Some Notes on the Computation of IRT Scale Scores, with Various and Sundry Comments on How It's Done, and What Can and Does Go Wrong

Psychology 840, February 2007

Reasons for this exercise:

1) The IRT may be useful in its own right for some.

2) Even if not (1), this is a real-life context for:

- Univariate ML and Bayes estimation with no closed-form solution, and
- The details of Newton-Raphson, which will provide a basis for consideration of potential difficulties with multi-parameter problems.

"Characterizing theta" in IRT models:

- Context: We assume that IRT item parameters are fixed-and-known, and consider the value of the latent variable the parameter subject to statistical estimation. That's not the real estimation problem in IRT, but it was the first.

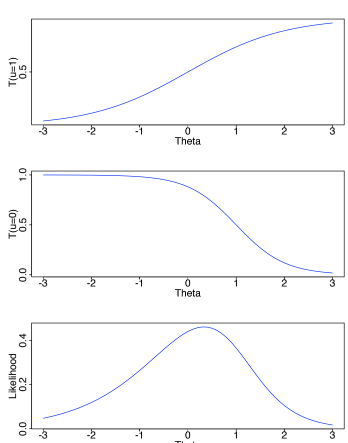
- We will consider, in historical order of precedence, which is inverse order of quality,
  - maximum likelihood (ML),
  - maximum *a posteriori* (MAP), and
  - expected *a posteriori* (EAP) estimation.

In this block, using R, we will consider only the two-parameter logistic (2PL) model, and two items, in some detail (about the math) and less detail (about the computer code).

In the next block, using C++, the code will include more complex IRT models, additional scoring schemes (an approximation to the response-pattern EAP, and direct computation of summed-score EAPs). There will be less math (which is in books).

IRT "test scoring" as considered by Lawley (1943), Lord (1952 and thereafter).

$$L = \prod_{i=1}^{nitems} T_i(u_i|\theta)$$



To find the maximum of the likelihood

$$L = \prod_{i=1}^{nitems} T_i(u_i|\theta)$$

"The drill": First, the loglikelihood

$$\ell = \sum_{i=1}^{nitems} \log T_i(u_i|\theta)$$

Then the partial derivative w.r.t. the parameter

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{nitems} \frac{1}{T_i(u_i|\theta)} \frac{\partial T_i(u_i|\theta)}{\partial \theta}$$

Must be zero at the value of the MLE

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{nitems} \frac{1}{T_i(u_i|\hat{\theta})} \frac{\partial T_i(u_i|\hat{\theta})}{\partial \theta} = 0$$

An aside about derivatives of logistics:

$$P = \frac{1}{1+\exp(-z)}$$

$$\frac{\partial P}{\partial z} = \frac{[1+\exp(-z)](0) - [-\exp(-z)](1)}{[1+\exp(-z)]^2}$$ quotient rule...

$$\frac{\partial P}{\partial z} = \frac{\exp(-z)}{[1+\exp(-z)]^2}$$ simplify...

$$\frac{\partial P}{\partial z} = \frac{1}{1+\exp(-z)} \frac{\exp(-z)}{1+\exp(-z)}$$ expand & ±1...

$$\frac{\partial P}{\partial z} = \frac{1}{1+\exp(-z)} \left(\frac{1+\exp(-z)}{1+\exp(-z)} - \frac{1}{1+\exp(-z)} \right)$$

$$\frac{\partial P}{\partial z} = P(1-P) = PQ$$

Can Mathematica do that?

For 2PL Trace Lines

For the correct/positive response:

$$T_i(1_i|\theta) = \frac{1}{1+\exp[-a_i(\theta - b_i)]}$$

$$\frac{\partial T_i(1_i|\theta)}{\partial \theta} = a_i T_i(1_i|\theta)[1 - T_i(1_i|\theta)]$$

$$\frac{\partial \log T_i(1_i|\theta)}{\partial \theta} = \frac{1}{T_i(1_i|\theta)} a_i T_i(1_i|\theta)[1 - T_i(1_i|\theta)]$$

$$\frac{\partial \log T_i(1_i|\theta)}{\partial \theta} = a_i[1 - T_i(1_i|\theta)]$$

For 2PL Trace Lines

For the incorrect/negative response:

$$T_i(0_i|\theta) = 1 - \frac{1}{1+\exp[-a_i(\theta - b_i)]}$$

$$\frac{\partial T_i(0_i|\theta)}{\partial \theta} = -a_i T_i(1_i|\theta)[1 - T_i(1_i|\theta)]$$

$$\frac{\partial \log T_i(0_i|\theta)}{\partial \theta} = -a_i \left(\frac{1}{1 - T_i(1_i|\theta)} \right) T_i(1_i|\theta)[1 - T_i(1_i|\theta)]$$

$$\frac{\partial \log T_i(0_i|\theta)}{\partial \theta} = -a_i T_i(1_i|\theta)$$

The normal equation, zero at the MLE:

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{nitems} \frac{1}{T_i(u_i|\hat{\theta})} \frac{\partial T_i(u_i|\hat{\theta})}{\partial \theta} = 0$$

For ***u*** = 1

$$\frac{\partial \log T_i(1_i|\theta)}{\partial \theta} = a_i[1 - T_i(1_i|\theta)]$$

For ***u*** = 0

$$\frac{\partial \log T_i(0_i|\theta)}{\partial \theta} = -a_i T_i(1_i|\theta)$$

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{nitems} a_i[u_i - T_i(1_i|\theta)]$$

(One cute version, of many)

(Makes it easy to do second derivative)

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{i=1}^{nitems} -a_i^2 T_i(1_i|\theta)[1 - T_i(1_i|\theta)]$$

Univariate Newton-Raphson

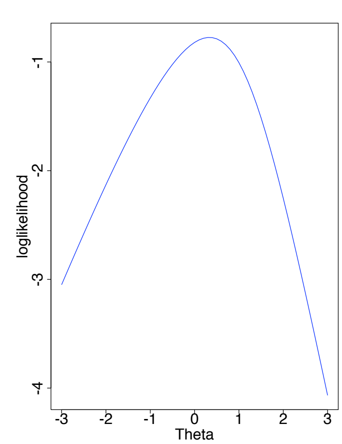
"Raphson was Newton's programmer"

$$\hat{\theta}_{next} = \hat{\theta}_{current} - \frac{\frac{\partial \ell}{\partial \theta}}{\frac{\partial^2 \ell}{\partial \theta^2}}$$

$$s.e._{\hat{\theta}} = \sqrt{\frac{1}{-\frac{\partial^2 \ell}{\partial \theta^2}}}$$

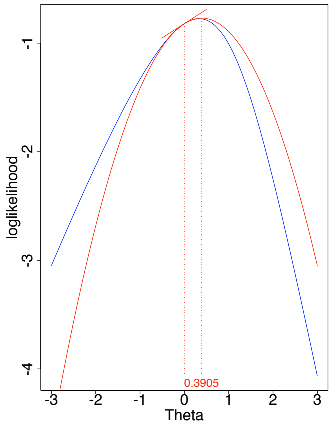
Our goal is to find the maximum of the likelihood

$$\ell = \sum_{i=1}^{nitems} \log T_i(u_i|\theta)$$



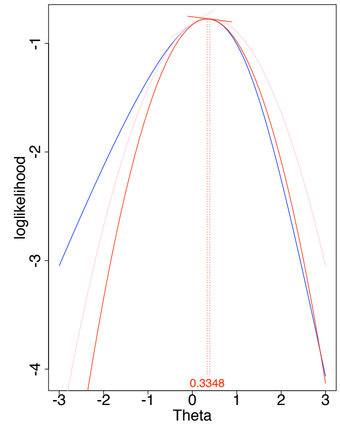
Newton-Raphson Iteration 1

$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
0.0	0.2615	-.6700



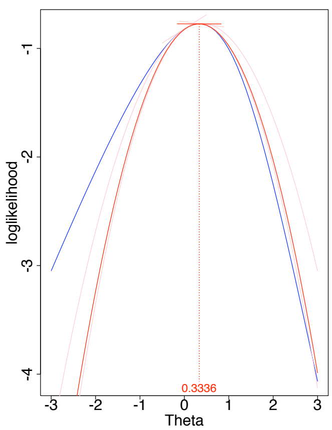
Newton-Raphson Iteration 2

$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
0.0	0.2615	-.6700
0.3905	-.0526	-.9450



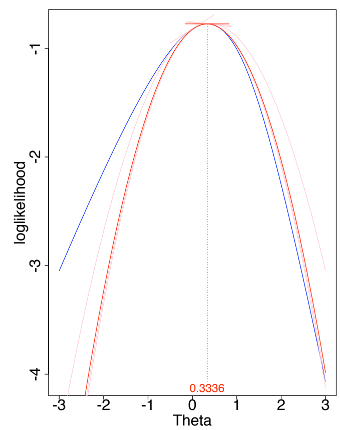
Newton-Raphson Iteration 3

$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
0.0	0.2615	-.6700
0.3905	-.0526	-.9450
0.3348	-.0011	-.9046



Newton-Raphson Iteration 4

$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
0.0	0.2615	-.6700
0.3905	-.0526	-.9450
0.3348	-.0011	-.9046
0.3336	-.0000	-.9037

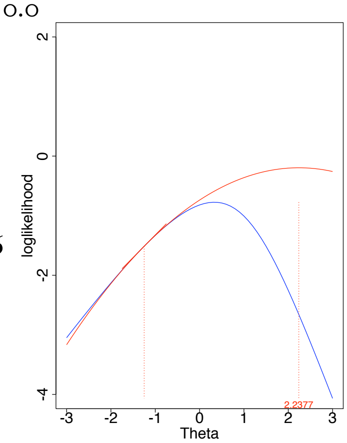


This works if the start value is “close enough”; otherwise not.

Start at -1.25, instead of 0.0

Newton-Raphson Iteration 1

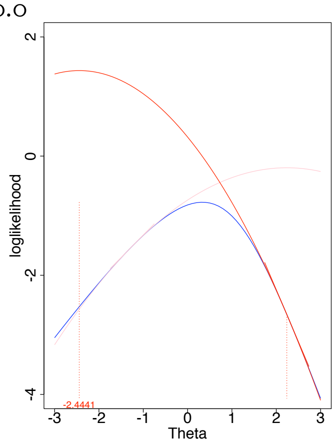
$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
-1.25	0.7553	-.2166



Start at -1.25, instead of 0.0

Newton-Raphson Iteration 2

$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
-1.25	0.7553	-.2166
2.2377	-1.7484	-.3734



"Helping" Newton-Raphson:

1) Bound the value of the second derivative away from zero (avoid divide-by-zero)*

2) Limit the step size*

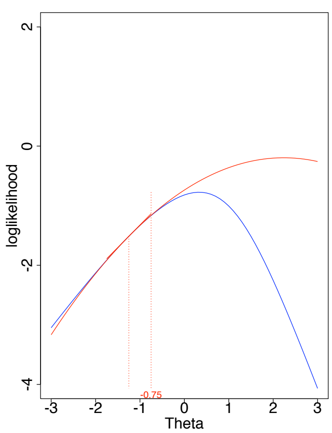
3) "Split the difference" or use the "method of false position" if it "switches sides" (first derivative changes sign) — keep it approaching from one side, to avoid divergence/oscillation.

* These require knowledge of the likely (or possible) size of the parameter.

Start at -1.25, fixes in

Newton-Raphson Iteration 1

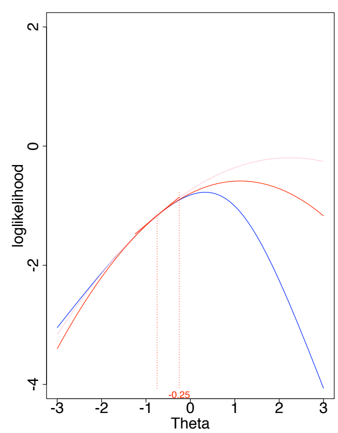
$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
-1.25	0.7553	-.2166



Start at -1.25, fixes in

Newton-Raphson Iteration 2

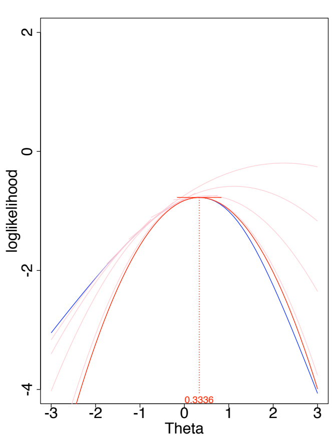
$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
-1.25	0.7553	-.2166
-0.75	0.6206	-.3317



Start at -1.25, fixes in

Newton-Raphson Iteration 6

$\hat{\theta}$	$\frac{\partial \ell}{\partial \theta}$	$\frac{\partial^2 \ell}{\partial \theta^2}$
-1.25	0.7553	-.2166
-0.75	0.6206	-.3317
-0.25	0.4105	-.5265
0.25	0.0730	-.8427
0.3366	-.0027	-.9059
0.3351	-.0014	-.9048



The "real" likelihood for IRT models includes a population distribution

$$L = \left(\prod_{i=1}^{nitems} T_i(u_i|\theta) \right) \phi(\theta)$$

$$\ell = \left(\sum_{i=1}^{nitems} \log T_i(u_i|\theta) \right) + \log[\phi(\theta)]$$

so

$$\frac{\partial \ell}{\partial \theta} = \left(\sum_{i=1}^{nitems} \frac{1}{T_i(u_i|\theta)} \frac{\partial T_i(u_i|\theta)}{\partial \theta} \right) + \frac{\partial \log[\phi(\theta)]}{\partial \theta}$$

is zero for the maximum a posteriori (MAP) estimate of theta

The derivatives for the standard normal population distribution are easy

$$\phi(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right)$$

$$\frac{\partial \log[\phi(\theta)]}{\partial\theta} = -\theta \qquad \frac{\partial^2 \log[\phi(\theta)]}{\partial\theta^2} = -1$$

so with a standard normal population distribution

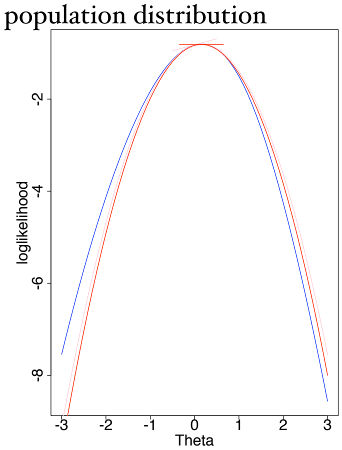
$$\frac{\partial\ell}{\partial\theta} = \left(\sum_{i=1}^{nitems} a_i[u_i - T_i(u_i|\theta)]\right) - \theta$$

$$\frac{\partial^2\ell}{\partial\theta^2} = \left(\sum_{i=1}^{nitems} -a_i^2 T_i(1_i|\theta)[1 - T_i(1_i|\theta)]\right) - 1$$

With a standard normal population distribution

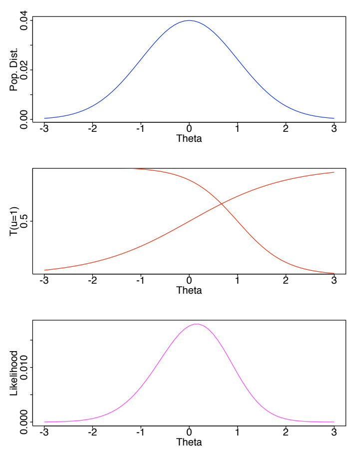
Newton-Raphson Iteration 4

$\hat{\theta}$	$\frac{\partial\ell}{\partial\theta}$	$\frac{\partial^2\ell}{\partial\theta^2}$
0.0	0.2616	-1.6700
0.1566	-.0081	-1.7757
0.1544	-.0041	-1.7741
0.1521	-.0000	-1.7725



Bock & Mislevy (1982) (re-)proposed the idea that an expected *a posteriori* (EAP; mean) estimate might be used instead of the MLE or MAP (mode).

$$L = \left(\prod_{i=1}^{nitems} T_i(u_i|\theta)\right)\phi(\theta)$$



If the likelihood (posterior if you call the population distribution a "prior") is

$$L = \left(\prod_{i=1}^{nitems} T_i(u_i|\theta)\right)\phi(\theta)$$

then the EAP is

$$EAP(\theta) = \bar{\theta} = \int_{-\infty}^{\infty}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta)\right)\phi(\theta)\right]\theta d\theta$$

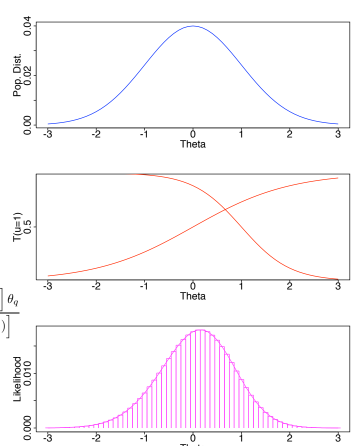
and the EAP and the posterior SD can be computed using quadrature as

$$EAP(\theta) = \bar{\theta} = \frac{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]\theta_q}{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]}$$

$$SD(\theta) = \sqrt{\frac{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right](\theta_q - \bar{\theta})^2}{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]}}$$

These computations require values of the trace lines only at a fixed set of quadrature points, so the use of more points per person can be traded (using memory storage) to compute the trace line values fewer times overall.

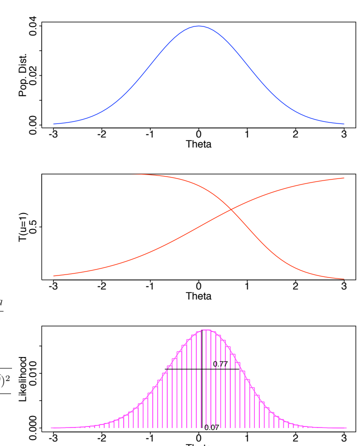
$$EAP(\theta) = \bar{\theta} = \frac{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]\theta_q}{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]}$$



Bock & co. prefer Gauss-Hermite quadrature. I think simple rectangular quadrature is more sensible.

$$EAP(\theta) = \bar{\theta} = \frac{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]\theta_q}{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]}$$

$$SD(\theta) = \sqrt{\frac{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right](\theta_q - \bar{\theta})^2}{\sum_{q=1}^{Q}\left[\left(\prod_{i=1}^{nitems} T_i(u_i|\theta_q)\right)\phi(\theta_q)\right]}}$$



Rectangular quadrature requires fewer points than one might think. Here 13 points gives the same answer to 2 decimal places (shown; actually 3 decimal places) as the 61 points of the preceding illustration.