2i}\right]\theta\right) \mathrm{d}\theta \\
&= \frac{1}{\mathrm{Z}} \prod_{i=1}^{p} \mathrm{b}_i \frac{1}{\sqrt{2\sum_{i=1}^{p} s_{1i} + 1}} \exp\left(\frac{1}{2} \frac{\left[\sum_{i=1}^{p} s_{2i}\right]^2}{\left[2\sum_{i=1}^{p} s_{1i} + 1\right]}\right).
\end{aligned}
$$

Rewriting the statistics in terms of accuracies and response times, combining exponents in the base measures b_i with that in the rightmost factor and simplifying the result leads us to the following expression of the manifest distribution

$$
p(\mathbf{x}, \mathbf{t}) = \frac{1}{\mathrm{Z}} \exp\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{t}\end{pmatrix}^{\mathsf{T}} \begin{pmatrix}\boldsymbol{\alpha} \odot \boldsymbol{\beta} \\ -\frac{1}{2}\boldsymbol{\beta} \odot \boldsymbol{\beta}\end{pmatrix} + \frac{1}{2t_+ + 2} \begin{pmatrix}\mathbf{x}\\ \mathbf{t}\end{pmatrix}^{\mathsf{T}} \begin{pmatrix}\boldsymbol{\alpha}\\ -\boldsymbol{\beta}\end{pmatrix} \begin{pmatrix}\boldsymbol{\alpha}\\ -\boldsymbol{\beta}\end{pmatrix}^{\mathsf{T}} \begin{pmatrix}\mathbf{x}\\ \mathbf{t}\end{pmatrix}\right) \omega(\mathbf{t}) \tag{13}
$$

where $t_+ = \sum_i t_i$, and $\omega(\mathbf{t})$ is a base measure,

$$
\omega(\mathbf{t}) = \frac{1}{\sqrt{t_+ + 1}} \prod_{i=1}^{p} \sum_{n=1}^{\infty} \sin\left(\frac{1}{2}\pi n\right) \exp\left(-\frac{1}{2\alpha_i^2}\pi^2 n^2 \left(t_i - T_{(er)}\right)\right).
$$

There are two important characteristics that can be observed from the manifest probability distribution of our version of the DM. The first observation is that the association between both accuracies and response times is scaled by the total time t_+ that is spent on the test. This implies smaller associations between accuracies and response times for pupils that take longer to complete the test, and larger associations for pupils that take less time to complete the test. In none of the three other manifest probability distributions that were considered here have we seen an influence of the total time that was spent on the test.

A second important characteristic that can be observed from the manifest distribution of our version of the DM concerns the association between accuracies and response time. The associations between accuracies and response times are scaled with the total test time, i.e.,

$$x_i\, t_i\, \frac{1}{1 + t_+}\, \frac{1}{2}\, \alpha_i \beta_i,$$

which implies that both accuracies and response times are associated to the (remaining) response times. There is only one way for accuracies to be independent from response times in the manifest distribution, which is when the associated discriminations are equal to zero. However, when this is the case the accuracies are not only independent of all of the response times, but they are also independent of the latent variable (ability) in the DM.

Discussion

In this paper we have characterized and studied the manifest probability distributions for accuracy and response times of three latent trait models: The Hierarchical model (HM) of van der Linden (2007), the Signed Residual Time model (SM) of Maris and van der Maas (2012), and the Drift Diffusion model (DM) as proposed by Tuerlinckx and De Boeck (2005). One major difference between the HM and DM on the one hand, and the SM on the other hand is in the relation they stipulate between accuracy and response times after having conditioned on the latent trait variables. Whereas the HM and DM assume or imply that accuracy and response time are independent conditional upon the latent variable, i.e.,

$$p(\mathbf{x}, \mathbf{t} \mid \zeta) = p(\mathbf{x} \mid \zeta) p(\mathbf{t} \mid \zeta),$$

this is not the case for the SM, where accuracy and response times are still correlated after conditioning on the latent variable.

One of our main findings is that accuracies and response times are dependent on each other in the manifest distributions for each of the investigated latent trait models. That is, the accuracy and response time of an item i are associated in the manifest distribution, a result that holds regardless of whether the underlying latent trait model assumes or implies conditional independence or not. Only for fringe cases have we found that accuracy and response time can become uncorrelated, such as when there is a zero correlation between speed and ability in the HM. Even though we found that the accuracies and response times are associated in each of the manifest distributions, we also observed fundamental differences in what function of response time was related to ability (response accuracy). For example, the log-transformed response times are modeled in the HM, the residual response times or *pseudo* response times are modeled in the SM, and in the DM the response times are modeled directly. This suggests an alternative direction of research on latent trait models for accuracy and response times, one that is less focused on the conditional independence property of the latent trait model and more focused on what response time function is best associated with ability (or response accuracy). A comparison of the manifest distributions that are found in this paper can help decide which function of response times is best associated to ability (or response accuracy).

This is not to say that the conditional independence property for the latent trait model does not have an impact in the characterization of its manifest distribution. In fact, latent trait models in the exponential family with the conditional independence property appear to generate manifest probability distributions that are known as Markov random fields (Kindermann and Snell, 1980), which are undirected graphical models with Markov properties (e.g., Lauritzen, 2004). Most importantly here is the property that the accuracy and response time of an item i are independent conditional upon the remaining accuracies and response times whenever the association between accuracy and response time of item i is equal to zero. This is an important and interesting property. The HM and SM for *pseudo* response times are examples of models with this property, even though we find that for these two models the association between accuracy and response time of an item i is equal to zero only in fringe cases in practice. In earlier work on the multidimensional IRT model we have found expressions of the manifest probability distribution with the Markov property that could express conditional independencies in more detail (Marsman, Maris, Bechger, and Glas, 2015; Epskamp et al., 2018; Marsman et al., 2018). This suggests that if we want to seriously model the dependency between accuracy and response time for a particular item given the remaining accuracies and response times, we have to consider higher-dimensional latent trait models.

Even though the manifest probability distributions of the DM and the SM for residual times are not Markov random fields[4], these manifest expressions are interesting in their own right. In the manifest expression for the DM, for example, the associations between accuracies and response times are a function of the total time the pupil has spent on the test, an aspect that is not being modeled in any of the other manifest expressions. This is one potential point of departure for modelling the impact of time restrictions on a test on the distribution of response accuracy. The manifest expression of the SM, on the other hand, provides a new and interesting way to view an old model, the Ising model. The Ising model is a undirected graphical model that is characterized by the following distribution,

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(\mathbf{x}^{\mathsf{T}}\boldsymbol{\mu} + \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{x}\right),$$

where $\mathbf{x}$ is a p-dimensional vector of $(0, 1)$ or $(-1, 1)$ variables x_i, $\boldsymbol{\mu}$ a $p-$dimensional vector of main effects, and $\boldsymbol{\Sigma}$ a $p \times p$ symmetric matrix of pairwise associations between variables. Whereas the pairwise associations are fixed effects in the Ising model, the manifest distribution of the SM indicates one way to model these associations as a random effect.

One famous conjecture from Holland (1990, p. 11) is that if there are large number of items on a test, and a smooth unidimensional IRT model (for accuracies) is used, the posterior distribution of the latent trait will be approximately normal. This conjecture has inspired several publications on the posterior normality of the latent trait in the context of IRT models for response accuracy (e.g., Chang and Stout, 1993; Chang, 1996; Zhang and Stout, 1997). An interesting conclusion that Holland (1990) deduced from this conjecture, in combination with the assumption that the log-likelihoods of the p items can be approximated using a p-variate normal with a rank one covariance

[4]Observe, however, that for both models the conditional distribution $p(\mathbf{x} \mid \mathbf{t})$ is a Markov random field, because they share the Markov property that when the association between the accuracy of an item i and an item j is equal to zero, these two variables are independent given the accuracy on the remaining items. Incidentally, for the manifest expressions of both the SM and the DM the associated conditional distribution is another instance of the Ising model.

matrix, is that the log of the manifest distribution of accuracy is approximately of quadratic form consisting of p main effects and a $p \times p$ matrix of associations that was of rank one. This enticed Holland (1990) to add a second conjecture that only two parameters can be consistently estimated per item. This idea points to interesting avenues of future research, such as the asymptotic posterior normality of the latent trait in the context of IRT models for response accuracy and response times. If it is reasonable to approximate the posterior of the latent trait (or the log-likelihood function) with a normal distribution, then we can use this approximation in combination with Corollary 1 or Theorem 2 to investigate the complexity of models for response accuracy and response times, and how model complexity is impacted by conditional independence property of the underlying latent trait model.

The latent variable distribution $g(\boldsymbol{\zeta})$ has allowed us to express the manifest probability distributions for a large class of latent trait models, but it also generated an unexpected parameter restriction in the manifest distribution of the HM, where we found that the variance of the speed variable ν_η^2 needed to be smaller than the smallest log-normal variance ϕ_i^{-1}. This parameter restriction follows from omitting the normalizing constants $\mathrm{Z}_i(\boldsymbol{\zeta})$ of the latent variable model in Eq. (7), which provides prior model structure. When a regular latent variable distribution is used —for example, a normal distribution on η— the model structure that is provided by the normalizing constants $\mathrm{Z}_i(\boldsymbol{\zeta})$ is integrated instead. Marsman et al. (2018) studied a similar scaling issue of the posterior distribution that results from using the latent variable distribution $g(\boldsymbol{\zeta})$ in the context of multidimensional IRT (see also Marsman, Waldorp, & Maris, 2017). Similar observations can be made for the posteriors and latent variable distributions that are generated by the latent variable model in Corollary 1 of Holland (1990), Corollary 1 of Hessen (2012) and Theorem 1 in Ip (2002).

The particular restriction that is imposed on the variance of the speed variable ν_η^2 in the HM is a rather strong restriction from a substantive point of view. From the perspective of the manifest distribution, however, it might be less of an issue since ν_η^2 is simply a scaling factor for the interactions between the log transformed response times and accuracies $\mathbf{x}$. That is, the manifest structure would not change when we absorbed ν_η^2 in the precisions and simply use the matrix of associations:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\alpha}\boldsymbol{\alpha}^\mathsf{T} & -\rho\boldsymbol{\alpha}\,\boldsymbol{\phi}^\mathsf{T} \\ -\rho\boldsymbol{\phi}\,\boldsymbol{\alpha}^\mathsf{T} & \boldsymbol{\phi}\,\boldsymbol{\phi}^\mathsf{T} \end{pmatrix}.$$

Alternatively, we may adopt a different base measure $\omega(\cdot)$ to remove the restriction.

References

Anderson, C., & Vermunt, J. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, *30*(1), 81–121. doi: 10.1111/0081-1750.00076

Anderson, C., & Yu, H.-T. (2007). Log-multiplicative association models as item response models. *Psychometrika*, *72*(1), 5–23. doi: 10.1007/s11336-005-1419-2

Bolsinova, M., De Boeck, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika*, *82*(4), 1126–1148. doi: 10.1007/s11336-016-9537-6

Bolsinova, M., & Maris, G. (2016). A test for conditional independence between response time and accuracy. *British Journal of Mathematical and Statistical Psychology*, *69*, 62–79. doi: 10.1111/bmsp.12059

Bolsinova, M., & Tijmstra, J. (2016). Posterior predictive checks for conditional independence between response time and accuracy. *Journal of Educational and Behavioural Statistics*, *41*(4), 123–145. doi: 10.3102/1076998616631746

Bolsinova, M., Tijmstra, J., & Molenaar, D. (2017). Response moderation models for conditional dependence between response time and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *70*(2), 257–279. doi: 10.1111/bmsp.12076

Chang, H. (1996). The asymptotic posterior normality of the latent trait for polytomous irt models. *Psychometrika*, *61*(3), 445–463. doi: 10.1007/BF02294549

Chang, H., & Stout, W. (1993). The asymptotic posterior normality of the latent trait in an irt model. *Psychometrika*, *58*(1), 37–52. doi: 10.1007/BF02294469

Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Robust measurement via a fused latent and graphical item response theory model. *Psychometrika*, *83*(3), 538–562. doi: 10.1007/s11336-018-9610-4

Cox, D., & Miller, H. (1970). *The theory of stochastic processes*. London, UK: Chapman & Hall.

Cressie, N., & Holland, P. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, *48*(1), 129–141. doi: 10.1007/BF02314681

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *The Annals of Statistics*, *3*(6), 1189–1242. doi: 10.1214/aos/1176343282

Efron, B. (1978). The geometry of exponential families. *The Annals of Statistics*, *6*(2), 362–376. doi: 10.1214/aos/1176344130

Epskamp, S., Maris, G., Waldorp, L., & Borsboom, D. (2018). Handbook of psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), (chap. Network Psychometrics). New York, NY: Wiley-Blackwell.

Fox, J.-P., Klein Entink, R., & van der Linden, W. (2007). Modeling of responses and response times with the package CIRT. *Journal of Statistical Software*, *20*(7). doi: 10.18637/jss.v020.i07

Hessen, D. (2012). Fitting and testing Conditional Multinormal Partial Credit Models. *Psychometrika*, *77*(4), 693–709. doi: 10.1007/s11336-012-9277-1

Holland, P. (1990). The Dutch Identity: A new tool for the study of item response models. *Psychometrika*, *55*(6), 5–18. doi: 10.1007/BF02294739

Ip, E. (2002). Locally dependent latent trait model and the Dutch Identity revisited. *Psychometrika*, *67*(3), 367–386. doi: 10.1007/BF02294990

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, *31*(1), 253–258. doi: 10.1007/BF02980577

Kac, M. (1968). Statistical physics: Phase transitions and superfluidity, vol. 1, Brandeis University summer institute in theoretical physics. In M. Chrétien, E. Gross, & S. Deser (Eds.), (pp. 241–305). New York: Gordon and Breach Science Publishers.

Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications* (Vol. 1). Providence, Rhode Island: American Mathematical Society.

Klein Entink, R., van der Linden, W., & Fox, J.-P. (2009). A Box-Cox normal model for response times. *British Journal of Mathematical and Statistical Psychology*, *62*(3), 621–640. doi: 10.1348/000711008X374126

Lauritzen, S. (2004). *Graphical models*. Oxford, UK.: Oxford University Press.

Lenz, W. (1920). Beiträge zum verständnis der magnetischen eigenschaften in festen körpern. *Physikalische Zeitschrift*, *21*, 613–615.

Maris, G., & van der Maas, H. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, *77*(4), 615–633. doi: 10.1007/S11336-012-9288-Y

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L., ... Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*(1), 15–35. doi: 10.1080/00273171.2017.1379379

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2015). Bayesian inference for low-rank Ising networks. *Scientific Reports*, *5*(9050). doi: 10.1038/srep09050

Marsman, M., Waldorp, L., & Maris, G. (2017). A note on large-scale logistic prediction: Using an approximate graphical model to deal with collinearity and missing data. *Behaviormetrika*, *44*(2), 513–534. doi: 10.1007/s41237-017-0024-x

McCullagh, P. (1994). Exponential mixtures and quadratic exponential families. *Biometrika*, *81*(4), 721–729. doi: 10.1093/biomet/81.4.721

Molenaar, D., Tuerlinckx, F., & van der Maas, H. (2015a). A bivariate generalized linear item response theory modeling framework to the analysis of responses and response times. *Multivariate Behavioural Research*, *50*(1), 56–74. doi: 10.1080/00273171.2014.962684

Molenaar, D., Tuerlinckx, F., & van der Maas, H. (2015b). A generalized linear factor model approach to the hierarchical framework for responses and response times. *British Journal of Mathematical and Statistical Psychology*, *68*(2), 197–219. doi: 10.1111/bmsp.12042

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. doi: 10.1037/0033-295X.85.2.59

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative poisson model. *Scandinavian Journal of Statistics*, *9*(1), 23–30.

Tuerlinckx, F., & De Boeck, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika*, *70*(4), 629–650. doi: 10.1007/s11336-000-0810-3

van der Linden, W. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*(3), 287–308. doi: 10.1007/s11336-006-1478-z

van der Linden, W. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioural Statistics*, *31*(2), 181–204. doi: 10.3102/10769986031002181

van der Linden, W., & Clas, C. (2010). Statistical tests of conditional independence between responses and/or response times on test items. *Psychometrika*, *75*(1), 120–139. doi: 10.1007/s11336-009-9129-9

van Rijn, P., & Ali, U. (2017a). A comparison of item response models for accuracy and speed of item responses with applications to adaptive testing. *British Journal of Mathematical and Statistical Psychology*, *70*(2), 317–345. doi: 10.1111/bmsp.12101

van Rijn, P., & Ali, U. (2017b). A generalized speed–accuracy response model for dichotomous items. *Psychometrika*, *83*(1), 109–131. doi: 10.1007/s11336-017-9590-9

Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, *71*(2), 262–286. doi: 10.1111/bmsp.12114

Zhang, J., & Stout, W. (1997). On holland's dutch identity conjecture. *Psychometrika*, *62*(3), 375–392. doi: 10.1007/BF02294557

Appendix

A The relation between Theorem 1 and the Dutch Identity

In this appendix we aim to clarify two things. First, we aim to clarify how our Theorem 1 relates to the Dutch Identity. Secondly, we aim to clarify how the choice of a multivariate normal posterior in the Dutch Identity relates to choosing a multivariate normal kernel $k(\boldsymbol{\zeta})$ in Theorem 1. Because a Dutch Identity has not been formulated for models of the exponential family form of Eq. (7), we start with formulating a Dutch Identity for models of this form.

Theorem 2. *Suppose that $p(\mathbf{x}, \mathbf{t})$ is of the form*

$$p(\mathbf{x}, \mathbf{t}) = \int_{\Omega_\zeta} p(\mathbf{x}, \mathbf{t} \mid \zeta)\, f(\zeta)\, d\boldsymbol{\zeta},$$

where Ω_ζ is the support of $\boldsymbol{\zeta}$, and $p(\mathbf{x}, \mathbf{t} \mid \zeta)$ is of the form of Eq. (7). Then for any vector $\mathbf{y} \in \Omega_\mathbf{x}$ and $\mathbf{w} \in \Omega_\mathbf{t}$, where $\Omega_\mathbf{x}$ and $\Omega_\mathbf{t}$ denote the support of $\mathbf{x}$ and $\mathbf{t}$, respectively, we have

$$\frac{p(\mathbf{x}, \mathbf{t})}{p(\mathbf{y}, \mathbf{w})} = \prod_{i=1}^{p} \frac{b_i(x_i, t_i)}{b_i(y_i, w_i)} \mathbb{E}_f \left(\exp\left(\sum_{i=1}^{p} [\mathbf{s}_i(x_i, t_i) - \mathbf{s}_i(y_i, w_i)]^\mathsf{T} \boldsymbol{\zeta} \right) \middle| \mathbf{S} = \sum_i \mathbf{s}_i(y_i, w_i) \right).$$

Proof.

$$\begin{aligned}
\frac{p(\mathbf{x},\,\mathbf{t})}{p(\mathbf{y},\,\mathbf{w})} &= \frac{\int_{\mathbb{R}} \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, \mathrm{b}_i(x_i,\, t_i)\, e^{[\mathbf{s}_i(x_i,\, t_i)^{\mathsf{T}}\boldsymbol{\zeta}]} f(\boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\zeta}}{\int_{\mathbb{R}} \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, \mathrm{b}_i(y_i,\, w_i)\, e^{[\mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}}\boldsymbol{\zeta}]} f(\boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\zeta}} \\
&= \prod_{i=1}^{p} \frac{\mathrm{b}_i(x_i,\, t_i)}{\mathrm{b}_i(y_i,\, w_i)} \int_{\mathbb{R}} \frac{\prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, e^{[\mathbf{s}_i(x_i,\, t_i)^{\mathsf{T}}\boldsymbol{\zeta}]} f(\boldsymbol{\zeta})}{\int_{\mathbb{R}} \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, e^{[\mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}}\boldsymbol{\zeta}]} f(\boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\zeta}}\, \mathrm{d}\boldsymbol{\zeta} \\
&= \prod_{i=1}^{p} \frac{\mathrm{b}_i(x_i,\, t_i)}{\mathrm{b}_i(y_i,\, w_i)} \int_{\mathbb{R}} \prod_{i=1}^{p} e^{\left[[\mathbf{s}_i(x_i,\, t_i)-\mathbf{s}_i(y_i,\, w_i)]^{\mathsf{T}}\boldsymbol{\zeta}\right]} \frac{\prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, e^{[\mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}}\boldsymbol{\zeta}]} f(\boldsymbol{\zeta})}{\int_{\mathbb{R}} \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, e^{[\mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}}\boldsymbol{\zeta}]} f(\boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\zeta}}\, \mathrm{d}\boldsymbol{\zeta} \\
&= \prod_{i=1}^{p} \frac{\mathrm{b}_i(x_i,\, t_i)}{\mathrm{b}_i(y_i,\, w_i)} \int_{\mathbb{R}} \prod_{i=1}^{p} e^{\left[[\mathbf{s}_i(x_i,\, t_i)-\mathbf{s}_i(y_i,\, w_i)]^{\mathsf{T}}\boldsymbol{\zeta}\right]} f\left(\boldsymbol{\zeta} \mid \mathbf{S} = \sum_i \mathbf{s}_i(y_i,\, w_i)\right) \mathrm{d}\boldsymbol{\zeta} \\
&= \prod_{i=1}^{p} \frac{\mathrm{b}_i(x_i,\, t_i)}{\mathrm{b}_i(y_i,\, w_i)} \mathbb{E}_f \left(\exp\left(\sum_{i=1}^{p} [\mathbf{s}_i(x_i,\, t_i) - \mathbf{s}_i(y_i,\, w_i)]^{\mathsf{T}}\, \boldsymbol{\zeta} \right) \middle| \mathbf{S} = \sum_i \mathbf{s}_i(y_i,\, w_i) \right).
\end{aligned}$$

□

It is easy to show that whenever $f(\boldsymbol{\zeta})$ is of the form of $g(\boldsymbol{\zeta})$ in Eq. (8), the expression for the ratio in Theorem 2 reduces to

$$\frac{p(\mathbf{x},\,\mathbf{t})}{p(\mathbf{y},\,\mathbf{w})} = \prod_{i=1}^{p} \frac{\mathrm{b}_i(x_i,\, t_i)}{\mathrm{b}_i(y_i,\, w_i)} \times \frac{\mathbb{E}_k\left(\exp\left(\sum_{i=1}^{p} \mathbf{s}_i(x_i,\, t_i)^{\mathsf{T}}\boldsymbol{\zeta}\right)\right)}{\mathbb{E}_k\left(\exp\left(\sum_{i=1}^{p} \mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}}\boldsymbol{\zeta}\right)\right)},$$

which was to be expected from Theorem 1.

One important result that follows from the Dutch Identity is that whenever the posterior $f(\theta \mid \mathbf{x},\,\mathbf{t}) = f(\theta \mid \mathbf{s})$ is chosen to be (multivariate) normal, we may express the marginals $p(\mathbf{x},\,\mathbf{t})$ in a form similar to that expressed in Corollary 1. This is shown, for example, in Corollary 1 of Holland (1990), Corollary 1 of Ip (2002) and Theorem 2 of Hessen (2012). For Theorem 1, however, it suffices to choose the kernel density $k(\boldsymbol{\zeta})$ to be (multivariate) normal instead. To see why this is the case, consider the posterior with a prior distribution $f(\theta) = g(\theta)$ of the form of Eq. (8),

$$\begin{aligned}
g(\theta \mid \mathbf{s}) &\propto \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, e^{[\mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}}\boldsymbol{\zeta}]} g(\boldsymbol{\zeta}) \\
&= \prod_{i=1}^{p} \frac{1}{\mathrm{Z}_i(\boldsymbol{\zeta})}\, e^{[\mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}}\boldsymbol{\zeta}]} \prod_{i=1}^{p} \mathrm{Z}_i(\boldsymbol{\zeta})\, k(\boldsymbol{\zeta}) \\
&= \exp\left(\left[\sum_{i=1}^{p} \mathbf{s}_i(y_i,\, w_i)^{\mathsf{T}} \right] \boldsymbol{\zeta} \right) k(\boldsymbol{\zeta}).
\end{aligned}$$

This is proportional to a multivariate normal distribution if and only if $k(\boldsymbol{\zeta})$ is a multivariate normal distribution. This shows why the assumption of posterior normality in combination with the Dutch Identity and the assumption of a normal kernel density in combination with our Theorem 1 lead to the same result.