PLSC 504

Introduction to Item Response Theory Models - I

November 28, 2017

Item Response Theory ("IRT")

- Origins in psychometrics / testing
- *Measurement* model – (typically) *no* $\mathbf{X}$
- *Unidimensional*
- *Discrete* responses $\mathbf{Y}$
- Equally descriptive and inferential

Basic Setup

$$Y^* = \text{ latent trait ("ability")}$$

$$Y = \text{observed measures}$$

- $i \in \{1, 2 \ldots N\}$ indexes *subjects* / *units*, and
- $j \in \{1, 2, \ldots J\}$ indexes *items* / *measures*.

$$Y_{ij} = \begin{cases} 0 & \text{if subject } i \text{ gets item } j \text{ "incorrect,"} \\ 1 & \text{if subject } i \text{ gets item } j \text{ "correct."} \end{cases}$$

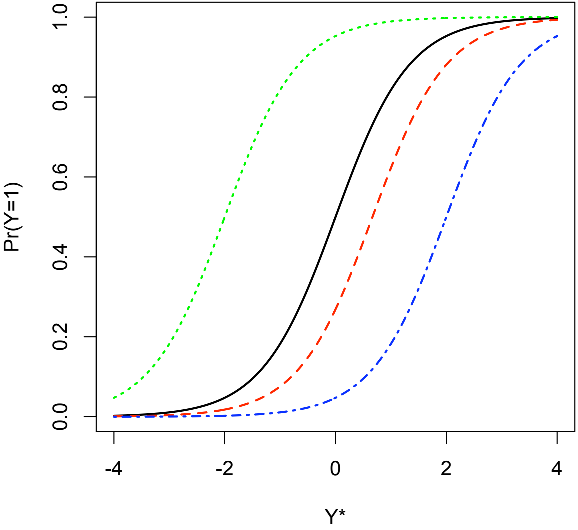
```
> head(SCOTUS,10)

   id Rehnquist Stevens OConnor Scalia Kennedy Souter Thomas Ginsburg Breyer
1   1         0       1       0      0       0      0      0        0      0
2   2         0       0       0      0       0      0      0        0      0
3   3         0       0       0      0       0      0      0        0      0
4   4         1       1       1      1       1      1      1        1      1
5   5         0       1       0      0       1      1      0        1      1
6   6         0       0       0      0       0      0     NA        0      0
7   7         1       1       1      0       1      1      0        1      1
8   8         0       1       0      0       0      0      0       NA      0
9   9         0       0       0      0       0      0      0        0      0
10 10         1       1       1      1       1      1      1        1      1
```

One-Parameter Logistic Model ("1PLM")

$$\Pr(Y_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$$

Here,

- θ_i = respondent *i*'s *ability*,
- β_j = item *j*'s *difficulty*.
- $\beta_j \equiv$ value of Y^* where $\Pr(Y_{ij} = 1) = 0.50$



1.0
0.8
0.6
0.4
0.2
0.0
Pr(Y=1)
-4
-2
0
2
4
Y*

1PLM

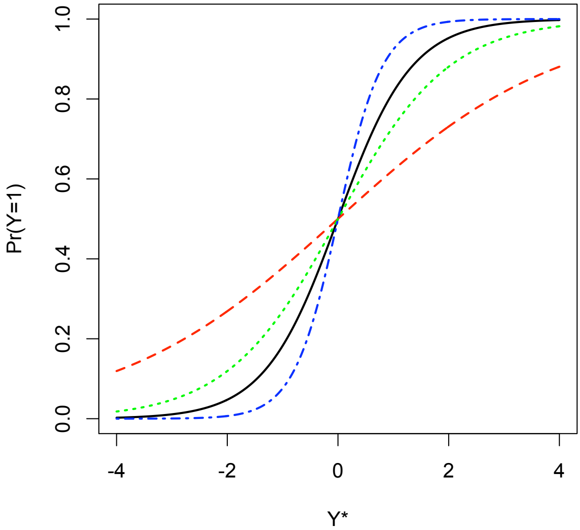
- a.k.a. “Rasch” model (Rasch 1960)
- Implicit “slope” $= 1.0$
- Implies items are equally “discriminating”
- If not...

Two-Parameter Logistic Model ("2PLM")

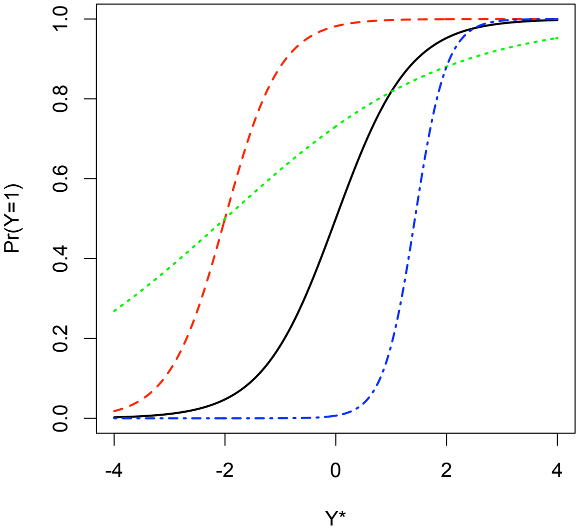
$$\Pr(Y_{ij} = 1) = \frac{\exp[\alpha_j(\theta_i - \beta_j)]}{1 + \exp[\alpha_j(\theta_i - \beta_j)]}$$

- θ_i = respondent *i*'s *ability*,
- β_j = item *j*'s *difficulty*,
- α_j = item *j*'s *discrimination*.

Identical Difficulty, Different Discrimination



Different Difficulty & Discrimination



2PLM

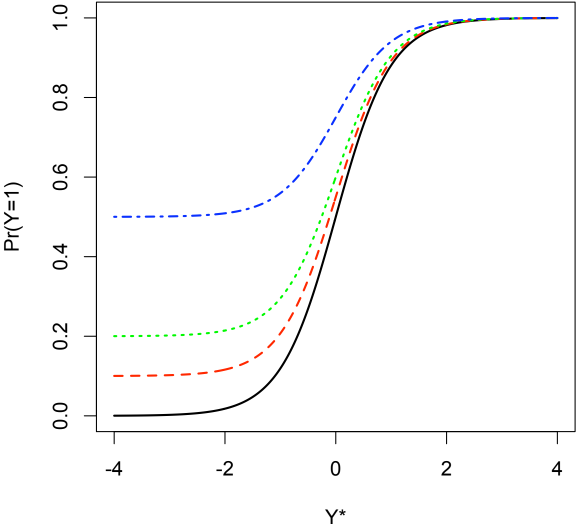
- Due to Birnbaum (1968)
- Similar to "typical" logit...
- Nests the 1PLM as a special case ($\alpha_j = 1 \forall j$)

Three-Parameter Logistic Model ("3PLM")

$$\Pr(Y_{ij} = 1) = \delta_j + (1 - \delta_j) \left\{ \frac{\exp[\alpha_j(\theta_i - \beta_j)]}{1 + \exp[\alpha_j(\theta_i - \beta_j)]} \right\}$$

- θ_i = respondent *i*'s *ability*,
- β_j = item *j*'s *difficulty*,
- α_j = item *j*'s *discrimination*.
- δ_j = *lower asymptote* of $\Pr(Y_{ij} = 1)$ (incorrectly: "guessing" parameter).

3PLM, Constant α & β, Varying δ



The Two Big Assumptions

- *Unidimensionality*
- *Local Item Independence* ("No LID"):

$$\mathrm{Cov}(Y_{ij}, Y_{ik} | \theta_i) = 0 \; \forall \; j \neq k$$

Estimation: Notation

$$P_{ij} = \Pr(Y_{ij} = 1),$$

$$\begin{aligned} Q_{ij} &= \Pr(Y_{ij} = 0) \\ &= 1 - \Pr(Y_{ij} = 1), \end{aligned}$$

$$\Psi = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_J \\ \alpha_1 \\ \vdots \\ \alpha_J \\ \delta_1 \\ \vdots \\ \delta_J \end{pmatrix}.$$

Estimation: Likelihoods

Known $\Psi = \alpha$, β, δ:

$$L(\mathbf{Y}|\Psi) = \prod_{j=1}^{J} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}}.$$

Known θ:

$$L(\mathbf{Y}|\theta) = \prod_{i=1}^{N} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}}.$$

Estimation: Likelihoods

$$L(\mathbf{Y}|\Psi,\theta) = \prod_{i=1}^{N}\prod_{j=1}^{J} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}}$$

$$\ln L(\mathbf{Y}|\Psi,\theta) = \sum_{i=1}^{N}\sum_{j=1}^{J} Y_{ij}\ln P_{ij} + (1-Y_{ij})Q_{ij}.$$

Parameterization

- $N + J$ parameters in the 1PLM,
- $N + 2J$ parameters in the 2PLM,
- $N + 3J$ parameters in the 3PLM.

But...

- NJ observations,
- Asymptotics as $N \to \infty$, $J \to \infty$...

Estimation: Conditional Likelihood

Total score is:

$$T_i = \sum_{j=1}^{J} Y_{ij} \ \in \ \{0, 1, ...J\}$$

$$L = \prod_{i=1}^{N} \frac{\exp[\alpha_j(\theta_t - \beta_j)]}{1 + \exp[\alpha_j(\theta_t - \beta_j)]}$$

θ_t are "score-group" parameters corresponding to the $J + 1$ possible values of T.

Estimation: Conditional Likelihood

- Equivalent to fitting a conditional logit model:

$$\Pr(Y_{ij} = 1) = \frac{\exp(\mathbf{Z}_{ij}\gamma)}{\sum_{j=1}^{J} \exp(\mathbf{Z}_{ij}\gamma)}$$

  with $\mathbf{Z}_{ij} =$ "item dummies."

- Useful only for 1PLM (since T_i is a sufficient statistic for θ_i).

Estimation: Marginal Likelihood

$$L(\mathbf{Y}|\Psi,\theta) = \prod_{i=1}^{N}\left[\int_{-\infty}^{\infty}\prod_{j=1}^{J} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}} d\theta\right]$$

- Analogous to "random effects"...
- Eliminates inconsistency as $N \to \infty$, *but*
- Requires *strong* exogeneity of θ and Ψ.

Estimation: Bayesian Approaches

- Place priors on θ, Ψ;
- Estimate via sampling from posteriors, via MCMC.
- Eliminates problems with $\hat{\alpha}$, $\hat{\beta}$, $\hat{\theta} = \infty$ (see below).
- Easily extensible to other circumstances (hierarchical/multilevel, etc.)

Identification

Two Issues:

- *Scale* invariance: $L(\hat{\Psi}) = L(\hat{\Psi} + c)$
- *Rotational* invariance: $L(\hat{\Psi}) = L(-\hat{\Psi})$

Fixes:

- Set one (arbitrary) $\beta_j = 0$, and another (arbitrary) $\beta_k > 0$, or
- Fix two θ_is at specific values.

Further (Potential) Concerns

- $Y_{ij} = 0/1 \; \forall \; i \; \rightarrow \; \beta_j = \pm\infty$.
- $Y_{ij} = 0/1 \; \forall \; j \; \rightarrow \; \theta_i = \pm\infty$.
- Separation / "empty cells" $\rightarrow \; \alpha_j = \pm\infty$.
- Problematic for joint and conditional approaches; more easily dealt with in the Bayesian framework.

Results

- Estimates of $\hat{\alpha}$s, $\hat{\beta}$s, and/or $\hat{\delta}$s, plus $\hat{\theta}$s
- Associated s.e.s / c.i.s
- "Scale-free" quantities of interest...