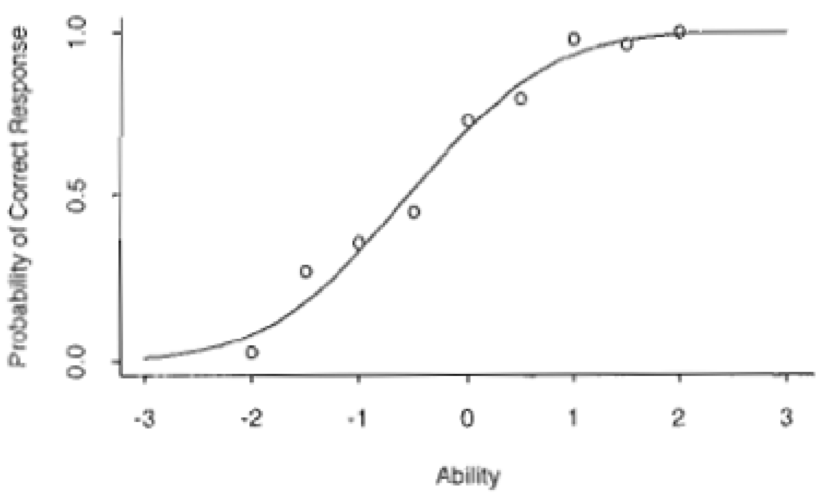
Lecture 5: Estimating the Parameter of Item Characteristic Curve

(Baker & Kim (2004), Chapter 2 / Embretson & Reise (2000), Chapters 4, 8)

Assume all examinee abilities are known. Because θ is continuous, we might think of splitting the θ scale up into k different groups: $\theta_1, \theta_2, \theta_3, ..., \theta_k$, so that we have $f_1, f_2, f_3, ..., f_k$ examinees in each group. For each group, we can compute the number of persons that answered the item correctly: $r_1, r_2, r_3, ..., r_k$. Then the proportion correct for the item at each ability level, $\theta_1, \theta_2, \theta_3, ..., \theta_k$, is $\frac{r_1}{f_1}, \frac{r_2}{f_2}, \frac{r_3}{f_3}, ..., \frac{r_k}{f_k}$. Because the only possible outcomes for an item are 0 or 1, and we also assume local independence, we can think of each of theses proportions as the outcome of n binomial experiment with mean = P_j and $n = f_j$. Also the true variance of scores at a fixed θ_j is $\frac{P(\theta_j)Q(\theta_j)}{f_j}$.

Our goal is to find the normal ogive or logistic model that best fits these proportions:



There are various ways of defining what we mean by a best fitting curve to this set of points. The most common is what is called a maximum likelihood solution. To determine maximum likelihood estimates (when abilities are known) we rely on the Newton-Raphson procedure.

Likelihood Function

The goal in maximum likelihood estimation is to find the parameter estimates for the normal ogive or logistic models that were most likely to have contributed to the observed data. Assuming R represents the complete set of responses observed for the test items, the likelihood of the observed data as a function of the unknown item parameters can be written as:

$$L(\theta) = \text{Prob}(R) = \prod_{j=1}^{k} \frac{f_j!}{r_j!(f_j - r_j)!} P_j^{r_j} Q_j^{f_j - r_j},$$

where P_j and Q_j denote the probability of correct and incorrect responses at θ_j. We can also evaluate the natural log of this expression, which has a strictly increasing relationship with the likelihood:

$$l(\theta) = \log(\text{Prob}(R)) = const + \sum_{j=1}^{k} r_j \log P_j + \sum_{j=1}^{k} (f_j - r_j) \log Q_j.$$

We want to find values for ζ and λ in the normal ogive or logistic models that make L as large as possible.

<u>Newton-Raphson Method</u>

The Newton-Raphson procedure is a common procedure for finding parameter estimates that maximize a continuous function. The basic idea is to conduct an iterative search to determine the location at which the derivative of the log-likelihood with respect to both parameters is equal to 0. In other words, we want to find the unique values of ζ and λ denoted $\hat{\zeta}$ and $\hat{\lambda}$ such that

$$\frac{\partial L}{\partial \zeta} = \sum_{j=1}^{k} f_j \frac{h_j^2}{P_j Q_j} \left[\frac{p_j - P_j}{h_j} \right] = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_{j=1}^{k} f_j \frac{h_j^2}{P_j Q_j} \left[\frac{p_j - P_j}{h_j} \right] \theta_j = 0$$

for the normal ogive model, and

$$\frac{\partial L}{\partial \zeta} = \sum_{j=1}^{k} f_j (p_j - P_j) = 0$$

$$\frac{\partial L}{\partial \lambda} = \sum_{j=1}^{k} f_j (p_j - P_j) \theta_j = 0$$

for the logistic model. The problem is that we can't evaluate these derivatives because the P_j s depend on the values of unknown item parameters and are nonlinear. So we use a Taylor expansion.

Let

$$\hat{\zeta}_{D+1} = \hat{\zeta}_D + \Delta\hat{\zeta}_D$$
$$\hat{\lambda}_{D+1} = \hat{\lambda}_D + \Delta\hat{\lambda}_D$$

represent how we intend to update our parameter estimates when moving from iteration D to iteration $D+1$. We want to determine $\hat{\zeta}_D$ and $\hat{\lambda}_D$.

Applying the Taylor expansion results in:

$$0 = \frac{\partial L}{\partial \zeta}(\zeta, \lambda) = \frac{\partial L}{\partial \zeta}(\hat{\zeta}_1, \hat{\lambda}_1) + \Delta\zeta_1 \frac{\partial^2 L}{\partial \zeta^2}(\hat{\zeta}_1, \hat{\lambda}_1) + \Delta\lambda_1 \frac{\partial^2 L}{\partial \zeta \partial \lambda}(\hat{\zeta}_1, \hat{\lambda}_1) + \text{higher order terms}$$
$$0 = \frac{\partial L}{\partial \lambda}(\zeta, \lambda) = \frac{\partial L}{\partial \lambda}(\hat{\zeta}_1, \hat{\lambda}_1) + \Delta\lambda_1 \frac{\partial^2 L}{\partial \lambda^2}(\hat{\zeta}_1, \hat{\lambda}_1) + \Delta\zeta_1 \frac{\partial^2 L}{\partial \zeta \partial \lambda}(\hat{\zeta}_1, \hat{\lambda}_1) + \text{higher order terms}$$

To simplify these expressions we can denote:

$$L_1 = \frac{\partial L}{\partial \zeta}(\hat{\zeta}_1, \hat{\lambda}_1) \qquad L_2 = \frac{\partial L}{\partial \lambda}(\hat{\zeta}_1, \hat{\lambda}_1)$$
$$L_{11} = \frac{\partial^2 L}{\partial \zeta^2}(\hat{\zeta}_1, \hat{\lambda}_1) \qquad L_{22} = \frac{\partial^2 L}{\partial \lambda^2}(\hat{\zeta}_1, \hat{\lambda}_1)$$
$$L_{12} = \frac{\partial^2 L}{\partial \zeta \partial \lambda}(\hat{\zeta}_1, \hat{\lambda}_1) \quad \text{and} \quad L_{21} = L_{12} = \frac{\partial^2 L}{\partial \zeta \partial \lambda}(\hat{\zeta}_1, \hat{\lambda}_1)$$

so that the two equations to be solved simultaneously are:

$$-L_1 = L_{11}\Delta\hat{\zeta} + L_{12}\Delta\hat{\lambda}$$
$$-L_2 = L_{21}\Delta\hat{\zeta} + L_{22}\Delta\hat{\lambda}$$

If we rewrite those two equations in the form of matrix algebra, a solution for $\Delta\hat{\zeta}$ and $\Delta\hat{\lambda}$ becomes:

$$\begin{bmatrix} \Delta\hat{\zeta} \\ \Delta\hat{\lambda} \end{bmatrix} = -\begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$$

so that our next estimates of $\hat{\zeta}$ and $\hat{\lambda}$ are therefore obtained as:

$$\begin{bmatrix} \hat{\zeta} \\ \hat{\lambda} \end{bmatrix}_{D+1} = \begin{bmatrix} \hat{\zeta} \\ \hat{\lambda} \end{bmatrix}_{D} - \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$$

Usually we stop iterating when the Δ s get very small, say less than .005.

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

This same Newton-Raphson procedure can be followed for both the normal ogive and logistic models. Of course the primary differences are how Newton-Raphson is implemented for these models and how the elements $L_1, L_2, L_{11}, L_{12}, L_{21}$, and L_{22} are computed. We've already shown how the L_1, L_2 values are computed for both the normal ogive and logistic models above. Baker & Kim (2004) works through expressions for all of the second derivatives for both models in Chapter 2. The primary advantage of the logistic model as opposed to the normal ogive is that computing all of these elements is much simpler. In the two-parameter logistic model, the second derivatives can be computed as:

$$\begin{aligned} L_{11} &= -\sum_j f_j W_j \\ L_{22} &= -\sum_j f_j W_j \theta_j^2 \\ L_{12} &= L_{21} = -\sum_j f_j W_j \theta_j \end{aligned}$$

where $W_j = P_j Q_j$. Sometimes the gradient elements are also written to include the W_j weights:

$$\begin{aligned} L_1 &= \sum_j f_j W_j \frac{(p_j - P_j)}{P_j Q_j} \\ L_2 &= \sum_j f_j W_j \frac{(p_j - P_j)}{P_j Q_j} \theta_j \end{aligned}$$

We can also derive standard errors for these parameters in both models. For the normal ogive and two-parameter logistic models, standard errors are computed as:

$$S^2_{\hat{\lambda}} = \frac{1}{\sum_{j=1}^{k} f_j W_j (\theta_j - \bar{\theta})^2} = S^2_{\hat{\alpha}}$$

$$S^2_{\hat{\zeta}} = \frac{1}{\sum_{j=1}^{k} f_j W_j} + \bar{\theta}^2 S^2_{\hat{\alpha}}$$

$$S^2_{\hat{\beta}} = \frac{1}{\alpha^2}\left[\frac{1}{\sum_{j=1}^{k} f_j W_j} + S^2_{\hat{\alpha}}(\hat{\beta} - \bar{\theta})^2\right].$$

Also, a chi-square goodness-of-fit test can be used to evaluate the fit of the model:

$$\chi^2 = \sum_{j=1}^{k} f_j W_j v_j^2,$$

where $v_j = \frac{p_j - P_j}{h_j}$ with $k-2$ degrees of freedom for the normal ogive model, and

$$\chi^2 = \sum_{j=1}^{k} f_j W_j v_j^2,$$

where $v_j = \frac{p_j - P_j}{P_j Q_j}$ with $k-2$ degrees of freedom for the logistic model.

We can follow the same general procedure when working with the three-parameter logistic model, except we need to find where the log-likelihood surface is maximized with respect to three parameters instead of two. As a result, the gradient vector becomes a 3×1 vector and the Hessian matrix a 3×3 matrix.

$$\begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix}_{D+1} = \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix}_{D} - \begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix}_{D}^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}$$

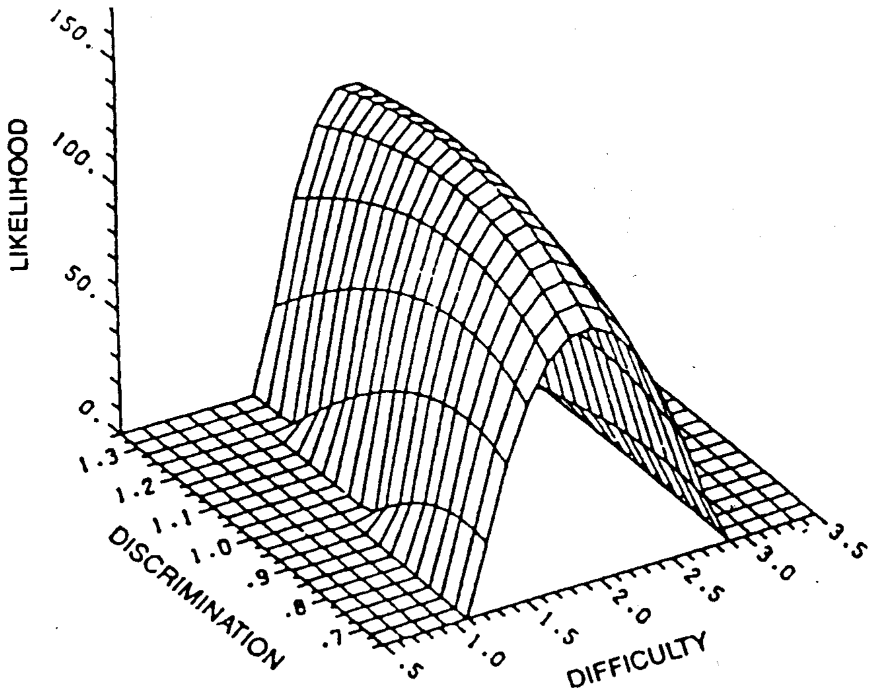
<u>Influence of the Weighting Coefficients</u>

	Normal model	Logistic model
	$\alpha = 1.0 \quad \beta = 0$	$\alpha = 1.702 \quad \beta = 0$

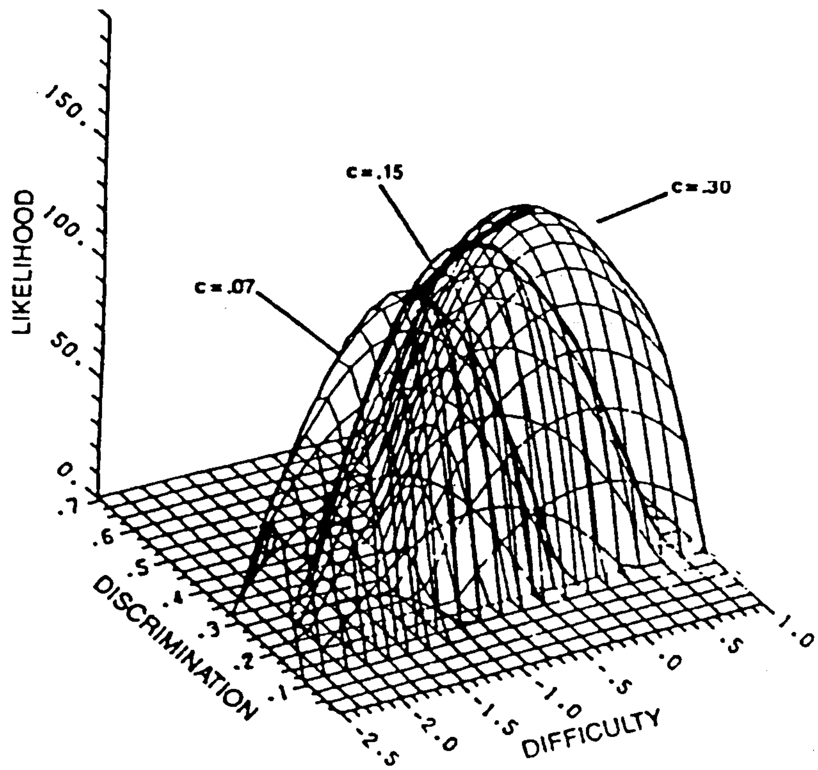
Ability score	P_j	h_j	W_j	P_j	W_j
-3.0	.0013	.0044	.0149	.0060	.0060
-2.5	.0062	.0175	.0497	.0140	.0138
-2.0	.0227	.0540	.1348	.0322	.0312
-1.5	.0668	.1295	.2690	.0722	.0670
-1.0	.1587	.2420	.4386	.1542	.1304
-.5	.3085	.3521	.5959	.2992	.2097
0	.5	.3989	.6365	.5000	.2500
.5	.6915	.3521	.5959	.7008	.2097
1.0	.8413	.2420	.4386	.8458	.1304
1.5	.9332	.1295	.2690	.9278	.0670
2.0	.9773	.0540	.1348	.9678	.0312
2.5	.9938	.0175	.0497	.9760	.0138
3.0	.9987	.0044	.0149	.9940	.0060

<u>Item Log-Likelihood Surface</u> (Baker & Kim (2004), p. 45 & p.54)

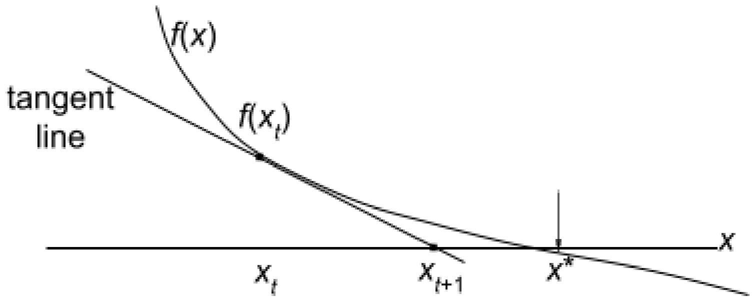
An example from the 2PL model



An example from the 3PL constructed for fixed α and β


LIKELIHOOD
150.
100.
50.
0.
c = .07
c = .15
c = .30
DISCRIMINATION
.7
.6
.5
.4
.3
.2
.1
DIFFICULTY
-2.5
-2.0
-1.5
-1.0
-.5
.0
.5
1.0

Finding the Zero of a Function: The Newton-Raphson Method



In calculus, the slope of $f(x_t)$ at x_t is given by the tangent line at x_t (i.e., $f'(x_t)$, the first derivative of f evaluated at x_t). In algebra, given two points the slope of a line can be computed using the formula: *rise*/*run*. In this setup, the two points are $(x_t, f(x_t))$ and $(x_{t+1}, 0)$. This gives us a *rise* of $f(x_t) - 0 = f(x_t)$ and a *run* of $x_t - x_{t+1}$.

Combining these results we get $f'(x_t) = \dfrac{f(x_t)}{x_t - x_{t+1}}$. Solving for x_{t+1}, we get $x_{t+1} = x_t - \dfrac{f(x_t)}{f'(x_t)}$.

So the next possible value of x, x_{t+1}, can be computed using current value x_t. Each successive approximation leads us closer to x^*. This method will converge to the zero of the function provided the initial guess x_0 is not too far from x^*, and $f(x)$ is well-behaved.