

DIF Primer

Josh Langfus

2024-02-05

Setup

Make sure the packages below are installed.

Read in the data

```
data_path <- "../data/test_data.csv"

dat <- read.csv(data_path)
```

Check that the data were read in successfully

```
head(dat)

##   sex gbi10d_01 gbi10d_02 gbi10d_03 gbi10d_04 gbi10d_05 gbi10d_06 gbi10d_07
## 1    0         2         3         2         1         1         2         2
## 2    0         2         3         2         3         1         1         2
## 3    1         3         0         2         3         3         3         3
## 4    0         1         0         1         3         2         0         2
## 5    0         2         2         2         3         1         1         1
## 6    0         2         1         1         1         1         0         0
##   gbi10d_08 gbi10d_09 gbi10d_10 gbi10m_01 gbi10m_02 gbi10m_03 gbi10m_04
## 1         2         1         1         2         3         3         2
## 2         2         2         2         0         0         0         0
## 3         3         0         1         3         1         3         1
## 4         3         1         2         0         2         2         0
## 5         3         1         1         1         0         1         1
## 6         0         0         0         0         0         0         0
##   gbi10m_05 gbi10m_06 gbi10m_07 gbi10m_08 gbi10m_09 gbi10m_10 raga_01 raga_02
## 1         2         3         2         2         2         3         1         1
## 2         0         0         1         0         0         2         2         0
## 3         2         2         3         3         3         1         0         0
## 4         0         2         2         0         1         1         4         4
## 5         0         0         1         0         0         1         2         2
## 6         0         0         0         0         0         0         2         2
##   raga_03 raga_04 raga_05 raga_06 raga_07 raga_08 raga_09 raga_10 raga_11
## 1         0         1         1         1         2         0         4         4         3
## 2         2         3         2         2         1         0         4         4         4
## 3         0         1         1         0         1         0         2         3         4
## 4         4         3         4         3         4         3         4         4         4
## 5         3         4         3         3         3         3         4         4         4
## 6         3         3         2         1         4         2         4         4         4
##   raga_12 raga_13 raga_14 raga_15 raga_16
## 1         2         4         2         4         2
```

## 2	4	4	4	4	4
## 3	3	2	3	4	1
## 4	4	4	4	4	4
## 5	4	4	4	4	4
## 6	4	4	4	4	4

You should see a `sex` column followed by 36 other columns that represent the item-level data.

About these data

The data represent responses from 411 parents about their children on three different scales. They also reported the child’s sex assigned at birth. The three scales are: the GBI10d (a depression measure), the GBI10m (measure of mania symptoms), and the RAGA (a measure of reactive aggression).

Except for the `sex` column, the column names tell you the name of the measure and the item number. So for example, the column titled `raga_08` contains item 8 from the RAGA; `gbi10m_03` has item 3 from the GBI10m. The two GBI scales have 10 items each; the RAGA has 16 items.

These scales have known factor structures. The two GBI scales are unidimensional – i.e., all ten items on each scale measure one “thing.” For the 10m, it’s mania symptoms and for the 10d it’s depression symptoms. The RAGA has a more complex factor structure. Although all the items measure reactive aggression to some extent, the first 8 items (`raga_01` to `raga_08`) measure “temper loss” and the latter 8 (`raga_09` to `raga_16`) measure “aggressive behavior.” The details here aren’t important for the current example; all you need to know is that there are two correlated factors.

A bit about DIF

The (very) basic goal of DIF is to find out if some of the items on your measure behave differently based on some grouping variable. Note that “behave differently” is doing a lot of work here, but I’m going to move past the details for the purposes of showing you how DIF works. To really understand what’s “behaving differently,” though, it’s important to understand how IRT works.

One challenge of doing DIF – maybe the central challenge – is that to know if *some* of the items are behaving differently, we need to have another set of items on the scale that we think are *not* behaving differently to provide a basis of comparison. These are called “anchor items.” How do you know which items aren’t DIF-y if the whole point is that we’re trying to figure that out? It’s a great question – one that has plagued the field since it’s inception.

There are two ways of looking at DIF. One way is to use a procedure called a “Mantel-Haenszel Test”, which helps establish whether there is potential DIF on any of the items. The second way is to fit an IRT model and use invariance testing. The examples below will show both.

The MH test matches respondents in both groups based on some other variable (by default, the total score on the test) and looks to see if there are significant differences in item performance in those strata. There are a lot of technical details and caveats which, again, I’ll ignore – but take a look at the Teresi et al. reading if you’re curious.

Doing the DIF - MH Test

For this example, we’re going to look at differential item function across the `sex` variable. We’ll do this in two stages. We’ll start with examining just the GBI10d items.

For simplicity, let’s split the data into three separate data frames, one for each measure, and each one also containing the `sex` column. This will make it easier to use the functions later on. Note that for the Sex variable, 0 = female and 1 = male (which were the only sexes reported in this sample). There was one person who did not respond to this question, and they have an NA value.

```
gbi10d <- dat %>%
  select(sex, starts_with("gbi10d"))

gbi10m <- dat %>%
  select(sex, starts_with("gbi10d"))

raga <- dat %>%
  select(sex, starts_with("raga"))
```

GBI 10-D MH test

In the function call below, we use `difMH()` from the `difR` package to run the MH test. As arguments, we give it the data frame with our items (and the grouping variable). Then we tell it which column contains the group variable (in this case, “sex”), and then we tell it which is the “focal” group – it doesn’t really matter, so I’ve chosen “0” at random, which is female.

```
difMH(gbi10d, "sex", "0")

##
## Detection of Differential Item Functioning using Mantel-Haenszel method
## with continuity correction and without item purification
##
## Results based on asymptotic inference
##
## Matching variable: test score
##
## No set of anchor items was provided
##
## No p-value adjustment for multiple comparisons
##
## Mantel-Haenszel Chi-square statistic:
##
##          Stat.  P-value
## gbi10d_01 0.1850 0.6672
## gbi10d_02 0.9588 0.3275
## gbi10d_03 0.3836 0.5357
## gbi10d_04 0.7255 0.3944
## gbi10d_05 1.8262 0.1766
## gbi10d_06 0.0126 0.9105
## gbi10d_07 0.0251 0.8741
## gbi10d_08 0.1925 0.6609
## gbi10d_09 0.3786 0.5383
## gbi10d_10 0.0033 0.9545
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Detection threshold: 3.8415 (significance level: 0.05)
##
## Items detected as DIF items: No DIF item detected
##
## Effect size (ETS Delta scale):
##
## Effect size code:
##  'A': negligible effect
```

```
## 'B': moderate effect
## 'C': large effect
##
##          alphaMH deltaMH
## gbi10d_01  1.2613 -0.5456 A
## gbi10d_02  1.1461 -0.3204 A
## gbi10d_03  1.4599 -0.8892 A
## gbi10d_04  1.3538 -0.7118 A
## gbi10d_05  1.2236 -0.4743 A
## gbi10d_06  1.0347 -0.0802 A
## gbi10d_07  1.1874 -0.4037 A
## gbi10d_08  1.2851 -0.5895 A
## gbi10d_09  1.0211 -0.0492 A
## gbi10d_10  0.7946  0.5402 A
##
## Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
## (for absolute values of 'deltaMH')
##
## Output was not captured!
```

The GBI 10d results suggest no DIF detected. Woohoo!

RAGA MH Test

Let's try with the RAGA

```
difMH(raga, "sex", "0", purify = TRUE)
```

```
##
## Detection of Differential Item Functioning using Mantel-Haenszel method
## with continuity correction and with item purification
##
## Results based on asymptotic inference
##
## Convergence reached after 3 iterations
##
## Matching variable: test score
##
## No set of anchor items was provided
##
## No p-value adjustment for multiple comparisons
##
## Mantel-Haenszel Chi-square statistic:
##
##          Stat.  P-value
## raga_01 0.0626 0.8024
## raga_02 0.0147 0.9034
## raga_03 1.2921 0.2557
## raga_04 3.6076 0.0575 .
## raga_05 0.0958 0.7569
## raga_06 0.2893 0.5907
## raga_07 0.0088 0.9252
## raga_08 0.0163 0.8983
## raga_09 0.0009 0.9757
## raga_10 0.0001 0.9921
```

```

## raga_11 0.3059 0.5802
## raga_12 0.0013 0.9711
## raga_13 0.6273 0.4283
## raga_14 8.3167 0.0039 **
## raga_15 0.0088 0.9251
## raga_16 2.0475 0.1525
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Detection threshold: 3.8415 (significance level: 0.05)
##
## Items detected as DIF items:
##
## raga_14
##
##
## Effect size (ETS Delta scale):
##
## Effect size code:
## 'A': negligible effect
## 'B': moderate effect
## 'C': large effect
##
##      alphaMH deltaMH
## raga_01  1.4432 -0.8620 A
## raga_02  1.7248 -1.2810 B
## raga_03  1.5551 -1.0376 B
## raga_04  1.2357 -0.4974 A
## raga_05  1.0033 -0.0078 A
## raga_06  1.0458 -0.1051 A
## raga_07  1.4569 -0.8843 A
## raga_08  1.2723 -0.5659 A
## raga_09  1.4061 -0.8010 A
## raga_10  0.9865  0.0320 A
## raga_11  0.8282  0.4429 A
## raga_12  0.9464  0.1295 A
## raga_13  0.9632  0.0880 A
## raga_14  2.2774 -1.9342 C
## raga_15  1.1214 -0.2693 A
## raga_16  1.6432 -1.1672 B
##
## Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
## (for absolute values of 'deltaMH')
##
## Output was not captured!

```

It looks like we might have DIF on 3 items. Let's explore that further using mirt.

```

## set the type of invariance we'll test
invariance <- c("slopes", "intercepts")

# Fit the multigroup model with sex as the grouping variable
fit1 <- multipleGroup(
  data = raga[2:17], ## need to cut out the sex column and give only items
  model = 1, # this is a shortcut for a 1-dimensional model

```

```

group = as.factor(raga$sex), # note, group has to be factor or character type
itemtype = "graded", # graded response model is appropriate for ordinal items
invariance = invariance,
verbose = TRUE
)

## NA values in group removed, along with associated rows in data

## Warning: data contains response patterns with only NAs

## Iteration: 1, Log-Lik: -6971.029, Max-Change: 2.83771Iteration: 2, Log-Lik: -6524.180, Max-Change: 1

## Uncomment this line to see the item loadings
# summary(fit1)

## Check for uniform DIF - just in the difficulty parameters
uniform <- DIF(fit1,
  which.par = c('d1','d2','d3','d4'),
  scheme = 'drop'
  # items2test = c("raga_02","raga_03","raga_14")
)

## NOTE: No hyper-parameters were estimated in the DIF model.
## For effective DIF testing, freeing the focal group hyper-parameters is recommended.

uniform

##      groups converged      AIC  SABIC      HQ      BIC      X2 df      p
## raga_01      0,1      TRUE   0.902   4.205   7.234 16.897   7.098 4 0.131
## raga_02      0,1      TRUE -17.420 -14.116 -11.087 -1.424 25.42 4 0
## raga_03      0,1      TRUE  -0.162   3.141   6.170 15.833   8.162 4 0.086
## raga_04      0,1      TRUE   3.320   6.624   9.653 19.316   4.68 4 0.322
## raga_05      0,1      TRUE   2.873   6.176   9.205 18.868   5.127 4 0.274
## raga_06      0,1      TRUE   3.018   6.322   9.351 19.014   4.982 4 0.289
## raga_07      0,1      TRUE   3.118   6.421   9.451 19.114   4.882 4 0.3
## raga_08      0,1      TRUE  -3.092   0.212   3.241 12.904 11.092 4 0.026
## raga_09      0,1      TRUE   7.003 10.306 13.335 22.998   0.997 4 0.91
## raga_10      0,1      TRUE  -3.747  -0.444   2.586 12.249 11.747 4 0.019
## raga_11      0,1      TRUE   5.174   8.477 11.506 21.169   2.826 4 0.587
## raga_12      0,1      TRUE -17.200 -13.896 -10.867 -1.204 25.2 4 0
## raga_13      0,1      TRUE   5.534   8.837 11.866 21.529   2.466 4 0.651
## raga_14      0,1      TRUE  -3.623  -0.320   2.710 12.373 11.623 4 0.02
## raga_15      0,1      TRUE  -2.282   1.021   4.050 13.713 10.282 4 0.036
## raga_16      0,1      TRUE  -0.212   3.092   6.121 15.784   8.212 4 0.084

nonuniform <- DIF(fit1,
  which.par = c('a1','d1','d2','d3','d4'),
  scheme = 'drop'
  # items2test = c("raga_02","raga_03","raga_14")
)

## NOTE: No hyper-parameters were estimated in the DIF model.
## For effective DIF testing, freeing the focal group hyper-parameters is recommended.

nonuniform

##      groups converged      AIC  SABIC      HQ      BIC      X2 df      p
## raga_01      0,1      TRUE   1.903   6.032   9.819 21.898   8.097 5 0.151
## raga_02      0,1      TRUE -16.118 -11.988 -8.202   3.877 26.118 5 0

```

```
## raga_03    0,1      TRUE    1.835    5.964    9.750    21.829    8.165    5 0.147
## raga_04    0,1      TRUE    5.084    9.213    13.000    25.079    4.916    5 0.426
## raga_05    0,1      TRUE    4.692    8.821    12.608    24.686    5.308    5 0.379
## raga_06    0,1      TRUE    3.802    7.931    11.718    23.797    6.198    5 0.287
## raga_07    0,1      TRUE    2.521    6.650    10.437    22.515    7.479    5 0.187
## raga_08    0,1      TRUE   -3.651    0.478    4.265    16.344    13.651    5 0.018
## raga_09    0,1      TRUE    8.830    12.959    16.745    28.824     1.17    5 0.948
## raga_10    0,1      TRUE   -1.751    2.378    6.165    18.244    11.751    5 0.038
## raga_11    0,1      TRUE    6.262    10.392    14.178    26.257    3.738    5 0.588
## raga_12    0,1      TRUE  -15.237  -11.108   -7.322    4.757    25.237    5 0
## raga_13    0,1      TRUE    5.558    9.687    13.474    25.553    4.442    5 0.488
## raga_14    0,1      TRUE   -1.650    2.479    6.266    18.345    11.65    5 0.04
## raga_15    0,1      TRUE   -1.576    2.553    6.340    18.419    11.576    5 0.041
## raga_16    0,1      TRUE    1.524    5.653    9.440    21.518    8.476    5 0.132
```

```
## with schema='drop', items with large p-values have DIF.
```

```
## fit a model allowing items 2,3, and 14 to vary across groups
```

```
fit2 <- multipleGroup(
  data = raga[2:17], ## need to cut out the sex column and give only items
  model = 1, # this is a shortcut for a 1-dimensional model
  group = as.factor(raga$sex), # note, group has to be factor or character type
  itemtype = "graded", # graded response model is appropriate for ordinal items
  ## Below, specify the names of items you want to KEEP INVARIANT
  ## I did this with a shorthand by referencing the column names, but you can just
  ## write out the names of the items in a vector, e.g. c("raga_01","raga_02")
  invariance = colnames(raga[2:17])[c(1,4:13,15,16)],
  verbose = TRUE
)
```

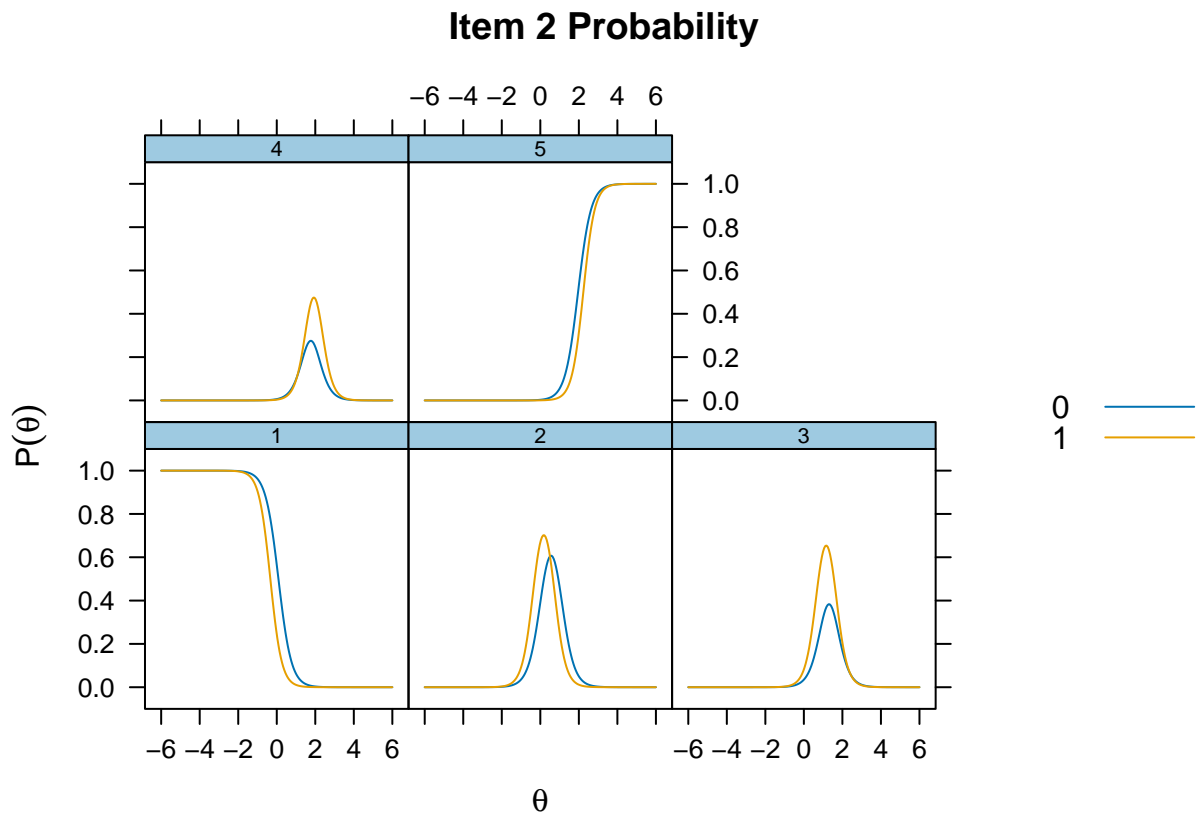
```
## NA values in group removed, along with associated rows in data
```

```
## Warning: data contains response patterns with only NAs
```

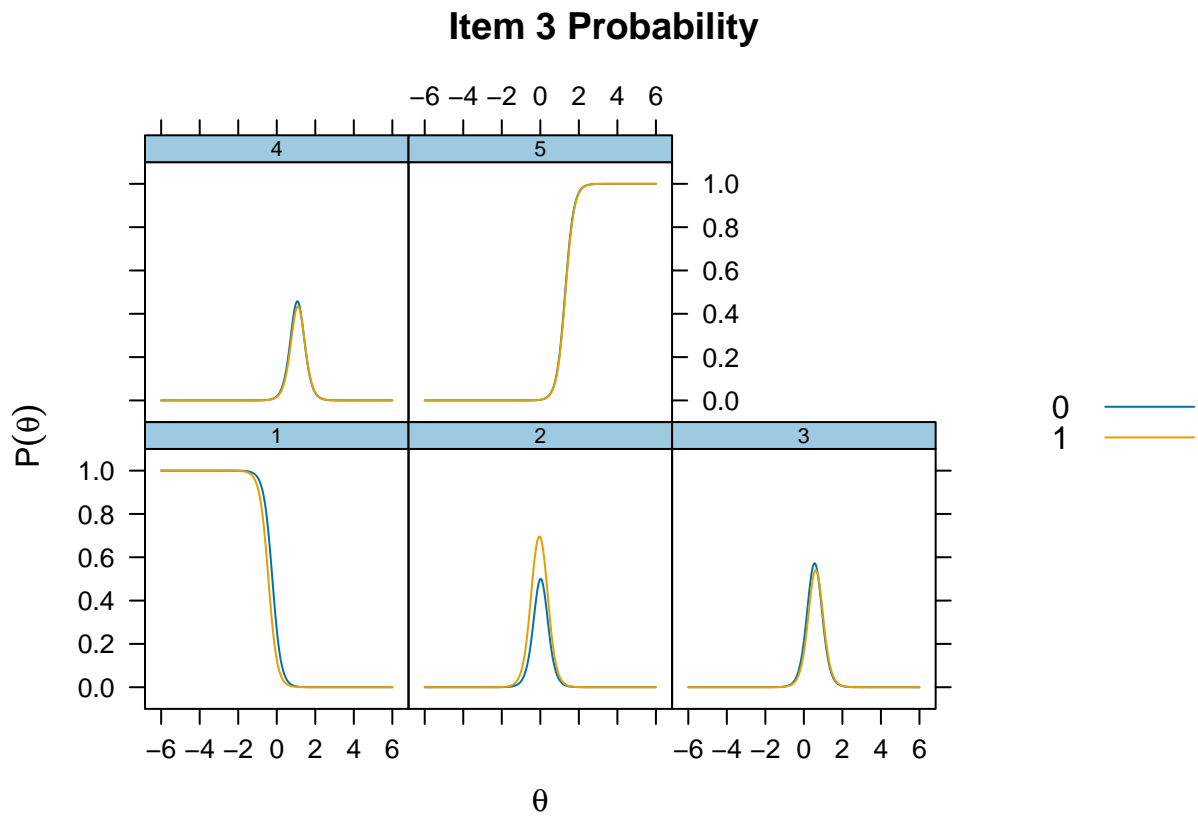
```
## Iteration: 1, Log-Lik: -6971.029, Max-Change: 3.78605Iteration: 2, Log-Lik: -6513.920, Max-Change: 1
```

The following plots show the DIF in the item traces

```
itemplot(fit2, item = 2)
```

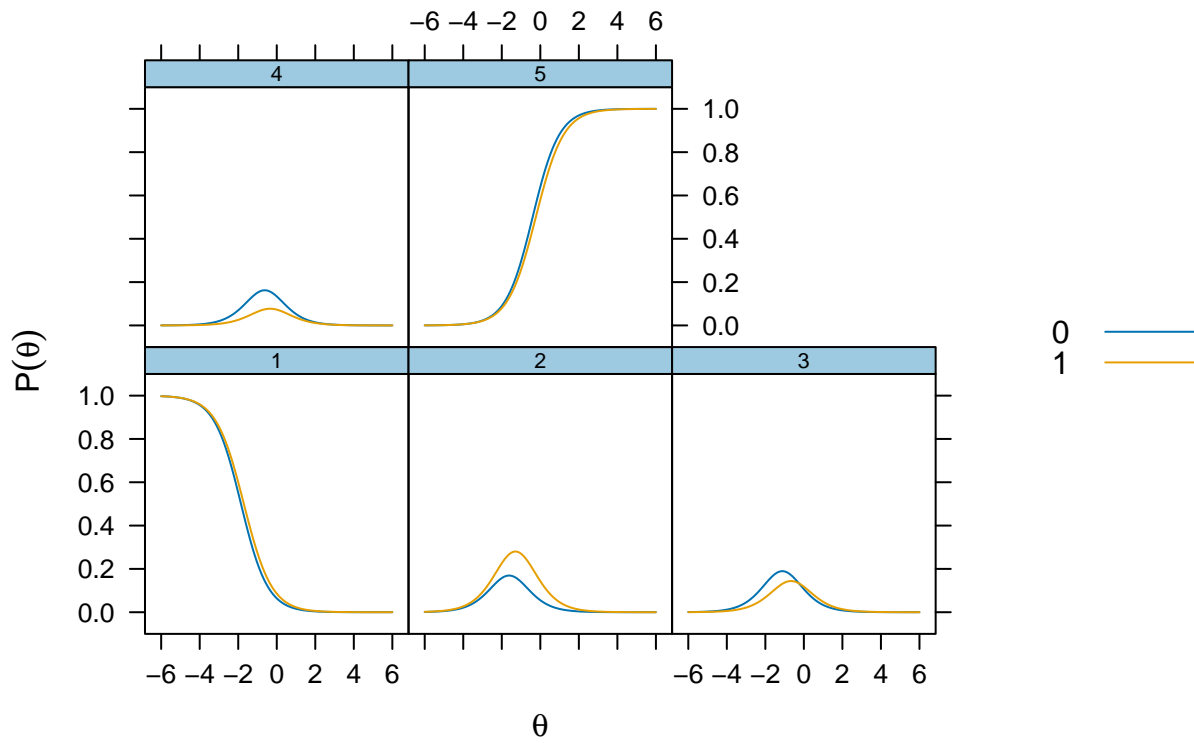


```
itemplot(fit2, item = 3)
```

```
itemplot(fit2, item = 14)
```

Item 14 Probability



Note that if we have anchor items, we can specify them in both the MH test and the mirt function above. For MH, we set them with the `anchor.items` argument; with the mirt DIF function, supply `items2test` with the items you think are NOT invariant.