Lecture 9: Estimating Ability Parameters Using Known Item Parameters / Item and Test Information

(Baker & Kim (2004), Chapter 3 & Embretson & Reise (2000), Chapter 7)

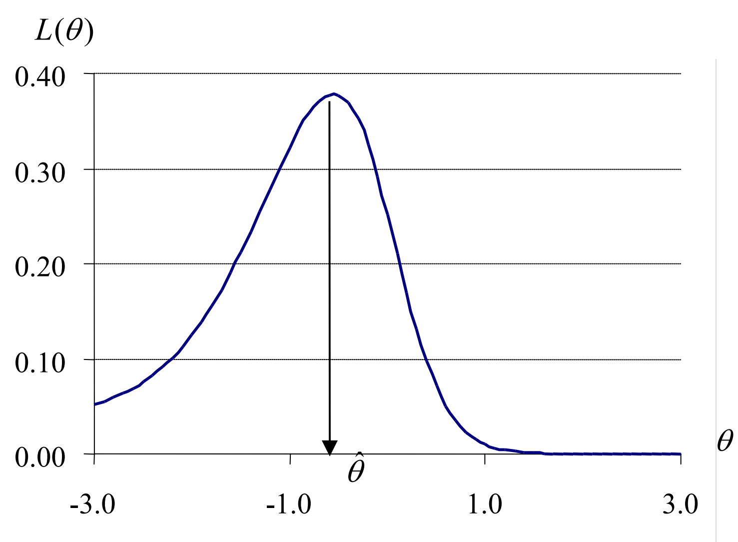
A second aspect of estimation requires determining ability estimates assuming the item parameters are known. To estimate examinee's latent ability via maximum likelihood estimation, three assumptions are made -- 1) The values of item parameters which are dichotomously scored are known (recall that we assume the examinee's ability known for item parameter estimation), 2) The examinees are independent and ability can be estimated one examinee at a time (examinee by examinee basis, similarly item by item basis for item parameter estimation in Chapter 2), and 3) All items in the test are modeled by the same IRT model (either normal ogive or logistic – 1PL, 2PL, or 3PL). Note that examinee's ability cannot be measured by a single item.

Let u_{ij} denote the item score for person j on item i, $u=1$ implies correct, while $u=0$ implies incorrect. A separate likelihood function (with respect to θ) can be computed for each person based on the item response pattern observed for that person $U_j = (u_{1j}, u_{2j}, ..., u_{nj})$:

$$\text{Prob}(U_j|\theta) = \prod_{i=1}^{n} P_i^{u_{ij}}(\theta_j) Q_i^{1-u_{ij}}(\theta_j)$$

If $P_i(\theta_{ij}) = P_{ij}$ and $Q_i(\theta_{ij}) = Q_{ij}$, then

$$L(\theta) = \text{Prob}(U_j|\theta) = \prod_{i=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}.$$

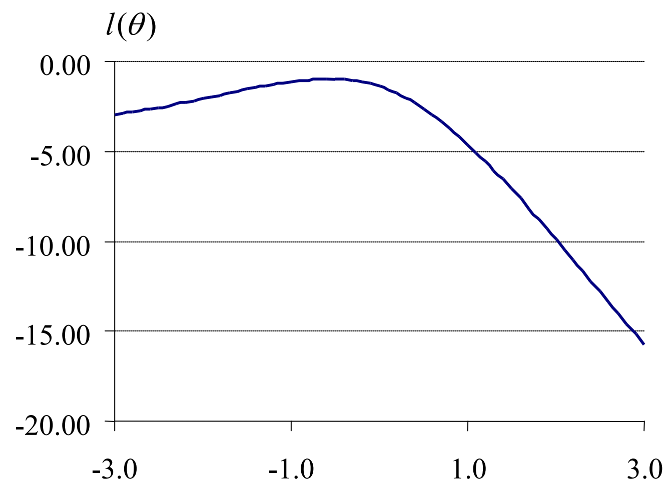


The natural log of this expression then becomes:

$$l(\theta) = \log L(\theta) = \log \text{Prob}(U_j|\theta) = \sum_{i=1}^{n}\left[u_{ij}\log P_{ij} + (1-u_{ij})\log Q_{ij}\right]$$

As for the item parameters, a maximum likelihood estimate of θ for an examinee having response pattern U_j is found by finding the location where the derivative of the log-likelihood function with respect to θ is 0. The derivative of the log-likelihood is given by:

$$\frac{\partial l}{\partial \theta_j} = \sum_{i=1}^{n} u_{ij}\frac{1}{P_{ij}}\frac{\partial P_{ij}}{\partial \theta_j} + \sum_{i=1}^{n}(1-u_{ij})\frac{1}{Q_{ij}}\frac{\partial Q_{ij}}{\partial \theta_j}.$$



To find the point where $\frac{\partial l}{\partial \theta_j} = 0$, we use the Newton-Raphson procedure, because the derivatives of the log-likelihood is nonlinear with respect to θ. As a result the Newton-Raphson updates are of the following form:

$$\left[\hat{\theta}_j\right]_{t+1} = \left[\hat{\theta}_j\right]_t - \left[\frac{\partial^2 l}{\partial \theta_j^2}\right]_t^{-1} \cdot \left[\frac{\partial l}{\partial \theta_j}\right]_t \qquad (1)$$

<u>Maximum Likelihood Estimation for the Normal Ogive and Logistic Models</u>

We just look at the relevant derivatives for the 2PL model, since the procedure is the same for the other models. Baker & Kim (2004) also presents derivations for the normal ogive and 3PL models.

We use the following two results to obtain the first and second partial derivatives:

$$\frac{\partial P_{ij}}{\partial \theta_j} = \lambda_i P_{ij} Q_{ij}; \quad \frac{\partial Q_{ij}}{\partial \theta_j} = -\lambda_i P_{ij} Q_{ij}$$

Then

$$\frac{\partial l}{\partial \theta_j} = \sum_{i=1}^{n} u_{ij} \frac{1}{P_{ij}} (\lambda_i P_{ij} Q_{ij}) + \sum_{i=1}^{n} (1-u_{ij}) \frac{1}{Q_{ij}} (-\lambda_i P_{ij} Q_{ij})$$

$$= \sum_{i=1}^{n} \lambda_i (u_{ij} - P_{ij}) = \sum_{i=1}^{n} \lambda_i u_{ij} - \sum_{i=1}^{n} \lambda_i P_{ij} = \sum_{i=1}^{n} \lambda_i W_{ij} \left(\frac{u_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right)$$

and

$$\frac{\partial^2 l}{\partial \theta_j^2} = \frac{\partial}{\partial \theta_j} [\sum_{i=1}^{n} \lambda_{ij} (u_{ij} - P_{ij})]$$

$$= \frac{\partial}{\partial \theta_j} [\sum_{i=1}^{n} \lambda_i u_{ij}] - \frac{\partial}{\partial \theta_j} [\sum_{i=1}^{n} \lambda_i P_{ij}]$$

$$= -\sum_{i=1}^{n} \lambda_i^2 P_{ij} Q_{ij} = -\sum_{i=1}^{n} \lambda_i^2 W_{ij}$$

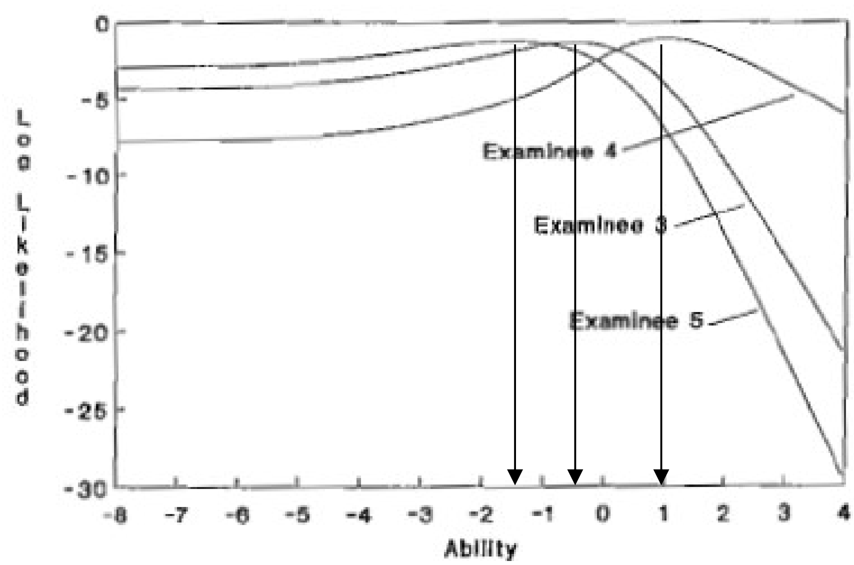
Note that the Urban-Muller weights $W_j = \dfrac{h_j^2}{P_j Q_j}$ for normal ogive model and $W_j = P_j Q_j$ for logistic model. Logistic model is preferred over the normal ogive model for computational convenience. The second derivative of *l* with respect to θ_j does not contain the observed data as was the case when estimating the item parameters. Substituting these first and second derivatives into the Newton-Raphson equation above (1), it solves for the value of $\hat{\theta}_j$ for each examinee iteratively.

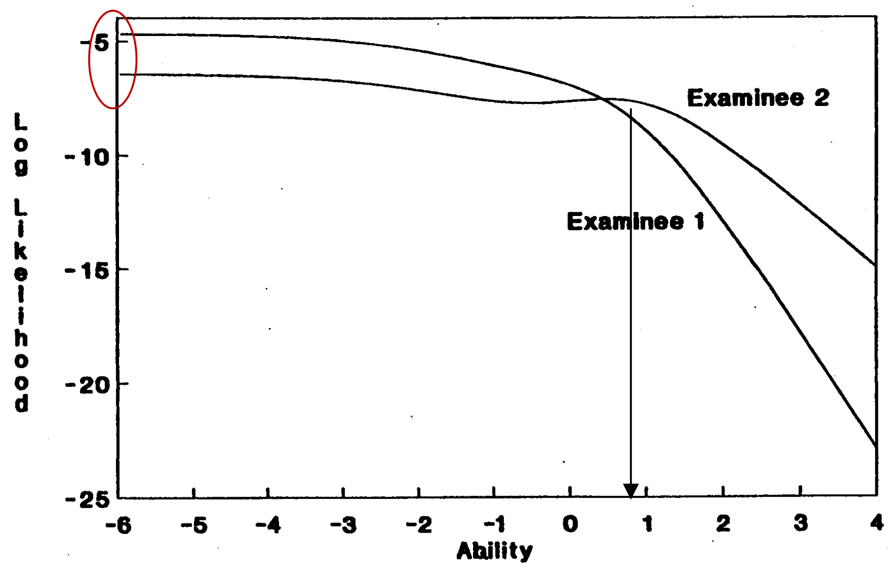
Once the maximum likelihood estimate is found, a standard error for $\hat{\theta}_j$ can be computed from the inverse of the second derivatives at that θ:

$$SE(\hat{\theta}_j) = \sqrt{S^2_{\hat{\theta}_j}} = \sqrt{\frac{1}{\sum_{i=1}^{n} \lambda_i^2 W_{ij}}}.$$

<u>Example:</u>

	Item Parameters			Examinee Item Responses				
Item	a_i	b_i	c_i	1	2	3	4	5
1	1.27	1.19	0.10	1	1	0	0	0
2	1.34	0.59	0.15	1	0	0	1	0
3	1.14	0.15	0.15	1	1	0	1	0
4	1.00	-0.59	0.20	0	0	1	1	0
5	0.67	-2.00	0.01	0	0	1	1	1





Item Information Function

Using the more traditional notation for the 2PL model (α s and β s as opposed to λ s and ζ s), the standard error is written as:

$$SE(\hat{\theta}_j) = \sqrt{S^2_{\hat{\theta}_j}} = \sqrt{\frac{1}{\sum_{i=1}^{n} \lambda_i^2 W_{ij}}} = \sqrt{\frac{1}{\sum_{i=1}^{n} \alpha_i^2 W_{ij}}}.$$

An attractive aspect of this quantity is that it is not dependent on the actual response pattern observed for the examinee. It is, however, dependent on θ and characteristics (i.e., item parameter estimates) of the specific test items administered to the examinee.

Rasch model: $$SE(\hat{\theta}_j) = \sqrt{\frac{1}{\sum_{i=1}^{n} W_{ij}}}$$

3PL: $$SE(\hat{\theta}_j) = \sqrt{\frac{1}{\sum_{i=1}^{n} \alpha_i^2 W_{ij} \left[\frac{P^*_{ij}}{P_{ij}}\right]^2}} \quad \text{where } P^*_{ij} = \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1 + \exp[\alpha_i(\theta_j - \beta_i)]}.$$

Because these standard errors are not dependent on the response pattern, we can, for existing test (assuming we know the item parameters), determine what the standard errors will be for examinees at varying levels of θ. In IRT, we report this through an *information function*, which is the square of the inverse of the standard error evaluated across all levels of θ.

An item information function is written as:

$$I_i(\theta) = \frac{[P_i^{'}(\theta)]^2}{P_i(\theta)Q_i(\theta)},$$

where $P_i^{'}(\theta) = \dfrac{\partial P_{ij}}{\partial \theta_j}$.

Notice that this is essentially just another way of writing the square of the inverse of the standard error formulas given above. In the normal ogive model this evaluates to

$$I_i(\theta) = \alpha_i^2 \frac{[h_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

while in the logistic models, it is computed as:

- 2PL: $I_i(\theta) = \alpha_i^2 P_i(\theta)Q_i(\theta)$
- Rasch: $I_i(\theta) = P_i(\theta)Q_i(\theta)$
- 3PL: $I_i(\theta) = \alpha_i^2 P_i(\theta)Q_i(\theta)\left[\dfrac{P_i^{*}(\theta)}{P_i(\theta)}\right]^2$

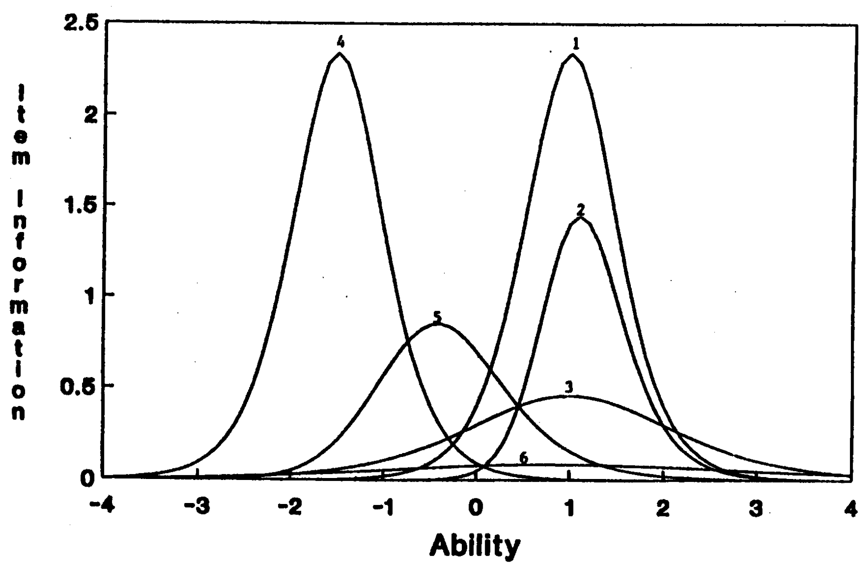
The maximum of each information function can also be uniquely expressed in terms of the item parameters:

- Normal Ogive $\max[I_i(\theta)] = \alpha_i^2 \dfrac{.3989^2}{(.5)(.5)} = .64\alpha_i^2$
- 2PL $\max[I_i(\theta)] = \alpha_i^2 (.5)(.5) = .25\alpha_i^2$
- Rasch $\max[I_i(\theta)] = (.5)(.5) = .25$
- 3PL $\max[I_i(\theta_{\max})]$, where $\theta_{\max} = \beta_i + \dfrac{1}{\alpha_i}\log\left[\dfrac{1+\sqrt{1+8c_i}}{2}\right]$.

Cramer (1946) shows that maximum likelihood estimates (like $\hat{\theta}$) have an asymptotically normal distribution, with mean = θ and variance = $1/I(\theta)$.

Some examples of item information functions are given below.

	Item Parameter		
Test item	b_i	a_i	c_i
1	1.00	1.80	0.00
2	1.00	1.80	0.25
3	1.00	0.80	0.00
4	-1.50	1.80	0.00
5	-0.50	1.20	0.10
6	0.50	0.40	0.15



Test information functions

A test information function is written as

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) = \sum_{i=1}^{n} \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

The attractive thing about the test information function is that it is a simple sum of the item information functions. As a result, it is very simple to evaluate what any one item will contribute to test information. This becomes useful in deciding how to construct a test. For example, assuming we have certain ability regions in which we want precise ability estimates, we can translate that goal into a target information function, and then add items to the test so as to closely match that function.