Categorical data with OpenMx

Mike Cheung

27 Apr 2019

Contents

Generate some data 1

A CFA with 6 items (information matrix is not positive definite) 2

A CFA with 5 items (information matrix is not positive definite) 4

A CFA with 4 items (information matrix is okay) 7

A CFA with 3 items (information matrix is okay) 9

Generate some data

```r
library(OpenMx)
```

```
## To take full advantage of multiple cores, use:
##   mxOption(NULL, 'Number of Threads', parallel::detectCores()) #now
##   Sys.setenv(OMP_NUM_THREADS=parallel::detectCores()) #before library(OpenMx)
```

```r
mxOption(NULL,'Number of Threads', parallel::detectCores()-4)

set.seed(109030)

N <- 500

## Factor loadings
lambda <- matrix(c(.4,.5,.6,.4,.5,.6), ncol=1)

## Model implied covariance matrix
sigma <- lambda %*% t(lambda)
diag(sigma) <- 1
sigma
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,] 1.00 0.20 0.24 0.16 0.20 0.24
## [2,] 0.20 1.00 0.30 0.20 0.25 0.30
## [3,] 0.24 0.30 1.00 0.24 0.30 0.36
## [4,] 0.16 0.20 0.24 1.00 0.20 0.24
## [5,] 0.20 0.25 0.30 0.20 1.00 0.30
## [6,] 0.24 0.30 0.36 0.24 0.30 1.00
```

```r
## Generate multivariate normal data
my.df <- mvtnorm::rmvnorm(N, sigma=sigma)

## Categorize into 0 and 1
my.df <- ifelse(my.df > 0 , yes=1, no=0)
```

```r
colnames(my.df) <- paste0("x", seq_len(nrow(lambda)))
my.df <- data.frame(my.df)

head(my.df)
```

```
##   x1 x2 x3 x4 x5 x6
## 1  1  1  1  1  1  1
## 2  0  0  1  1  0  0
## 3  0  1  0  0  0  0
## 4  0  0  1  0  0  0
## 5  1  0  0  0  0  0
## 6  0  0  0  1  0  0
```

```r
apply(my.df, 2, mean)
```

```
##    x1    x2    x3    x4    x5    x6
## 0.492 0.468 0.520 0.482 0.504 0.510
```

```r
## Convert them into mxFactor
for (i in seq_len(ncol(my.df)) ) {
  my.df[, i] <- mxFactor(my.df[, i], levels=c(0,1))
}
```

A CFA with 6 items (information matrix is not positive definite)

```r
dataRaw <- mxData( observed=my.df, type="raw" )

# residual variances
resVars <- mxPath( from=c("x1","x2","x3","x4","x5","x6"), arrows=2,
                   free=FALSE, values=c(1,1,1,1,1,1),
                   labels=c("e1","e2","e3","e4","e5","e6") )
# latent variance
latVar <- mxPath( from="F1", arrows=2,
                  free=FALSE, values=1, labels ="varF1" )
# factor loadings
facLoads <- mxPath( from="F1", to=c("x1","x2","x3","x4","x5","x6"), arrows=1,
                    free=c(TRUE,TRUE,TRUE,TRUE,TRUE,TRUE),
                    values=c(.5,.5,.5,.5,.5,.5), labels=c("l1","l2","l3","l4","l5","l6") )
# means
means <- mxPath( from="one", to=c("x1","x2","x3","x4","x5","x6","F1"), arrows=1,
                 free=FALSE, values=0,
                 labels=c("meanx1","meanx2","meanx3","meanx4","meanx5","meanx6","meanF") )
# thresholds
thresholds <- mxThreshold(vars=c("x1","x2","x3","x4","x5","x6"), nThresh=c(1,1,1,1,1,1),
                          free=TRUE, values=c(0,0,0,0,0,0),
                          labels=c("var1_th1","var2_th1","var3_th1","var4_th1","var5_th1","var6_th1") )

mxmodel <- mxModel("A CFA with 5 items", type="RAM",
                   manifestVars=c("x1","x2","x3","x4","x5","x6"), latentVars="F1",
                   dataRaw, resVars, latVar, facLoads, means, thresholds)

mxFit <- mxTryHardOrdinal(mxmodel)
```

```
##
## Beginning initial fit attempt
##
## Beginning fit attempt 1 of at maximum 10 extra tries
##
##  Lowest minimum so far:  4019.56783020441
##
## Beginning fit attempt 2 of at maximum 10 extra tries
##
##  Lowest minimum so far:  4019.56783020438
##
## Beginning fit attempt 3 of at maximum 10 extra tries
##
## Beginning fit attempt 4 of at maximum 10 extra tries
##
## Beginning fit attempt 5 of at maximum 10 extra tries
##
## Beginning fit attempt 6 of at maximum 10 extra tries
##
## Beginning fit attempt 7 of at maximum 10 extra tries
##
## Beginning fit attempt 8 of at maximum 10 extra tries
##
## Beginning fit attempt 9 of at maximum 10 extra tries
##
## Beginning fit attempt 10 of at maximum 10 extra tries
##
## Retry limit reached
##
## Solution found
## Final run, for Hessian and/or standard errors and/or confidence intervals
## Warning messages generated from final run for Hessian/SEs/CIs
##
## Solution found!  Final fit=4019.5678 (started at 4033.6632)  (11 attempt(s): 11 valid, 0 errors)
##
## Start values from best fit:
## 0.460246136453502,0.588964802462443,0.762590608462285,0.381922080280775,0.567983568683796,0.63900355
```

```r
summary(mxFit)
```

```
## Summary of A CFA with 5 items
##
## free parameters:
##        name     matrix row col    Estimate  Std.Error A
## 1        l1          A  x1  F1  0.46024614 0.09981399
```

```
## 2        l2          A  x2  F1  0.58896480 0.10597864
## 3        l3          A  x3  F1  0.76259061         NA
## 4        l4          A  x4  F1  0.38192208 0.08614565
## 5        l5          A  x5  F1  0.56798357 0.12085151
## 6        l6          A  x6  F1  0.63900356 0.12409376
## 7  var1_th1 Thresholds   1  x1  0.02154671 0.03093973
## 8  var2_th1 Thresholds   1  x2  0.09466997 0.06287470
## 9  var3_th1 Thresholds   1  x3 -0.06207406         NA !
## 10 var4_th1 Thresholds   1  x4  0.04864538         NA !
## 11 var5_th1 Thresholds   1  x5 -0.01101681 0.02639954
## 12 var6_th1 Thresholds   1  x6 -0.03019225 0.06815998
##
## Model Statistics:
##               |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
##        Model:             12                   2988              4019.568
##    Saturated:             27                   2973                    NA
## Independence:             12                   2988                    NA
## Number of observations/statistics: 500/3000
##
##
## ** Information matrix is not positive definite (not at a candidate optimum).
##   Be suspicious of these results. At minimum, do not trust the standard errors.
##
## Information Criteria:
##       |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:      -1956.432               4043.568                 4044.208
## BIC:     -14549.681               4094.143                 4056.054
## To get additional fit indices, see help(mxRefModels)
## timestamp: 2019-04-27 18:18:55
## Wall clock time: 2.040484 secs
## optimizer:  CSOLNP
## OpenMx version number: 2.12.2
## Need help?  See help(mxSummary)
```

A CFA with 5 items (information matrix is not positive definite)

```
## Drop the last item
dataRaw <- mxData( observed=my.df[, -6], type="raw" )

# residual variances
resVars <- mxPath( from=c("x1","x2","x3","x4","x5"), arrows=2,
                   free=FALSE, values=c(1,1,1,1,1),
                   labels=c("e1","e2","e3","e4","e5") )
# latent variance
latVar <- mxPath( from="F1", arrows=2,
                  free=FALSE, values=1, labels ="varF1" )
# factor loadings
facLoads <- mxPath( from="F1", to=c("x1","x2","x3","x4","x5"), arrows=1,
                    free=c(TRUE,TRUE,TRUE,TRUE,TRUE),
                    values=c(.5,.5,.5,.5,.5), labels=c("l1","l2","l3","l4","l5") )
# means
means <- mxPath( from="one", to=c("x1","x2","x3","x4","x5","F1"), arrows=1,
```

```r
                 free=FALSE, values=0,
                 labels=c("meanx1","meanx2","meanx3","meanx4","meanx5","meanF") )
# thresholds
thresholds <- mxThreshold(vars=c("x1","x2","x3","x4","x5"), nThresh=c(1,1,1,1,1),
                          free=TRUE, values=c(0,0,0,0,0),
                          labels=c("var1_th1","var2_th1","var3_th1","var4_th1","var5_th1") )

mxmodel <- mxModel("A CFA with 5 items", type="RAM",
                   manifestVars=c("x1","x2","x3","x4","x5"), latentVars="F1",
                   dataRaw, resVars, latVar, facLoads, means, thresholds)

mxFit <- mxTryHardOrdinal(mxmodel)
```

```
##
## Beginning initial fit attempt
##
## Beginning fit attempt 1 of at maximum 10 extra tries
##
##  Lowest minimum so far:  3372.33493831405
##
## Beginning fit attempt 2 of at maximum 10 extra tries
##
##  Lowest minimum so far:  3372.334938314
##
## Beginning fit attempt 3 of at maximum 10 extra tries
##
## Beginning fit attempt 4 of at maximum 10 extra tries
##
## Beginning fit attempt 5 of at maximum 10 extra tries
##
##  Lowest minimum so far:  3372.33452876697
##
## Beginning fit attempt 6 of at maximum 10 extra tries
##
##  Fit attempt worse than current best:  3372.33493831408 vs 3372.33452876697
##
## Beginning fit attempt 7 of at maximum 10 extra tries
##
##  Fit attempt worse than current best:  3372.33475860832 vs 3372.33452876697
##
## Beginning fit attempt 8 of at maximum 10 extra tries
##
##  Fit attempt worse than current best:  3372.33493831401 vs 3372.33452876697
##
## Beginning fit attempt 9 of at maximum 10 extra tries
```

```
##
##  Lowest minimum so far:  3372.33415011019
##
## Beginning fit attempt 10 of at maximum 10 extra tries
##
##  Fit attempt worse than current best:  3372.3348781707 vs 3372.33415011019
##
## Retry limit reached
##
## Solution found
## Final run, for Hessian and/or standard errors and/or confidence intervals
## Warning messages generated from final run for Hessian/SEs/CIs
##
## Solution found!  Final fit=3372.3342 (started at 3383.2652)  (11 attempt(s): 11 valid, 0 errors)
##
## Start values from best fit:
## 0.483284837694139,0.507297682740633,0.846951255931905,0.397116456113947,0.546392742796154,0.02217159
```

```r
summary(mxFit)
```

```
## Summary of A CFA with 5 items
##
## free parameters:
##        name     matrix row col    Estimate   Std.Error A
## 1        l1          A  x1  F1  0.48328484 0.035122722
## 2        l2          A  x2  F1  0.50729768          NA
## 3        l3          A  x3  F1  0.84695126 0.082156177
## 4        l4          A  x4  F1  0.39711646 0.060050648
## 5        l5          A  x5  F1  0.54639274 0.057927549
## 6  var1_th1 Thresholds   1  x1  0.02217159 0.051384164
## 7  var2_th1 Thresholds   1  x2  0.09157961 0.009516114
## 8  var3_th1 Thresholds   1  x3 -0.06418268          NA !
## 9  var4_th1 Thresholds   1  x4  0.04881181          NA !
## 10 var5_th1 Thresholds   1  x5 -0.01020358 0.039806891
##
## Model Statistics:
##                |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
##        Model:             10                   2490              3372.334
##    Saturated:             20                   2480                    NA
## Independence:             10                   2490                    NA
## Number of observations/statistics: 500/2500
##
##
## ** Information matrix is not positive definite (not at a candidate optimum).
##   Be suspicious of these results. At minimum, do not trust the standard errors.
##
## Information Criteria:
##       |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:      -1607.666               3392.334                 3392.784
## BIC:     -12102.040               3434.480                 3402.740
```

```
## To get additional fit indices, see help(mxRefModels)
## timestamp: 2019-04-27 18:18:59
## Wall clock time: 0.9087341 secs
## optimizer:  CSOLNP
## OpenMx version number: 2.12.2
## Need help?  See help(mxSummary)
```

A CFA with 4 items (information matrix is okay)

```
## Drop the last two items
dataRaw <- mxData( observed=my.df[, -c(5,6)], type="raw" )

# residual variances
resVars <- mxPath( from=c("x1","x2","x3","x4"), arrows=2,
                   free=FALSE, values=c(1,1,1,1),
                   labels=c("e1","e2","e3","e4") )
# latent variance
latVar <- mxPath( from="F1", arrows=2,
                  free=FALSE, values=1, labels ="varF1" )
# factor loadings
facLoads <- mxPath( from="F1", to=c("x1","x2","x3","x4"), arrows=1,
                    free=c(TRUE,TRUE,TRUE,TRUE),
                    values=c(.5,.5,.5,.5), labels=c("l1","l2","l3","l4") )
# means
means <- mxPath( from="one", to=c("x1","x2","x3","x4","F1"), arrows=1,
                 free=FALSE, values=0,
                 labels=c("meanx1","meanx2","meanx3","meanx4","meanF") )
# thresholds
thresholds <- mxThreshold(vars=c("x1","x2","x3","x4"), nThresh=c(1,1,1,1),
                          free=TRUE, values=c(0,0,0,0),
                          labels=c("var1_th1","var2_th1","var3_th1","var4_th1") )

mxmodel <- mxModel("A CFA with 4 items", type="RAM",
                   manifestVars=c("x1","x2","x3","x4"), latentVars="F1",
                   dataRaw, resVars, latVar, facLoads, means, thresholds)

mxFit <- mxTryHardOrdinal(mxmodel)
```

```
##
## Beginning initial fit attempt

##
## Beginning fit attempt 1 of at maximum 10 extra tries

##
##  Lowest minimum so far:  2710.67499489066

##
## Beginning fit attempt 2 of at maximum 10 extra tries

##
##  Lowest minimum so far:  2710.67499489044

##
## Beginning fit attempt 3 of at maximum 10 extra tries
```

```
##
##  Lowest minimum so far:  2710.67499489037
##
## Beginning fit attempt 4 of at maximum 10 extra tries
##
##  Lowest minimum so far:  2710.67499489035
##
## Beginning fit attempt 5 of at maximum 10 extra tries
##
## Beginning fit attempt 6 of at maximum 10 extra tries
##
## Beginning fit attempt 7 of at maximum 10 extra tries
##
## Beginning fit attempt 8 of at maximum 10 extra tries
##
## Beginning fit attempt 9 of at maximum 10 extra tries
##
##  Lowest minimum so far:  2710.6749948903
##
## Beginning fit attempt 10 of at maximum 10 extra tries
##
## Retry limit reached
##
## Solution found
## Final run, for Hessian and/or standard errors and/or confidence intervals
##
## Solution found!  Final fit=2710.675 (started at 2721.2996)  (11 attempt(s): 11 valid, 0 errors)
##
## Start values from best fit:
## 0.485100903008894,0.396523139794714,1.03180807737273,0.414974449943737,0.0218463093013718,0.08653693
```

```r
summary(mxFit)
```

```
## Summary of A CFA with 4 items
##
## free parameters:
##       name     matrix row col    Estimate  Std.Error A
## 1       l1          A  x1  F1  0.48510090 0.12375783
## 2       l2          A  x2  F1  0.39652314 0.10722393
## 3       l3          A  x3  F1  1.03180808 0.35368056
## 4       l4          A  x4  F1  0.41497445 0.10896717
## 5 var1_th1 Thresholds   1  x1  0.02184631 0.06423925
## 6 var2_th1 Thresholds   1  x2  0.08653693 0.06180739
## 7 var3_th1 Thresholds   1  x3 -0.07225153 0.10967451
## 8 var4_th1 Thresholds   1  x4  0.04904185 0.08199430
##
## Model Statistics:
```

```
##                 |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
##        Model:              8                   1992              2710.675
##    Saturated:             14                   1986                    NA
## Independence:              8                   1992                    NA
## Number of observations/statistics: 500/2000
##
## Information Criteria:
##       |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:      -1273.325               2726.675                 2726.968
## BIC:      -9668.824               2760.392                 2734.999
## To get additional fit indices, see help(mxRefModels)
## timestamp: 2019-04-27 18:19:01
## Wall clock time: 0.1876481 secs
## optimizer:  CSOLNP
## OpenMx version number: 2.12.2
## Need help?  See help(mxSummary)
```

A CFA with 3 items (information matrix is okay)

```r
## Drop the last three items
dataRaw <- mxData( observed=my.df[, -c(4,5,6)], type="raw" )

# residual variances
resVars <- mxPath( from=c("x1","x2","x3"), arrows=2,
                   free=FALSE, values=c(1,1,1),
                   labels=c("e1","e2","e3") )
# latent variance
latVar <- mxPath( from="F1", arrows=2,
                  free=FALSE, values=1, labels ="varF1" )
# factor loadings
facLoads <- mxPath( from="F1", to=c("x1","x2","x3"), arrows=1,
                    free=c(TRUE,TRUE,TRUE),
                    values=c(.5,.5,.5), labels=c("l1","l2","l3") )
# means
means <- mxPath( from="one", to=c("x1","x2","x3","F1"), arrows=1,
                 free=FALSE, values=0,
                 labels=c("meanx1","meanx2","meanx3","meanF") )
# thresholds
thresholds <- mxThreshold(vars=c("x1","x2","x3"), nThresh=c(1,1,1),
                          free=TRUE, values=c(0,0,0),
                          labels=c("var1_th1","var2_th1","var3_th1") )

mxmodel <- mxModel("A CFA with 3 items", type="RAM",
                   manifestVars=c("x1","x2","x3"), latentVars="F1",
                   dataRaw, resVars, latVar, facLoads, means, thresholds)

mxFit <- mxTryHardOrdinal(mxmodel)
```

```
##
## Beginning initial fit attempt
```

```
##
## Beginning fit attempt 1 of at maximum 10 extra tries
```

```
##
##  Lowest minimum so far:  2038.23547946313
##
## Beginning fit attempt 2 of at maximum 10 extra tries
##
##  Lowest minimum so far:  2038.23547946311
##
## Beginning fit attempt 3 of at maximum 10 extra tries
##
##  Lowest minimum so far:  2038.23547946311
##
## Beginning fit attempt 4 of at maximum 10 extra tries
##
## Beginning fit attempt 5 of at maximum 10 extra tries
##
##  Lowest minimum so far:  2038.2354794631
##
## Beginning fit attempt 6 of at maximum 10 extra tries
##
##  Lowest minimum so far:  2038.2354794631
##
## Beginning fit attempt 7 of at maximum 10 extra tries
##
## Beginning fit attempt 8 of at maximum 10 extra tries
##
## Beginning fit attempt 9 of at maximum 10 extra tries
##
## Beginning fit attempt 10 of at maximum 10 extra tries
##
## Retry limit reached
##
## Solution found
## Final run, for Hessian and/or standard errors and/or confidence intervals
##
## Solution found!  Final fit=2038.2355 (started at 2045.4743)  (11 attempt(s): 11 valid, 0 errors)
##
## Start values from best fit:
## 0.505006483397195,0.419048594560313,0.952198217509894,0.0221994438803565,0.0868869279436971,-0.06997
```

```r
summary(mxFit)
```

```
## Summary of A CFA with 3 items
##
## free parameters:
##       name     matrix row col    Estimate  Std.Error A
```

```
## 1       l1          A  x1  F1  0.50500648 0.15621983
## 2       l2          A  x2  F1  0.41904859 0.12696523
## 3       l3          A  x3  F1  0.95219822 0.41547215
## 4 var1_th1 Thresholds   1  x1  0.02219944 0.06279637
## 5 var2_th1 Thresholds   1  x2  0.08688693 0.06094178
## 6 var3_th1 Thresholds   1  x3 -0.06997877 0.07877013
##
## Model Statistics:
##               |  Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
##        Model:             6                   1494             2038.235
##    Saturated:             9                   1491                   NA
## Independence:             6                   1494                   NA
## Number of observations/statistics: 500/1500
##
## Information Criteria:
##       |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:      -949.7645               2050.235                 2050.406
## BIC:     -7246.3890               2075.523                 2056.479
## To get additional fit indices, see help(mxRefModels)
## timestamp: 2019-04-27 18:19:01
## Wall clock time: 0.07354212 secs
## optimizer:  CSOLNP
## OpenMx version number: 2.12.2
## Need help?  See help(mxSummary)
```

```r
sessionInfo()
```

```
## R version 3.5.3 (2019-03-11)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Linux Mint 19.1
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/libopenblasp-r0.2.20.so
##
## locale:
##  [1] LC_CTYPE=en_SG.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_SG.UTF-8        LC_COLLATE=en_SG.UTF-8
##  [5] LC_MONETARY=en_SG.UTF-8    LC_MESSAGES=en_SG.UTF-8
##  [7] LC_PAPER=en_SG.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_SG.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] OpenMx_2.12.2
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.1      mvtnorm_1.0-10  lattice_0.20-38 digest_0.6.18
##  [5] MASS_7.3-51.3   grid_3.5.3      magrittr_1.5    evaluate_0.13
##  [9] stringi_1.4.3   Matrix_1.2-17   rmarkdown_1.12  tools_3.5.3
## [13] stringr_1.4.0   xfun_0.6        yaml_2.2.0      parallel_3.5.3
## [17] compiler_3.5.3  htmltools_0.3.6 knitr_1.22
```