Supplemental material for "Finding archetypal patterns for binary questionnaires"

Ismael Cabero[1] and Irene Epifanio[2]

June 2020

The material contained herein is supplementary to the article named in the title and published in SORT-Statistics and Operations Research Transactions Volume 44(1).

[1] Departamento de Didáctica de las Matemáticas, Universidad Internacional de La Rioja, Spain.
Email: ismael.cabero@unir.net
[2] Departament de Matemàtiques-IMAC-IF, Universitat Jaume I, Castelló 12071, Spain. Email: epifanio@uji.es

FADA with kernel and parametric IRF estimates

As mentioned in the manuscript, we can use other approaches to estimate IRFs. One common approach is to estimate IRFs parametrically, i.e. curves are modelled by a set of parameters that are estimated, as in the three-parameter logistic (3PL) model. Another important approach is not to assume any mathematical form and to estimate IRFs nonparametrically, for example by kernel smoothing (Ramsay, 1991). Ramsay (1997) and Rossi et al. (2002) defend the flexibility of nonparametric methods compared with the restrictions of parametric methods. In any case, we can apply ADA to the IRFs estimated by any method selected by the researcher. We compare here the results using these two estimation methods.

Figure 1 shows IRFs estimated by kernel smoothing with the R package **KernSmoothIRT** (Mazza et al., 2014) and the 3PL model with the R package **irtoys** (Partchev and Maris, 2017; Rizopoulos, 2006). Note that the estimates are quite different, also if we compare them with those in Figure 8 of the paper. On the one hand, parametric models are not as flexible as nonparametric methods (remember the previous analysis about the variation in the upper asymptote for the second PC component). The possible shapes of the 3PL model estimates are restricted by the functional form. On the other hand, although kernel smoothing makes it possible to represent the data well, there is too much local curvilinearity, i.e. they are not as smooth as the estimated IRFs in the previous section. With kernel and 3PL model estimation methods, we have the estimates of IRFs in a series of points and we can apply ADA to obtain the functional archetypoids. Note that in FDA we can also work point-wise with discretized functions to a fine grid. Nevertheless, as the input data are different, we can expect variation in the archetypoids obtained for the different estimation methods.

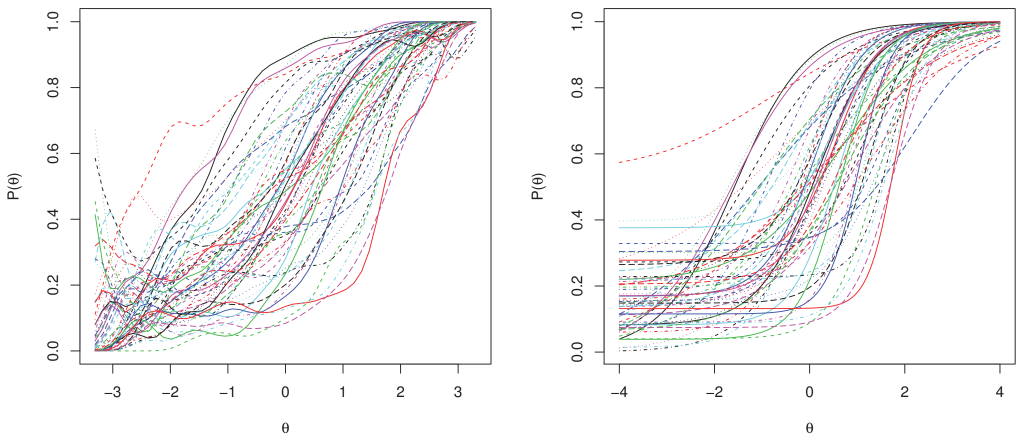

Figure 1: *Estimated IRFs by kernel smoothing (left-hand panel) and the 3PL model (right-hand panel) for the ACT math exam.*

Figure 2 shows the PC scores together with the functions with extreme scores for each component, as computed by the R package **KernSmoothIRT**. Note the strange estimate for item 3 in the lower end of the ability range, and the noisy estimates for the other items. The 4 archetypoids derived from kernel smoothing and 3PL model estimates are shown in Figure 3. Note that none of the archetypoids coincides with the extreme PC scores. Two of the archetypoids coincide in all three estimation methods: item 2 and item 60, which are typical of the easiest and hardest items, respectively. The other two archetypoids in each case show a similar positive slope, but begin at different values of θ.

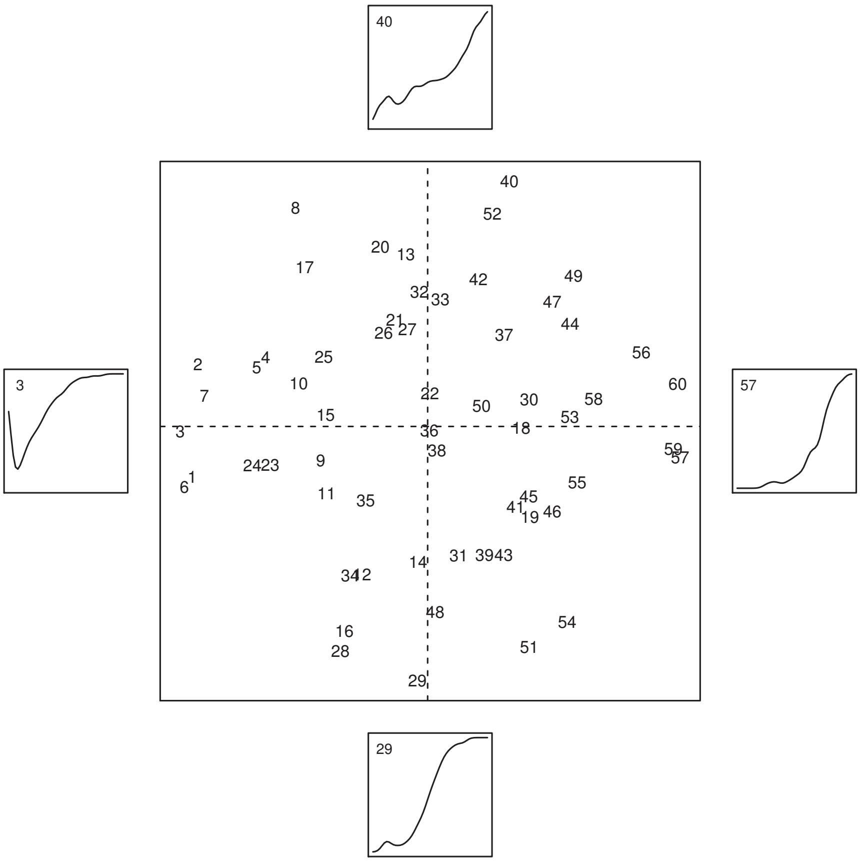

Figure 2: *First two principal components for the ACT math exam with kernel smoothing. In the interior plot, numbers are the identifiers of the items. The small plots show the estimated IRFs for the most extreme items for each principal component.*

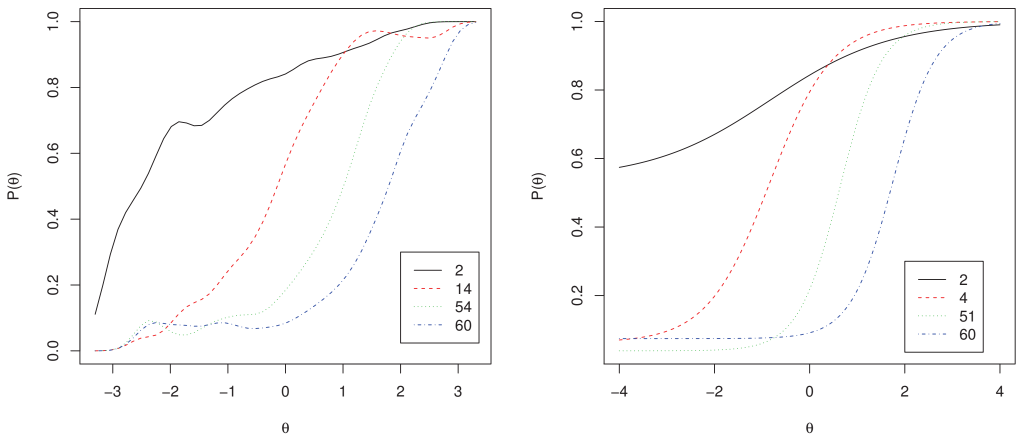

Figure 3: *The four archetypoids for the estimated IRFs by kernel smoothing (left-hand panel) and the 3PL model (right-hand panel) for the ACT math exam. See the legend inside each plot.*

In summary, we can apply ADA to different ways of estimating IRFs. Obviously, depending on the input, ADA, or any other method, returns different solutions. So we must be cautious with the estimation method we use.

References

Mazza, A., Punzo, A. and McGuire, B. (2014). KernSmoothIRT: An R package for kernel smoothing in item response theory. *Journal of Statistical Software*, 58, 1–34.

Partchev, I. and Maris, G. (2017). *irtoys: A Collection of Functions Related to Item Response Theory (IRT)*. R package version 0.2.1.

Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56, 611–630.

Ramsay, J. O. (1997). *A Functional Approach to Modeling Test Data*, pp. 381–394. New York, NY: Springer New York.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response analysis. *Journal of Statistical Software, Articles*, 17, 1–25.

Rossi, N., Wang, X. and Ramsay, J. O. (2002). Nonparametric item response function estimates with the EM algorithm. *Journal of Educational and Behavioral Statistics*, 27, 291–317.