

Lecture 3: IRT models – Normal Ogive vs. Logistic Models

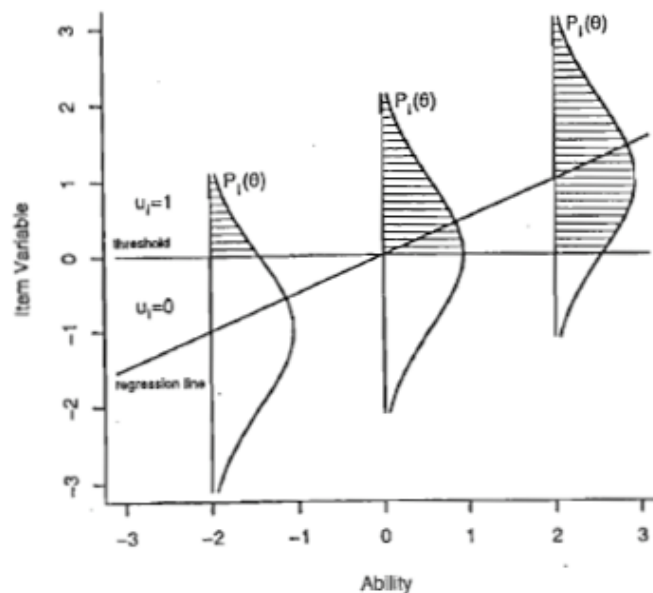
(Baker & Kim (2004), Chapter 1 & Embretson & Reise (2000), Chapters 3-4)

IRT models assume a particular functional relationship between the latent trait or ability measured by the test, denoted θ , and the probability of correct response to the item, $P_i(\theta)$. The latent ability scale for θ ranges from $-\infty$ to $+\infty$.

- **Normal Ogive models**

The normal ogive shape is typically justified as follows:

- For any item, an item variable Γ_i , representing a propensity to answer the item correctly, is a linear function of θ .
- At each θ_j , there exists a conditional distribution of Γ_i that is normal with mean $\mu_j|\theta_j$ and variance $\sigma_j^2|\theta_j$.
- The variance $\sigma_j^2|\theta_j$ is assumed constant over all θ .
- Each item is also associated with a critical value γ_i , such that if $\Gamma_i > \gamma_i$, the item is answered correctly, and if not the item is answered incorrectly. The probability of correct response at a given θ is determined by the area under the normal curve above γ_i at that θ .



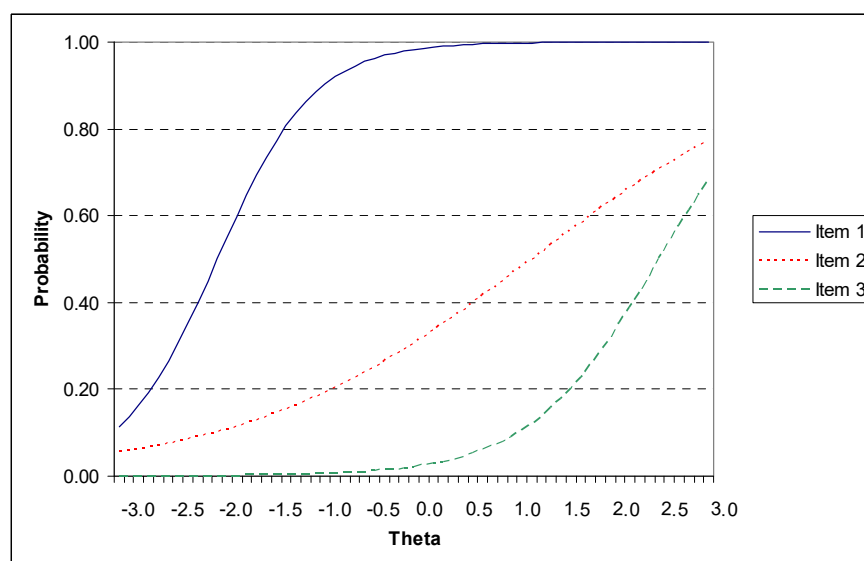
In a normal ogive model in IRT, parameters analogous to the slope of the regression line and threshold γ_i control the slope and location of the item characteristic curve (ICC). The cumulative normal evaluated at a point θ_j is then computed as:

$$\Phi(Z) = \int_{-Z_{ij} = -(\theta_j - \mu_i) / \sigma_i}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dz.$$

The normal ogive model uses $\beta_i = \mu_i$ and $\alpha_i = 1/\sigma_i$ as parameters defining the form of the normal ogive ICC. Then the probability of correct response to an item i for a person having ability level θ_j is written as: $P_i(\theta_j) = \int_{-Z_{ij} = -\alpha_i(\theta_j - \beta_i)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dz.$

Example:

Item	1	2	3
α_i	1.2	0.4	0.9
β_i	-2.0	1.2	2.5



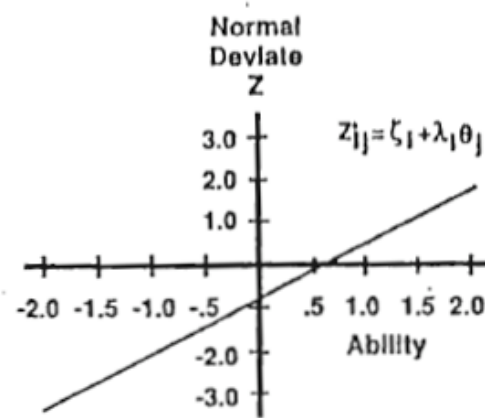
Notice that the β_i parameter, which controls the location of the ICC, also corresponds to the θ value at which exactly 50% of the examinees answer the item correctly. The α_i parameter controls the slope of the ICC. In fact, when θ is close to being equal to β_i , the ICC is nearly linear, with slope equal to $\alpha_i / \sqrt{2\pi}$ (Lord & Novick, 1968, p. 368).

With respect to the normal distribution, $\alpha_i = 1/\sigma_i$ and $\beta_i = \mu_i$, so that

$$Z_{ij} = \frac{\theta_j - \mu_i}{\sigma_i} = \alpha_i(\theta_j - \beta_i).$$

Example: Item characteristic curve based on Normal Ogive Model, $\alpha_i = 1.2$ and $\beta_i = 0.6$ (Baker & Kim, 2004, p.10)

θ_j	$\theta_j - \beta_i$	Z_{ij}	$P_i(\theta_j)$
-2.0	-2.6	-3.12	.0003
-1.5	-2.1	-2.52	.0059
-1.0	-1.6	-1.92	.0274
-.5	-1.1	-1.32	.0934
0	-.6	-.72	.2358
.5	-.1	-.12	.4522
1.0	+.4	+.48	.6844
1.5	+.9	+1.08	.8599
2.0	+1.4	+1.68	.9535



This can be advantageous in thinking about estimation as we all have familiarity with how linear regression functions can be estimated.

Logistic Models

A logistic model has the following general form:

$$P_i(\theta_j) = \frac{\exp[Z_{ij}]}{1 + \exp[Z_{ij}]},$$

where $Z_{ij} = \alpha_i^*(\theta_j - \beta_i)$. In logistic models, Z_{ij} is referred to as the logit, β_i is a location (difficulty) parameter, and α_i^* is the discrimination or scale parameter equal to the reciprocal of the standard deviation of the logistic function.

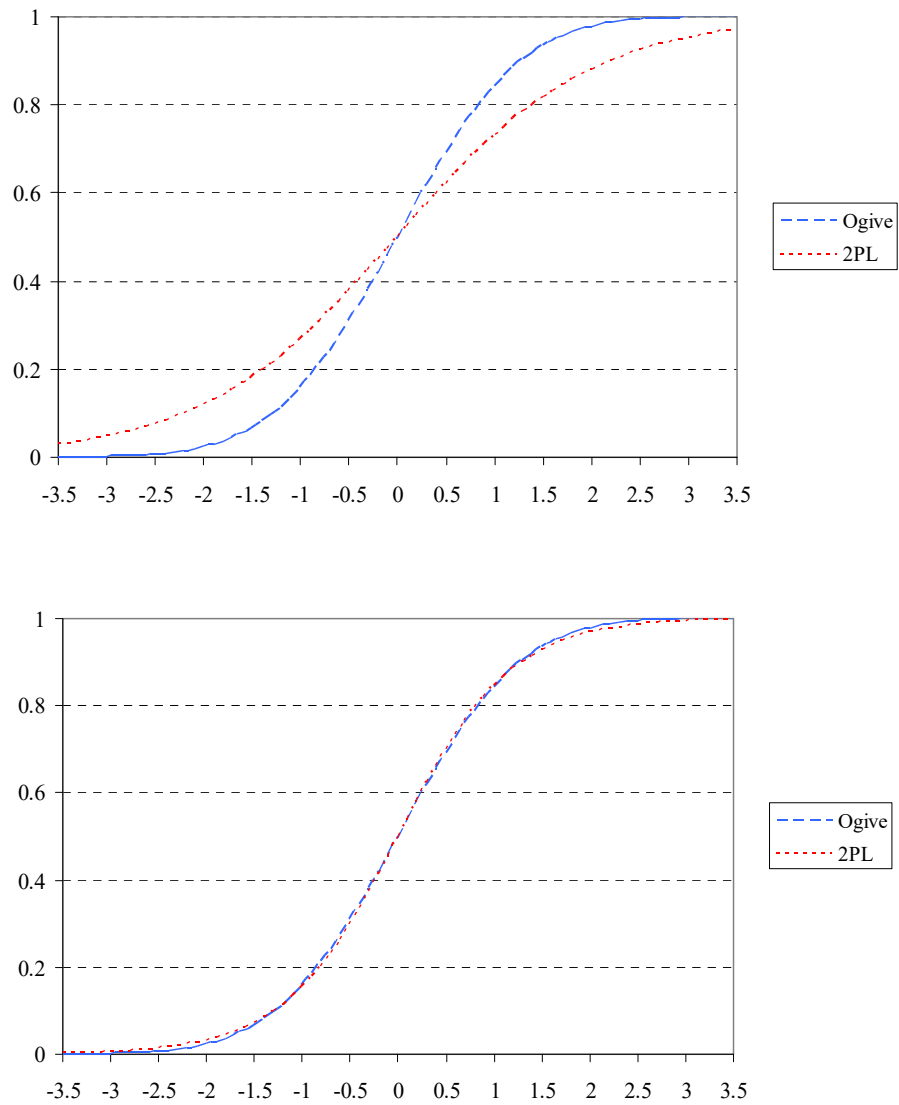
Example: Item characteristic curve based on Logistic Model, $\alpha_i^* = 1.2$ and $\beta_i = 0.6$ (p.15)

θ_j	$\theta_j - \beta_i$	Z_{ij}	$P_i(\theta_j)$
-2.0	-2.6	-3.12	.0423
-1.5	-2.1	-2.52	.0745
-1.0	-1.6	-1.92	.1279
-.5	-1.1	-1.32	.2108
0	-.6	-.72	.3274
.5	-.1	-.12	.4700
1.0	+.4	+.48	.6178
1.5	+.9	+1.08	.7465
2.0	+1.4	+1.68	.8429

The normal ogive and logistic parameterizations are very similar when the logit in the logistic model is multiplied by $D = 1.702$. In fact, the following relationship can be established:

$$|\Phi(Z_{ij}) - \Psi(1.702Z_{ij})| < .01, \text{ for } -\infty \leq \theta_j \leq \infty,$$

where Ψ denotes the logistic function and Φ the normal ogive function. Equivalently, the discrimination parameter in the logistic model can be multiplied by 1.702 to result in a close match to the normal ogive model ($\alpha = 1.702\alpha^*$).



An attractive feature of the logistic model is that it is in closed form, and thus does not require evaluating an integral. The logistic function is also associated with the log-odds of getting the item correct:

$$\ln \left[\frac{P_i(\theta_j)}{Q_i(\theta_j)} \right] = \alpha_i(\theta_j - \beta_i)$$

As with the normal deviate in the normal ogive model, we refer to $\alpha_i(\theta_j - \beta_i)$ as the logistic deviate when considering the logistic model. Unlike the normal deviate, the logistic deviate has a non-linear association with θ .

The model above is more specifically called the two parameter logistic (2PL) model, and is attributed to Birnbaum (1968):

$$P_i(\theta_j) = \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1 + \exp[\alpha_i(\theta_j - \beta_i)]}.$$

Rasch or one parameter logistic (1PL) model

$$P_i(\theta_j) = \frac{\exp[(\theta_j - \beta_i)]}{1 + \exp[(\theta_j - \beta_i)]}$$

Rasch model has nice statistical properties which make it easier to estimate than the 2PL. The main problem in IRT estimation is to estimate the item (structural) parameters in the presence of unknown examinee ability (incidental) parameters. The Rasch model is attractive because it has a simple sufficient statistic for examinee ability – the *total test score*. As a result, test scores can substitute for the latent abilities when estimating the item parameters. Another attractive feature of the Rasch model is that it possesses a property called ***specific objectivity*** (i.e., item-free & examinee-free measure).

Specific objectivity implies that “1) comparisons between persons are invariant over the specific items used to measure them, and 2) comparisons between items are invariant over the specific persons used to calibrate them.”

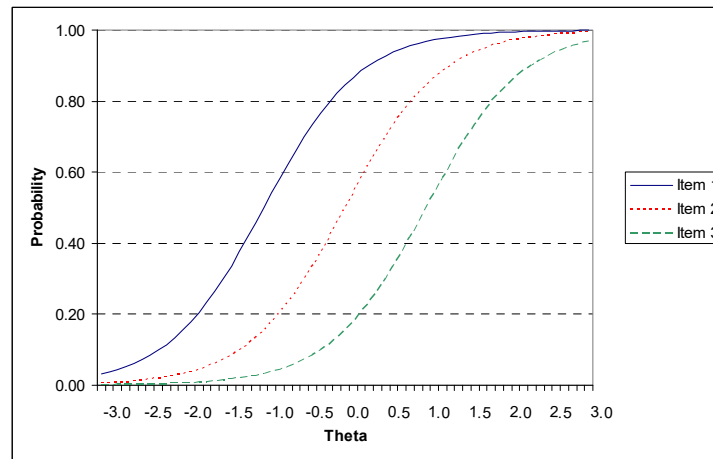
$$1) \ln \left[\frac{P_i(\theta_1)}{Q_i(\theta_1)} \right] - \ln \left[\frac{P_i(\theta_2)}{Q_i(\theta_2)} \right] = \theta_1 - \theta_2.$$

$$2) \ln \left[\frac{P_1(\theta_j)}{Q_1(\theta_j)} \right] - \ln \left[\frac{P_2(\theta_j)}{Q_2(\theta_j)} \right] = \beta_2 - \beta_1.$$

The primary disadvantage of the Rasch model relative to the 2PL is that it is not able to account for items that may vary with respect to their discriminating power.

Example:

Item	1	2	3
β_i	-1.0	0.0	1.0



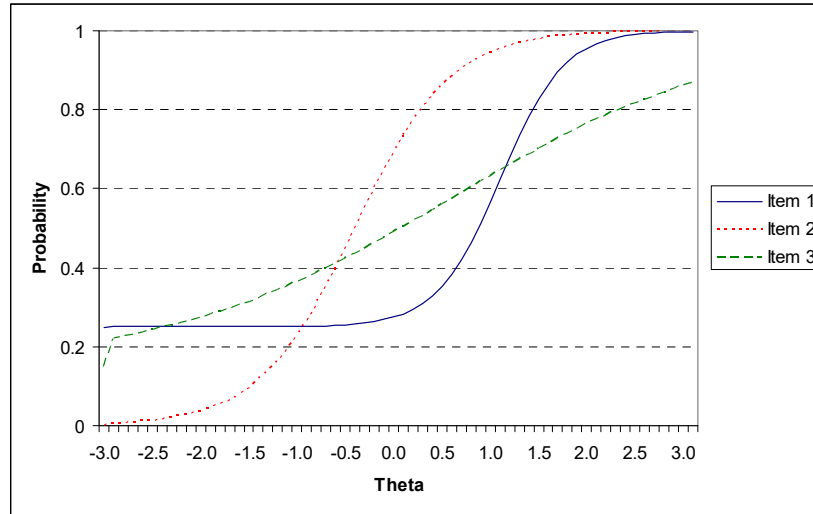
Three parameter logistic model (3PL) – Birnbaum’s Three parameter model

$$P_i(\theta_j) = c_i + (1 - c_i) \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1 + \exp[\alpha_i(\theta_j - \beta_i)]}$$

The parameter c_i is sometimes called a “guessing” parameter, but more generally defines the lower asymptote of the ICC for the 3PL. Although attractive from a practical standpoint for multiple choice items, the guessing parameter is difficult to estimate with any accuracy. Large standard errors will be observed whenever the difficulty and discrimination parameters assume extreme values.

Example:

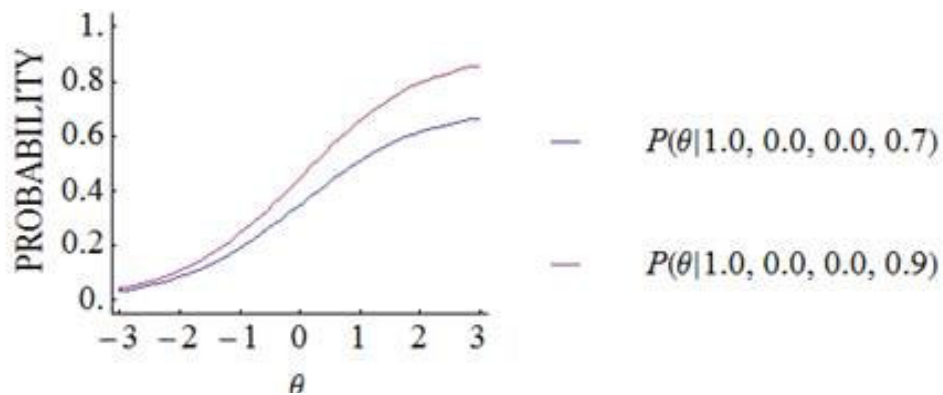
Test Item	Item Parameter		
	β_i	α_i	c_i
1	1.00	1.80	0.25
2	-0.50	1.20	0.00
3	0.50	0.40	0.15

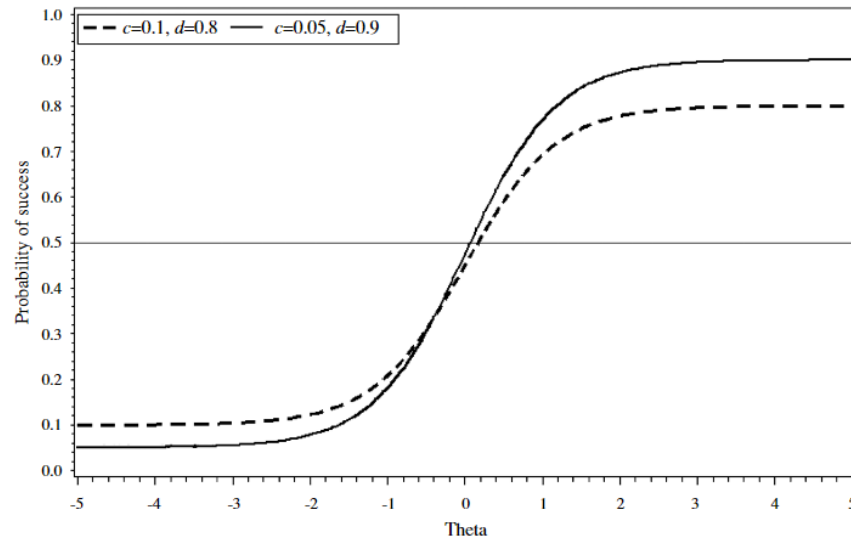


Four parameter logistic (4PL) model

Barton and Lord (1981) introduced an upper asymptote parameter, expressed by the lowercase d , into the 3PL model, resulting in the 4PL model:

$$P_i(\theta_j) = c_i + \frac{(d - c_i)}{1 + \exp[-\alpha_i(\theta_j - \beta_i)]} \quad (0 \leq c \leq d \leq 1)$$



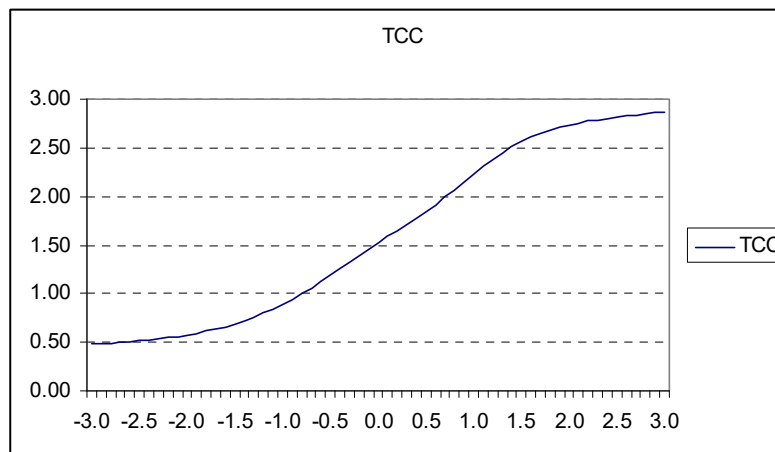


Test characteristic curves (TCCs)

Even though IRT models performances at the item level, we can look at the relationship between test scores and θ :

$$T(\theta_j) = \sum_i^N P_i(\theta_j).$$

The form of TCC is dependent on the parameters of the items that comprise the test. The curve represents the expected test scores conditional on θ_j . The shape of the TCC will generally not be normal ogive or logistic in form like the ICCs.



Exercise: Plot ICCs and TCC of the following items.

Test	Item parameter
------	----------------

