The latent factor structure of child development

Anonymous Cogsci Submission

Abstract

TO DO

Keywords: child development; milestones; IRT; cross-validation;

Introduction

to do

Data

A child's development can be thought of as the set of developmental milestones that they have reached at a particular point in time. This conceptualization results in data with the same structure as the item response data common to educational measurement. In education, item response data is most typically students responding to test items (i.e., questions) and, in the dichotomous case, getting each question either correct or incorrect. In the context of child development, the child is the "student," and each developmental milestone is the "item."

We use Kinedu, a Mexico-based child development app, as a source for this type of data. When parents first start using the Kinedu app, they are asked a series of questions about which developmental milestones their child has reached. We consider the 1946 children between 2 and 55 months of age whose parents responded to all 414 of the developmental milestones. Each developmental miletone on Kinedu is mapped to a milestone group: physical, cognitive, linguistic, or social & emotional. Table 1 shows the number of developmental milestones in each group along with an example milestone translated to English.

Table 1: Developmental milestone groups and examples

Group	Count	Milestone
Physical	180	Stands on their toes
Cognitive	100	Finds objects on the floor
Linguistic	75	Babbles to imitate conversations
Social & Emotional	59	Complains when play is interrupted

Figure 1 shows the age (in months) and number of developmental milestones for each child. At 12 months old, most children have reached about 200 developmental milestones. At 24 months old, most children have reached about 300 developmental milestones. Finally, at 48 months old, most children have reached about 375 of the 414 developmental milestones.

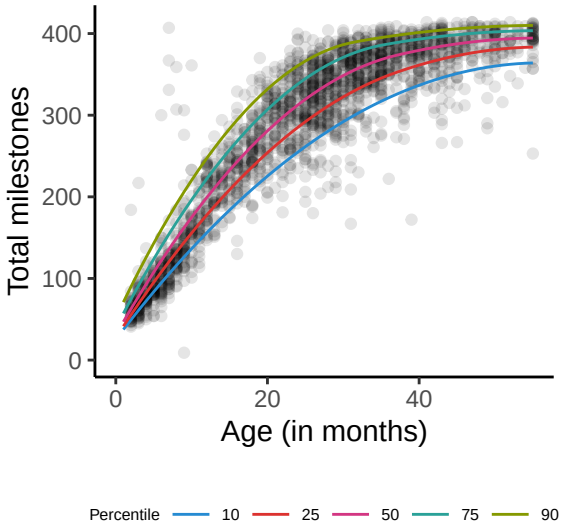

Figure 1: Number of milestones by age

probably should describe where the percentile curves come from

Empirical assessment of the dimensionality of child development

We frame the assessment of the dimensionality of child development as a model comparison question.

Models

Item response theory offers a suite of models with which to model item response data. We adopt the notation used in Chalmers & others (2012). Let $i = 1, \ldots, I$ represent the distinct children and $j = 1, \ldots, J$ the developmental milestones. The Kinedu item response data is stored in a matrix, y, where element y_{ij} denotes if the ith child has or has not achieved the jth developmental milestone as reported by their parent/guardian. Each model represents the ith child's development using m latent factors $\boldsymbol{\theta}_i = (\theta_1, \ldots, \theta_m)$. The jth milestone's discriminations (i.e. slopes) $\boldsymbol{a_j} = (a_1, \ldots, a_m)$ capture the latent factor loadings onto that milestone.

We fit four two-parametric logistic (2PL) models where a child's development is represented by $m = 1$, $m = 2$, $m = 3$, and $m = 4$ latent factors. Hereafter, we, for example, refer to a 2PL model with $m = 4$ latent factors as a 4F 2PL model.

According to the 2PL model, the probability of a child having achieved a developmental milestone is

$$P(y_{ij} = 1|\boldsymbol{\theta_i}, \boldsymbol{a_j}, b_j) = \sigma(\boldsymbol{a_j}^\top \boldsymbol{\theta_i} + b_j)$$

where b_j is the milestone easiness (i.e. intercept) and $\sigma(x) = \frac{e^x}{e^x+1}$ is the standard logistic function. We also fit a 1F Rasch model where each of the discriminations is fixed to 1.

The 2PL models learn the latent factor structure entirely from the data, making them exploratory. The bifactor model offers an alternative specification where each milestone loads onto a general factor θ_0 and a specific factor θ_s (Cai, Yang, & Hansen, 2011). The assignment of each developmental milestone to its specific factor is an opportunity to specify the latent factor structure, making the model confirmatory as opposed to exploratory. We map each milestone to its specific factor according to the four developmental milestone groups shown in Table 1. For the bifactor model, the probability of a child having achieved a developmental milestone is

$$P(y_{ij} = 1|\theta_0, \theta_s, a_0, a_s) = \sigma(a_0\theta_0 + a_s\theta_s + b_j).$$

Model comparison

Model comparison in IRT typically uses information criterion such as AIC and BIC (Maydeu-Olivares, 2013). However, these methods are not guaranteed to work with modest sample sizes or misspecification (McDonald & Mok, 1995). Instead, we prefer a marginalized version of cross-validation. In essence, we partition the data into folds based on the children (i.e. the rows of the item response matrix). Then for each fold, we estimate the item parameters using all but that fold, and calculate the likelihood of that fold by integrating over $g(\theta)$.

Mathematically and following notation similar to Vehtari, Gelman, & Gabry (2017), we partition the data into K subsets $y^{(k)}$ for $k = 1, \ldots, K$. Each model is fit separately to each training set $y^{(-k)}$ yielding item parameter estimates which we compactly denote $\Psi_j^{(-k)}$. The predictive (i.e. out-of-sample or cross-validated) likelihood of $y^{(k)}$ is

$$p(y^{(k)}|y^{(-k)}) = \prod_{i \in i^{(k)}}^{I} \int_\theta \prod_{j=1}^{J} \hat{\Pr}(y_{ij}^{(k)}|\Psi_j^{(-k)}, \theta) g(\theta) d\theta.$$

The ultimate quantity of interest for each model is the log predictive likelihood for the entire item response matrix, which is defined as

$$\text{lpl } y = \sum_{k=1}^{K} \log p(y^{(k)}|y^{(-k)}).$$

Results

fascinating that dropping 1 month olds changed the winner from 3F to 4F

Table 2 shows the number of parameters, the in-sample log likelihood (which neccessarily increases with more parameters), and the lpl y defined in the model comparison section. The 4F 2PL model performs best which is evidence that child development between the ages of 2 and 55 months follows a multidimensional path.

what else to add here?

Computing is done in R (R Core Team, 2019), model fitting in the R package mirt (Chalmers & others, 2012), and data wrangling/visualization in the set of R packages known as the tidyverse (Wickham, 2017).

Table 2 can be made clearer

Table 2: Model performance: The 4F 2PL performs best

model	npars	in_log_lik	lpl y (out-of-sample)
1F Rasch	415	-254983.9	-255443.5
1F 2PL	828	-222073.0	-223072.0
2F 2PL	1241	-212957.9	-214498.2
Bifactor	1242	-210030.2	-211682.2
3F 2PL	1653	-208887.6	-210943.4
4F 2PL	2064	-208124.1	-210793.5

Understanding the latent factor structure

section is written as if 3F is winner may need to edit to 4F

To understand each of the three factors in the best performing model, we fit the model to the the full dataset. We then estimate the factor loadings (i.e. discriminations or slopes) using a varimax rotation. The varimax rotation results in orthogonal and, therefore, more interpretable factors (Kaiser, 1959). Figure 2 shows the distribution of factor loadings for each group on each of the three factors. The first factor load mainly on cognitive and linguistic milestones. The second factor is a combination of each of the groups with the strongest loadings on the physical and social & emotional milestones. The third mainly load positively on linguistic milestones and, interestingly, negatively on physical milestones.

We also estimate the factor scores for each child using expected a posteriori (EAP) with a three dimensional standard normal distribution (Embretson & Reise, 2013). Figure 3 shows the relationship between age and factor score for each factor. The first factor, perhaps unsurprisingly, has a high correlation (r = 0.90) with age. The second factor has a strong association with age from 2 to 16 months but thereafter is unrelated to age. By and large, the third factor does not have any association with age.

Dimensionality across the age-span

For all of the data, we've shown evidence that 4-factors performs best. But is this latent factor dimensionality consistent across age-span? For example, perhaps for very young children 1-factor is sufficient and then later on 2- and then 3-factors become valuable. We take two approaches to assessing the dimensionality of child development across the age-span. First, we examine the performance of each of the models by age. Second, we partition the data by age and use

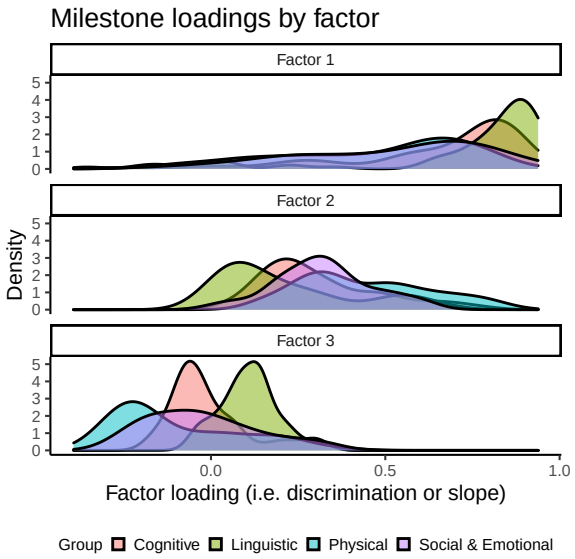

Figure 2: Factor loadings by group

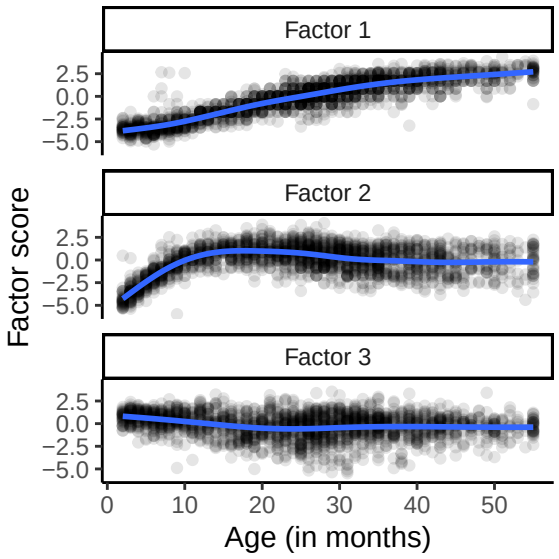

Figure 3: The first factor is highly associated with age

the same cross-validation procedure to find the best fitting model in each partition.

Full model

figure currently shows up at end of paper

Figure 4 shows . . .

Age-partitioned models

Discussion

to do

Acknowledgements

We'd like to thank Kinedu for providing the data that made this research possible.

References

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, *16*(3), 221.

Chalmers, R. P., & others. (2012). Mirt: A multidimensional item response theory package for the r environment. *Journal of Statistical Software*, *48*(6), 1–29.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Kaiser, H. F. (1959). Computer program for varimax rotation in factor analysis. *Educational and Psychological Measurement*, *19*(3), 413–420.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives*, *11*(3), 71–101.

McDonald, R. P., & Mok, M. M.-C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research*, *30*(1), 23–40.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `https://www.R-project.org/`

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, *27*(5), 1413–1432.

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from `https://CRAN.R-project.org/package=tidyverse`

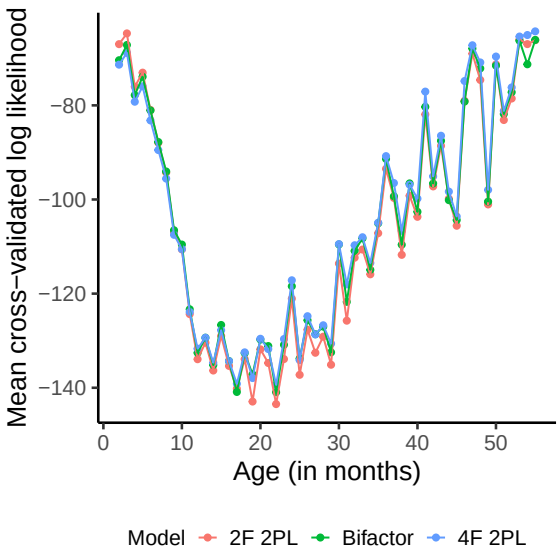

Figure 4: More flexible models perform better at older ages