2PL Demo

Adam Kurth

2024-01-19

Understanding MIRT

2PL Model Simulate data using Monte Carlo

```
# person trait level, e.g., math skill
theta <- rnorm(n = 1000, mean = 0, sd = 1)

# Item parameters (a, b, c) for 4 items
a <- c(0.5, 1, 1.5, 2)
b <- c(0, -1, 1, 0)
c <- c(0.2, 0.15, 0.25, 0.2)

n.persons <- length(theta)
n.items <- length(a)
response.data <- matrix(NA, n.persons, n.items)

for (i in 1:n.persons) {
    for (j in 1:n.items) {
        p <- c[j] + (1 - c[j]) / (1 + exp(-(a[j] * theta[i] - b[j]))) #2PL model
        u <- runif(n = 1, min = 0, max = 1)
        if (u < p) {
            response.data[i, j] <- 1
        } else {
            response.data[i, j] <- 0
        }
    }
}
colnames(response.data) <- c("I1", "I2", "I3", "I4")
```

- a, b, c: item parameters.
- θ, `theta`: personal trait level.
- p: probability of positive response (gets answer right)
- u: random number between 0 and 1, (uniformly distributed)
- if $u < p$ then `response.data = 1`: individual gets answer correct.
- if $u > p$ then `response.data = 0`: individual gets answer wrong.

```
start_time = Sys.time() #start time
mirt.out <- mirt::mirt(data = response.data, model = 1, itemtype = "2PL", storeEMhistory=TRUE) # 2PL mo
```

```
## Iteration: 1, Log-Lik: -2596.883, Max-Change: 0.07289Iteration: 2, Log-Lik: -2595.099, Max-Change: 0
```

```r
end_time = Sys.time() # end time
time_to_conv = end_time - start_time
time_to_conv
```

```
## Time difference of 0.1188409 secs
```

- mirt.out <- mirt::mirt() function fits a maximum likelihood (posterier) factor analysis model to the data (dichotomus/polytomous)

Output of `mirt.out`:

- 'M-step' = indicates that the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm used for optimization in M-step if the EM algorithm.
- 'EM acceleration' = indicates that ramsay's acceleration method is used to speed up the convergence of the EM algorithm.
- 'Number of rectangular quadrature' = the number of quadrature points used in the numerical integration of the likelihood function.
- 'Latent density type' = the distribution of the latent trait (F1 in this case)
- 'log-likelihood' = how well the model fits the data
- 'Estimated parameters' = number of estimated parameters
- 'AIC' = Akaike information criterion (measure of model fit, takes into number of parameters)
- BIC/SABIC = Bayesian information criterion/Sample-size Adjusted BIC (measure of model fit, takes into number of parameters)
- `G2 (3) = 1.1, p = 0.7772` : results of the likelihood ratio test comparing the fitted model against saturated model
- `RMSEA = 0, CFI = NaN, TLI = NaN`: additional measures of model fit, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI)