Logit Difficulty as a Linear Combination

If Prince John has assigned us, with a severe penalty for failure, the problem of determining the proficiency of all the archers in the realm subject to the constraint that many of the contestants could not or would not appear on the same field, a reasonable set of agents might be shooting arrows at targets. Each target would have an inherent difficulty that could be estimated readily with a suitable field trial, data analysis, and software package.

It may, however, save us time and trouble to think about the difficulty of each task, not as a vague amalgam of unspecified characteristics, but as a specific linear combination of more basic components (and an error term) that describe what makes a target difficult to hit. Important components might include, for example, size of the bull's-eye, distance from the archer, movement of the target, wind, and elevation.

We will begin once again with the basic form for dichotomously scored items, which you are probably tired of looking at. The probability of archer *v* succeeding on target *i* is a simple function of the logit "distance" between the archer and the target:

6. $$P = \frac{B_v}{B_v + \Delta_i} = \frac{e^{\beta_v - \delta_i}}{e^{\beta_v - \delta_i} + 1},$$ where β_v and δ_i are the logit versions of the ability and difficulty parameters and equal to the natural logs of B_v and Δ_i respectively.

In Fischer's (1973) *Linear Logistic Test Model* (*LLTM*), the difficulties are restricted somewhat to be linear combinations of more basic parameters.

7. $$\delta_i = w_{ik}\gamma_k + e_i,$$ γ_k are the basic parameters and w_{ik} are the known coefficients for this item.

The decomposition can be inflicted on any of the Rasch models mentioned above, but for the dichotomous case, the probability of observing a response of *1* is:

8. $$p(x_{vi} = 1 \mid \beta_v, \delta_i) = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}} = \frac{e^{\beta_v - \sum_{k=1}^{p} w_{ik}\gamma_k}}{1 + e^{\beta_v - \sum_{k=1}^{p} w_{ik}\gamma_k}}.$$

While, at first blush, this may not seem to make things easier, typically there are far fewer identifiable components than possible targets in the world so it can be a very parsimonious description of the tasks.

At the risk of stretching the archery analogy too far, we'll propose three two-level components for the difficulty of a target: 122 cm. or 60 cm. diameter targets, 30 m. or 90 m. distance, and stationary or swinging. The eight distinct target templates can then be described with four parameters as:

9. $$[\delta_i] = \left[\sum_{j=0}^{3} w_{ij}\eta_j\right] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \eta_0 = (base = 122cm.size, 30m.range, still) \\ \eta_1 = (size = 60cm) \\ \eta_2 = (range = 90m) \\ \eta_3 = (swinging) \end{bmatrix}$$

LLTM was proposed, formalized, and implemented by Gerhardt Fischer et al (1973, 1977, 1987, 1995). The idea originated from consideration of the cognitive operations required to solve math test problems. Scheiblechner (1972) decomposed the items into seven operations (negation, disjunction, conjunction, …). The item's difficulty was then expressed as a linear combination of these basic operations. Decomposing the tasks into basic operations wasn't original; using Rasch methods to derive *sufficient statistics* and *sample-freed* estimators was.

The design matrix for this situation does not simply indicate the presence of the condition, but indicates the number of times each operation is required in the problem's solution. For example, a possible decomposition of a specific set of items might include:

10. $$[\delta_i] = \left[\sum_{j=1}^{7} w_{ij}\eta_j\right] = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 2 & 1 & 0 & 0 & 0 \\ \ldots \end{bmatrix} \begin{bmatrix} \eta_1 = (negation) \\ \eta_2 = (disjunction \\ \eta_3 = (conjunction) \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \end{bmatrix}$$

Typically, the number of possible operations, *p*, is much smaller than the number of items so can provide a very parsimonious description of the items. It can be the basis of a fuller understanding of what makes an item difficult and suggests the possibility of generating new items at precise levels of difficulty. While the estimation of the basic parameters can be handled with least squares and the Pair algorithm we are so taken with, it would not be in keeping with the philosophy of Fischer and the Viennese school to use anything less than full conditional maximum likelihood (*CML*). Nor is there much reason not to use it, given the current power of computing and the availability of open-source software (Mair & Hatzinger, 2007). When *CML* is used, likelihood ratio tests are naturally available for a variety of interesting hypotheses.

By properly constructing the design matrix $\{w_i\}$, the model has also been applied to study the effects of item position (Hahne, 2008; Hohensinn et al., 2008; Kubinger, 2008) and less intuitively, the measurement of change (Fischer & Molenaar, 1995). In all likelihood, its greatest impacts should be in the decomposition and modelling of item difficulties (Gorin & Embretson, (2006); Newstead, et al., 2006), assessment engineering (Gierl & Haladyna, 2012; Luecht, 2013), and more or less automated item generation Newstead, et al., 2006; Poinstingl, 2008; Sonnleitner, 2008).

The *Linear Logistic Test Model* is a very powerful and very significant addition to the Rasch extended family; it still pays homage to Rasch's Specific Objectivity, sufficient statistics, and sample-freed estimators. While Fischer and his disciples like to call *LLTM* an *extended* Rasch Model (hence, the *e* in the *R-package*, *eRm* (Mair and Hatzinger, 2007),) *LLTM* is a *Rasch Model* in the sense I have used the term. It fits nicely under the general Rasch umbrella while blanketing almost the entire Rasch family (Fischer, 1973, 1995) unless I have finally mixed too many metaphors.

Poisson Counts

The Poisson form is the oldest child in the Rasch family; Rasch used it in the 1950's to analyze oral reading after observing the number of words read and the number of errors made. The distribution is often presented in introductory probability courses as the distribution of rare events. A standard example is the number of defects in a bolt of cloth, which is roughly analogous to errors in writing or reading text. The probability of finding a defect at any given spot on the cloth is small; all spots are equally likely candidates for a defect; and there is no upper limit on the number of spots or the number of defects. Similarly for oral reading, the probability of misreading any word is small and all words are assumed equally likely to be misread. Perhaps more realistically, because all the probabilities are very small, there is no practical difference among them.

The basic model is:

11. $$p\{a_{vi}\} = e^{-\lambda_{vi}} \frac{\lambda_{vi}^{a_{vi}}}{a_{vi}!}$$

where a_{vi} is the count observed, and, in our case, $\lambda_{vi} = \beta_v \varepsilon_i$, where, in Rasch's original study, β_v is the *proficiency* of the person at reading aloud, and ε_i is the *easiness* of the passage to be read.

This expression yields a familiar looking estimation equation for the relationship of two passages:

12. $$\hat{\varepsilon}_i - \hat{\varepsilon}_j = \ln(a_{vi} / a_{vj}).$$

As a Rasch model, the Poisson has been successfully applied to counts of errors made in oral reading (Rasch, 1960), of errors of various types in written essays (Andrich, 1973), of number of words read in a given time, of time taken to complete a task, of points scored in various games, and variety of cases where the score was a count of events with no definite upper limit. Generally, the number of *trials* (e.g., words that might be read or points that might be scored) is large compared to the number of *events* (e.g., words actually read, mistakes made, or points scored). The details are in Rasch (1960), Andrich (1973, 1988), and Smith & Smith (2004). The advent of computerized scoring of extended response items might lead to renewed popularity of the Poisson Rasch, because computers are really patient when counting things. Many types of counts could be collected for any examinee work sample and reported diagnostically. And then combined, or not, into a summary being assured that the sum of Poissons is still a Poisson.