Estimating Discrete Latent Variable Models Using Amortized Variational Inference

Karel Veldkamp, Dylan Molenaar, Raoul Grasman

Application: Mixture IRT

- Three-dimensional IRT model
- Intercepts vary across latent classes
- 28 Items
- N=10.000

Application: Mixture IRT

- Three-dimensional IRT model
- Intercepts vary across latent classes
- 28 Items
- N=10.000

	3D	**10D**
MPLUS	812	–
AVI	52	78

Table 1: Runtime in seconds

Today

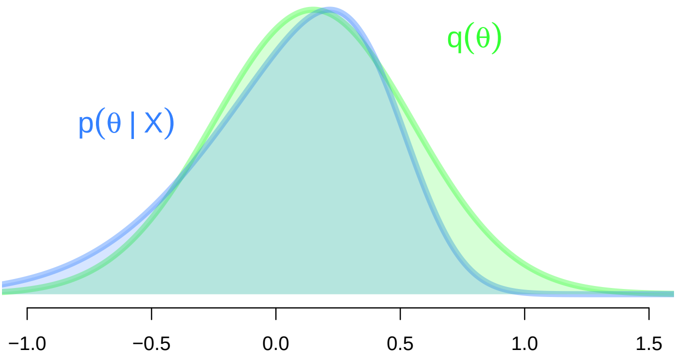
- Amortized variational inference
- Discrete Latent variable models
- Simulation 1: LCA and GDINA
- Simulation 2: Mixture MIRT
- Application: Narcisissm

Traditional estimation of latent variable models

- EM algorithm
  - Integrating over the latent variables
- Stochastic optimization
- MCMC

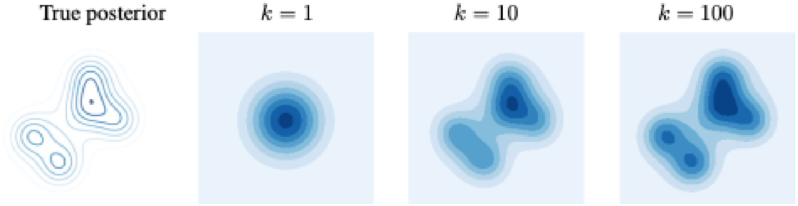
variational inference

- Approximate $p(\theta|X)$ with $q(\theta|X)$
- $\hat{q}(\boldsymbol{\theta}_i \,|\, \mathbf{x}_i) = \arg\min_{q(\boldsymbol{\theta}|\mathbf{x})\in\mathcal{F}} \mathrm{KL}[q(\boldsymbol{\theta}_i|\mathbf{x}_i)||p(\boldsymbol{\theta}_i|\mathbf{x}_i)]$
- $\mathcal{F}$: parametric family of distributions



Importance weighting

- $\mathcal{F}$ might be too simple to properly approximate the true posterior
- Taking multiple importance weighted samples of the posterior
- Approaches MML when $S \to \infty$



Cremer, C., Morris, Q., & Duvenaud, D. (2017). Reinterpreting importance-weighted autoencoders. arXiv preprint arXiv:1704.02916.

Amortised variational inference (AVI)

- Learn an inference function that maps response pattern to posterior parameters
- Function parameters are fixed across observations
- Sample from posterior
- Train model parameters and inference function parmaeters jointly using gradient descent

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes.

AVI for IRT

- Has been used to estimate MIRT parameters in many contexts

Interpretable variational autoencoders for cognitive models
M Curi, GA Converse, J Hajewski… - 2019 international joint …, 2019 - ieeexplore.ieee.org
… The relation between **VAE** and **MIRT** models has not been previously shown in the literature and is also an original contribution of this work. The decoder is an ML2P model under three …
Save Cite Cited by 52 Related articles All 3 versions
[PDF] ieee.org
UvA-linker: full text

A deep learning algorithm for high-dimensional exploratory item factor analysis
CJ **Urban**, DJ **Bauer** - Psychometrika, 2021 - cambridge.org
Marginal maximum likelihood (MML) estimation is the preferred approach to fitting item response theory models in psychometrics due to the MML estimator's consistency, normality, and …
Save Cite Cited by 48 Related articles All 19 versions Web of Science: 19
[PDF] cambridge.org
UvA-linker: full text

Estimating three-and four-parameter **MIRT** models with importance-weighted sampling enhanced variational auto-encoder
T Liu, C Wang, G Xu - Frontiers in Psychology, 2022 - frontiersin.org
… (IWVAE); to make the section self-contained, we also provide an overview of **VAE** and IWAE; important tricks for handling missing data as well as improving numerical stability are also …
Save Cite Cited by 15 Related articles All 7 versions Web of Science: 8
[PDF] frontiersin.org
UvA-linker: full text

Estimation of multidimensional item response theory models with correlated latent variables using variational autoencoders
G Converse, M Curi, S Oliveira, J Templin - Machine learning, 2021 - Springer
… matrix in a **VAE** requires a novel **VAE** architecture, which can … In addition, we show that the ML2P-**VAE** method is capable of … **MIRT** models (or models that are roughly equivalent to **MIRT** …
Save Cite Cited by 27 Related articles All 8 versions
[PDF] springer.com
UvA-linker: full text

Discrete latent variable models

- Gradient to the inference model
- Reparameterization
  - Normal distribution:
    - $z = \mu + \sigma\varepsilon$, $\varepsilon \sim N(0, 1)$
  - Categorical distribution
    - $z = \arg\max_{k \in 1,\dots,K}[log \pi_k + g_k]$, $g_k \sim Gumbel(0, 1)$

Concrete Distribution

- Replace the argmax operation with a continuous relaxation
  - $z_c = \frac{\exp((\log \alpha_c + g_c)/\lambda)}{\sum_i =1^n \exp((\log \alpha_i + g_i)/\lambda)}$
- Used for machine learning tasks
  - image recognition
  - structured output prediction
  - semi-supervised learning

Maddison, C. J., Mnih, A., & Teh, Y. W. (2016). The concrete distribution: A continuous relaxation of discrete random variables. arXiv preprint arXiv:1611.00712.

Jang, E., Gu, S., & Poole, B. (2016). Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144.

Simulation studies

- Two simulation studies to compare AVI and MML
  - Simple Discrete LVMs
    - LCA
    - GDINA
  - Hybrid LVMS
    - Mixture IRT

Simulation 1: LCA and GDINA

- LCA

$$p(\mathbf{Y} = \mathbf{y}) = \sum_{k}^{K} p(C = k) \prod_{i}^{I} P(Y_i = y_i | C = k)$$

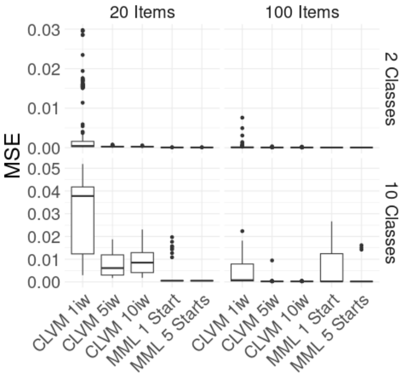
- GDINA

$$p(Y_i = y_i) = \sum_{k=1}^{K_j} \delta_{jk} \alpha_k \sum_{k=1}^{K_j - 1} \sum_{k'=k+1}^{K_j} \delta_{jkk'} \alpha_k \alpha_{k'}.$$

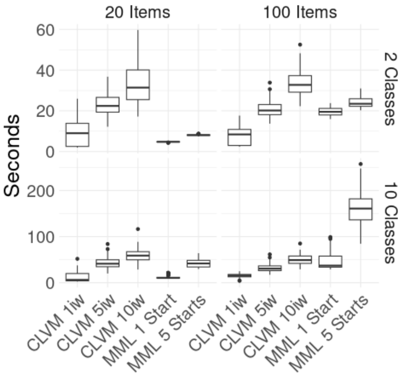
Simulation 1: LCA and GDINA

- 20 or 100 items
- 2 or 10 classes/attributes
- N $=$ 10.000
- 100 replications

LCA Results

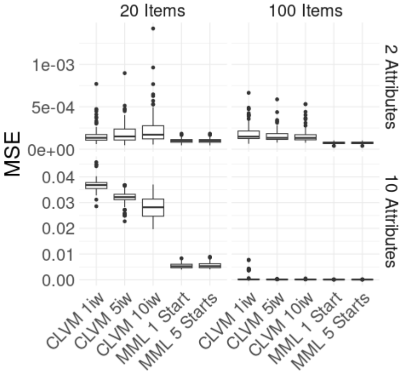


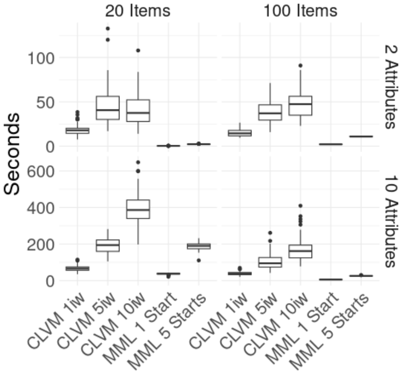
20 Items
100 Items
2 Classes
10 Classes
MSE
0.03
0.02
0.01
0.00
0.05
0.04
0.03
0.02
0.01
0.00
CLVM 1iw
CLVM 5iw
CLVM 10iw
MML 1 Start
MML 5 Starts



20 Items
100 Items
2 Classes
10 Classes
Seconds
60
40
20
0
200
100
0
CLVM 1iw
CLVM 5iw
CLVM 10iw
MML 1 Start
MML 5 Starts

GDINA Results





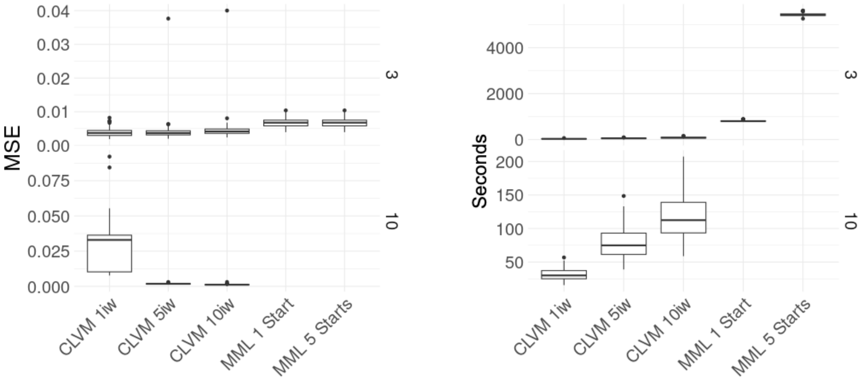
Simulation 2: Mixture MIRT

$$P(X_j = 1 \mid \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \cdot \frac{\exp(\mathbf{a}_j^\top \boldsymbol{\theta} + b_{jk})}{1 + \exp(\mathbf{a}_j^\top \boldsymbol{\theta} + b_{jk})}$$

Simulation 2: Mixture MIRT

- 3 Dimensional model with 28 items
- 10 Dimensional model with 110 items
- Intercept vary across two classes

Results



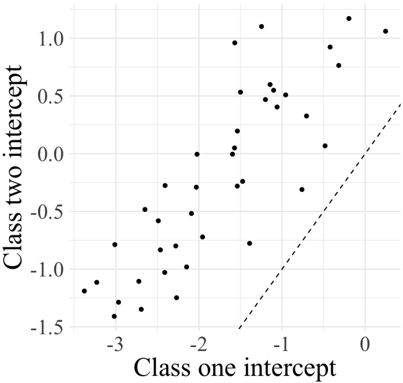
MSE
0.04
0.03
0.02
0.01
0.00
0.075
0.050
0.025
0.000
3
10
CLVM 1iw
CLVM 5iw
CLVM 10iw
MML 1 Start
MML 5 Starts
Seconds
4000
2000
0
200
150
100
50
3
10
CLVM 1iw
CLVM 5iw
CLVM 10iw
MML 1 Start
MML 5 Starts

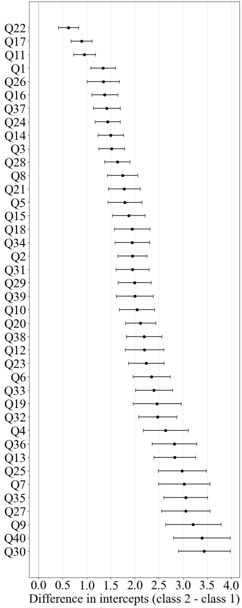
Application: Narcisissm

- Narcissism Personality Inventory
- 40 Binary forced choice Items
- 11,243 response
- 7 Subscales, simple structure

Results

- Intercepts
- Bootstrapped standard errors





Results

- Males are more likely to be in the narcissistic class
- People in the narcissistic class are younger

Grijalva, E., Newman, D. A., Tay, L., Donnellan, M. B., Harms, P. D., Robins, R. W., & Yan, T. (2015). Gender differences in narcissism: a meta-analytic review. Psychological bulletin, 141 (2), 261.

Weidmann, R., Chopik, W. J., Ackerman, R. A., Allroggen, M., Bianchi, E. C., Brecheen, C., et al. (2023). Age and gender differences in narcissism: A comprehensive study across eight measures and over 250,000 participants. Journal of personality and social psychology, 124 (6), 1277.

Try it yourself

- Github repository
- Fit models easily

```
git clone https://github.com/KarelVeldkamp/Discrete_VAEs/
cd Discrete_VAEs
pip3 install -r requirements.txt
python3 fit_model.py MIXIRT [data_path] [n_classes] [Q_path]
```