The EM Algorithm for MMLE, Step by Step

EDUC 8720: Item Parameter Estimation

Derek Briggs

2026-02-15

The EM Algorithm: Overview

The EM algorithm alternates between two steps:

Step	What it does	Analogy
E-step	Uses current item parameter guesses to figure out how examinees are distributed across the ability scale	"If these were the true item parameters, where would each examinee likely fall?"
M-step	Uses that ability distribution to re-estimate item parameters	"Given where we think examinees fall, what item parameters best fit the responses?"

We repeat until the estimates stop changing.

The Data: LSAT-6

- 1000 examinees, 5 dichotomous items, 32 unique response patterns

8 Most Common Response Patterns

Pattern	Freq
11111	298
11011	173
10011	81
10111	80
11101	61
11001	56
10001	29
10101	28

Initial Setup

2PL model with intercept/slope parameterization:

$$P_i(\theta) = \frac{1}{1 + \exp(-(d_i + a_i\theta))}$$

Starting values: all intercepts $d_i = 0$, all slopes $a_i = 1$

Quadrature: 10 nodes from -4 to $+4$ with weights from the standard normal

These starting values treat all items as identical — the *data* will differentiate them.

E-Step 1: Likelihood at Each Node

For each response pattern l at each quadrature node X_k:

$$L_l(X_k) = \prod_{i=1}^{n} P_i(X_k)^{u_{li}} Q_i(X_k)^{1-u_{li}}$$

"How probable is this response pattern if the examinee's ability were exactly X_k?"

Pattern 22 {1,0,1,0,1}: Likelihood at each node

X_k	L_l(X_k)
-4.0000	0.00000561
-3.1111	0.00007111
-2.2222	0.00076087
-1.3333	0.00568600
-0.4444	0.02214009
0.4444	0.03452866

E-Step 2: Marginal Probability

Sum over nodes, weighted by quadrature weights:

$$P_l = \sum_{k=1}^{q} L_l(X_k) \cdot A(X_k)$$

This is the **marginal likelihood** of pattern l — the probability of seeing that pattern averaged across all ability levels.

Pattern	Freq	Marginal Prob
11111	298	0.0955352
11011	173	0.0360619
10011	81	0.0224199
10111	80	0.0360619
11101	61	0.0360619
11001	56	0.0224199

E-Step 3: Posterior Probabilities

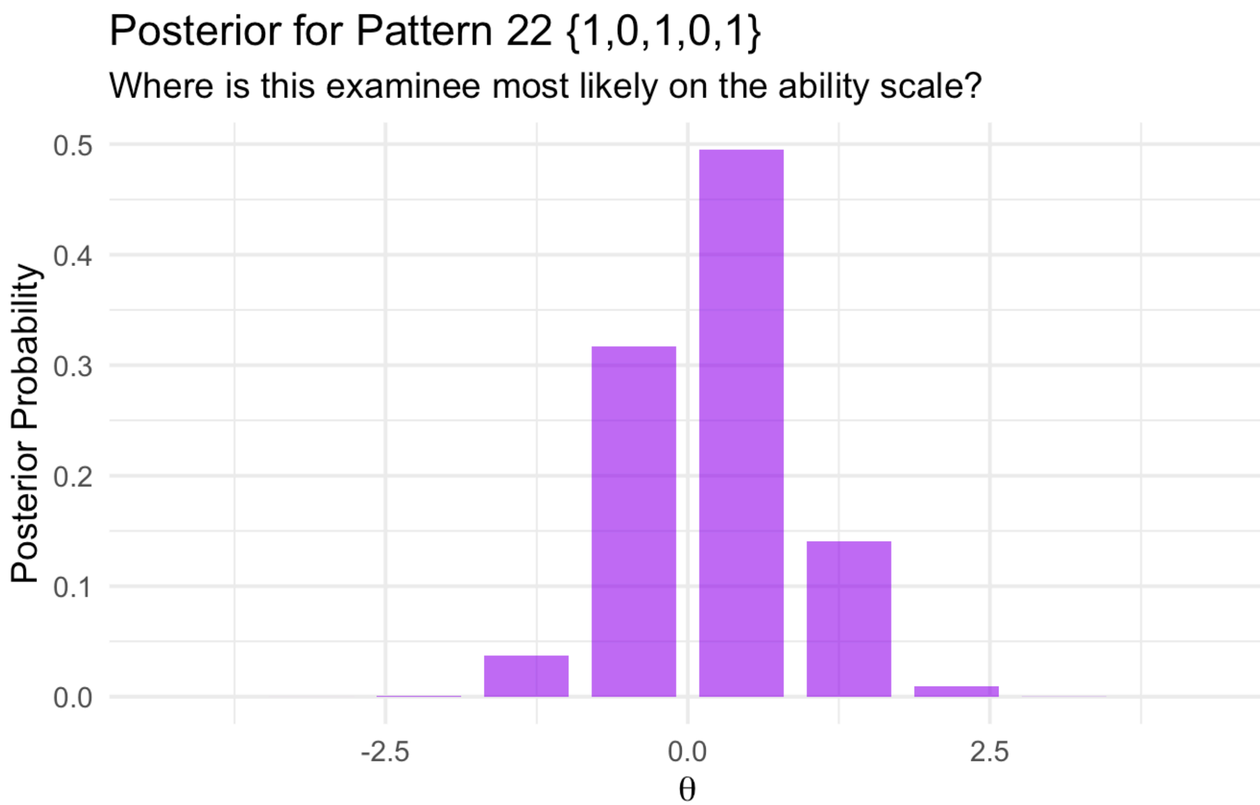
Apply Bayes' theorem at each node:

$$P(X_k \mid \mathbf{u}_l) = \frac{L_l(X_k) \cdot A(X_k)}{P_l}$$

Pattern 22 {1,0,1,0,1}; P_l = 0.02241993

X_k	$L_l(X_k)$	$A(X_k)$	$L \times A$	Posterior
-4.0000	0.00000561	0.000119	0.00000000	0.00000003
-3.1111	0.00007111	0.002805	0.00000020	0.00000890
-2.2222	0.00076087	0.030020	0.00002284	0.00101879
-1.3333	0.00568600	0.145800	0.00082902	0.03697690
-0.4444	0.02214009	0.321300	0.00711361	0.31728951
0.4444	0.03452866	0.321300	0.01109406	0.49483018
1.3333	0.02157010	0.145800	0.00314492	0.14027339
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

E-Step 3: Posterior (Visual)



E-Step 4: The "Artificial Data"

Aggregate across all examinees (weighted by pattern frequency f_l):

$$\bar{n}_k = \sum_{l=1}^{s} f_l \cdot P(X_k \mid \mathbf{u}_l) \quad \text{(expected examinees at node } k\text{)}$$

$$\bar{r}_{ik} = \sum_{l=1}^{s} u_{li} \cdot f_l \cdot P(X_k \mid \mathbf{u}_l) \quad \text{(expected correct on item } i \text{ at node } k\text{)}$$

These play the same role as the observed frequencies in the JMLE estimation equations.

E-Step 4: Artificial Data Table

Artificial data from E-Step (Cycle 1)

X_k	$\bar{n}_k$	$\bar{r}_{1k}$	$\bar{r}_{2k}$	$\bar{r}_{3k}$	$\bar{r}_{4k}$	$\bar{r}_{5k}$
-4.0000	0.0	0.0	0.0	0.0	0.0	0.0
-3.1111	0.1	0.0	0.0	0.0	0.0	0.0
-2.2222	2.7	1.4	0.4	0.2	0.5	1.0
-1.3333	31.1	22.1	9.6	4.9	11.7	18.0
-0.4444	182.0	155.1	91.1	56.3	105.5	137.5
0.4444	402.3	374.1	278.7	206.0	304.1	351.3
1.3333	295.8	286.4	249.0	211.1	259.6	278.2
2.2222	77.6	76.5	72.0	66.5	73.3	75.5
3.1111	8.1	8.0	7.8	7.6	7.9	8.0
4.0000	0.4	0.4	0.4	0.3	0.4	0.4

Why Do the $\bar{r}_{ik}$ Columns Differ?

We started with **identical** parameters for all 5 items, yet the $\bar{r}_{ik}$ values differ.

- $\bar{n}_k$ depends only on the posterior — **same for all items**
- $\bar{r}_{ik}$ multiplies by u_{li} (the actual response) — only correct responses contribute

The raw data break the symmetry:

```
##   Item 1: 0.924 proportion correct
##    Item 2: 0.709 proportion correct
##    Item 3: 0.553 proportion correct
##    Item 4: 0.763 proportion correct
##    Item 5: 0.87 proportion correct
```

Items with higher proportion correct $\longrightarrow$ larger $\bar{r}_{ik}$ values.

The M-Step: Estimating Item Parameters

Treat the artificial data as known and solve (one item at a time):

$$\sum_{k=1}^{q} \bar{n}_k \left[\frac{\bar{r}_{ik}}{\bar{n}_k} - P_i(X_k) \right] = 0 \quad (\text{for } d_i)$$

$$\sum_{k=1}^{q} \bar{n}_k \left[\frac{\bar{r}_{ik}}{\bar{n}_k} - P_i(X_k) \right] X_k = 0 \quad (\text{for } a_i)$$

Logic: Find the d_i and a_i that make the residuals (observed − expected) sum to zero. The X_k multiplier in the second equation captures the *trend* across ability (slope).

Solve iteratively using **Newton-Raphson** (same as in JMLE).

M-Step: Results After Cycle 1

Item Parameter Estimates After 1 EM Cycle

Item	Intercept (d_i)	Slope (a_i)	Difficulty (b_i)
1	2.1659	0.9438	-2.2949
2	0.4102	0.9410	-0.4360
3	-0.3787	0.9697	0.3905
4	0.7263	0.9322	-0.7791
5	1.5321	0.9198	-1.6658

Already, the items have differentiated! Items with higher proportion correct get larger (more negative) difficulty estimates.

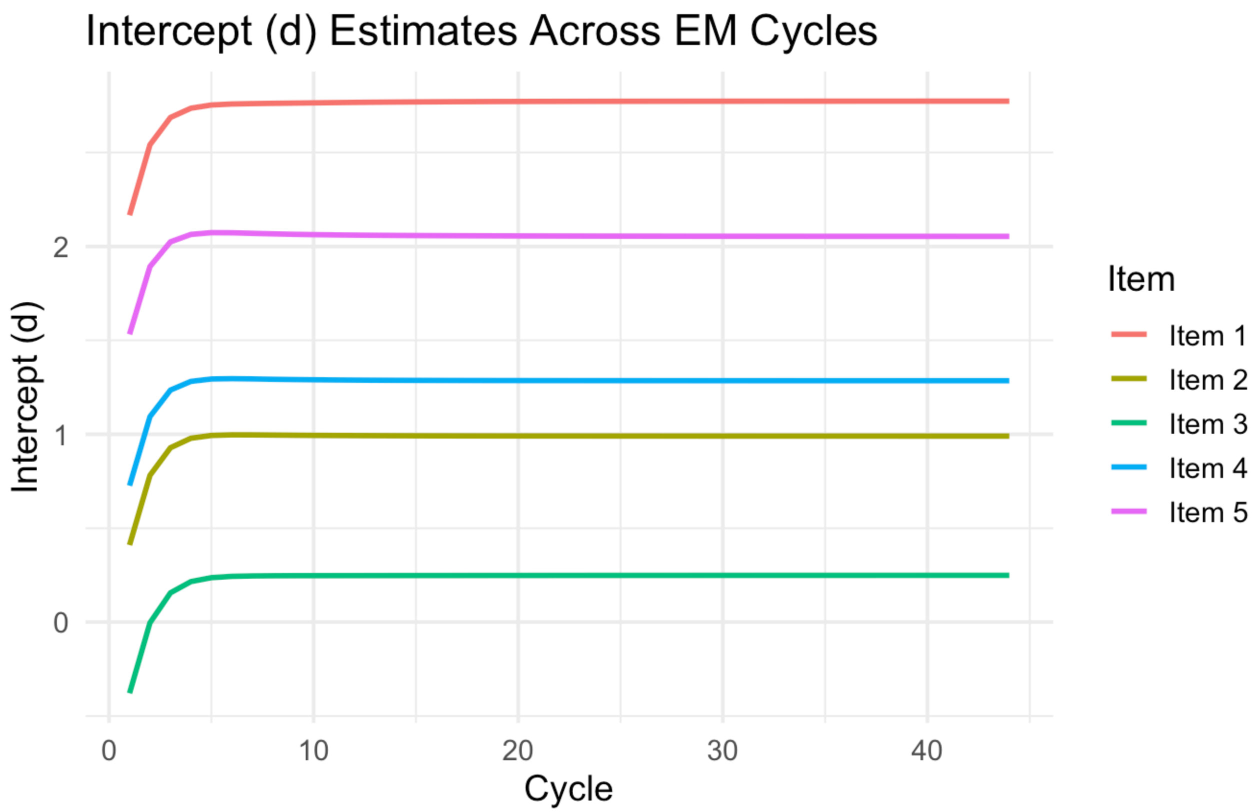
The Full EM Loop

Now iterate: use these new estimates as input to the next E-step, repeat.

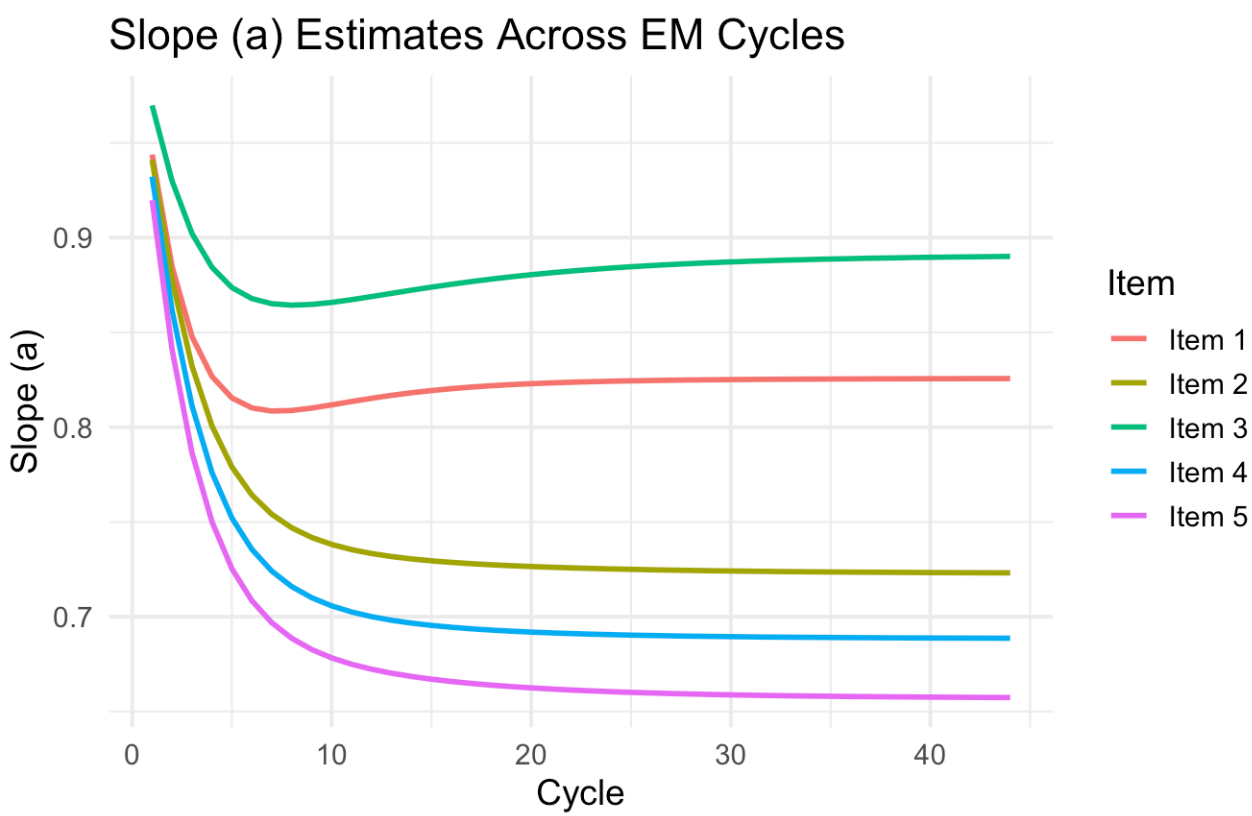
Final Estimates After 44 EM Cycles

Item	Intercept (d_i)	Slope (a_i)	Difficulty (b_i)
1	2.7732	0.8256	-3.3588
2	0.9903	0.7231	-1.3695
3	0.2491	0.8901	-0.2798
4	1.2848	0.6887	-1.8657
5	2.0535	0.6573	-3.1239

Watching EM Converge



Watching EM Converge (cont.)



EM converges **slowly** — parameters change rapidly early, then creep toward final values.

Comparison with `mirt`

Our MMLE/EM vs. mirt Package

Item	Our a	mirt a	Our b	mirt b
1	0.826	0.825	-3.359	-3.361
2	0.723	0.723	-1.369	-1.370
3	0.890	0.890	-0.280	-0.280
4	0.689	0.689	-1.866	-1.866
5	0.657	0.658	-3.124	-3.123

Small differences reflect implementation details: `mirt` uses 61 quadrature points (vs. our 10) and Bock-Aitkin acceleration.

EM Algorithm Summary

```
1. INITIALIZE: Starting values for all item parameters

2. E-STEP:
   → Compute likelihood at each node for each response pattern
   → Apply Bayes' theorem to get posterior at each node
   → Aggregate into "artificial data" (n-bar, r-bar)

3. M-STEP:
   → For each item separately:
     Treat artificial data as known, solve for d and a
     using Newton-Raphson

4. CONVERGED? If yes → stop. If no → go to step 2.
```

Key Takeaways

- **JMLE** estimates every person's θ along with item parameters $\longrightarrow$ inconsistent estimates (Neyman-Scott problem)
- **MMLE** integrates out θ by assuming a population distribution $\longrightarrow$ consistent item parameter estimates
- **EM** handles the circular dependency: the E-step infers the ability distribution, the M-step re-estimates item parameters, repeat
- MMLE produces **item parameters only** — person θ estimates require a separate step (MLE, MAP, or EAP)