We formulate computer adaptive testing as an optimization problem. We have as our dataset the responses of m test-takers to n questions.

Let $D_{ij} = 1$ indicate that tester j answered question i correctly

We model the test taking task with a 3 parameter logistic model

$$p_i(\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

where Θ_j is the ability level of tester j

b_i is the difficulty of item i

a_i is the discrimination of item i

and c_i represents pseudo-guessing, an asymptotic minimum, i.e. $p_i(\infty) = c_i$.

We wish to find $\vec{\theta}$, $\vec{a}$, $\vec{b}$, and $\vec{c}$ such that the likelihood of D is maximized

$$\mathcal{L}(\vec{\theta}, \vec{a}, \vec{b}, \vec{c} \mid D) =$$

$$\prod_{i,j} D_{ij} * p_i(\theta_j) + \neg D_{ij} * [1 - p_i(\theta_j)]$$

We find the partial derivatives with respect to the parameters

The derivative of the standard logistic function $f(x) = \frac{1}{1+e^{-x}}$ is $f(x)f(-x)$.

Let $g(\theta_j, a_i, b_i) = a_i(\theta_j - b_i)$

Then via the chain rule

$$\partial \mathcal{L}_{ij} / \partial \theta_j =$$
$$D_{ij}(1-c_i) f(g(\theta_j, a_i, b_i)) *$$
$$f(-g(\theta_j, a_i, b_i)) a_i -$$
$$\neg D_{ij}(1-c_i) f(g(\theta_j, a_i, b_i)) *$$
$$f(-g(\theta_j, a_i, b_i)) a_i$$

$$\partial \mathcal{L}_{ij} / \partial b_i = -\partial \mathcal{L}_{ij} / \partial \theta_i$$

$$\partial \mathcal{L}_{ij} / \partial a_i =$$
$$\sum_j D_{ij}(1-c_i) f(g(\theta_j, a_i, b_i)) *$$
$$f(-g(\theta_j, a_i, b_i))(\theta_j - b_i) -$$
$$\neg D_{ij}(1-c_i) f(g(\theta_j, a_i, b_i)) *$$
$$f(-g(\theta_j, a_i, b_i))(\theta_j - b_i)$$

$$\partial \mathcal{L}_{ij} / \partial c_i - D_{ij}(1 - f(g(\ldots)))$$
$$- \neg D_{ij}(1 - f(g(\ldots)))$$

We use these partial derivatives to find a local maximum of the likehihood function

Once a_i, b_i, and c_i are estimated for all test items, given a new test-taker

with an initial guessed aptitute θ and correct response vector $\vec{d}$ we find θ' that maximizes the probability of $\vec{d}$. In the event that $\vec{d}$ is all 1's or all 0's (corresponding to a train of responses that is all correct or all incorrect then $\theta' = \theta - f$, where f is a hyper-parameter

To recommend an item given a

initial guess of θ we use a maximum information gain criteria, i.e., we choose the item with an estimated correct probability of correct response that is closest to 0.5

Lastly, we can simulate adaptive testing relative to the test set to determine the expected error after a given number of questions

Two stopping criteria come to mind, we can terminate the adaptive test when the expected error shrinks sufficently or we may terminate when $\Delta\theta$ shrinks sufficently.

We use $E[\vec{\theta}]$ from the training data as the initial guess for a new test-taker.