Extending the inferential capability of a generalised partial credit model

Sujit Sahu

UNIVERSITY OF Southampton

http://www.soton.ac.uk/~sks/

Co-authors: Mark Bass, Carla Sabariego, Alarcos Cieza, Carolina Fellinghauer and Somnath Chatterji

IWSM 2017

Extending the inferential capability of a generalised partial credit model

Sujit Sahu

UNIVERSITY OF Southampton

http://www.soton.ac.uk/∼sks/

Co-authors: Mark Bass, Carla Sabariego, Alarcos Cieza, Carolina Fellinghauer and Somnath Chatterji

IWSM 2017

Outline

- What are Generalised Partial Credit Models (GPCM)?
- Why are we interested in GPCMs?
- What are our contributions?
- Summary discussion

Outline

- What are Generalised Partial Credit Models (GPCM)?
- Why are we interested in GPCMs?
- What are our contributions?
- Summary discussion

Outline

- What are Generalised Partial Credit Models (GPCM)?
- Why are we interested in GPCMs?
- What are our contributions?
- Summary discussion

Outline

- What are Generalised Partial Credit Models (GPCM)?
- Why are we interested in GPCMs?
- What are our contributions?
- Summary discussion

What are Generalised Partial Credit Models (GPCM)?

- Wider Context: Item Response Theory
- Appropriate for data set up where: Individuals are replying to items, e.g.
- (1) examinees answering multiple choice quesions
  (2) patients reporting their health status on different items
- In (1) we record a binary outcome (0 = incorrect & 1 = correct)
- In (2) we record a multinomial result: 1-5 say where 1: very good health to 5: very poor health.
- There are many other examples: Students rating their professors.

What are Generalised Partial Credit Models (GPCM)?

- Wider Context: Item Response Theory
- Appropriate for data set up where: Individuals are replying to items, e.g.
- (1) examinees answering multiple choice quesions
  (2) patients reporting their health status on different items
- In (1) we record a binary outcome (0 = incorrect & 1 = correct)
- In (2) we record a multinomial result: 1-5 say where 1: very good health to 5: very poor health.
- There are many other examples: Students rating their professors.

What are Generalised Partial Credit Models (GPCM)?

- Wider Context: Item Response Theory
- Appropriate for data set up where: Individuals are replying to items, e.g.
- (1) examinees answering multiple choice quesions
  (2) patients reporting their health status on different items
- In (1) we record a binary outcome (0 = incorrect & 1 = correct)
- In (2) we record a multinomial result: 1-5 say where 1: very good health to 5: very poor health.
- There are many other examples: Students rating their professors.

What are Generalised Partial Credit Models (GPCM)?

- Wider Context: Item Response Theory
- Appropriate for data set up where: Individuals are replying to items, e.g.
- (1) examinees answering multiple choice quesions
(2) patients reporting their health status on different items
- In (1) we record a binary outcome (0 = incorrect & 1 = correct)
- In (2) we record a multinomial result: 1-5 say where 1: very good health to 5: very poor health.
- There are many other examples: Students rating their professors.

What are Generalised Partial Credit Models (GPCM)?

- Wider Context: Item Response Theory
- Appropriate for data set up where: Individuals are replying to items, e.g.
- (1) examinees answering multiple choice quesions
  (2) patients reporting their health status on different items
- In (1) we record a binary outcome (0 = incorrect & 1 = correct)
- In (2) we record a multinomial result: 1-5 say where 1: very good health to 5: very poor health.
- There are many other examples: Students rating their professors.

What are Generalised Partial Credit Models (GPCM)?

- Wider Context: Item Response Theory
- Appropriate for data set up where: Individuals are replying to items, e.g.
- (1) examinees answering multiple choice quesions
  (2) patients reporting their health status on different items
- In (1) we record a binary outcome (0 = incorrect & 1 = correct)
- In (2) we record a multinomial result: 1-5 say where 1: very good health to 5: very poor health.
- There are many other examples: Students rating their professors.

Model ingredients (parameters)

- How well an item discriminates between different examinees (students)?
  - Known as item discrimination parameters. Notation: α or *a*.
  - Typical range from 0 to 2, where higher is better.
  - One such parameter for each item j, $j = 1, \ldots, J$.
- How well an item judges the level of examinees for which the item is appropriate?
  - Known as the difficulty parameters. Notation: β or *b*.
  - Typical range from –3 to 3. Average 0.
  - One such parameter For each item j, $j = 1, \ldots, J$.
- How can we estimate the un-observed ability levels of the examinees?
  - Known as the latent trait parameters. Notation θ.
  - Typical range from –3 to 3.
  - One such parameter for each examinee i, $i = 1, \ldots, I$.
- Ended up having parameters α_j,β_j and θ_i.

Model ingredients (parameters)

- How well an item discriminates between different examinees (students)?
  - Known as item discrimination parameters. Notation: α or a.
  - Typical range from 0 to 2, where higher is better.
  - One such parameter for each item j, $j = 1, \ldots, J$.
- How well an item judges the level of examinees for which the item is appropriate?
  - Known as the difficulty parameters. Notation: β or b.
  - Typical range from –3 to 3. Average 0.
  - One such parameter For each item j, $j = 1, \ldots, J$.
- How can we estimate the un-observed ability levels of the examinees?
  - Known as the latent trait parameters. Notation θ.
  - Typical range from –3 to 3.
  - One such parameter for each examinee i, $i = 1, \ldots, I$.
- Ended up having parameters α_j,β_j and θ_i.

Model ingredients (parameters)

- How well an item discriminates between different examinees (students)?
  - Known as item discrimination parameters. Notation: α or a.
  - Typical range from 0 to 2, where higher is better.
  - One such parameter for each item j, $j = 1, \ldots, J$.

- How well an item judges the level of examinees for which the item is appropriate?
  - Known as the difficulty parameters. Notation: β or b.
  - Typical range from –3 to 3. Average 0.
  - One such parameter For each item j, $j = 1, \ldots, J$.

- How can we estimate the un-observed ability levels of the examinees?
  - Known as the latent trait parameters. Notation θ.
  - Typical range from –3 to 3.
  - One such parameter for each examinee i, $i = 1, \ldots, I$.

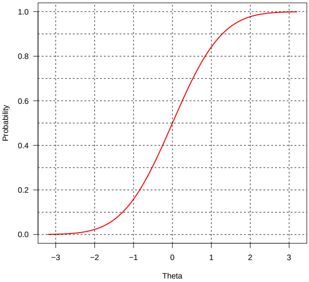
- Ended up having parameters α_j, β_j and θ_i.

Model ingredients (parameters)

- How well an item discriminates between different examinees (students)?
  - Known as item discrimination parameters. Notation: α or *a*.
  - Typical range from 0 to 2, where higher is better.
  - One such parameter for each item j, $j = 1, \ldots, J$.
- How well an item judges the level of examinees for which the item is appropriate?
  - Known as the difficulty parameters. Notation: β or *b*.
  - Typical range from –3 to 3. Average 0.
  - One such parameter For each item j, $j = 1, \ldots, J$.
- How can we estimate the un-observed ability levels of the examinees?
  - Known as the latent trait parameters. Notation θ.
  - Typical range from –3 to 3.
  - One such parameter for each examinee i, $i = 1, \ldots, I$.
- Ended up having parameters α_j, β_j and θ_i.

Models in Item Response Theory

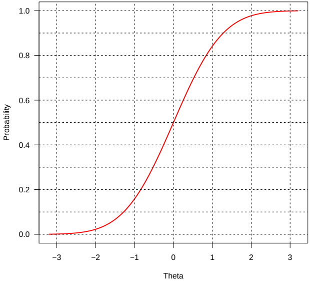
- Consider binary outcome for each item.
- Objective is to model the probability of correct response, p_{ij}, as a function of the parameters α_j, β_j and θ_i.
- Typically, an item characteristic curve (ICC) is visualised. $p_{ij} = F\left(\alpha_j(\theta_i - \beta_j)\right)$ where $F(\cdot)$ is a cumulative density function (cdf).



- ICC: $\alpha_j = 1$, $\beta_j = 0$ and F is the N(0,1) cdf.

Models in Item Response Theory

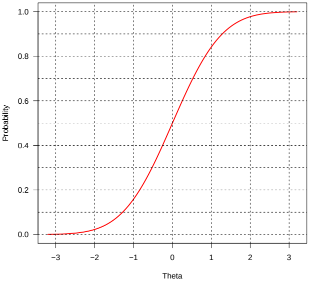
- Consider binary outcome for each item.
- Objective is to model the probability of correct response, p_{ij}, as a function of the parameters α_j, β_j and θ_i.
- Typically, an item characteristic curve (ICC) is visualised. $p_{ij} = F\left(\alpha_j(\theta_i - \beta_j)\right)$ where $F(\cdot)$ is a cumulative density function (cdf).



- ICC: $\alpha_j = 1$, $\beta_j = 0$ and F is the N(0,1) cdf.

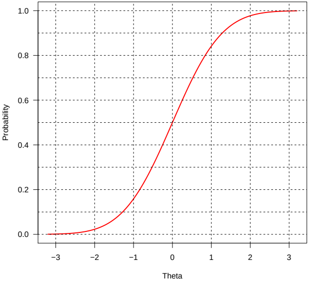
Models in Item Response Theory

- Consider binary outcome for each item.
- Objective is to model the probability of correct response, p_{ij}, as a function of the parameters α_j, β_j and θ_i.
- Typically, an item characteristic curve (ICC) is visualised. $p_{ij} = F\left(\alpha_j(\theta_i - \beta_j)\right)$ where $F(\cdot)$ is a cumulative density function (cdf).



- ICC: $\alpha_j = 1$, $\beta_j = 0$ and F is the N(0,1) cdf.

- Consider binary outcome for each item.
- Objective is to model the probability of correct response, p_{ij}, as a function of the parameters α_j, β_j and θ_i.
- Typically, an item characteristic curve (ICC) is visualised. $p_{ij} = F\left(\alpha_j(\theta_i - \beta_j)\right)$ where $F(\cdot)$ is a cumulative density function (cdf).



- ICC: $\alpha_j = 1$, $\beta_j = 0$ and F is the N(0,1) cdf.

Moving from IRT models to GPCM

- Dealing with multinomial response for each item.
- The jth item has K_j possibilities; $K_j = 5$ for the health status example.
- Let Y_{ij} denote the response of the ith individual on the jth item.
- The GPCM, Muraki (1992), is given by:

$$\Pr\left(Y_{ij} = y_{ij}|\theta_i, \alpha_j, \boldsymbol{\beta}_j\right) = \frac{\exp\left\{\sum_{h=1}^{y_{ij}} \alpha_j(\theta_i - \beta_{jh})\right\}}{\sum_{k=1}^{K_j} \exp\left\{\sum_{h=1}^{k} \alpha_j(\theta_i - \beta_{jh})\right\}}, \quad (1)$$

- the observed y_{ij} can take any value between 1 and K_j.
- This assumes $F(\cdot)$ to be the cdf of the logistic distribution.
- The item difficulty parameters β_j have been replaced by β_{jh}.

Our contributions: 3 inferential extensions

1. Estimation of all the parameters accounting for the covariate effects.
   - How do we separate out the effects of the covariates ($\mathbf{x}$)?
   - Where in the GPCM shall we insert $\mathbf{x}_i^T \gamma$?
2. How can we rank the J items so that a subset can be chosen subject to a given level of information content?
   - This will enable us to reduce the number of items.
3. How can we predict the latent ability scores θ of new individuals whose data are observed after model fitting has already been performed?
   - We do not want to re-fit the models with the new data.
   - Would like to develop a computational Bayesian device.
4. We perform all the extensions in a unified Bayesian modelling framework.

Our contributions: 3 inferential extensions

1. Estimation of all the parameters accounting for the covariate effects.
   - How do we separate out the effects of the covariates ($\mathbf{x}$)?
   - Where in the GPCM shall we insert $\mathbf{x}_i^T\gamma$?
2. How can we rank the J items so that a subset can be chosen subject to a given level of information content?
   - This will enable us to reduce the number of items.
3. How can we predict the latent ability scores θ of new individuals whose data are observed after model fitting has already been performed?
   - We do not want to re-fit the models with the new data.
   - Would like to develop a computational Bayesian device.
4. We perform all the extensions in a unified Bayesian modelling framework.

Our contributions: 3 inferential extensions

1. Estimation of all the parameters accounting for the covariate effects.
   - How do we separate out the effects of the covariates ($\mathbf{x}$)?
   - Where in the GPCM shall we insert $\mathbf{x}_i^T \gamma$?
2. How can we rank the J items so that a subset can be chosen subject to a given level of information content?
   - This will enable us to reduce the number of items.
3. How can we predict the latent ability scores θ of new individuals whose data are observed after model fitting has already been performed?
   - We do not want to re-fit the models with the new data.
   - Would like to develop a computational Bayesian device.
4. We perform all the extensions in a unified Bayesian modelling framework.

Our contributions: 3 inferential extensions

1. Estimation of all the parameters accounting for the covariate effects.
   - How do we separate out the effects of the covariates ($\mathbf{x}$)?
   - Where in the GPCM shall we insert $\mathbf{x}_i^T \gamma$?
2. How can we rank the J items so that a subset can be chosen subject to a given level of information content?
   - This will enable us to reduce the number of items.
3. How can we predict the latent ability scores θ of new individuals whose data are observed after model fitting has already been performed?
   - We do not want to re-fit the models with the new data.
   - Would like to develop a computational Bayesian device.
4. We perform all the extensions in a unified Bayesian modelling framework.

Our extended model is:

$$\Pr\left(Y_{ij} = y_{ij} | \theta_i, \alpha_j, \boldsymbol{\beta}_j, \boldsymbol{\gamma}\right) = \frac{1}{G_{ij}} \exp\left\{\sum_{h=1}^{y_{ij}} \alpha_j(\theta_i - \beta_{jh}) + I(y_{ij} > 1)\mathbf{x}_i^T \boldsymbol{\gamma}\right\}$$

where $I(y_{ij} > 1)$ is the indicator variable taking the value 1 if $y_{ij} > 1$ and 0 otherwise and

$$G_{ij} = \sum_{k=1}^{K_j} \exp\left\{\sum_{h=1}^{k} \alpha_j(\theta_i - \beta_{jh}) + I(k > 1)\mathbf{x}_i^T \boldsymbol{\gamma}\right\}.$$

- The regression term $\mathbf{x}_i^T \boldsymbol{\gamma}$ is excluded from the reference category ($j = 1$), achieved by $I(k > 1)$.
- Otherwise, it cancels in the ratio.

Why and how to make the model Bayesian?

- Difficult to estimate all the parameters.
- Difficult to make inference without re-course to step-wise procedures.
- Difficult to predict the ability scores for the new people.

Solution: Adopt Bayesian computation.

- Assume priors: $\theta_i \sim N(0, 1)$.
- Assume $\alpha_j \sim N(m_\alpha, s_\alpha^2)I(\alpha > 0)$
- $\beta_{jh} \sim N(m_\beta, s_\beta^2)$ for all j and $h > 1$.
- $\beta_{j1} = 0$ for the reference category (identifiability constraint).
- For weak prior distributions, we take $m_\alpha = m_\beta = 0$ and $s_\alpha^2 = s_\beta^2 = 10$.

Item ordering using Fisher Information

- Define the expected Fisher information for θ_i from the jth item:

$$I_{ij}(\theta_i, \xi) = -E\left[\frac{\partial^2}{\partial\theta_i^2} \log L(\boldsymbol{\theta}_n, \xi; \mathbf{y})\right],$$

- Adopting Bayesian methods. Integrate the parameters ξ.

$$I_{ij}(\theta_i) \equiv E\left(I_{ij}(\theta_i, \xi)|\mathbf{y}\right) = \int I_{ij}(\theta_i, \xi)\pi(\xi|\mathbf{y})d\xi.$$

- Use MCMC samples $\theta_i^{(\ell)}$ and $\xi^{(\ell)}$) for $\ell = 1, \ldots, L$ to do the integration!

$$\hat{I}_{ij} \equiv \frac{1}{L}\sum_{\ell=1}^{L} I_{ij}(\theta_i^{(\ell)}, \xi^{(\ell)}).$$

- Finally

$$\hat{I}_j = \sum_{i=1}^{n} \hat{I}_{ij}.$$

Predicting the ability scores

- Aim is to predict the latent ability score for new individuals whose data are observed after model fitting.
- Use the Bayesian predictive distribution:

$$\pi(\theta_{n+1}|\mathbf{y}, \mathbf{y}_{n+1}) = \int \pi(\theta_{n+1}|\xi, \mathbf{y}, \mathbf{y}_{n+1})\pi(\xi|\mathbf{y}, \mathbf{y}_{n+1})d\xi.$$

- But sampling from the full posterior, $\pi(\xi|\mathbf{y}, \mathbf{y}_{n+1})$, requires re-fitting.
- We want to avoid re-fitting!
- Our proposal is to approximate the full posterior $\pi(\xi|\mathbf{y}, \mathbf{y}_{n+1})$ by the fitted posterior $\pi(\xi|\mathbf{y})$ with n data points.
- We then use a Metropolis-Hastings step to sample from the predictive distribution: $\pi(\theta_{n+1}|\xi^{(\ell)}, \mathbf{y}, \mathbf{y}_{n+1})$.
- The details are in the paper.

The data set

- Model Disability Survey (MDS) is a standardized instrument for data collection on disability.
- It provides comprehensive and systematic documentation on all aspects of health and functioning in a country.
- We will analyse a data set collected from Sri Lanka by the World Health Organisation.
- Data from 3000 individuals on 17 items on capacity levels.
- Respondents were asked to select one of $K = 5$ categories, ranging from 1, "no difficulty" to 5, "extreme difficulty" in capacity levels.
- Item descriptions follow.

Item descriptions

Table: Item description and number of respondents in the 5 categories. Questions begin with “How much difficulty do you have...”

Item	Description	Category 1	2	3	4	5	Total
1	seeing	1449	499	421	408	202	2979
2	hearing	2668	140	70	56	29	2963
3	walking or climbing	2050	316	223	228	181	2998
4	remembering or concentrating	2295	352	155	131	67	3000
5	washing all over or dressing	2708	188	32	35	37	3000
6	communicating	2781	132	37	28	22	3000
7	using hands and fingers	2543	223	107	86	41	3000
8	sleeping	2423	255	132	117	72	2999
9	shortness of breath	2530	224	125	89	32	3000
10	doing household tasks	2419	287	127	88	79	3000
11	providing care for others	2345	263	127	116	149	3000
12	joining community activities	2435	266	107	95	94	2997
13	feeling sad, low or depressed?	2294	384	174	93	52	2997
14	feeling worried or nervous	2304	366	162	107	54	2993
15	getting along with close people	2476	280	132	67	45	3000
16	coping with everything	2409	338	119	80	52	2998
17	bodily aches or pain	1792	491	328	258	126	2995

Our data set

- Out of 3000 individuals, 1791 (59.7%) are females.
- Other covariate information: Age, sex and income, categorised into 5 levels.
- Boxplot of the total scores of the respondents from the 17 items (minimum=17, maximum=85)

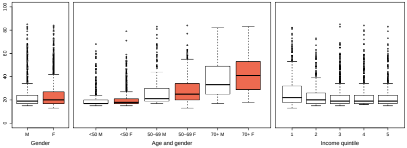


Figure: Left panel: by gender; middle panel by gender and age groups and right panel by income groups.

Results

Table: Parameter estimates for the regression coefficients. The first quintile of the 5 level factor income is the baseline, $\gamma_{\text{Income2}}, \ldots, \gamma_{\text{Income5}}$ are the incremental effects for the upper income quintiles.

Parameter	Estimate	95% CI
γ_{Age}	0.793	(0.745, 0.836)
γ_{Female}	0.104	(0.018, 0.200)
γ_{Income2}	0.150	(–0.005, 0.305)
γ_{Income3}	–0.071	(–0.185, 0.040)
γ_{Income4}	–0.074	(–0.208, 0.060)
γ_{Income5}	–0.184	(–0.321, –0.057)

- All effects are significant.
- Hence the covariates must be included.

Effect of including the covariates

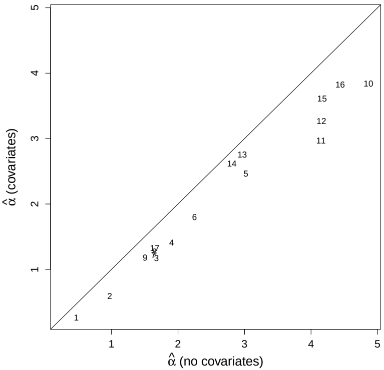


Figure: Estimates of α from the GPCM with and without including the covariates. Item numbers are used as the ploting symbols.

- Item discriminatory parameter estimates become smaller.
- As the covariates explain a portion of the total variability.

Results from item ordering

Table: FIC = Fisher Information Criterion. Items have been added in order of importance.

Step	Item added	FIC	% of FIC	Cumulative %
1	10 (household tasks)	2.76	12.68	12.68
2	11 (providing care)	2.54	11.65	24.33
3	16 (coping with everything)	2.46	11.28	35.61
4	12 (joining community)	2.39	10.96	46.57
5	15 (getting along)	2.15	9.85	56.43
6	13 (feeling sad)	1.78	8.18	64.60
7	14 (feeling worried)	1.72	7.88	72.48
8	17 (bodily aches)	1.04	4.76	77.24
9	3 (walking)	0.92	4.25	81.49
10	5 (washing or dressing)	0.91	4.16	85.65
11	4 (remembering)	0.82	3.75	89.39
12	8 (sleeping)	0.68	3.11	92.51
13	7 (using hands)	0.50	2.28	94.79
14	9 (shortness of breath)	0.47	2.14	96.93
15	6 (communicating)	0.45	2.07	99.00
16	2 (hearing)	0.12	0.53	99.50
17	1 (seeing)	0.10	0.47	100.00

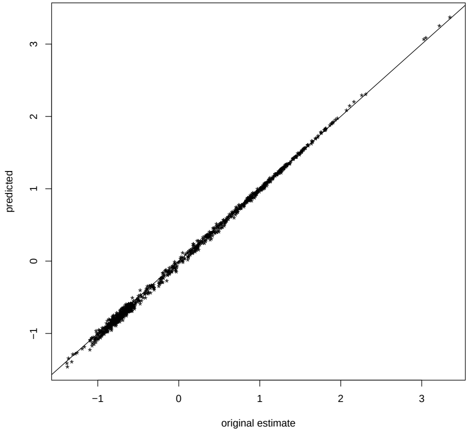


Figure: Scatter plot of the predicted θ_i for a new set of 1000 presumed new respondents against the original estimates obtained using the full data set including the new respondents.

Discussion

- Presented full Bayesian methods for making inference with the GPCMs.
- Achieved three inferential extensions for the GPCMs.
- By including the covariates in the model itself we are able to account for (rather than adjust for) the effects of the covariates.
- The unified model frees us from ad-hoc step-wise inference procedures which use the data several times without tracking the correct amount of uncertainty.
- All the assumptions in the modelling are also explicit, which enables their scrutiny!
- Easy implementation using publicly available software packages, such as WinBUGS.