Variational Inference for the Partial Credit Model

Paolo Frazzetto, Nicoló Navarin, Alessandro Sperduti

The 4th Italian Conference on Big Data and Data Science - ITADATA 2025

September 9-11, 2025
Turin, Italy

UNIVERSITÀ DEGLI STUDI DI PADOVA

- **Core Problem:** Measuring latent human traits (e.g., ability, personality) is fundamental in education and social sciences.

The Challenge: Scaling Psychometric Analysis

- **Core Problem:** Measuring latent human traits (e.g., ability, personality) is fundamental in education and social sciences.
- **Key Model:** The Partial Credit Model (PCM) is essential for analyzing ordered data like Likert scales or multi-step problems.

The Challenge: Scaling Psychometric Analysis

- **Core Problem:** Measuring latent human traits (e.g., ability, personality) is fundamental in education and social sciences.
- **Key Model:** The Partial Credit Model (PCM) is essential for analyzing ordered data like Likert scales or multi-step problems.
- **The Bottleneck:** The rise of massive datasets from digital assessments and MOOCs presents a major computational challenge.

- **Core Problem:** Measuring latent human traits (e.g., ability, personality) is fundamental in education and social sciences.
- **Key Model:** The Partial Credit Model (PCM) is essential for analyzing ordered data like Likert scales or multi-step problems.
- **The Bottleneck:** The rise of massive datasets from digital assessments and MOOCs presents a major computational challenge.

Scalability Crisis

Traditional methods like Marginal Maximum Likelihood (MML) via the EM algorithm or Bayesian MCMC are robust but **do not scale**. They become prohibitively slow with large numbers of respondents (N) and items (I).

The Challenge: Scaling Psychometric Analysis

- **Core Problem:** Measuring latent human traits (e.g., ability, personality) is fundamental in education and social sciences.
- **Key Model:** The Partial Credit Model (PCM) is essential for analyzing ordered data like Likert scales or multi-step problems.
- **The Bottleneck:** The rise of massive datasets from digital assessments and MOOCs presents a major computational challenge.

Scalability Crisis

Traditional methods like Marginal Maximum Likelihood (MML) via the EM algorithm or Bayesian MCMC are robust but **do not scale**. They become prohibitively slow with large numbers of respondents (N) and items (I).

How can we perform fast, scalable, and principled inference for complex psychometric models like the PCM?

Our Contribution: The VA-PCM Framework

We introduce the **VA-PCM**, a novel Variational Autoencoder framework for the Partial Credit Model.

Our Contribution: The VA-PCM Framework

We introduce the **VA-PCM**, a novel Variational Autoencoder framework for the Partial Credit Model.

Our key contributions are:

1. **A Scalable Generative Model:** We formulate the PCM within a deep generative framework, enabling fast, amortized variational inference.

We introduce the **VA-PCM**, a novel Variational Autoencoder framework for the Partial Credit Model.

Our key contributions are:

1. **A Scalable Generative Model:** We formulate the PCM within a deep generative framework, enabling fast, amortized variational inference.
2. **Principled Covariate Integration:** We incorporate respondent covariates (X) into the model to enhance the precision of ability (Z) estimation.

We introduce the **VA-PCM**, a novel Variational Autoencoder framework for the Partial Credit Model.

Our key contributions are:

1. **A Scalable Generative Model:** We formulate the PCM within a deep generative framework, enabling fast, amortized variational inference.
2. **Principled Covariate Integration:** We incorporate respondent covariates (X) into the model to enhance the precision of ability (Z) estimation.
3. **Psychometrically-Informed Inference:** We design a novel prior and inference structure that guarantees the ordering of item difficulty thresholds ($\Psi_{i1} < \Psi_{i2} < \cdots < \Psi_{im}$) by construction, ensuring interpretable parameters.

Example: Turin is the best city in Italy.

Background: Likert Scales

Example: Turin is the best city in Italy.

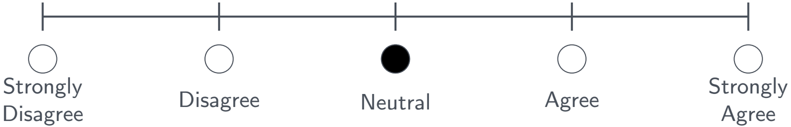


Figure: Example of a Likert scale with 5 values.

Background: The Partial Credit Model (PCM)

- An extension of the Rasch model for items with multiple, ordered response categories ($r \in \{0, \ldots, m\}$).
- **Example:** Perfect for 5-point Likert scales.
- **Core Idea:** The probability of a respondent n choosing category r for item i depends on their latent ability (Z_n) and a set of item "step difficulty" thresholds ($\boldsymbol{\Psi}_i$).

- An extension of the Rasch model for items with multiple, ordered response categories ($r \in \{0, \ldots, m\}$).
- **Example:** Perfect for 5-point Likert scales.
- **Core Idea:** The probability of a respondent n choosing category r for item i depends on their latent ability (Z_n) and a set of item "step difficulty" thresholds ($\boldsymbol{\Psi}_i$).

PCM Formula

$$p(R_{ni} = r | Z_n, \boldsymbol{\Psi}_i) = \frac{\exp\left(\sum_{k=0}^{r} (Z_n - \Psi_{ik})\right)}{\sum_{h=0}^{m} \exp\left(\sum_{k=0}^{h} (Z_n - \Psi_{ik})\right)}$$

- An extension of the Rasch model for items with multiple, ordered response categories ($r \in \{0, \ldots, m\}$).
- **Example:** Perfect for 5-point Likert scales.
- **Core Idea:** The probability of a respondent n choosing category r for item i depends on their latent ability (Z_n) and a set of item "step difficulty" thresholds ($\boldsymbol{\Psi}_i$).

PCM Formula

$$p(R_{ni} = r | Z_n, \boldsymbol{\Psi}_i) = \frac{\exp\left(\sum_{k=\mathbf{0}}^{r} (Z_n - \Psi_{ik})\right)}{\sum_{h=\mathbf{0}}^{m} \exp\left(\sum_{k=\mathbf{0}}^{h} (Z_n - \Psi_{ik})\right)}$$

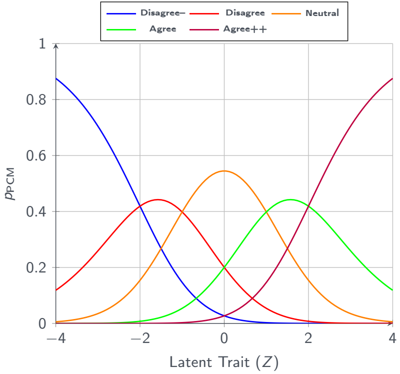
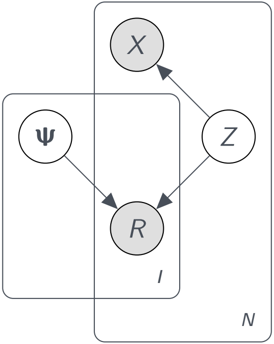


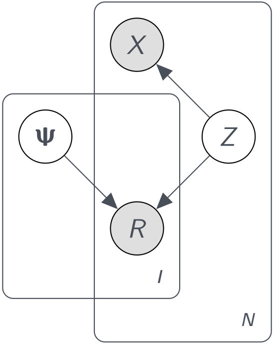
Figure: As ability Z increases, the probability of selecting higher categories increases.

Our generative story defines a joint probability distribution over all variables.



Università degli Studi di Padova

Our generative story defines a joint probability distribution over all variables.



Model Factorization

$$p(R, X, Z, \boldsymbol{\Psi}) = \underbrace{p(R|Z, \boldsymbol{\Psi})}_{\text{PCM Likelihood}} \mid \underbrace{p(X|Z)}_{\text{Covariate Model}} \mid \underbrace{p(Z)}_{\text{Ability Prior}} \mid \underbrace{p(\boldsymbol{\Psi})}_{\text{Item Prior}}$$

- I : Item amount;
- N : Respondents amount;
- R : Responses;
- Z : Hidden Ability or Trait;
- X : Covariate features;
- Ψ : Item Difficulty.

Key Innovation: The Prior on Item Thresholds $p(\boldsymbol{\Psi})$

- **Problem:** Standard priors (e.g., Gaussian) do not enforce the crucial ordering constraint: $\Psi_{i1} < \Psi_{i2} < \cdots < \Psi_{im}$.

Key Innovation: The Prior on Item Thresholds $p(\boldsymbol{\Psi})$

- **Problem:** Standard priors (e.g., Gaussian) do not enforce the crucial ordering constraint: $\Psi_{i1} < \Psi_{i2} < \cdots < \Psi_{im}$.
- **Our Solution:** A Stick-Breaking Process.
  1. We place a **Dirichlet prior** on the *proportions* of the latent scale: $\Delta\boldsymbol{\Psi} \sim \mathrm{Dir}(\boldsymbol{\alpha})$.
  2. These proportions are deterministically transformed into a set of **guaranteed ordered** thresholds $\boldsymbol{\Psi}$.

- **The Goal:** Compute the posterior $p(Z, \mathbf{\Psi}|R, X)$, which is intractable.

- **The Goal:** Compute the posterior $p(Z, \mathbf{\Psi}|R, X)$, which is intractable.
- **The Method:** Approximate it with a simpler, tractable distribution $q(Z, \mathbf{\Psi}|R, X)$.

- **The Goal:** Compute the posterior $p(Z, \mathbf{\Psi}|R, X)$, which is intractable.
- **The Method:** Approximate it with a simpler, tractable distribution $q(Z, \mathbf{\Psi}|R, X)$.

Mean-Field Approximation

We assume the approximate posterior factorizes:
$q(Z, \mathbf{\Psi}|R, X) \approx q_Z(Z|R, X)\, q_\Psi(\mathbf{\Psi}|R)$

Amortized Variational Inference

- **The Goal:** Compute the posterior $p(Z, \boldsymbol{\Psi} | R, X)$, which is intractable.
- **The Method:** Approximate it with a simpler, tractable distribution $q(Z, \boldsymbol{\Psi} | R, X)$.

Mean-Field Approximation

We assume the approximate posterior factorizes:
$q(Z, \boldsymbol{\Psi} | R, X) \approx q_Z(Z | R, X)\, q_\Psi(\boldsymbol{\Psi} | R)$

Amortization using Neural Networks (The "Encoder")

- **Ability Inference:** $\mathrm{NN}_Z(R_n, X_n) \rightarrow$ parameters of $q_Z(Z_n | \dots)$
- **Item Inference:** $\mathrm{NN}_\Psi(R_{\cdot i}) \rightarrow$ parameters of $q_\Psi(\boldsymbol{\Psi}_i | \dots)$

Amortized Variational Inference

- **The Goal:** Compute the posterior $p(Z, \mathbf{\Psi}|R, X)$, which is intractable.
- **The Method:** Approximate it with a simpler, tractable distribution $q(Z, \mathbf{\Psi}|R, X)$.

Mean-Field Approximation

We assume the approximate posterior factorizes:
$q(Z, \mathbf{\Psi}|R, X) \approx q_Z(Z|R, X)\, q_\Psi(\mathbf{\Psi}|R)$

Amortization using Neural Networks (The "Encoder")

- **Ability Inference:** $\mathrm{NN}_Z(R_n, X_n) \to$ parameters of $q_Z(Z_n|\dots)$
- **Item Inference:** $\mathrm{NN}_\Psi(R_{\cdot i}) \to$ parameters of $q_\Psi(\mathbf{\Psi}_i|\dots)$

Optimization: Maximize the Evidence Lower Bound (ELBO)

$$\mathcal{L}_{\text{ELBO}} = \underbrace{\mathbb{E}_q[\log p(R|Z, \mathbf{\Psi})]}_{\text{Response Reconstruction}} + \underbrace{\mathbb{E}_q[\log p(X|Z)]}_{\text{Covariate Reconstruction}} \underbrace{- \mathrm{KL}(q_Z||p_Z) - \mathrm{KL}(q_\Psi||p_\Psi)}_{\text{Regularization via Priors}}$$

Proof-of-Concept: Experimental Setup

UNIVERSITÀ DEGLI STUDI DI PADOVA

Data

- We use synthetic data to ensure a known ground truth for validation.
- **Config:**
  - $N = 1000$ respondents
  - $I = 10$ items
  - $m = 5$ categories/item

Proof-of-Concept: Experimental Setup

Data

- We use synthetic data to ensure a known ground truth for validation.
- **Config:**
  - $N = 1000$ respondents
  - $I = 10$ items
  - $m = 5$ categories/item

Implementation

- PyTorch
- Pyro (Deep Probabilistic Programming Language)

Proof-of-Concept: Experimental Setup

Data

- We use synthetic data to ensure a known ground truth for validation.
- **Config:**
  - $N = 1000$ respondents
  - $I = 10$ items
  - $m = 5$ categories/item

Implementation

- PyTorch
- Pyro (Deep Probabilistic Programming Language)

Model Architecture

- **Encoder (Inference):** NNs map responses and covariates to latent posterior parameters.
- **Decoder (Generative):** Samples of Z and $\boldsymbol{\Psi}$ are used to reconstruct responses (PCM) and covariates.

Proof-of-Concept: Experimental Setup

Data

- We use synthetic data to ensure a known ground truth for validation.
- **Config:**
  - $N = 1000$ respondents
  - $I = 10$ items
  - $m = 5$ categories/item

Implementation

- PyTorch
- Pyro (Deep Probabilistic Programming Language)

Model Architecture

- **Encoder (Inference):** NNs map responses and covariates to latent posterior parameters.
- **Decoder (Generative):** Samples of Z and $\boldsymbol{\Psi}$ are used to reconstruct responses (PCM) and covariates.

Training

- Adam optimizer.
- We use KL Annealing to prevent posterior collapse and ensure stable convergence.

Can our model recover the true parameters used to generate the data?

Can our model recover the true parameters used to generate the data?

Table: Parameter recovery metrics on the preliminary synthetic dataset.

Parameter	Correlation	RMSE	R^2 Score
Ability (Z)	0.344	0.884	-0.007
Thresholds ($\boldsymbol{\Psi}$)	0.750	0.742	0.498

Can our model recover the true parameters used to generate the data?

Table: Parameter recovery metrics on the preliminary synthetic dataset.

Parameter	**Correlation**	**RMSE**	**R^2 Score**
Ability (Z)	0.344	0.884	-0.007
Thresholds ($\boldsymbol{\Psi}$)	0.750	0.742	0.498

Success!

Excellent item parameter recovery. The strong correlation (0.75) and R^2 score show the model correctly learns the underlying structure of the items.

Can our model recover the true parameters used to generate the data?

Table: Parameter recovery metrics on the preliminary synthetic dataset.

Parameter	Correlation	RMSE	R^2 Score
Ability (Z)	0.344	0.884	-0.007
Thresholds ($\boldsymbol{\Psi}$)	0.750	0.742	0.498

Success!

Excellent item parameter recovery. The strong correlation (0.75) and R^2 score show the model correctly learns the underlying structure of the items.

Expected Challenge

Latent ability recovery is poor. This is a well-known difficulty in IRT, especially with short test lengths (only 10 items). There is not enough data per person for a precise estimate.

We introduced **VA-PCM**, a scalable and principled deep generative framework for the Partial Credit Model.

We introduced **VA-PCM**, a scalable and principled deep generative framework for the Partial Credit Model.

- Our model successfully integrates respondent covariates (X) to aid the inference of latent abilities (Z).

Conclusion & Summary

We introduced **VA-PCM**, a scalable and principled deep generative framework for the Partial Credit Model.

- Our model successfully integrates respondent covariates (X) to aid the inference of latent abilities (Z).
- A key innovation is a stick-breaking prior that guarantees psychometrically valid, ordered item parameters ($\boldsymbol{\Psi}$) by construction.

We introduced **VA-PCM**, a scalable and principled deep generative framework for the Partial Credit Model.

- Our model successfully integrates respondent covariates (X) to aid the inference of latent abilities (Z).
- A key innovation is a stick-breaking prior that guarantees psychometrically valid, ordered item parameters ($\boldsymbol{\Psi}$) by construction.
- Preliminary proof-of-concept results are promising: the framework demonstrates a strong ability to recover item parameters from synthetic data, validating the core mechanism.

More Extensive Validation

- Test on larger synthetic and real-world datasets (e.g., PISA, TIMSS).

More Extensive Validation

- Test on larger synthetic and real-world datasets (e.g., PISA, TIMSS).
- Rigorous benchmarking against traditional MML-EM and MCMC methods on accuracy and speed.

More Extensive Validation

- Test on larger synthetic and real-world datasets (e.g., PISA, TIMSS).
- Rigorous benchmarking against traditional MML-EM and MCMC methods on accuracy and speed.

Model Extensions

- Handle multi-dimensional latent traits ($Z \in \mathbb{R}^d$) to capture more complex cognitive structures.

Future Work

More Extensive Validation

- Test on larger synthetic and real-world datasets (e.g., PISA, TIMSS).
- Rigorous benchmarking against traditional MML-EM and MCMC methods on accuracy and speed.

Model Extensions

- Handle multi-dimensional latent traits ($Z \in \mathbb{R}^d$) to capture more complex cognitive structures.
- Explore more advanced neural architectures (e.g., Normalizing Flows) to improve the variational approximation.

Future Work

More Extensive Validation

- Test on larger synthetic and real-world datasets (e.g., PISA, TIMSS).
- Rigorous benchmarking against traditional MML-EM and MCMC methods on accuracy and speed.

Model Extensions

- Handle multi-dimensional latent traits ($Z \in \mathbb{R}^d$) to capture more complex cognitive structures.
- Explore more advanced neural architectures (e.g., Normalizing Flows) to improve the variational approximation.

This work aims to bridge the gap between traditional psychometrics and modern deep generative modeling.

Thank you!

Thank you!

Contact & Acknowledgments

- **Contact:** `paolo.frazzetto@unipd.it`

- **Acknowledgments:** We gratefully acknowledge the support of the PNRR project "Future AI Research - Spoke 2 - SimboliG" and of Amajor SpA.

Questions?

Thank you for your attention :-)

Emma Meneghini: emma.meneghini@studenti.unipd.it
Paolo Frazzetto: paolo.frazzetto@unipd.it
Nicolò Navarin: nicolo.navarin@unipd.it