What Rasch did: the mathematical underpinnings of the Rasch model.

Alex McKee, PhD.

9th Annual UK Rasch User Group Meeting, 20/03/2015

PEARSON

Our course

- Initial conceptualisation
- Separation of parameters
- Specific objectivity
- Connections to IRT (Item Response Theory)

PEARSON

Initial concepts – *ability and difficulty*.

$0 < \eta_j < \infty$ - the **ability** of person *j.*

$0 < \delta_i < \infty$ - the **difficulty** of item *i.*

$$\eta_1 = 2\eta_2$$

$$\delta_1 = 2\delta_2$$

When above is true:

$$\frac{\eta_1}{\delta_1} = \frac{\eta_2}{\delta_2}$$

and in general

$$\frac{\eta_1}{\delta_1} = \frac{K\eta_2}{K\delta_2} = \frac{\eta_2}{\delta_2}$$

PEARSON

Initial concepts – *situational parameter*

$$\xi_{ij} = \frac{\eta_j}{\delta_i}$$ - *situational parameter*

or

$$\xi_{ij} = \eta_j \epsilon_i$$

where ϵ_i is an item easiness parameter. Hence

$$\delta_i = \frac{1}{\epsilon_i}$$

is an item difficulty parameter.

Initial concepts – *the choice of the logistic model*

$$f(\xi_{ij}) = \frac{\xi_{ij}}{1+\xi_{ij}}$$

$$P(u_{ij}=1|\xi_{ij}) = \frac{\xi_{ij}}{1+\xi_{ij}}$$

$$P(u_{ij}=0|\xi_{ij}) = 1 - \frac{\xi_{ij}}{1+\xi_{ij}} = \frac{1+\xi_{ij}}{1+\xi_{ij}} - \frac{\xi_{ij}}{1+\xi_{ij}} = \frac{1}{1+\xi_{ij}}.$$

PEARSON

Initial concepts – *the choice of the logistic model*

Substituting

$$\xi_{ij} = \frac{\eta_j}{\delta_i}$$

$$P(u_{ij} = 1|\eta_j, \delta_j) = \frac{\frac{\eta_j}{\delta_i}}{1 + \frac{\eta_j}{\delta_i}} = \frac{\eta_j}{\delta_i + \eta_j}$$

$$P(u_{ij}|\eta_j, \delta_j) = \frac{\left(\frac{\eta_j}{\delta_i}\right)^{u_{ij}}}{1 + \frac{\eta_j}{\delta_i}}$$

$$P(u_{ij} = 0|\eta_j, \delta_j) = \frac{1}{1 + \frac{\eta_j}{\delta_i}} = \frac{\delta_i}{\delta_i + \eta_j}$$

PEARSON

Initial concepts – *the odds of a correct response*

$$\frac{P(u_{ij}=1|\eta_j,\delta_i)}{P(u_{ij}=0|\eta_j,\delta_i)} = \frac{\frac{\eta_j}{\delta_i+\eta_j}}{\frac{\delta_i}{\delta_i+\eta_j}} = \frac{\eta_j}{\delta_i} = \xi_{ij}$$

Recall that the statistical odds in favour of an event is the ratio of the probability in favour to probability against.

PEARSON

Separation of parameters

- Two underlying assumptions of the Rasch model.
- Develop the formula for the probability of a response vector given a particular total score *r*.
- Calculate the conditional probability of a pattern of item responses (u_{ij}) given that their sum is *r.*
- Show the sufficiency of the total score.
- Generalise the results to many test scores.

Definition: A statistic is sufficient with respect to a statistical model and its associated unknown parameter if "no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter" Fisher, R.A. (1922). "On the mathematical foundations of theoretical statistics".Philosophical Transactions of the Royal Society A 222: 309–368.doi:10.1098/rsta.1922.0009. JFM 48.1280.02. JSTOR 91208.

PEARSON

Separation of parameters – two assumptions of the Rasch model

1. The probability of correct response of examinee *j* to a dichotomously scored item *i* is given by

$$P(u_{ij} = 1 | \eta_j, \delta_i) = \frac{\xi_{ij}}{1 + \xi_{ij}}.$$

2. Given the values of the parameters, all answers are stochastically independent (Rasch, 1966, p.50).

Definition: Stochastic independence – the occurrence of one event does not affect the probability of the other.

PEARSON

Separation of parameters - formula for the probability of a response vector given a particular total score *r*.

$$P(u_{1j}, u_{2j}, \cdots u_{nj} | \eta_j, \delta_1, \delta_2, \cdots \delta_n)$$
$$= P(u_{1j}|\eta_j, \delta_1) P(u_{2j}|\eta_j, \delta_2) \cdots P(u_{nj}|\eta_j, \delta_n)$$

$$P(u_{ij}|\eta_j, \delta_j) = \frac{\left(\frac{\eta_j}{\delta_i}\right)^{u_{ij}}}{1 + \left(\frac{\eta_j}{\delta_i}\right)}, \quad \delta_i = \frac{1}{\epsilon_i}, \quad P(u_{ij}|\eta_j, \epsilon_i) = \frac{(\eta_j \epsilon_i)^{u_{ij}}}{1 + (\eta_j \epsilon_i)}$$

Note: In general, if $E_1, E_2, E_3 \cdots E_n$ *are n independent events having respective probabilities* $p_1, p_2, p_3 \cdots p_n$*, then the probability of occurrence of* E_1 *and* E_2 *and* E_3 *and* $\cdots E_n$ *is* $p_1 p_2 p_3 \cdots p_n$

PEARSON

Separation of parameters - formula for the probability of a response vector given a particular total score *r*.

$$P(u_{ij}|\eta_j, \epsilon_i) = \frac{(\eta_j \epsilon_i)^{u_{ij}}}{1 + (\eta_j \epsilon_i)}$$

$$P(u_{1j}, u_{2j} \cdots u_{nj}|\eta_j, \epsilon_1, \epsilon_2, \cdots \epsilon_n) = \frac{\prod_{i=1}^{n} \eta_j^{u_{ij}} \epsilon_i^{u_{ij}}}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

$$P(u_{1j}, u_{2j} \cdots u_{nj}|\eta_j, \epsilon_1, \epsilon_2, \cdots \epsilon_n) = \frac{\eta_j^{u_{.j}} \prod_{i=1}^{n} \epsilon_i^{u_{ij}}}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

PEARSON

Separation of parameters - formula for the probability of a response vector given a particular total score *r*.

For a total test score of zero

	ID1	ID2
Item1	0	1
Item 2	0	0
Item 3	0	0

$$P(u_{ij} = 0 | \eta_j, \epsilon_i) = \frac{(\eta_j \epsilon_i)^0}{1 + \eta_j \epsilon_i}$$

$$P(u_{.j} = 0 | \eta_j, \epsilon_1, \epsilon_2, \epsilon_3, \cdots \epsilon_n) = \prod_{i=1}^{n} \frac{1}{1 + \eta_j \epsilon_i}$$

PEARSON

	ID1	ID2
Item1	0	1
Item 2	0	0
Item 3	0	0

$$P(u_{1j} = 1 | \eta_j, \epsilon_1, \epsilon_2, \cdots \epsilon_n) = \frac{\eta_j \epsilon_1}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

$$P(u_{2j} = 1 | \eta_j, \epsilon_1, \epsilon_2, \cdots \epsilon_n) = \frac{\eta_j \epsilon_2}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

$$\vdots$$

$$P(u_{nj} = 1 | \eta_j, \epsilon_1, \epsilon_2, \cdots \epsilon_n) = \frac{\eta_j \epsilon_n}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

$$P(u_{.j} = 1 | \eta_j, \epsilon_1, \epsilon_2, \cdots \epsilon_n) = \frac{\eta_j \sum_{i=1}^{n} \epsilon_i}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

PEARSON

Separation of parameters - formula for the probability of a response vector given a particular total score *r.*

For a test score of two:

$$P(u_{.j} = 2 | \eta_j, \epsilon_i, \epsilon_2, \epsilon_3 \cdots \epsilon_n) = \frac{\eta_j^2(\epsilon_1\epsilon_2 + \epsilon_1\epsilon_3 + \cdots \epsilon_{n-1}\epsilon_n)}{\prod_{i=1}^{n}(1 + \eta_j\epsilon_i)}$$

$$P(u_{.j} = r | \eta_j, \epsilon_i, \epsilon_2, \epsilon_3 \cdots \epsilon_n) = \frac{\eta_j^r \gamma_r}{\prod_{i=1}^{n}(1 + \eta_j\epsilon_i)}$$

$$\gamma_r = \gamma(r, \epsilon_1, \epsilon_2, \cdots, \epsilon_n) = \sum_{(u_{ij})}^{r} \left(\prod_{i=1}^{n} \epsilon_i^{u_{ij}} \right)$$

Where $\sum_{(u_{ij})}^{r}$ indicates that the number of terms in the sum is a function of $\binom{n}{r}$ for a given value of r (i.e sum only combinations that give r) and (u_{ij}) denotes an item response vector for person j that yields a test score of r.

PEARSON

Calculate the conditional probability of a pattern of item responses given that their sum is *r.*

So far we know:

$$P(u_{1j}, u_{2j} \cdots u_{nj} | \eta_j, \epsilon_1, \epsilon_2, \cdots \epsilon_n) = \frac{\eta_j^{u_{.j}} \prod_{i=1}^{n} \epsilon_i^{u_{ij}}}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

$$P(u_{.j} = r | \eta_j, \epsilon_i, \epsilon_2, \epsilon_3 \cdots \epsilon_n) = \frac{\eta_j^{r} \gamma_r}{\prod_{i=1}^{n}(1 + \eta_j \epsilon_i)}$$

$$P\left((u_{ij}) | u_{.j} = r, \epsilon_1, \epsilon_2 \cdots \epsilon_n\right) = \frac{P\left((u_{ij}) | \eta_j, \epsilon_1, \epsilon_2 \cdots, \epsilon_n\right)}{P\left(u_{.j} = r | \eta_j, \epsilon_1, \epsilon_2 \cdots, \epsilon_n\right)}$$

$$\frac{\prod_{i=1}^{n} \epsilon_i^{u_{ij}}}{\gamma_r} = \frac{\prod_{i=1}^{n} \epsilon_i^{u_{ij}}}{\sum_{(u_{ij})}^{r} \left(\prod_{i=1}^{n} \epsilon_i^{u_{ij}}\right)}$$

← Does not depend on ability

PEARSON

Separation of parameters: Generalise the results to many test scores.

So far, all persons have the same test score, what about when there are many different test scores?

$$P(u_{ij}|\eta_j, \epsilon_i) = \frac{(\eta_j \epsilon_i)^{u_{ij}}}{1 + (\eta_j \epsilon_i)}$$

Assumption 2 - Given the values of the parameters, all answers are stochastically independent (Rasch, 1966, p.50).

$$P\left([u_{ij}]|(\eta_j),(\epsilon_i)\right) = \prod_{j=1}^{N}\prod_{i=1}^{n} P(u_{ij}|\eta_j, \epsilon_i)$$

Separation of parameters: Generalise the results to many test scores.

$$P\left([u_{ij}] | (\eta_j), (\epsilon_i)\right) = \prod_{j=1}^{N} \prod_{i=1}^{n} P(u_{ij} | \eta_j, \epsilon_i)$$

$$= \frac{\prod_{j=1}^{N} \prod_{i=1}^{n} (\eta_j \epsilon_i)^{u_{ij}}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + \eta_j \epsilon_i)}$$

Since ability is constant relative to subscript *i* and easiness is a constant relative to subscript *j*...

$$P([u_{ij}] | (\eta_j), (\epsilon_i)) = \frac{\prod_{j=1}^{N} \eta_j^{u_{.j}} \prod_{i=1}^{n} \epsilon_i^{u_{i.}}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + \eta_j \epsilon_i)} = \frac{\prod_{j=1}^{N} \eta_j^{r_j} \prod_{i=1}^{n} \epsilon_i^{s_i}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + \eta_j \epsilon_i)}$$

PEARSON

Separation of parameters: Generalise the results to many test scores.

$$P([u_{ij}]|(\eta_j),(\epsilon_i)) = \frac{\prod_{j=1}^{N}\eta_j^{u_{.j}}\prod_{i=1}^{n}\epsilon_i^{u_{i.}}}{\prod_{j=1}^{N}\prod_{i=1}^{n}(1+\eta_j\epsilon_i)} = \frac{\prod_{j=1}^{N}\eta_j^{r_j}\prod_{i=1}^{n}\epsilon_i^{s_i}}{\prod_{j=1}^{N}\prod_{i=1}^{n}(1+\eta_j\epsilon_i)}$$

$$P(u.j = r_j, (u_i. = s_i)|(\eta_j),(\epsilon_i))$$

$$= \frac{\prod_{j=1}^{N}\eta_j^{r_j}\prod_{i=1}^{n}\epsilon_i^{s_i}}{\prod_{j=1}^{N}\prod_{i=1}^{n}(1+n_j\epsilon_i)}[rs]$$

$$\frac{1}{[rs]}$$

Divide the top equation by the bottom equation to get:

The probability of obtaining the item response matrix conditional on a particular set of row and column marginal totals

Separation of parameters: Generalise the results to many test scores.

Therefore, once the totals have been recorded, any further statement as regards which of the items were answered correctly by which persons is, according to our model, useless as a source of information about the parameters…Thus, the row totals and the column totals are not only suitable for estimating the parameters: They imply every possible statement about the parameters that can be made on the basis of the observations. – Rasch (1966b, p. 101)

Consequently, the person total scores are ***sufficient estimators*** of the ability and the item total scores are ***sufficient estimators*** of the item easiness (Kim and Baker 2004).

Separation of parameters: Generalise the results to many test scores.

$$P(u.j = r_j, (u_i. = s_i) | (\eta_j), (\epsilon_i))$$

$$= \frac{\prod_{j=1}^{N} \eta_j^{r_j} \prod_{i=1}^{n} \epsilon_i^{s_i}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + n_j \epsilon_i)} [rs] \qquad \gamma(E) = \sum_{(s_i)}^{s} [rs] \prod_{i=1}^{n} \epsilon_i^{s_i}$$

$$P((u.j = r_j) | (\eta_j), (\epsilon_i)) = \frac{\gamma(E) \prod_{j=1}^{N} \eta_j^{r_j}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + \eta_j \epsilon_i)}$$

Where $\sum_{(s_i)}^{s}$ indicates summation over all item vectors yielding an item score of s_i as well as over all possible values of s.

PEARSON

Separation of parameters: Generalise the results to many test scores.

$$P(u.j = r_j, (u_i. = s_i) | (\eta_j), (\epsilon_i)) \quad \gamma(E) = \sum_{(s_i)}^{s} [rs] \prod_{i=1}^{n} \epsilon_i^{s_i}$$

$$= \frac{\prod_{j=1}^{N} \eta_j^{r_j} \prod_{i=1}^{n} \epsilon_i^{s_i}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + n_j \epsilon_i)} [rs]$$

$$P((u_{.j} = r_j) | (\eta_j), (\epsilon_i)) = \frac{\gamma(E) \prod_{i=1}^{N} \eta_j^{r_j}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + \eta_j \epsilon_i)}$$

$$P((u_i. = s_i) | (\eta_j), (\epsilon_i)) = \frac{\gamma(E_1) \prod_{i=1}^{n} \epsilon_i^{s_i}}{\prod_{j=1}^{N} \prod_{i=1}^{n} (1 + \eta_j \epsilon_i)}$$

PEARSON

Separation of parameters: Generalise the results to many test scores.

$$P((u_{.j} = r_j)|(u_i = s_i), (\eta_j)) = \frac{[rs]\prod_{j=1}^{N} \eta_j^{r_j}}{\gamma(E_1)}$$

$$P((u_{i.} = s_i)|(u_{.j} = r_j), (\epsilon_i)) = \frac{[rs]\prod_{i=1}^{n} \epsilon_i^{s_i}}{\gamma(E)}$$

PEARSON

Specific objectivity

*The **comparison** of any two subjects can be carried out in such a way that no parameters are involved other than those of the two subjects....Similarly, any two stimuli can be compared independently of all other parameters than those of the two stimuli, the parameters of all other stimuli as well the parameters of the subjects having been replaced with observable numbers. It is suggested that comparisons carried out under such circumstances be designated as specific objective. – Rasch(1966b, p.104-105)*

PEARSON

Specific objectivity

“…a comparison is specifically objective if it depends exclusively on properties residing in the objects and is invariant with respect to the instruments by means of which the comparison is made” – Scheiblechner (1977)

Wright 1968 – “item-free person calibration”

The comparison of the difficulty of two items should not depend on the ability levels of the groups used to measure the difficulty of the items.

Specific Objectivity

Objective measurement is the repetition of a unit amount that maintains its size, within an allowable range of error, no matter which instrument, intended to measure the variable of interest, is used and no matter who or what relevant person or thing is measured. – Rasch.org

Underlying the whole issue of specific objectivity is a basic property of all IRT models, namely, that they establish a common metric for the item and ability parameters. Because of this, ***it is clear that specific objectivity is essentially a restatement of the invariance principle common to all IRT models***. Seock-Ho Kim and Baker,2004

PEARSON

Specific Objectivity

Upon examining the Rasch model from a classical theory point of view, Whitely and Dawis (1974) concluded:
Applying the Rasch model to typical trait data does not necessarily yield objective measurement, **since some of the claimed advantages of applying the model depend directly upon the characteristics of the item pool** rather than the model. For an item pool to fully possess objective measurement, a set of rigorous conditions must be met.

Wright 1977 – above is correct but when the conditions are met the Rasch model will provide objective measurement.

PEARSON

Relationship to IRT models

$$\log \eta_j = \theta_j, \qquad -\infty < \theta_j < \infty$$

$$log \delta_i = \beta_i, \qquad -\infty < \beta_i < \infty$$

$$\eta_j = e^{\theta_j} \qquad \delta_i = e^{\beta_i}$$

$$P(u_{ij}|\eta_j, \delta_j) = \frac{\left(\frac{\eta_j}{\delta_i}\right)^{u_{ij}}}{1 + \left(\frac{\eta_j}{\delta_i}\right)}$$

PEARSON

Relationship to IRT

$$P(u_{ij}|\eta_j, \delta_j) = \frac{\left(\frac{\eta_j}{\delta_i}\right)^{u_{ij}}}{1 + \left(\frac{\eta_j}{\delta_i}\right)}$$

$$P(u_{ij}|\theta_j, \beta_j) = \frac{(e^{\theta_j} e^{-\beta_i})^{u_{ij}}}{1 + e^{\theta_j} e^{-\beta_i}}$$

$$= \frac{(e^{\theta_j - \beta_i})^{u_{ij}}}{1 + e^{(\theta_j - \beta_i)}}$$

← Two parameter IRT model with discrimination at unity.

PEARSON

Relationship to IRT

$$\frac{P(u_{ij}=1|\eta_j,\delta_i)}{P(u_{ij}=0|\eta_j,\delta_i)} = \frac{\frac{\eta_j}{\delta_i+\eta_j}}{\frac{\delta_i}{\delta_i+\eta_j}} = \frac{\eta_j}{\delta_i} = \xi_{ij}$$

$$\xi_{ij} = \frac{P(u_{ij}=1)}{P(u_{ij}=0)} = \frac{\frac{1}{1+e^{-(\theta_j-\beta_i)}}}{\frac{e^{-(\theta_j-\beta_i)}}{1+e^{-(\theta_j-\beta_i)}}} = e^{(\theta_j-\beta_i)}$$

$$log\xi_{ij} = \theta_j - \beta_i$$

Summary

-Rasch model gives the probability of a correct response given the ability of the person and the difficulty of the item.

-Ability and item difficulty can be computed separately.

-Under requisite conditions the model gives a common metric that can be used for comparisons.

-There is an IRT formulation of the model.

-Units of measurement within the IRT formulation are the log of the odds (logits).

PEARSON