Bayesian Joint Modeling of Response Times with Dynamic Latent Ability

Xiaojing Wang

Department of Statistics

University of Connecticut

Joint work with Abhisek Saha and Dipak K. Dey

Presentation at BMCA in honor of Professor Kuo Lynn
May 12, 2018

Outline

Item Response Theory (IRT) Models

IRT models are frequently used in modeling dichotomous data from measurement testing, since they allow:

- assessing the 'abilities' of examinees.
- studying the effectiveness of different test items.

Typical One-Parameter IRT Models

$$\Pr(X_{ij} = 1 \mid \theta_i, d_j) = \mathrm{F}(\theta_i - d_j),$$

where θ_i: the i-th person's 'ability', d_j: the j-th test item's difficulty, X_{ij}: the correctness of the j-th test item taken by the i-th person.

- Rasch Model:

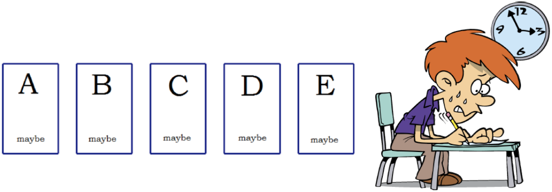
$$\Pr(X_{ij} = 1 \mid \theta_i, d_j) = \frac{\exp(\theta_i - d_j)}{1 + \exp(\theta_i - d_j)}.$$

- Normal Ogive or Probit Model:

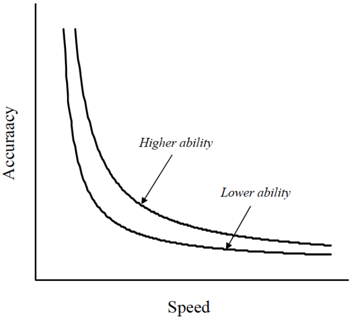
$$\Pr(X_{ij} = 1 \mid \theta_i, d_j) = \Phi(\theta_i - d_j).$$

Response Times on a Test Item

In usual cases, inferences about latent traits of test takers have been mainly based on their responses to the items while the time taken to complete an item has been often ignored. In computerized testing, the response time of items can be collected at no additional cost.



Speed-accuracy tradeoff



- The shape of the curve is entirely arbitrary;
- The only thing implied by the tradeoff is a monotonically decreasing relation between speed and ability.

Item Difficulty and Time Intensity

Item 1	Item 2
375	375
229	229
——— +	58
	79
	——— +

- Item 2 involves a longer series of cognitive operations and requires more time than item 1.
- The test taker's ability is challenged by the nature of the operations involved in it.

Thissen's model and Its Variations

The response time R_{ij}, that is, the time that the individual i spends on the jth item, can usually be modeled as:

$$\log R_{ij} = \mu + \nu_i + \tau_j + \beta L(|\theta_i - d_j|) + \zeta_{ij},$$

where

- μ is the overall mean parameter;
- ν_i and τ_j are interpreted as the "slowness parameter" for the ith person and the jth item;
- $L(|\theta_i - d_j|)$ is a function of the distance between the ability and item difficulty;
- β is a slope parameter in the regression;

In Thissen's model, $L(|\theta_i - d_l|) = -(\theta_i - d_l)$, which indicates the distance between the ability and the item difficulty has a monotone relationship with the response time.

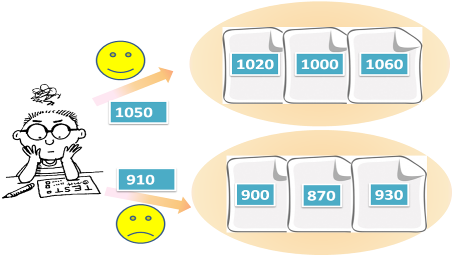
Dependences Between Response Times and Responses

- Clearly, to jointly model response times and responses, it will maximize the information to infer the person's ability θ_i and the item difficulty d_j;
- There are several literatures exist in the discussion of jointly modeling item responses and response times, such as Ferrando and Lorenzo-Seva (2007), van der Linden et al. (2010), Ranger and Kuhn (2012) and Ranger (2013).
- However, all existing models are based on an one-time exam for each test taker.

MetaMetrics Data

The data considered has been collected from a school district in Mississippi. The data has thousands of students registered over five years in a Computer Adaptive Instruction and Testing program conducted by MetaMetrics Inc.

Computer Adaptive Instruction and Testing (CAIT)



Typical One-Parameter IRT Models

$$\Pr(X_{ij} = 1 \mid \theta_i, d_j) = \mathrm{F}(\theta_i - d_j).$$

Three Major Features of MetaMetrics Data Structure

1. Longitudinal observations at variable and irregular time points;
2. Potential dependence among test items;
3. Uncertainty associated with each item difficulty within the test, though the ensemble mean of the test items is given.

Basic Notation

- $i = 1, \cdots, n$: indicates the person,
- $t = 1, \cdots, T_i$: means distinctive test day for the i-th person,
- $s = 1, \cdots, S_{i,t}$: signifies the replicated tests within a distinctive t-th day for i-th person,
- $j = 1, \cdots, K_{i,t,s}$: expresses the item numbers for each test,
- $X_{i,t,s,j}$: corresponds to the correctness of the answer of the j-th item in the s-th test on the t-th day taken by the i-th person.
- $d_{i,t,s,j}$: represents the difficult level of the j-th item in the s-th test at the t-th day taken by the i-th person.
- $R_{i,t,s}$: denotes the response time to sth test in day t that the ith individual spent on.

The Observation Equations of Item Responses

$$\Pr(X_{i,t,s,j} = 1 \mid \theta_{i,t}, a_{i,t,s}, \varphi_{i,t}, \eta_{i,t,s}, \epsilon_{i,t,s,j}) = \mathrm{F}(\theta_{i,t} - d_{i,t,s,j} + \varphi_{i,t} + \eta_{i,t,s}),$$

- $\theta_{i,t}$: the i-th person's ability on the t-th day.
- $d_{i,t,s,j}$: $d_{i,t,s,j} = a_{i,t,s} + \epsilon_{i,t,s,j}$ with ensemble mean $a_{i,t,s}$ and $\epsilon_{i,t,s,j} \sim \mathcal{N}(0, \sigma^2)$ with known σ.
- $\varphi_{i,t} \sim \mathcal{N}(0, \delta_i^{-1})$: daily random effect for the i-th person taking test at the t-th day with unknown δ_i.
- $\eta_{i,t,s}$: test random effect for the i-th person taking test at the t-th day for the s-th test and $\eta_{i,t} \sim \mathcal{N}_{S_{i,t}}(0, \tau_i^{-1}\boldsymbol{I} \mid \sum_{s=1}^{S_{i,t}} \eta_{i,t,s} = 0)$ with $\eta_{i,t} = (\eta_{i,t,1}, \ldots, \eta_{i,t,S_{i,t}})'$ and unknown individual τ_i.
- Let $\mathrm{F}(\cdot)$ be the logistic cdf.

The Observation Equations of Response Times

$$\log(R_{i,t,s}) = \mu_i - \nu_{i,t} + \beta L(|\theta_{i,t} - a_{i,t,s}|) + \zeta_{i,t,s},$$

- μ_i reflects the general response time for i-th respondent;
- $\nu_{i,t}$ implies the speed of the respondent i at the t-th day with $\nu_{i,t} \sim \mathcal{N}(0, \kappa_i^{-1})$;
- $|\theta_{i,t} - a_{i,t,s}|$ indicates the distance between the ability on the test difficulty and $L(\cdot)$ is a function to characterize the relationship between this distance and the response time;
- β is a regression coefficient to adjust the influence of the distance function to the response time;
- $\zeta_{i,t,s} \sim \mathcal{N}(0, \varrho^{-1})$ is a residual term.

The System Equation

$$\theta_{i,t} = \theta_{i,t-1} + c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^{+} + w_{i,t}$$

It models the dynamic change of one's ability through three terms:

1. by depending on one's ability at the previous time point, $\theta_{i,t-1}$.
2. via a parametric growth model, to be discussed.
3. by involving a random component $w_{i,t}$ to represent the change in the i-th person's ability on the t-th day, where $w_{i,t}$ is assumed to be $\mathcal{N}(0, \phi^{-1}\Delta_{i,t})$ with $\Delta_{i,t}$ presenting the time lapse between the person's t-th testing day and $(t-1)$-th testing day and ϕ is unknown.

The System Equation in DIR Models (Continued)

$$\theta_{i,t} = \theta_{i,t-1} + c_i(1 - \rho\theta_{i,t-1})\Delta^{+}_{i,t} + w_{i,t}$$

The parametric growth model arises from the following considerations:

1. c_i is the average growth rate of the i-th person's ability over time.
2. $\Delta^{+}_{i,t} = \min\{\Delta_{i,t}, \Delta_{T_{\max}}\}$, truncated at pre-specified $\Delta_{T_{\max}}$ ($\Delta_{T_{\max}} = 14$) to reflect likely vacation time where learning may not be happening.
3. $-\rho\theta_{i,t-1}$ is a "correction factor", slowing down the effect of the linear growth as the ability level becomes larger.

Summary of the Joint RT-DIR Model

To sum up, the Joint RT-DIR Models are constructed via two stages as follows:

$$
\begin{aligned}
\text{System equation:} \quad & \theta_{i,t} = \theta_{i,t-1} + c_i(1-\rho\theta_{i,t-1})\Delta^{+}_{i,t} + w_{i,t}\,, \\
\text{Observation equations:} \quad & \Pr(X_{i,t,s,j}=1 \mid \theta_{i,t}, a_{i,t,s}, \varphi_{i,t}, \eta_{i,t,s}, \epsilon_{i,t,s,j}) \\
= \quad & \frac{\exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \epsilon_{i,t,s,j})}{1+\exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \epsilon_{i,t,s,j})}, \\
\log(R_{i,t,s}) \quad = \quad & \mu_i - \nu_{i,t} + \beta L(|\theta_{i,t} - a_{i,t,s}|) + \zeta_{i,t,s}.
\end{aligned}
$$

Here $\epsilon_{i,t,s,j} \sim \mathcal{N}(0,\sigma^2)$, $\varphi_{i,t} \sim \mathcal{N}(0,\delta_i^{-1})$, $w_{i,t} \sim \mathcal{N}(0,\phi^{-1}\Delta_{i,t})$, $\eta_{i,t} \sim \mathcal{N}_{S_{i,t}}(0,\tau_i^{-1}\mathbf{I} \mid \sum_{s=1}^{S_{i,t}} \eta_{i,t,s}=0)$, $\zeta_{i,t,s} \sim \mathcal{N}(0,\varrho^{-1})$ and $\nu_{i,t} \sim \mathcal{N}(0,\kappa_i^{-1})$. Here, $L(|\theta_{i,t} - a_{i,t,s}|)$ can either equal to monotone relationship, i.e., $(\theta_{i,t} - a_{i,t,s})$ or inverted U shape, i.e., $|\theta_{i,t} - a_{i,t,s}|$.

A Simulation Example

$$
\begin{aligned}
\text{System equation:} \quad & \theta_{i,t} = \theta_{i,t-1} + c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^{+} + w_{i,t}\,, \\
\text{Observation equation:} \quad & \Pr(X_{i,t,s,j} = 1 \mid \theta_{i,t}, a_{i,t,s}, \varphi_{i,t}, \eta_{i,t,s}, \epsilon_{i,t,s,j}) \\
= \quad & \frac{\exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \epsilon_{i,t,s,j})}{1 + \exp(\theta_{i,t} - a_{i,t,s} + \eta_{i,t,s} + \epsilon_{i,t,s,j})}, \\
\log(R_{i,t,s}) \quad = \quad & \mu_i - \nu_{i,t} + \beta \mid \theta_{i,t} - a_{i,t,s} \mid + \zeta_{i,t,s}.
\end{aligned}
$$

Let

$$
\begin{aligned}
& T_i = 50, S_{i,t} = 4, \text{for } i = 1, \cdots, 10, t = 1, \cdots, 50, \\
& K_{i,t,s} = 10, \text{for } i = 1, \cdots, 10, t = 1, \cdots, 50, s = 1, \cdots, 4, \\
& \Delta_{i,t} = 10 + t, t = 1, \cdots, T_i/2, \Delta_{i,t} = t - 10, t = T_i/2, \cdots, T_i, \\
& \phi = 1/0.0281^2, \rho = 0.12, \sigma = 0.73, \beta = -0.17, \varrho = 1.25.
\end{aligned}
$$

Values of Unknowns Used in the Simulation

Table: Values of unknowns used in the simulation

i	c	δ	τ	κ	μ
1	0.0055	2.0408	4	2.3256	1.6
2	0.0065	1.3333	3.1250	1.5873	1.47
3	0.0026	1.8182	4.3478	1.6949	1
4	0.0037	1.2346	2.7027	0.5495	1.92
5	0.0061	1.5873	3.7037	1.2658	1.45
6	0.0047	1	2.8571	0.9346	1.73
7	0.0035	2.2222	4	1.3889	1.5
8	0.0043	1.0526	2.2222	1.8182	1.35
9	0.0039	1.1494	9.0909	2.7027	0.81
10	0.0015	2	4.5455	1.2195	1.23

Simulation Results

A comparison of ability estimates between DIR-RT and DIR models is shown below:

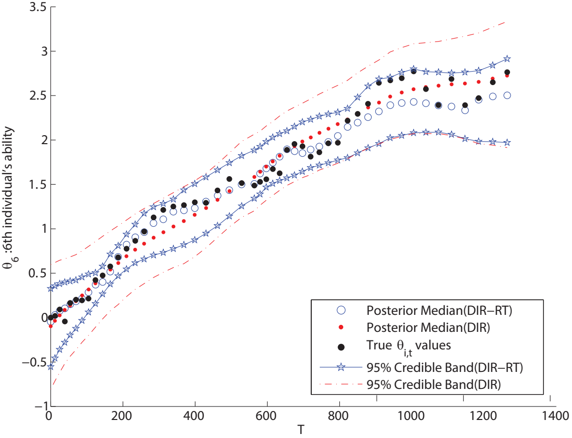


Figure: 6th individual

MetaMetric Testbed Application

For illustration purpose, we randomly select a sample of 25 individuals from MetaMetric testbed, where different characteristics for each student are shown in the table below.

Table: Characteristics of the first 3 individuals randomly sampled from the EdSphere data

	Total Tests	Days	Max. Tests/Days	Range of Items/Test	Max. Gap	Initial Grade
No.1	150	74	9	4-22	79	4
No.2	203	128	15	6-24	107	2
No.3	211	107	9	5-24	79	3

The Ability Growth of θ_{10} for Two Different Linkages

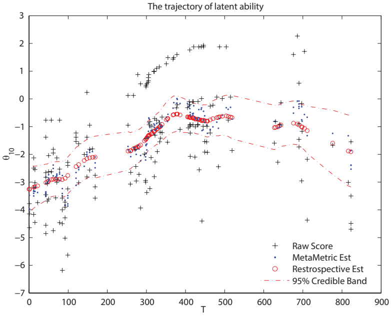


Figure: Monotone Linkage

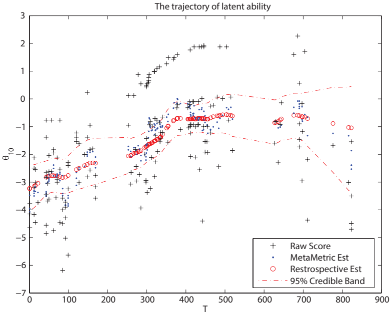


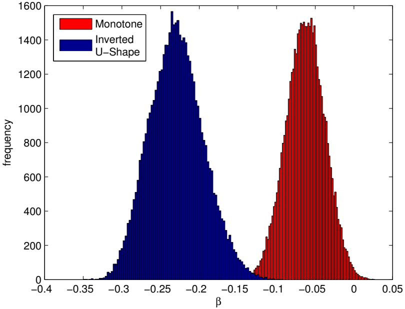
Figure: Inverted U-shaped Linkage

Comparison of Two Linkages

The regression slope β plays a key role in controlling the influence of the ability-difficulty distance function to the response time.

Using Lindley's Method

We are interested in testing $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$.



Model(β)	Inverted U-shape
PM	-0.2305
95% CI	$(-0.2940, -0.1571)$
99% CI	$(-0.3105, -0.1345)$

Model(β)	Monotone
PM	-0.0627
95% CI	$(-0.1125, -0.0137)$
99% CI	$(-0.1317, 0.0003)$

Using Partial DIC

Inverted U-shape: 5661.4 versus Monotone: 5775.3.

Retrospective Estimation of Ability Growth

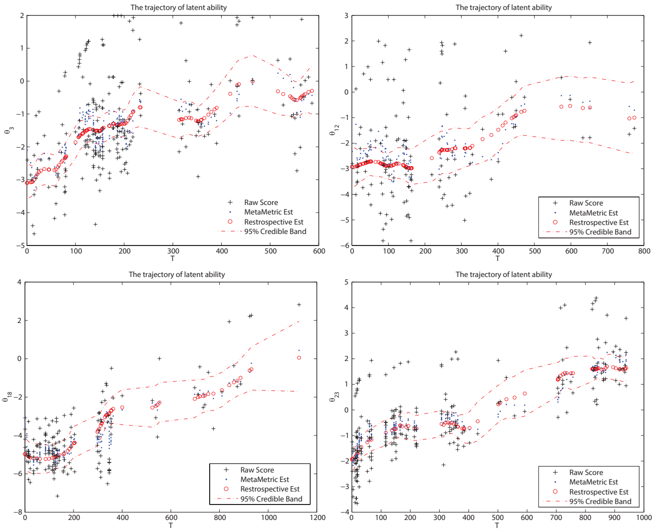
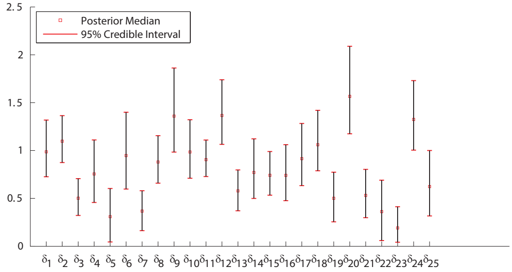
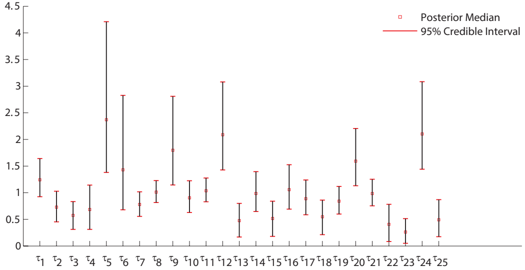


Figure: The posterior summary of the ability growth for θ_3, θ_{12}, θ_{18} and θ_{23}.

Results of $\delta_i^{-1/2}$'s and $\tau_i^{-1/2}$





Summary of Joint RT-DIR Models

- Our proposed DIR-RT models can jointly model the observations of response times and item responses with sharing ability parameters and can accommodate the complex longitudinal data observed at individually-varying and irregularly-spaced time points.
- From our simulation study, we have noticed that incorporation of response time into the item response model in the analysis of longitudinal data has both significantly improved the precision and reduced the bias for the ability estimation.
- Our analysis is the first of its kind to conduct empirical studies on the choice of linkage function to describe the relationship between the ability-difficulty and response time in a joint modeling of item responses and response times for longitudinal data in educational testing.

Future Work

- Many extensions of current DIR-RT models are possible, such asextensions to two-parameter and three-parameter DIR-RT models.
- we consider using either model-based or distance-based clustering methods to analyze the psychological behaviors of students reflected in the patterns shown in the growth trajectories.
- we will improve the efficiency of our computation by developing big data schemes to make the parallel computing possible.

Acknowledgement

MetaMetrics®

High Road Learning

Learning Personalized. Anytime. Anywhere.

Thank you!

xiaojing.wang@uconn.edu