PLSC 597: Modern Measurement

Miscellaneous Topics

March 1, 2018

Various Topics...

- IRT Variants
- Measurement Via Random Forests
- Confirmatory Factor Analysis
  (`## Not run`)

IRT: Ordinal Responses

Items are K-category ordinal responses $Y_{ij} \in \{1, ...K\}$.

Samejima's (1969) "Graded Response Model" ("GRM")

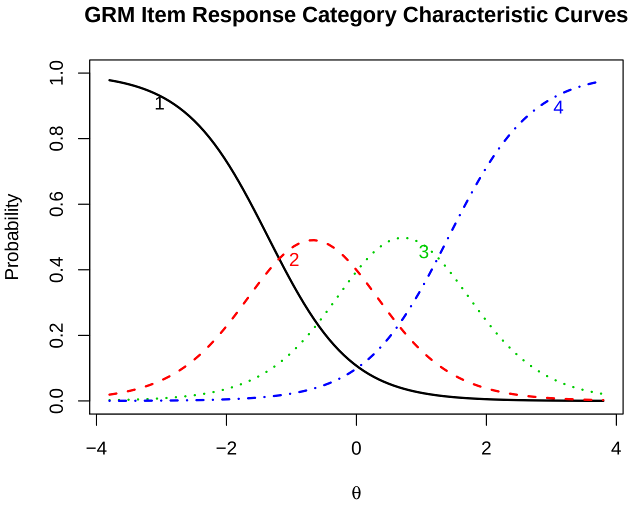
The cumulative probability is:

$$\Pr(Y_{ij} > k | \alpha_j, \beta_{jk}, \theta_i) \equiv \Lambda_{jk}(\theta) = \frac{\exp[\alpha_j(\theta_i - \beta_{jk})]}{1 + \exp[\alpha_j(\theta_i - \beta_{jk})]}$$

Response probabilities:

$$\Pr(Y_{ij} = k | \alpha_j, \beta_{jk}, \theta_i) = \begin{cases} 1 - \Lambda_{j1}(\theta) \text{ for } k = 1 \\ \Lambda_{jk}(\theta) - \Lambda_{jk+1}(\theta) \text{ for } k \in \{2 ... K - 1\} \\ \Lambda_{jK}(\theta) \text{ for } k = K \end{cases}$$

ICCs: Graded Response Model



Other Variants: MGRM

"Modified Graded Response Model" (Muraki 1990):

- Common thresholds τ_k across items...
- Each item has one difficulty parameter / "intercept" β_j
- Formally:

$$\Pr(Y_{ij} > k|\alpha_j, \beta_j, \theta_i, \tau_k) \equiv \Lambda_{jk}(\theta) = \frac{\exp[\alpha_j(\theta_i - \beta_j + \tau_k)]}{1 + \exp[\alpha_j(\theta_i - \beta_j + \tau_k)]}$$

- These can be especially useful for Likert-type scales, where all response categories are identical (e.g., "Strongly Disagree," "Disagree," "Neither Agree nor Disagree," "Agree," "Strongly Agree")

Other Variants: PCM

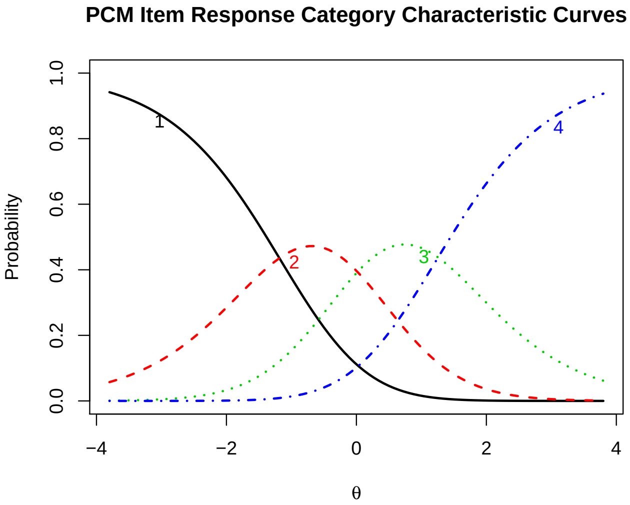
"Partial Credit Model" (Masters 1982):

- Probabilities for each category are defined as ratios of latent values divided by sums of latent values across all possible outcomes.
- Example of Thissen and Steinberg's "divide by total" models.
- Formally:

$$\Pr(Y_{ij} = k|\beta_{jt}, \theta_i) \equiv \Lambda_{jk}(\theta) = \frac{\exp \sum_{t=1}^{k}(\theta_i - \beta_{jt})}{\sum_{r=1}^{k} \exp \sum_{t=1}^{k}(\theta_i - \beta_{jt})}$$

- These are very similar to the GRM, except that in the PCM there is not necessarily a value for ability where $Pr(Y_{ij} = k)$ is highest for every k.

ICCs: Partial Credit Model



The Generalized PCM

- Note that the PCM assumes common discrimination ("slopes") across items, a la the 1PLM.
- Muraki (1992): Generalized by allowing different discrimination by different items
- Formally:

$$\Pr(Y_{ij} = k|\alpha_j, \beta_{jt}, \theta_i) \equiv \Lambda_{jk}(\theta) = \frac{\exp \sum_{t=1}^{k} \alpha_j(\theta_i - \beta_{jt})}{\sum_{r=1}^{k} \exp \sum_{t=1}^{k} \alpha_j(\theta_i - \beta_{jt})}$$

- GPCM : PCM :: 2PLM : 1PLM

POLITY IV Example

POLITY IV (http://www.systemicpeace.org/inscr/p4manualv2016.pdf) measures institutional characteristics of countries between 1800 and 2016:

- Regulation of Executive Recruitment (XRREG): 1="Unregulated," 2="Designational," 3="Regulated"
- Competitiveness of Executive Recruitment (XRCOMP): 1="Selection," 2="Transitional," 3="Election"
- Openness of Executive Recruitment (XROPEN): 1="Closed," 2="Dual: Designation," 3="Dual: Election," 4="Open"
- Executive Constraints (XCONST): 1="Unlimited Authority," (2) 3="Slight to Moderate Limitation," (4) 5="Substantial Limitations," (6) 7="Parity or Subordination"
- Regulation of Participation (PARREG): 1="Unregulated," 2="Multiple Identity," 3="Sectarian," 4="Restricted," 5="Regulated"
- Competitiveness of Participation (PARCOMP): 0="NA," 1="Repressed," 2="Suppressed," 3="Factional," 4="Transitional," 5="Competitive"

The "POLITY score" is a variant of a sum score of these components that ranges from -10 (fully autocratic) to 10 (fully democratic); see the manual for details.

POLITY IV (1800-2016)

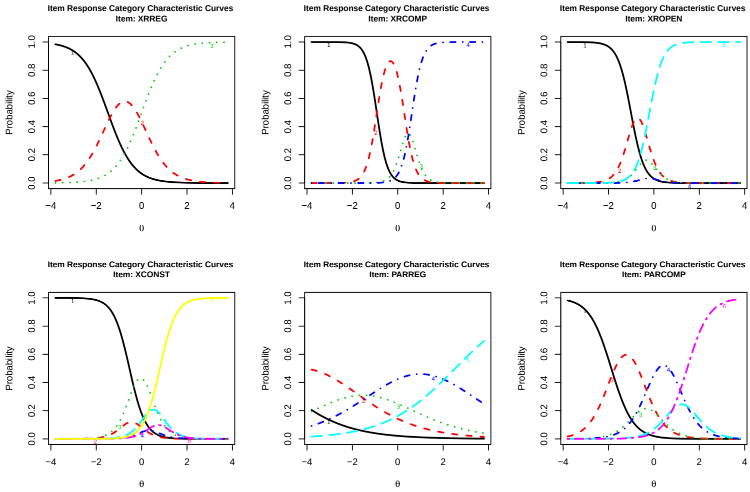
```
> describe(PIRT)
            vars     n    mean    sd median trimmed  mad  min  max range  skew kurtosis   se
Country*       1 16456   94.59 55.41     95    94.4 69.7    1  194   193  0.02    -1.19 0.43
Year           2 16456 1939.36 59.89   1958  1944.4 60.8 1800 2016   216 -0.60    -0.88 0.47
POLITYscore    3 16456   -0.54  7.16     -3    -0.7  8.9  -10   10    20  0.26    -1.48 0.06
XRREG          4 16456    2.47  0.66      3     2.6  0.0    1    3     2 -0.86    -0.37 0.01
XRCOMP         5 16456    1.60  1.05      1     1.6  1.5    0    3     3  0.25    -1.32 0.01
XROPEN         6 16456    2.90  1.54      4     3.1  0.0    0    4     4 -0.85    -1.01 0.01
XCONST         7 16456    3.79  2.41      3     3.7  3.0    1    7     6  0.22    -1.54 0.02
PARREG         8 16456    3.52  1.04      4     3.6  1.5    1    5     4 -0.39    -0.48 0.01
PARCOMP        9 16456    2.66  1.48      3     2.6  1.5    0    5     5  0.13    -1.03 0.01
```

```
> cor(PIRT[,3:9])
            POLITYscore XRREG XRCOMP XROPEN XCONST PARREG PARCOMP
POLITYscore      1.0000  0.36  0.834   0.49  0.911 0.0085    0.82
XRREG            0.3560  1.00  0.648   0.23  0.429 0.1480    0.32
XRCOMP           0.8339  0.65  1.000   0.69  0.771 0.0910    0.64
XROPEN           0.4949  0.23  0.686   1.00  0.506 0.0599    0.37
XCONST           0.9112  0.43  0.771   0.51  1.000 0.0826    0.69
PARREG           0.0085  0.15  0.091   0.06  0.083 1.0000    0.26
PARCOMP          0.8171  0.32  0.638   0.37  0.690 0.2578    1.00
```

An Alternative POLITY

```
> POLITY.GRM <- grm(PIRT[,4:9])
> POLITY.GRM

Call:
grm(data = PIRT[, 4:9])

Coefficients:
$XRREG
Extrmt1  Extrmt2   Dscrmn
 -1.489    0.003    1.775

$XRCOMP
Extrmt1  Extrmt2  Extrmt3   Dscrmn
  -0.93     0.29     0.63     4.33

$XROPEN
Extrmt1  Extrmt2  Extrmt3  Extrmt4   Dscrmn
  -1.01    -0.43    -0.23    -0.20     3.42

$XCONST
Extrmt1  Extrmt2  Extrmt3  Extrmt4  Extrmt5  Extrmt6   Dscrmn
  -0.54    -0.37     0.28     0.36     0.66     0.80     2.83

$PARREG
Extrmt1  Extrmt2  Extrmt3  Extrmt4   Dscrmn
  -5.89    -2.51    -0.54     2.53     0.65

$PARCOMP
Extrmt1  Extrmt2  Extrmt3  Extrmt4  Extrmt5   Dscrmn
  -1.88    -0.55    -0.13     0.98     1.47     2.08

Log.Lik: -110333
```

POLITY ICCs

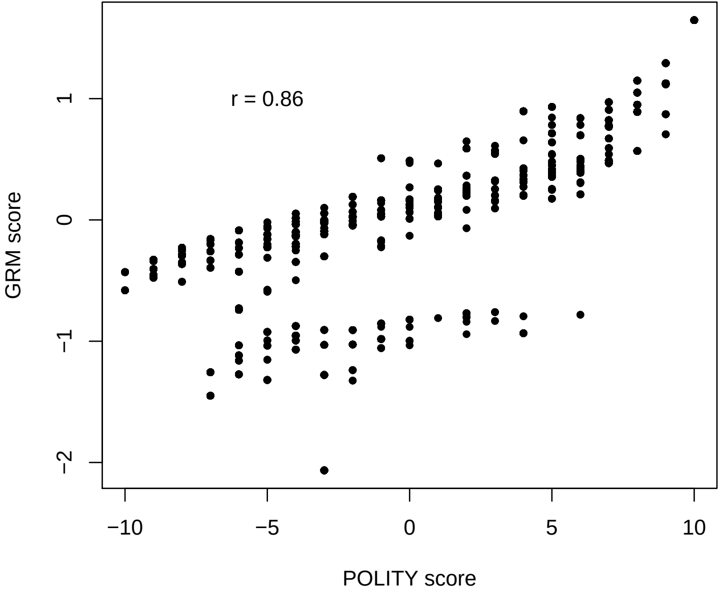


Item Response Category Characteristic Curves
Item: XRREG
Item Response Category Characteristic Curves
Item: XRCOMP
Item Response Category Characteristic Curves
Item: XROPEN
Item Response Category Characteristic Curves
Item: XCONST
Item Response Category Characteristic Curves
Item: PARREG
Item Response Category Characteristic Curves
Item: PARCOMP
Probability
θ

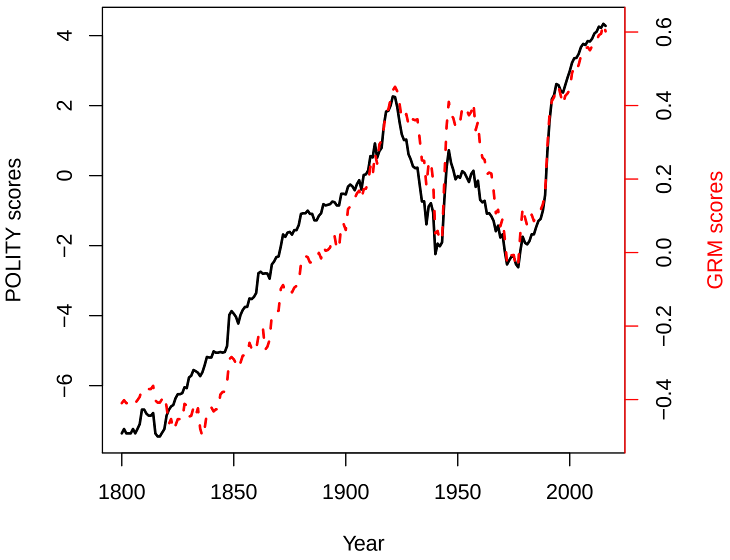
Alternative POLITY Scores...

```
> factor.scores(POLITY.GRM,robust.se=TRUE)

Call:
grm(data = PIRT[, 4:9])

Scoring Method: Empirical Bayes

Factor-Scores for observed response patterns:
   XRREG XRCOMP XROPEN XCONST PARREG PARCOMP  Obs     Exp     z1 se.z1
1      1      1      1      1      1       1  106  35.326 -2.065  0.56
2      1      1      1      1      2       4    3  11.830 -1.325  0.36
3      1      1      1      1      3       3   39  17.023 -1.320  0.35
4      1      1      1      1      3       4  270  13.115 -1.278  0.35
5      1      1      1      1      4       2  472 121.987 -1.449  0.37
6      1      1      1      1      4       3   60  16.076 -1.273  0.34
7      1      1      1      2      2       4    5   0.499 -1.057  0.29
.
.
.
```

POLITY comparison



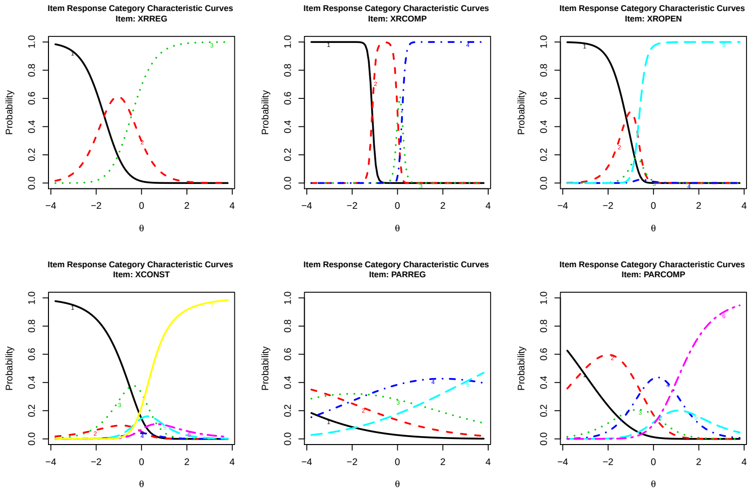
Mean POLITY and GRM Scores, 1800-2016



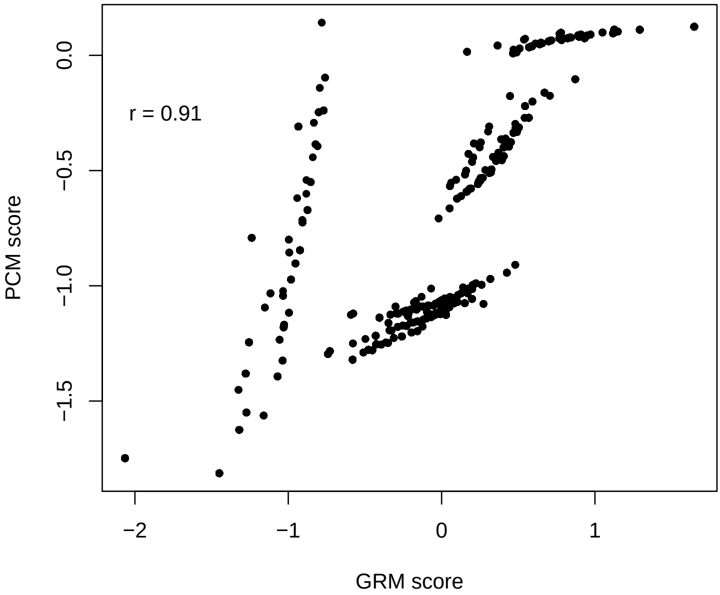
Same, Using GPCM

```
> POLITY.PCM <- gpcm(PIRT[,4:9],start.val="random",
+                    control=list(iter.qN=1000,
+                                 optimizer="optim"))
> POLITY.PCM

Call:
gpcm(data = PIRT[, 4:9], start.val = "random")

Coefficients:
$XRREG
Catgr.1  Catgr.2   Dscrmn
  -1.57    -0.54     2.23

$XRCOMP
Catgr.1  Catgr.2  Catgr.3   Dscrmn
 -1.116   -0.118    0.054   11.774

$XROPEN
Catgr.1  Catgr.2  Catgr.3  Catgr.4   Dscrmn
  -1.13    -0.52    -0.24    -1.38     4.02

$XCONST
Catgr.1  Catgr.2  Catgr.3  Catgr.4  Catgr.5  Catgr.6   Dscrmn
   1.36    -2.56     2.54    -1.98     1.02    -1.80     0.81

$PARREG
Catgr.1  Catgr.2  Catgr.3  Catgr.4   Dscrmn
   -5.8     -2.8     -1.2      2.6      0.3

$PARCOMP
Catgr.1  Catgr.2  Catgr.3  Catgr.4  Catgr.5   Dscrmn
 -2.911   -0.078   -1.132    1.162    0.113    0.941

Log.Lik: -108715
```

GPCM ICCs


Item Response Category Characteristic Curves
Item: XRREG
Item Response Category Characteristic Curves
Item: XRCOMP
Item Response Category Characteristic Curves
Item: XROPEN
Item Response Category Characteristic Curves
Item: XCONST
Item Response Category Characteristic Curves
Item: PARREG
Item Response Category Characteristic Curves
Item: PARCOMP
Probability
θ

Comparing GRM and GPCM Scores



Ordinal IRT Models in R

- `ltm` package
  - `grm`: Graded Response Model
  - `gpcm`: Generalized Partial Credit Model (with `constraints` for PCM)
- `eRm` ("Extended Rasch Modeling") package
  - `PCM`: Partial Credit Model
  - `RSM`: Rating Scale Model (a parsimonious partial credit model)
- `PP` ("Person Parameter Estimation") package
  - `PP_gpcm`: Generalized Partial Credit Model
- Others (e.g., `mirt`)...

References

- Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika 43:561-573.
- Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika 37:29-51.
- Embretson, S. E., and Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. Psychometrika 47:149-174.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. Applied Psychological Measurement 14:59-71.
- Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. Applied Psychological Measurement 16:159-176.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika monograph (vol. 17).

Dimensionality...

- Key point: All items measure the same underlying trait.
- Not (usually) known, or (conclusively) knowable in practice.
- Many, many tests...
- Most common: Drasgow and Lissak (1983)
  - *If the data are not unidimensional, the second eigenvalue of the (tetrachoric) correlation matrix of the items will be greater than would be the case under the (assumed) unidimensional model.*
  - The latter is simulated via Monte Carlo, using the item parameters from the estimated (unidimensional) model.

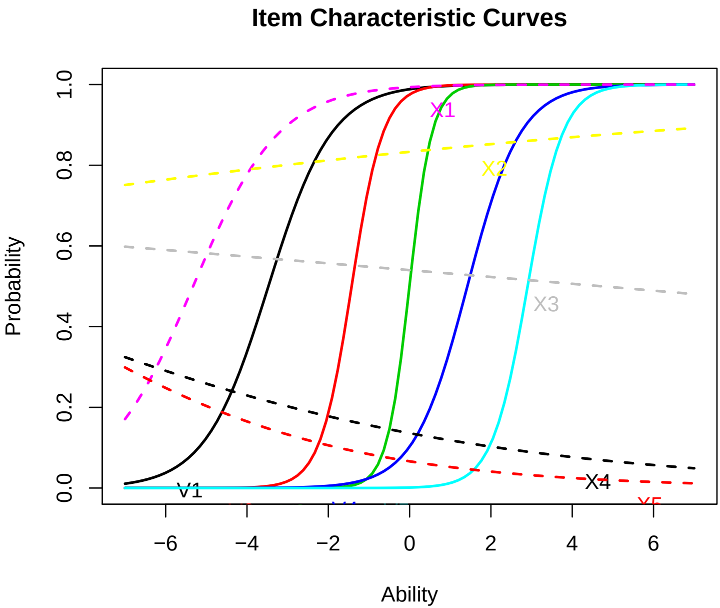
Assessing Dimensionality: A Simulation

```
> N <- 1000
> K <- 5
> set.seed(7222009)
> Discrims <- runif(K,1,3)
> D1 <- sim.irt(nvar=K,n=N,low=-3,high=3,
+               a=Discrims,c=0,d=NULL,mu=0,sd=1,
+               mod="logistic")
> Discrims <- runif(K,1,3)
> D2 <- sim.irt(nvar=K,n=N,low=-3,high=3,
+               a=Discrims,c=0,d=NULL,mu=0,sd=1,
+               mod="logistic")
> Data<-cbind(D1$items,D2$items)
> colnames(Data)<-c("V1","V2","V3","V4","V5",
+                   "X1","X2","X3","X4","X5")

> tetrachoric(Data)

Call: tetrachoric(x = Data)
tetrachoric correlation
   V1    V2    V3    V4    V5    X1    X2    X3    X4    X5
V1  1.00
V2  0.37  1.00
V3  0.45  0.71  1.00
V4  0.40  0.54  0.57  1.00
V5 -0.09  0.21  0.56  0.44  1.00
X1  0.28  0.35  0.20  0.25 -0.23  1.00
X2 -0.17  0.04  0.05  0.11 -0.29  0.53  1.00
X3 -0.05  0.00 -0.04  0.06 -0.04  0.47  0.33  1.00
X4 -0.24 -0.12 -0.07  0.02  0.05  0.20  0.47  0.47  1.00
X5  0.10 -0.10 -0.16  0.04  0.04  0.07  0.19  0.42  0.40  1.00
```

Dimensionality Simulation (continued)

```
> Model<-ltm(Data~z1)
> summary(Model)

Coefficients:
              value std.err  z.vals
Dffclt.V1    -3.473   0.607  -5.725
Dffclt.V2    -1.426   0.123 -11.618
Dffclt.V3    -0.004   0.044  -0.094
Dffclt.V4     1.420   0.152   9.326
Dffclt.V5     2.903   0.487   5.962
Dffclt.X1    -5.325   1.973  -2.699
Dffclt.X2   -22.324  33.133  -0.674
Dffclt.X3     4.727  11.388   0.415
Dffclt.X4   -11.601   8.394  -1.382
Dffclt.X5   -10.329   6.301  -1.639
Dscrmn.V1     1.281   0.311   4.117
Dscrmn.V2     2.582   0.602   4.289
Dscrmn.V3     3.592   1.584   2.268
Dscrmn.V4     1.526   0.265   5.766
Dscrmn.V5     2.302   0.871   2.645
Dscrmn.X1     0.944   0.415   2.276
Dscrmn.X2     0.072   0.107   0.673
Dscrmn.X3    -0.034   0.081  -0.421
Dscrmn.X4    -0.159   0.116  -1.372
Dscrmn.X5    -0.256   0.159  -1.609

Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.012
quasi-Newton: BFGS
```

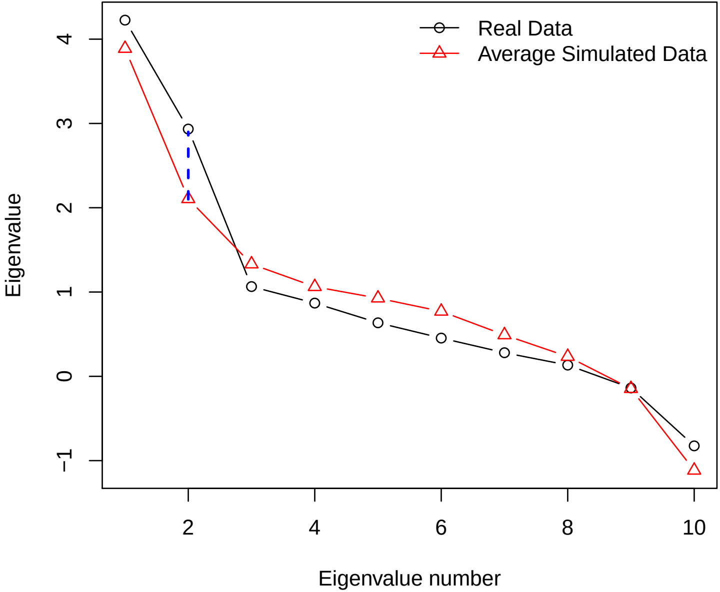
ICCs: Dimensionality Simulation



Dimensionality Test

```
> UDTest<-unidimTest(Model)
> UDTest

Unidimensionality Check using Modified Parallel Analysis

Call:
ltm(formula = Data ~ z1)

Matrix of tertachoric correlations
       V1     V2     V3    V4     V5   X1     X2     X3     X4     X5
V1  1.000  0.373  0.446 0.936  0.871 0.28 -0.170 -0.047 -0.238  0.099
V2  0.373  1.000  0.712 0.541  0.921 0.35  0.038 -0.003 -0.118 -0.099
V3  0.446  0.712  1.000 0.571  0.963 0.20  0.048 -0.038 -0.066 -0.156
V4  0.936  0.541  0.571 1.000  0.436 0.93  0.114  0.065  0.023  0.036
V5  0.871  0.921  0.963 0.436  1.000 0.85 -0.294 -0.039  0.049  0.042
X1  0.278  0.351  0.195 0.925  0.852 1.00  0.533  0.468  0.939  0.886
X2 -0.170  0.038  0.048 0.114 -0.294 0.53  1.000  0.328  0.466  0.189
X3 -0.047 -0.003 -0.038 0.065 -0.039 0.47  0.328  1.000  0.467  0.419
X4 -0.238 -0.118 -0.066 0.023  0.049 0.94  0.466  0.467  1.000  0.405
X5  0.099 -0.099 -0.156 0.036  0.042 0.89  0.189  0.419  0.405  1.000

Alternative hypothesis: the second eigenvalue of the observed data is substantially larger
than the second eigenvalue of data under the assumed IRT model

Second eigenvalue in the observed data: 2.9
Average of second eigenvalues in Monte Carlo samples: 2.1
Monte Carlo samples: 100
p-value: 0.02
```

Dimensionality Test: Plot



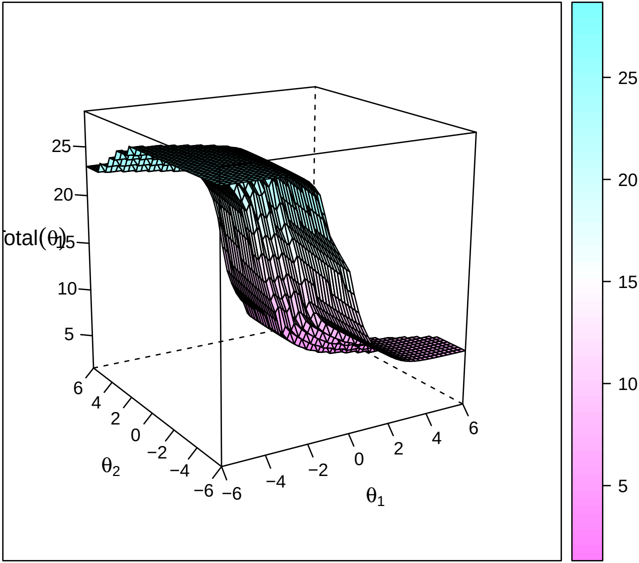
Multidimensional IRT using mirt

```
> SIM.MIRT <- mirt(Data,2,itemtype="2PL")
Iteration: 267, Log-Lik: -3311.943, Max-Change: 0.00010

> summary(SIM.MIRT)

Rotation:  oblimin

Rotated factor loadings:

        F1      F2    h2
V1 -0.1307 -0.6137 0.380
V2  0.0237 -0.8434 0.715
V3  0.0148 -0.8830 0.782
V4  0.1711 -0.6655 0.492
V5 -0.0719 -0.8061 0.645
X1  0.8929 -0.2220 0.881
X2  0.5884  0.0389 0.344
X3  0.6099  0.0987 0.371
X4  0.7879  0.1692 0.626
X5  0.5878  0.1880 0.362

Rotated SS loadings:  2.5 3.1

Factor correlations:

       F1     F2
F1  1.000 -0.087
F2 -0.087  1.000
```

Multidimensional IRT: POLITY Example

```
> POLITY.MIRT <- mirt(PIRT[,4:9],2,itemtype="graded")
Iteration: 236, Log-Lik: -106186.835, Max-Change: 0.00008

> summary(POLITY.MIRT)

Rotation:  oblimin

Rotated factor loadings:

              F1      F2    h2
XRREG   -0.73092 -0.1381 0.618
XRCOMP  -1.02349  0.0827 1.000
XROPEN  -0.95170 -0.0674 0.952
XCONST  -0.87641  0.0667 0.735
PARREG  -0.00742 -0.9976 1.000
PARCOMP -0.73345 -0.0777 0.581

Rotated SS loadings:  3.8 1

Factor correlations:

     F1   F2
F1 1.00 0.32
F2 0.32 1.00
```

Expected Total Score vs. Two Dimensions



Predictive Models as Measurement Models

Challenges to Measurement:

- Sparse / Missing Data
- Overfitting
- Subjectivity

Challenges to Prediction:

- Sparse / Missing Data
- Overfitting
- Subjectivity

Predictive Models I: Classification Trees

- Intuition:
  - For outcome Y and covariates $\mathbf{X}$, split by X_k that minimizes classification error
  - Repeat ad nauseam...
- See Breiman et al. (1984); Rokach and Maimon (2008)
- Applications: Kastellec (2010), others
- Issues: Overfitting $\rightarrow$ subjective pruning

Predictive Models II: Random Forests

- Intuition:
  - Grow lots of random classification trees on Y, each using $\mathbf{X}_m < \mathbf{X}$ covariates
  - Let each "vote" on how to classify each observation
  - The forest chooses the classification with the most votes

- Variable Importance: Assess how much prediction accuracy decreases when we:
  - Omit (do not split on) a variable X_k ("Mean Decrease in the Gini"), or
  - Randomly permute the values of X_k and then re-grow the forest(s) ("Mean Decrease in Accuracy")

- See Breiman (2001); Berk (2010)

Random Forests: Advantages

- Strong predictive ability
- ~~Subjectivity~~
- No Overfitting
- Fast, Robust Handling of Missingness
- Availability of Diagnostics

Random Forests: Measurement

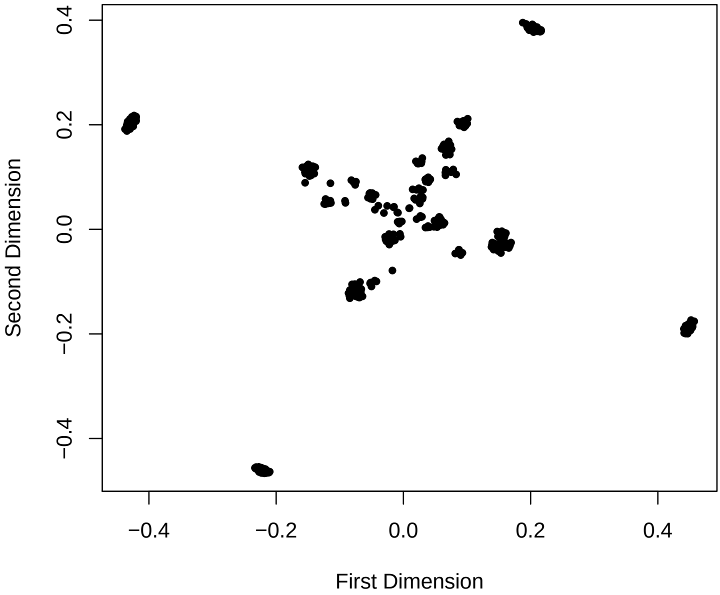
- "Unsupervised" (Shi and Horvath 2006):
  - Sample $X'_{1i}...X_{1N}$ from X_1; repeat for $X_2...X_k$ ("synthetic" data)
  - Run $\mathbf{X}$, $\mathbf{X}'$ through random forests
  - Assess "proximity" $\sim$ proportion of "votes" putting two observations in the same class

- Apply conventional Euclidean scaling methods to proximity results...

RF Example: Simulated MIRT Data

```
> source("Code/FunctionsRFclustering.txt")
>
> no.forests <- 20 # N of forests
> no.trees <- 500 # N of trees per forest
>
> distRF <- RFdist(Data,mtry1=3,no.trees,no.forests,addcl1=T,
+                  addcl2=F,imp=T, oob.prox1=T)
>
> # MDS on the results:
> cmd1<-cmdscale(as.dist(distRF$cl1),2)
>
> # Plotting...
```

Euclidean MDS on Random Forest Distances



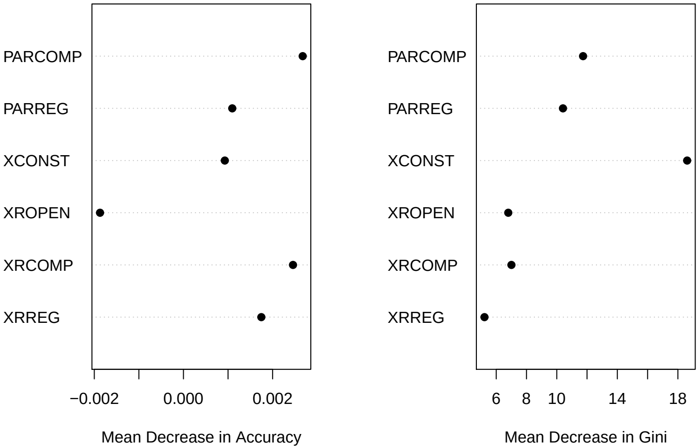
RF Example: POLITY Components

```
> RF.POLITY <- randomForest(x=PIRT[,4:9],ntree=no.trees,
+                           importance=TRUE,proximity=TRUE)
>
> # MDS and plotting:
>
> POLITY.MDS <-cmdscale(as.dist(RF.POLITY$proximity),2)
```

Euclidean MDS on POLITY Random Forests

(plot goes here...)

Variable Importance: POLITY



Possible Future Work

- Cross-Validation of Approaches
- Benchmarking
- Sensitivity Testing (missingness, etc.)
- Expand RF approach to FA/PCA/IRT...