BCH Two-step Auxiliary Variable Integration with `MplusAutomation`

Adding covariates and distal outcome variables to mixture models

Adam Garber

9/8/2021

This `R` tutorial automates the BCH two-step axiliary variable procedure (Bolk, Croon, Hagenaars, 2004) using the `MplusAutomation` package (Hallquist & Wiley, 2018) to estimate models and extract relevant parameters. To learn more about auxiliary variable integration methods and why multi-step methods are necessary for producing un-biased estimates see Asparouhov & Muthén (2014). The name of this mehtod, BCH, stands for Bolck, Croon, & Hagenaars, the authors who developed this method (Bolk, Croon, Hagenaars, 2004).

Data source & applied example. This tutorial utilizes the public-use data repository named the *Longitudinal Survey of American Youth* (LSAY; Miller et al., 1992). The applied example used in this tutorial is based off the example presented in the seminal chapter on mixture modeling by Katherine Masyn (2013). This tutorial contains the 9 math indicators from this original study as well as two auxiliary variables. See the following book chapter for details about the applied example:

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), The Oxford Handbook of Quantitative Methods: Vol. 2. Statistical analysis (pp. 551–611). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199934898.013.0025

If using this tutorial to automate the 3-step approach it would be greatly appreciated if you cite this resource:

Garber, A. C. (2021). BCH Two-step Auxiliary Variable Integration Using MplusAutomation. Retrieved from https://psyarxiv.com/wmfcj

---

Associated Github repository:

https://github.com/garberadamc/BCH-Method

Load packages

```
library(MplusAutomation)  # a conduit between R & Mplus
library(here)             # to locate or send files within the Rproject folder
library(gt)               # for pretty tables
library(tidyverse)        # for everything else...
```

Read in `LSAY` data

```
lsay_data <- read_csv(here("data", "lsay_subset.csv")) %>%
  mutate(female = gender - 1)
```

LCA Indicators & Auxiliary Variables: Math Attitudes Example[1]

Name	Variable Description
7th Grade Student Math Attitudes	
enjoy	I enjoy math.
good	I am good at math.
undrstnd	I ussually understand what we are doing in math.
nervous	Doing math often makes me nervous or upset
scared	I often get scared when I open my math book and see a page of problems.
useful	Math is useful in everyday problems.
logical	Math helps a person think logically.
job	It is important to know math to get a good job.
adult	I will use math in many ways as an adult.
Auxiliary Variables	
female	Self-reported student gender (0=Male, 1=Female).
math_irt	Standardized IRT math test score reported in 9th grade (two years distal).

[1]Note. The example presented replicates the model found originally in Masyn (2013). All data is from the public-use dataset, the Longitudinal Survey of American Youth (LSAY; Miller et al., 1992)

"Manual BCH Two-step" Auxiliary Variable Integration Method

Step 1 - Estimate the unconditional model with all covariate & distal outcome variables mentioned in the auxiliary statement.

```
m_step1  <- mplusObject(
  TITLE = "Step1_bch_automation - LSAY",
  VARIABLE =
   "categorical = enjoy-adult;

    usevar = enjoy-adult;

    classes = c(5);

    !!! NOTE: All auxiliary variables to be considered in the final model should be listed here !!!
    auxiliary =
    math_irt female;",

  ANALYSIS =
   "estimator = mlr;
    type = mixture;
    starts = 500 100;",

  SAVEDATA =
   "File=bch_lsay_092021.dat;
    save=bchweights;
    format=free;
    Missflag= 9999;",

  PLOT =
   "type = plot3;
    series = enjoy-adult(*);",

  usevariables = colnames(lsay_data),
  rdata = lsay_data)

m_step1_fit <- mplusModeler(m_step1,
                 dataout=here("bch_mplus", "step1_bch.dat"),
                 modelout=here("bch_mplus", "step1_bch.inp") ,
                 check=TRUE, run = TRUE, hashfilename = FALSE)
```

Extract saved data from the step 1 unconditional model.

Extract saved data from the step 1 model `mplusObject` named "m_step1_fit"

```
savedata <- as.data.frame(m_step1_fit[["results"]]
                                      [["savedata"]])
```

Rename the column in savedata for "C" and change to "N"

```
colnames(savedata)[colnames(savedata)=="C"] <- "N"
```

Step 2 - Estimate the model with auxiliary variables using BCH weights

Example demonstrated is a moderation model with covariate & distal outcome.

Specification details:

- This example contains one distal outcome variable (`math_irt`) and one binary covariate (`female`).
- Under each class-specific statement (e.g., `%C#1%`) the distal outcome is mentioned to estimate the intercept mean (in square brackets) & variance parameters.
- Moderation is specified by mentioning the `"outcome ON covariate;"` syntax under each of the class-specific statements.
- Note that the binary covariate is centered so that reported distal means (intercepts) are estimated at the weighted average of `female`.

```
m_step2  <- mplusObject(
  TITLE = "Step2_bch_automation LSAY",

  VARIABLE =
 "usevar = BCHW1-BCHW5 math_irt female;
  missing are all (9999);
  classes = c(5);
  training = BCHW1-BCHW5(bch); ",

  DEFINE =
 "center female (grandmean);",

  ANALYSIS =
 "estimator = mlr;
  type = mixture;
  starts = 0;",

  MODEL =
 "!!! DISTAL = math_irt  !!!
  !!! MODERATOR = female !!!

  %OVERALL%

  c on female;
  math_irt on female;
  math_irt;

  %C#1%

  [math_irt] (m1);
  math_irt;
  math_irt on female (s1);

  %C#2%
```

```
 [math_irt] (m2);
 math_irt;
 math_irt on female (s2);

 %C#3%

 [math_irt] (m3);
 math_irt;
 math_irt on female (s3);

 %C#4%

 [math_irt] (m4);
 math_irt;
 math_irt on female (s4);

 %C#5%

 [math_irt] (m5);
 math_irt;
 math_irt on female (s5); ",

 MODELCONSTRAINT =
"New (diff12 diff13 diff14 diff15
      diff23 diff24 diff25
      diff34 diff35 diff45
      slope12 slope13 slope14 slope15
      slope23 slope24 slope25
      slope34 slope35 slope45);

 diff12 = m1-m2;   diff24 = m2-m4;
 diff13 = m1-m3;   diff25 = m2-m5;
 diff14 = m1-m4;   diff34 = m3-m4;
 diff15 = m1-m5;   diff35 = m3-m5;
 diff23 = m2-m3;   diff45 = m4-m5;

 slope12 = s1-s2;   slope24 = s2-s4;
 slope13 = s1-s3;   slope25 = s2-s5;
 slope14 = s1-s4;   slope34 = s3-s4;
 slope15 = s1-s5;   slope35 = s3-s5;
 slope23 = s2-s3;   slope45 = s4-s5;",

 MODELTEST =
 ## NOTE: Only a single Wald test can be conducted per model run. Therefore,
 ## this example requires running separate models for each omnibus test (e.g.,
 ## 2 models; 1 outcome and 1 slope coefficient). This can be done by
 ## commenting out all but one test and then making multiple input/output files.

"m1=m2;       !!! Distal outcome omnibus Wald test for `math_irt` !!!
 m2=m3;
 m3=m4;
 m4=m5;
```

```
  !s1=s2;      !!! Slope difference omnibus Wald test `math_irt on female` !!!
  !s2=s3;
  !s3=s4;
  !s4=s5; ",

  usevariables = colnames(savedata),
  rdata = savedata)

m_step2_fit <- mplusModeler(m_step2,
                 dataout=here("bch_mplus", "step2_bch_moderation.dat"),
                 modelout=here("bch_mplus", "step2_bch_moderation.inp"),
                 check=TRUE, run = TRUE, hashfilename = FALSE)
```

References

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. Political analysis, 12(1), 3-27.

Hallquist, Michael N., and Joshua F. Wiley. 2018. "MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus." Structural Equation Modeling, 1–18. https://doi.org/10.1080/10705511.2017.1402334.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural equation modeling: A multidisciplinary Journal, 14(4), 535-569.

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), The Oxford Handbook of Quantitative Methods: Vol. 2. Statis- tical analysis (pp. 551–611). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199934898.013.0025

R Core Team. 2019.R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham H et al., (2019). "Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. doi: 10.21105/joss.01686.

For more examples using `MplusAutomation`:

https://garberadamc.github.io/project-site/
