A probabilistic graphical model for joint modeling of item response and process

Xiang Liu

Educational Testing Service

November 19, 2020

Background

- Various process data becomes increasingly available, thanks to the improved digital assessment platforms.
- NAEP digital assessments.
- In addition to the traditional item responses, how do we make sense of the process data in large scale?

Example: calculator use

- On-screen calculator for math items.
- The behaviors captured in the process data may relate to both item response and proficiency.
- The additional information from process data could prove to be useful in improving measurement precision as well as leading to better understanding the cognitive process of students.
- Develop a modeling approach that could simultaneously consider item response and process data.
- 2017 NAEP math assessment release block - grade 8.

Data summary

- The block consists of 19 items.
- Over $30,000$ students.
- A random sample of $3,000$ students.
- Some items have few calculator use, i.e. less than 30.
- I used 15 items.

Figure: calculator use by item

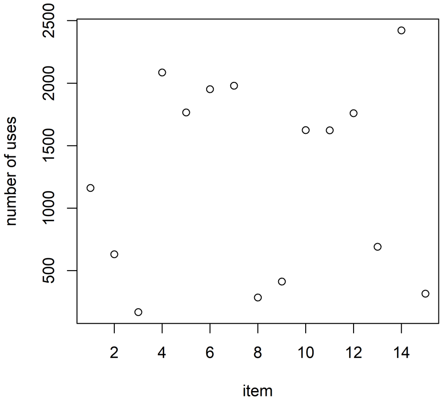


Figure: calculator use by student

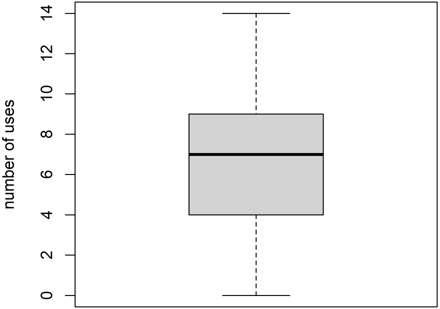


Figure: number of correct answers by item

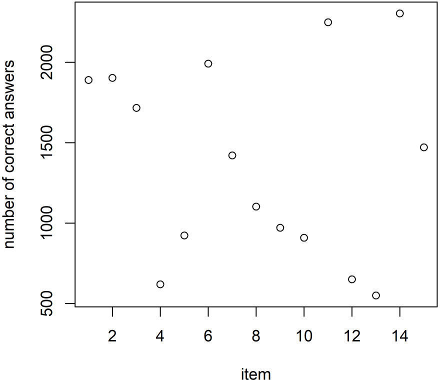


Figure: number of correct answers by student

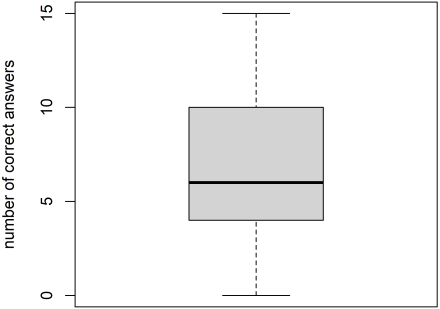
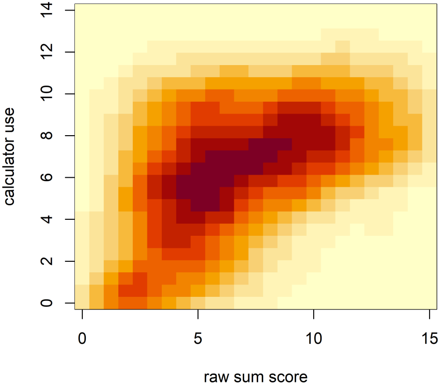


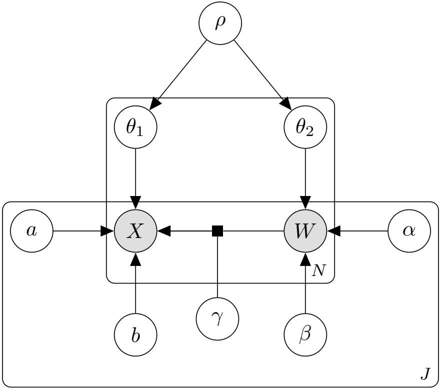
Figure: 2D density of raw score and calculator use



What do we model?

- Varying levels of item difficulty
- Varying levels of calculator attractiveness
- Varying levels of proficiency
- Varying levels of propensity of using the calculator
- Correlation between proficiency and calculator propensity
- Varying levels of helpfulness of using a calculator (conditional proficiency)

A probabilistic graphical model



Conditional independence assumptions

The probability of a calculator use is given by

$$P(W_{ij} = 1 | \theta_{i2}, \alpha_j, \beta_j) = \frac{\exp[\alpha_j(\theta_{i2} - \beta_j)]}{1 + \exp[\alpha_j(\theta_{i2} - \beta_j)]}.$$

The probability of a correct response is given by

$$P(X_{ij} = 1 | \theta_{i1}, a_j, b_j, w_{ij}, \gamma_j) = \frac{\exp[a_j(\theta_{i1} - b_j + w_{ij}\gamma_j)]}{1 + \exp[a_j(\theta_{i1} - b_j + w_{ij}\gamma_j)]}.$$

$$\boldsymbol{\theta}_i \sim \Phi\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

Marginal probability of joint item response and calculator use

Marginalizing $\boldsymbol{\theta}_i$ over the multivariate normal distribution function Φ, the joint probability of observing $\boldsymbol{X}_i = (x_{i1}, x_{i2}, \ldots, x_{iJ})$ and $\boldsymbol{W}_i = (w_{i1}, w_{i2}, \ldots, w_{iJ})$ is

$$
\begin{aligned}
& P(\boldsymbol{X}_i = \boldsymbol{x}_i, \boldsymbol{W}_i = \boldsymbol{w}_i | \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \rho) \\
& = \iint_{\boldsymbol{\theta}_i \in \mathbb{R}^2} \prod_{j=1}^{J} P(X_{ij} = x_{ij} | \theta_{i1}, a_j, b_j, w_{ij}, \gamma_j) \\
& P(W_{ij} = w_{ij} | \theta_{i2}, \alpha_j, \beta_j) d\Phi(\boldsymbol{\theta}_i; \rho)
\end{aligned} \tag{1}
$$

Parameter estimation

- Observing item response $\boldsymbol{x}$ and calculator use $\boldsymbol{w}$, we need to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{b}$, γ, and ρ.
- Direct optimization of the marginal log-likelihood in $5 * J + 1$-dimensional space is difficult and requires working with

$$\sum_{i=1}^{N} \log \iint_{\boldsymbol{\theta}_i \in \mathbb{R}^2} \phi(\boldsymbol{\theta}_i; \rho) \prod_{j=1}^{J} P(X_{ij} = x_{ij} | \theta_{i1}, a_j, b_j, w_{ij}, \gamma_j)$$
$$P(W_{ij} = w_{ij} | \theta_{i2}, \alpha_j, \beta_j) d\boldsymbol{\theta}_i$$

- An Expectation-Maximization (EM) algorithm.

Complete data log-likelihood

The complete data log-likelihood of the model parameters is

$$\log L = \sum_{j=1}^{J} l_j(\alpha_j, \beta_j) + \sum_{j=1}^{J} l_j(a_j, b_j, \gamma_j) + \sum_{i=1}^{N} \log \phi(\boldsymbol{\theta}_i; \rho), \quad (2)$$

where

$$l_j(\alpha_j, \beta_j) = \sum_{i=1}^{N} w_{ij} \log P_{ij} + (1 - w_{ij}) \log Q_{ij}, \quad (3)$$

and

$$l_j(a_j, b_j, \gamma_j) = \sum_{i=1}^{N} x_{ij} \log \dot{P}_{ij} + (1 - x_{ij}) \log \dot{Q}_{ij}. \quad (4)$$

Gradients

$$\nabla l_j(\alpha_j, \beta_j) = \begin{pmatrix} \frac{\partial l_j}{\partial \alpha_j} \\ \\ \frac{\partial l_j}{\partial \beta_j} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N}(w_{ij} - P_{ij})(\theta_{i2} - \beta_j) \\ \\ -\sum_{i=1}^{N}(w_{ij} - P_{ij})\alpha_j \end{pmatrix}, \quad (5)$$

and

$$\nabla l_j(a_j, b_j, \gamma_j) = \begin{pmatrix} \frac{\partial l_j}{\partial a_j} \\ \\ \frac{\partial l_j}{\partial b_j} \\ \\ \frac{\partial l_j}{\partial \gamma_j} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N}(x_{ij} - P_{ij})(\theta_{i1} - b_j + w_{ij}\gamma_j) \\ \\ -\sum_{i=1}^{N}(x_{ij} - P_{ij})a_j \\ \\ \sum_{i=1}^{N}(x_{ij} - P_{ij})a_j w_{ij} \end{pmatrix}. \quad (6)$$

$$\frac{\partial l(\rho)}{\partial \rho} = N\rho^3 - \left(\sum_{i=1}^{N} \theta_{i1}\theta_{i2} \right) \rho^2 - \left(N - \sum_{i=1}^{N} (\theta_{i1}^2 + \theta_{i2}^2) \right) \rho - \sum_{i=1}^{N} \theta_{i1}\theta_{i2} \tag{7}$$

EM algorithm

Start with some initial guess $\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{a}^{(0)}, \boldsymbol{b}^{(0)}, \boldsymbol{\gamma}^{(0)}, \rho^{(0)}$. At the Kth iteration:

- Update posterior distribution for each $\boldsymbol{\theta}_i$ given the current parameter estimates.
- Update parameter estimates by solving

$$E_{\boldsymbol{\theta}|\ldots}[\nabla l_j(\alpha_j, \beta_j)] = 0, \tag{8}$$

$$E_{\boldsymbol{\theta}|\ldots}[\nabla l_j(a_j, b_j, \gamma_j)] = 0, \tag{9}$$

and

$$E_{\boldsymbol{\theta}|\ldots}[\frac{\partial l(\rho)}{\partial \rho}] = 0. \tag{10}$$

- For derivative based methods to solve the systems of equations (8) and (9), we need to compute the Jacobian matrix of the gradients $\boldsymbol{J}_{\nabla l_j(\alpha_j,\beta_j)}$ and $\boldsymbol{J}_{\nabla l_j(a_j,b_j,\gamma_j)}$
- The third order polynomial in (10) can be solved analytically.

Covariance Matrix of the estimator

- Invert the negative of the Hessian matrix of the marginal log-likelihood, $-\boldsymbol{H}^{-1}$. Doable but slow.
- Alternatively, we can use the missing information principal,

$$-l''(\zeta) = E_{\boldsymbol{\theta}|\ldots}[-l_c''(\zeta)] - E\left[-\frac{\partial^2 \log f(\boldsymbol{\theta}|\zeta)}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}^T}\right] \tag{11}$$

Results: item discriminations

Table

item	estimate	s.e.
1	1.254	0.068
2	1.233	0.066
3	1.992	0.100
4	0.583	0.057
5	1.654	0.095
6	1.420	0.082
7	1.120	0.062
8	1.027	0.057
9	0.990	0.060
10	1.560	0.088
11	1.375	0.079
12	1.752	0.103
13	0.923	0.063
14	1.504	0.090
15	1.869	0.092

Results: item difficulties

Table

item	estimate	s.e.
1	-0.563	0.050
2	-0.639	0.048
3	-0.267	0.031
4	3.421	0.388
5	1.455	0.107
6	-0.528	0.054
7	-0.551	0.066
8	0.558	0.051
9	1.025	0.072
10	0.783	0.073
11	-0.970	0.061
12	1.473	0.104
13	2.168	0.133
14	-0.486	0.071
15	0.035	0.032

Results: calculator discriminations

Table

item	estimate	s.e.
1	0.835	0.054
2	0.632	0.058
3	0.179	0.090
4	1.713	0.086
5	1.945	0.094
6	2.474	0.127
7	1.549	0.077
8	0.524	0.075
9	1.742	0.116
10	2.469	0.127
11	0.749	0.050
12	2.181	0.107
13	0.678	0.057
14	1.238	0.072
15	1.112	0.089

Results: calculator unattractiveness

Table

item	estimate	s.e.
1	0.650	0.060
2	2.282	0.193
3	15.859	7.907
4	-0.702	0.039
5	-0.273	0.031
6	-0.444	0.030
7	-0.590	0.039
8	4.523	0.605
9	1.551	0.066
10	-0.097	0.028
11	-0.230	0.057
12	-0.251	0.030
13	1.962	0.154
14	-1.469	0.072
15	2.330	0.144

Results: calculator effect

Table

item	estimate	s.e.
1	-0.017	0.073
2	-0.259	0.087
3	-0.191	0.107
4	1.200	0.268
5	0.973	0.109
6	0.229	0.077
7	-1.017	0.081
8	-1.113	0.182
9	0.785	0.142
10	0.027	0.077
11	0.186	0.074
12	0.468	0.098
13	0.981	0.133
14	0.830	0.095
15	0.159	0.091

Results: correlation between proficiency and calculator propensity

The correlation ρ is estimated to be 0.6186 with a standard error of 0.0166

Assessment of model fit

- How do we know our model fits the data?
- Absolute fit and relative model fit.
- Log-likelihood ratio test against an independent model (no correlation or any calculator effect).

$$LRT = -2\log\left(\frac{L_{ind}(\hat{\omega_{ind}})}{L(\hat{\omega})}\right) = -2(l_{ind}(\hat{\omega_{ind}}) - l(\hat{\omega})) \tag{12}$$

- The log-likelihood for the independent model is -45425.3, and the log-likelihood for the joint model is -44635.4. But we need to fit $16 = 15 + 1$ more parameters.
- $LRT = 1579.805$ with $df = 16$. $P - val < 0.0001$.
- The joint model very likely fits better.

Testing overall model fit

- The model induces a multinomial distribution, $\boldsymbol{\pi}$, for response/process patterns $\boldsymbol{X}$, $\boldsymbol{W}$.
- Intuitively, we can test the residuals between the model induced multinomial probabilities, π, and the observed probabilities p. For example, $\chi^2 = 2N\sum_{c=1}^{C}(\pi_c - p_c)^2/\pi_c$, for $C = K^J$ where K is the number of possible patterns for each item. In this case, $K = 4$.
- But the $J-$dimensional contingency table has $K^J = 1,073,741,824$ cells! We don't (will never) have enough data.
- We need a method to test the sparse contingency table.

Limited information fit statistics

- Instead of testing K^J cells of $\boldsymbol{\pi}$, we can decompose the contingency table into lower order marginal subtables.
- $\dot{\boldsymbol{\pi}}_r$ is the vector of the rth way marginal probabilities with $s_i = \binom{J}{r}(K-1)^r$ cells.
- $\boldsymbol{\pi}_r = (\dot{\boldsymbol{\pi}}_1', \dot{\boldsymbol{\pi}}_2', \ldots, \dot{\boldsymbol{\pi}}_r')$, marginal probabilities up to order r. We can test residuals $\boldsymbol{e}_r = \boldsymbol{p}_r - \boldsymbol{\pi}_r$.
- Marginal residuals have an asymptotic normal distribution, $\sqrt{N}\hat{\boldsymbol{e}}_r \xrightarrow{d} N(0, \Sigma_r)$.
- Consider test statistic $T_r = N\hat{\boldsymbol{e}}_r'\hat{W}\hat{\boldsymbol{e}}_r$, e.g. $\hat{W} = \boldsymbol{I}$.

P-value

- T_r asymptotically has a mixture chi-square distribution.
- We can use $A\chi_v + B$ to approximate. moment matching and solve for A, B, v.