Assessing uncertainty in posterior intercepts from random effect models

Jochen Einbeck and Nick Sofroniou

Durham University
WARWICK
THE UNIVERSITY OF WARWICK

December 17, 2017

Motivating application: PIAAC survey of adult skills

PIAAC: 'Programme for the International Assessment of Adult Competencies'

- carried out from 01/08/2011 to 31/03/2012 by the OECD in 24 countries (or sub–country entities);
- designed to assess the proficiency of adults in the key competencies of literacy, numeracy, and problem–solving in technology–rich environments.

For this work, we focus on literacy, being defined as

'the ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential'.

Objective:

- Rank countries by literacy competency.
- Identify which countries can/cannot be distinguished in terms of their ranking (due to uncertainty).

Data from OECD 'PIAAC data explorer'.

OECD
BETTER POLICIES FOR BETTER LIVES

On the raw scale, literacy is measured by a continuous score on a scale from 0 (worst) to 500 (best).

- PIAAC explorer categorizes these scores into six subcategories.

Our response will be a dichotomized variable indicating ‘people reaching level 3 or above’‘

- ‘level 2 and below’ considered as low–skilled.
- corresponds to key policy marker used to demarcate poor basic skills in the complementary PISA survey carried out at 15–years of age (Eurostat, 2016).

Covariates:

- gender;
- a factor for age (16–24, 25–34, 35–44, 45–54, and 55+);
- many others: employment status, reading habits (but only three covariates can be extracted at a time);

Two–level problem:

- Upper level: 24 countries (i)
- Lower level: 2×5 age/gender subcategories (j)

Logistic regression model with random effect:

$$\text{logit}(\mu_{ij}) \equiv \log \frac{\mu_{ij}}{1-\mu_{ij}} = x_{ij}^T\beta + z_i,$$

where

- $y_{ij} \sim \text{Bin}(n_{ij}, \mu_{ij})/n_{ij}$;
- n_{ij} 'effective sample size' for subcategory j in country i (Sofroniou et al, 2008);
- $\mu_{ij} = E(y_{ij}|z_i)$,
- $g(\cdot)$ is the (unspecified) density of the z_i.

Interested in z_i!

Model fitting via NPML:

Approximate marginal likelihood by finite mixture,

$$L = \prod_{i=1}^{n} \int f(y_i|z_i,\beta) g(z)\, dz_i = \prod_{i=1}^{n} \sum_{k=1}^{K} \overbrace{f(y_i|z_k,\beta)}^{f_{ik}} \pi_k$$

where $y_i = (y_{i,1}, \ldots y_{i,10})^T$ and $f(y_i|z_i,\beta) = \prod_{j=1}^{10} f(y_{ij}|z_i,\beta)$.
β, z_k π_k can be estimated via EM.
As by–product, obtain 'posterior probability that case i stems from component k'

$$w_{ik} = \frac{\pi_k f_{ik}}{\sum_{\ell} \pi_{\ell} f_{i\ell}}$$

(Aitkin, 1999)

The mean of the posterior distribution $z_i|y_i$ can be estimated via 'Empirical Bayes Predictions'

$$\tilde{z}_i = \sum_{k=1}^{K} w_{ik}\hat{z}_k$$

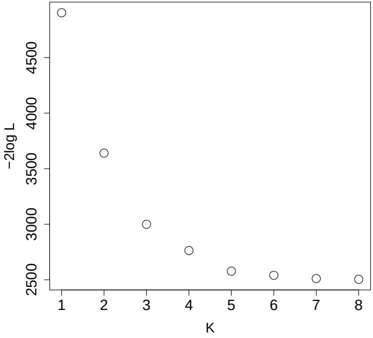
Assuming the w_{ik} to be 'constants',

$$\text{Var}(\tilde{z}_i) = \sum_{k=1}^{K} w_{ik}^2 \text{Var}(\hat{z}_k) + \sum_{j \neq k} w_{ij} w_{ik} \text{Cov}(\hat{z}_j, \hat{z}_k)$$

Note that $\sum_k w_{ik} = 1$, so the covariance part will usually be very small.

The number of mass points is increased until $-2\log L$ shows little change.

For the full age*gender interaction model, we employ $K = 6$ mass points:

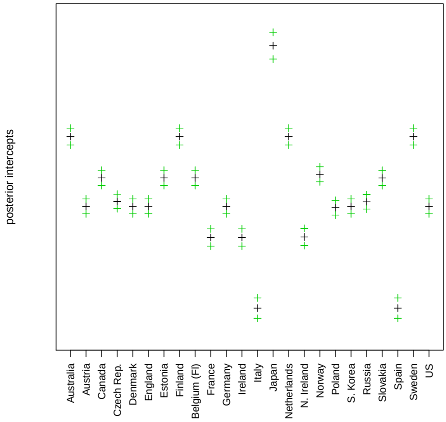


Estimated w_{ik}, $\hat{z}_k$, $\tilde{z}_i$, $SE(\tilde{z}_i)$ (excerpt):

k	1	2	3	4	5	6	$\tilde{z}_i$	$SE(\tilde{z}_i)$
Australia	0	0	0	0	1	0	0.7054	0.0294
Austria	0	0.0002	0.9998	0	0	0	0.2275	0.0260
Canada	0	0	0	1	0	0	0.4223	0.0268
Czech Republic	0	0	0.8269	0.1731	0	0	0.2613	0.0252
...								
England	0	0	0.9988	0.0012	0	0	0.2278	0.0260
Japan	0	0	0	0	0	1	1.3282	0.0468
...								
$\hat{z}_k$	-0.4696	0.0143	0.2276	0.4223	0.7054	1.3282		

A high value of $\tilde{z}_i$ corresponds to high (mean) literacy competency in a country.

Posterior intercepts with 95% confidence intervals



- These are $\tilde{z}_i \pm 1.96\sqrt{\text{Var}(\tilde{z}_i)}$, with the analytical expression used for the latter.
- Does ignore uncertainty in the allocation of observations to mass points.

- For all i (countries),

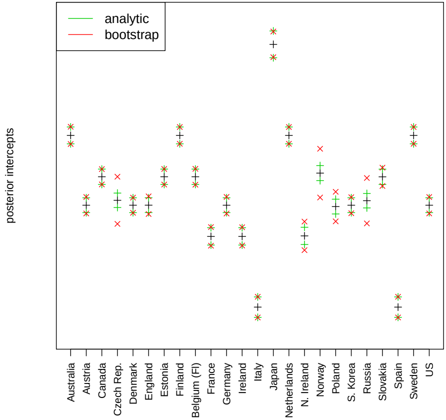
  (i) from the set of mass points $\hat{z}_1, \ldots, \hat{z}_k$ draw a point $\check{z}_i$ with probability w_{ik};

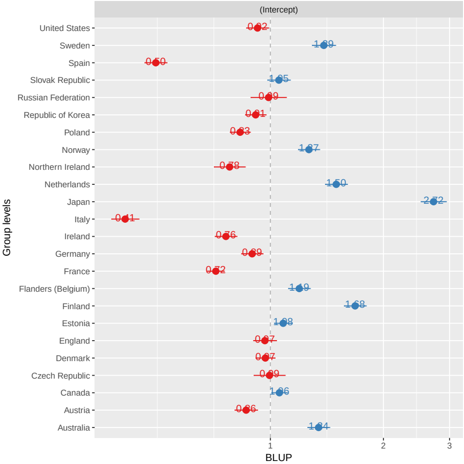
  (ii) for all j (subcategories), generate new $\check{y}_{ij} \sim \text{Bin}(n_{ij}, \check{\mu}_{ij})/n_{ij}$, where

$$\check{\mu}_{ij} = \frac{\exp\{x_{ij}^T \hat{\beta} + \check{z}_i\}}{1 + \exp\{x_{ij}^T \hat{\beta} + \check{z}_i\}}$$

- Having generated a complete set (y_{ij}), re–fit the model, yielding a new set of posterior intercepts.
- Repeat this procedure M times and then take standard deviations of the M posterior intercepts, for each i.

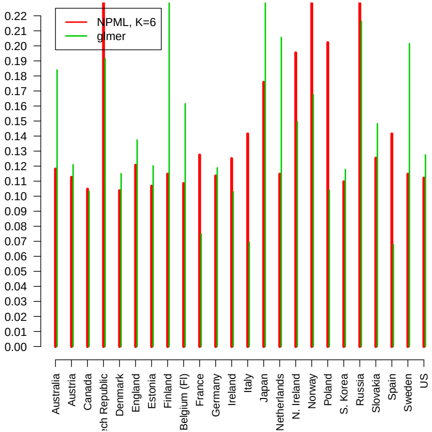
Bootstrapped and analytical intervals





(using function `sjp.glmer`)

Comparing lengths of intervals



Mean interval lengths:

- NPML (K=6, bootstrap): 0.15231
- glmer: 0.15257

We are still assuming the w_{ik} (which initialize the bootstrap) to be 'constant'.

Possible solution: Posterior likelihood (Aitkin et al, 2014)

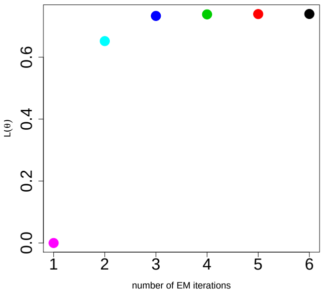
- Let $\theta = (\mu_k, z_k, \beta)$.
  Draw samples $\hat{\theta}^{[m]}$ from $L(\theta)$ and plug into w_{ik}.

- Requires full MCMC...

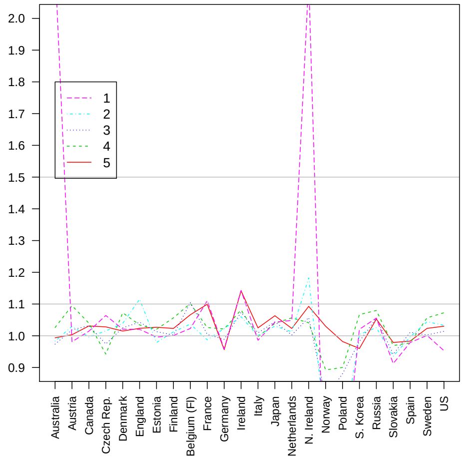
Use existing EM process path.

- Say, in EM iteration $s = 1, \ldots, S$, we have obtained parameter estimates $\hat{\theta}^{[s]}$ with associated weight matrices $W^{[s]}$ and likelihoods $L^{[s]} \equiv L(\hat{\theta}^{[s]})$.
- Hence, we possess S ‘draws’ from $L(\theta)$, including the final iteration, corresponding to MLE $\hat{\theta}^{[s]} \equiv \hat{\theta}$.
- While these S draws do not represent the correct shape of $L(\theta)$, the matrices $W^{[s]}$ can still be used to assess the *sensitivity* of the NPML–Bootstrap to imprecision in the ${w_{ik}}'s$.

$L(\hat{\theta}^{[s]})$ versus s.



Interval length $w_{ik}^{[s]}$ relative to $w_{ik} \equiv w_{ik}^{[6]}$.



Why not replace the random effect by country indicators

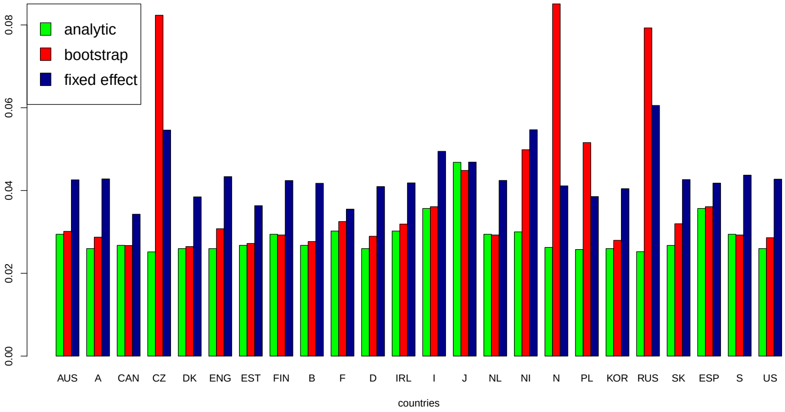
$$\text{logit}(\mu_{ij}) = \beta_0 + x_{ij}^T\beta + \sum_{j>1} \gamma_j \mathbf{1}_{\{\text{country}=j\}}.$$

Rankings based on $\hat{\gamma}_j$ and associated standard errors immediately available!

However, note: The quantity of interest is $\hat{\beta}_0 + \hat{\gamma}_j$, with

$$SE(\hat{\beta}_0 + \hat{\gamma}_j) = \sqrt{\text{Var}(\hat{\beta}_0) + \text{Var}(\hat{\gamma}_j) + 2\text{Cov}(\hat{\beta}_0, \hat{\gamma}_j)}.$$

Interval length



- Posterior intercepts constitute a way of producing continuous 'scores', which are adjusted to covariates, and are less volatile than the raw rates, and hence can be conveniently be employed for ranking problems.
- Bootstrapped confidence intervals appear to be a sufficiently reliable tool to assess the uncertainty in posterior intercepts, and hence the separability of the rankings.
- Discrete random effects lead to grouped posterior random effects, which could suggest the presence of latent (omitted) categorical explanatory variables, and may have policy relevance.

Aitkin, M. (1999). A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models. *Biometrics*, **55**, 117–128.

Aitkin, M., Duy, V. and Francis, B. (2014). Statistical Modelling of the group structure for social networks. *Social Networks*, **38**, 74–87.

Eurostat (2016). Smarter, greener, more inclusive? Indicators to support the Europe 2020 strategy. *Publications Office of the European Union*, European Commission, Luxembourg.

Sofroniou, N., Hoad, D. and Einbeck, J. (2008). Leage tables for literacy survey data based on random effect models. In: Eilers, P (Ed.), 23rd IWSM, Utrecht, 402–405.