Item Parameter Estimation I

Bock & Lieberman (1970)
Bock & Aitkin (1981)

The plan:

I. Using R
  a) Bock-Lieberman ML, Normal Ogive
  b) Bock-Lieberman ML, 2PL
  c) Bock-Aitkin "EM," 2PL
II. Using C++
  a) Bock-Aitkin "EM," 3PL, Graded model
III. Using R
  a) Albert MCMC, Normal Ogive

Stouffer & Toby (1951) Data, 4 Questions

Response Pattern	Frequency	Response Pattern	Frequency
1111	42	0110	4
1110	23	0101	2
1101	6	0011	1
1011	6	1000	38
0111	1	0100	9
1100	24	0010	6
1010	25	0001	2
1001	7	0000	20

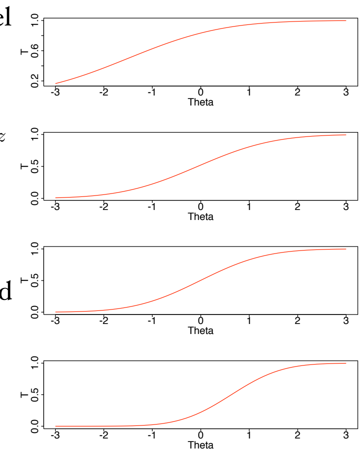
The normal ogive model for the positive (1) item response is

$$T_i(1_i|\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_i\theta - c_i} e^{\frac{-z^2}{2}} dz$$

or more compactly

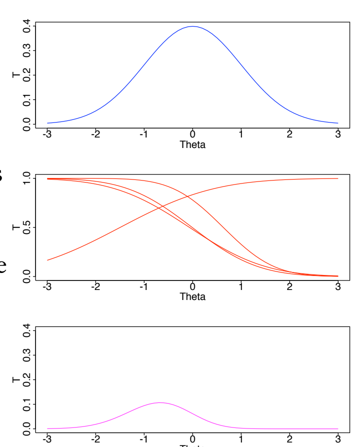
$$T_i(1_i|\theta) = \Phi[a_i\theta - c_i]$$

These are the four fitted curves for the Stouffer-Toby data—the goal.



For any response pattern the probability is computed like this one (this is 1000; 38 respondents): The population distribution is shown in blue, the trace lines in red, and the posterior in magenta. The area of the posterior is the probability of the response pattern.

$$P_{\mathbf{k}} = \int_{-\infty}^{\infty} \left(\prod_{i=1}^{nitems} T_i(k_i|\theta) \right) \phi(\theta) d\theta$$



There are 16 response patterns **k** for four items; for each:

$$P_{\mathbf{k}} = \int_{-\infty}^{\infty} \left(\prod_{i=1}^{nitems} T_i(k_i|\theta) \right) \phi(\theta) d\theta$$

The likelihood for the observed frequencies $r_{\mathbf{k}}$ for all 16 patterns is:

$$L \sim \prod_{\mathbf{k}} P_{\mathbf{k}}^{r_{\mathbf{k}}}$$

The loglikelihood is:

$$\ell = \sum_{\mathbf{k}} r_{\mathbf{k}} \log(P_{\mathbf{k}})$$

We seek the ML estimates of the parameters (a and c, buried in T, above).

Estimation can be implemented in R with just that much information, with the integration approximated by rectangular quadrature (sums of heights at points) as it was to compute EAPs.

Bock & Lieberman (1970) did not have automagic minimizers with built-in numerical derivatives.

So they needed the first and second derivatives of

$$\ell = \sum_{\mathbf{k}} r_{\mathbf{k}} \log(P_{\mathbf{k}})$$

in which the item parameters are buried in

$$P_{\mathbf{k}} = \int_{-\infty}^{\infty} \left(\prod_{i=1}^{nitems} T_i(k_i|\theta) \right) \phi(\theta) d\theta$$

specifically, in

$$T_i(1_i|\theta) = \Phi[a_i\theta - c_i]$$

After quite a bit of stuff, and ignoring side-trips, they arrived at

$$\frac{\partial P_{\mathbf{k}}}{\partial u_i} = (-1)^{k_i+1} \int_{-\infty}^{\infty} \frac{\partial g_i}{\partial u_i} \phi(g_i) \left(\prod_{h=1, h\neq i}^{nitems} T_h(k_h|\theta) \right) \phi(\theta) d\theta$$

in which u is either a or c, $g_i = a_i\theta - c_i$, and $\frac{\partial g_i}{\partial u_i}$ is either θ or -1, depending on whether a or c replaces u.

That all goes into

$$\frac{\partial \ell}{\partial a_i} = N \sum_{\mathbf{k}} \frac{p_{\mathbf{k}}}{P_{\mathbf{k}}} \frac{\partial P_{\mathbf{k}}}{\partial a_i} \quad \text{and} \quad \frac{\partial \ell}{\partial c_i} = N \sum_{\mathbf{k}} \frac{p_{\mathbf{k}}}{P_{\mathbf{k}}} \frac{\partial P_{\mathbf{k}}}{\partial c_i}$$

in which p is the observed proportion for **k**.

Bock & Lieberman (1970) suggest approximating the matrix of second derivatives with

$$\frac{\partial^2 \ell}{\partial u_i \partial v_i} \cong -N \sum_{\mathbf{k}} \frac{1}{P_{\mathbf{k}}} \frac{\partial P_{\mathbf{k}}}{\partial u_i} \frac{\partial P_{\mathbf{k}}}{\partial v_i}$$

We can make this work with Newton-Raphson directly, but not with R's minimizer?

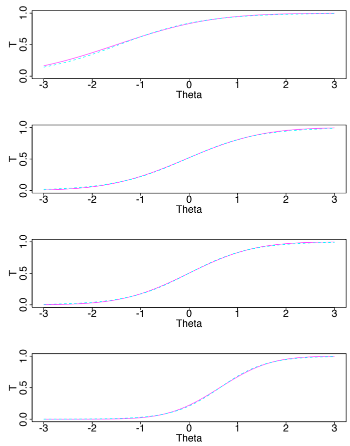
Comparing the information matrices, R's numerical hessian (red) vs. Bock & Lieberman's SSCP approximation (blue):

	a1	c1	a2	c2	a3	c3	a4	c4
a1	40 41							
c1	33 32	84 83						
a2	1 0	-3 -2	25 26					
c2	-4 -4	-10 -10	3 2	97 97				
a3	-1 1	-2 -2	3 2	-1 -1	18 18			
c3	-4 -4	-11 -11	-1 -1	-17 -17	1 0	87 87		
a4	2 1	2 2	1 2	6 6	3 2	6 6	15 15	
c4	-2 -2	-9 -9	2 2	-17 -17	2 2	-18 -19	-21 -22	62 62

The logistic model ("2PL") can be substituted for the normal ogive with minimal changes.

In the R, with the change in functions (normal to logistic) we also change to slope-threshold parameterization.

Normal ogive
(magenta, solid curves)
vs.
Logistic
(cyan, dashed curves)
for the Stouffer-Toby
data



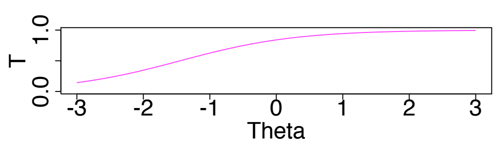
Direct maximization of the likelihood does not scale to (many) tens of items and some multiple of (many) tens of item parameters.

Bock & Aitkin (1981) proposed a solution to this problem.

Bock & Aitkin (1981) in pictures—
repeat EM-EM-EM... until convergence:
E-step (for each item) creates pseudo-data table:

Quadrature points:	-3.0	-2.5	...	0.0	0.5	...	3.0
r1	0.07	0.43	...	36.14	32.93	...	0.53
r0	0.37	1.32	...	7.45	3.67	...	0.001

M-step (for each item) fits the trace line(s):



The quadrature-based computation of the probability of each response pattern is:

$$P_{\mathbf{k}} = \sum_{q=1}^{Q} \left(\prod_{i=1}^{nitems} T_i(k_i|\theta_q) \right) \phi(\theta_q)$$

The expected frequency correct/positive/1 at each quadrature point is:

$$r_{i1q} = \frac{\sum_{\mathbf{k}} r_{\mathbf{k}} k_i \left(\prod_{i=1}^{nitems} T_i(k_i|\theta_q) \right) \phi(\theta_q)}{P_{\mathbf{k}}}$$

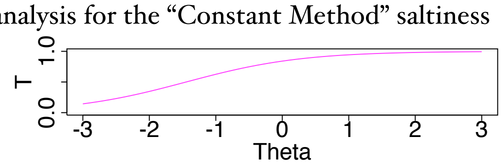
and the expected frequency for the r0 row is:

$$r_{i0q} = \frac{\sum_{\mathbf{k}} r_{\mathbf{k}} (1-k_i) \left(\prod_{i=1}^{nitems} T_i(k_i|\theta_q) \right) \phi(\theta_q)}{P_{\mathbf{k}}}$$

Once the pseudo-data table is created, like

Quadrature points:	-3.0	-2.5	...	0.0	0.5	...	3.0
r1	0.07	0.43	...	36.14	32.93	...	0.53
r0	0.37	1.32	...	7.45	3.67	...	0.001

the M-step (for each item) fits the trace line(s) using the pseudo-data table as real data, and ML. Such a problem is formally identical to probit or logit analysis for the "Constant Method" saltiness data.



Standard errors are not "free," however.

The quadrature-based computation of the probability of each response pattern is:

$$P_{\mathbf{k}} = \sum_{q=1}^{Q} \left(\prod_{i=1}^{nitems} T_i(k_i|\theta_q) \right) \phi(\theta_q)$$

The expected frequency correct/positive/1 at each quadrature point is:

$$r_{i1q} = \frac{\sum_{\mathbf{k}} r_{\mathbf{k}} k_i \left(\prod_{i=1}^{nitems} T_i(k_i|\theta_q) \right) \phi(\theta_q)}{P_{\mathbf{k}}}$$

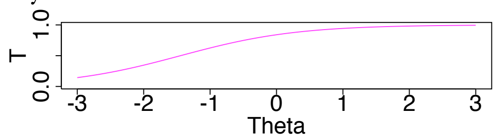
and the expected frequency for the r0 row is:

$$r_{i0q} = \frac{\sum_{\mathbf{k}} r_{\mathbf{k}} (1-k_i) \left(\prod_{i=1}^{nitems} T_i(k_i|\theta_q) \right) \phi(\theta_q)}{P_{\mathbf{k}}}$$

Once the pseudo-data table is created, like

Quadrature points:	-3.0	-2.5	...	0.0	0.5	...	3.0
r1	0.07	0.43	...	36.14	32.93	...	0.53
r0	0.37	1.32	...	7.45	3.67	...	0.001

the M-step (for each item) fits the trace line(s) using the pseudo-data table as real data, and ML. Such a problem is formally identical to probit or logit analysis for the "Constant Method" saltiness data.



Standard errors are not "free," however.