Item Response Theory
“Modern Psychometrics”

Christophe Lalanne
ch.lalanne@gmail.com

February, 2010

Summary

❝ abstract here. . . ❞

Outline

1. Overview of IRT models
2. The Rasch model
3. Extension of the Rasch model
4. Models for polytomous items
5. Non-parametric IRT models

Various functions used throughout this chapter were collated in the package **Psychomisc**.

History

1950 Gulliksen : “Classical” theory for test construction
1960 Rasch : One-parameter measurement model (item difficulty)
1968 Birnbaum : More general model with three paraleters (item difficulty, discrimination, and “guessing”)
1972 Bock : Generalization of MRIs for polytomous items
1973 Fischer : LLTM (linear logistic latent trait model)
1980 Whitely-Embretson : MLTM (multicomponent latent trait model)
1984 Embretson : GLTM (general component latent trait model)

But see [9, chap. 15] and [5, chap. 2] for an overview and a generalized conceptualisation of IRT.

Overview of IRT models

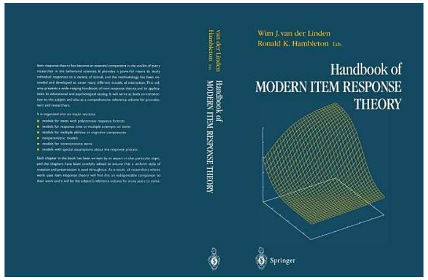
For dichotomous items, we have the Rasch model (1-PL) [10] and the 2- [3] and 3-PL models.

For polytomous ordinal data, there is the graded response model [11], the rating scale model [2], and the partial credit model [8, 14].

With polytomous nominal data, we usually rely on the nominal Bock model [3, 4].

For a more complete discussion about this classification, please refer to e.g. [12] or [13].

Overview of IRT models (Con't)



What is a Rasch Model?

As quoted by J. Rost [5, p. 27],

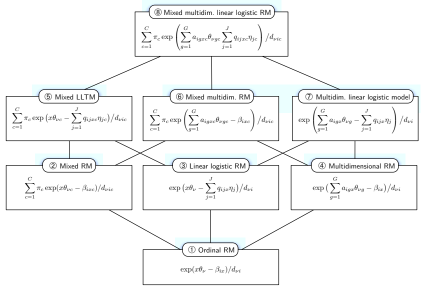
> a Rasch model is an item response model aimed at measuring one or more quantitative latent variables on a metric level of measurement, and that has the properties of sufficiency, separability, specific objectivity, and latent additivity.

There are several important keywords in the above definition. But, as recalled by J. Rost, definition of the Rasch model may differ whether we refer to the US or european litterature.

The Rasch Model: Key concepts

- *metric level of measurement*:
- *sufficiency*
- *separability*
- *specific objectivity*
- *latent additivity*

Overview of Item Response Models



J. Rost, [5, p. 32], Fig. 2.1: *A hierarchy of generalized Rasch models.*

The Rasch Model: Statistical formulation

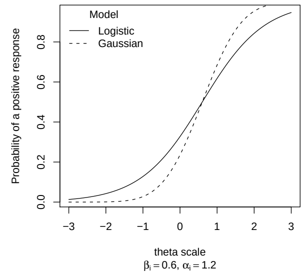
Using the 2-PL as our base model, we may expressed the probability of giving a positive (or 'correct') response as

$$P_i(\theta) = \Pr(\alpha_i^*, \beta_i, \theta) = \Psi(Z_i) = \frac{e^{Z_i}}{1+e^{Z_i}}, \tag{1}$$

where $Z_i = \alpha_i^*(\theta - \beta_i)$ denotes the logit. Here, β_i is the difficulty parameter, i.e. the point on the θ-scale where $P_i(\theta) = 0.5$. The scale (or discrimination) parameter, α_i, is the inverse of the SD of the logistic function.

The 1-PL (aka Rasch) model is easily derived from Equation 1 by imposing $\alpha_i^* = 1$ (intersection-free item curves).

Logistic 'ogive' model



Logistic vs. gaussian model

As shown below, logistic and gaussian function may be used for item response curve, and their first moment are identical, i.e. $\beta_i = \mu_i$. Only the slope differs between the two, due to the difference in the expression of the variance.

```
x <- seq(-3,3,by=.01)
beta <- 0.6
alpha <- 1.2
z <- alpha*(x-beta)
ylogist <- 1/(1+exp(-z))
ynorm <- pnorm(x,mean=beta,sd=1/alpha)
```


Logistic vs. gaussian model (Con't)

For a distribution $\mathcal{N}(0;1)$ (the so-called 'probit'), the first two moments of the corresponding logistic function are 0 and $\pi^2/3$: The slope of the logistic curve is lower compared to the gaussian case ($\alpha_i = 1/\sigma_i$).

A correction factor of 1.702 [7] may be applied so as to correct the scale parameter of the logistic function, if necessary. In fact, it can be shown that

$$|\,\Phi(Z_i) - \Psi(1.702 \times Z_i)\,| < 0.01, \quad \text{for } -\infty < \theta < \infty$$

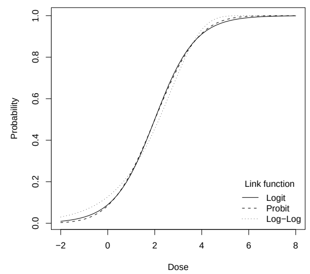
Most software use the gaussian model because of its relative simplicity, and when using posterior density (e.g. in approach relying on the computation of the marginal likelihood).

Logistic vs. gaussian model (Con't)

A (complementary) log-log function may also be used and it is easily shown that all three link functions mainly differ in the tail of their distribution, with larger ones in the non-gaussian case. J. Faraway [6] provides an illustration of this point:

```
data(bliss,package="faraway"); attach(bliss)
ilogit <- function(x) exp(x)/(1+exp(x))
m.logit <- glm(cbind(dead,alive)~conc, family=binomial)
m.probit <- glm(cbind(dead,alive)~conc,family=binomial(link=probit))
m.clog <- glm(cbind(dead,alive)~conc,family=binomial(link=cloglog))
x <- seq(-2,8,0.2)
pl <- ilogit(m.logit$coef[1]+m.logit$coef[2]*x)
pp <- pnorm(m.probit$coef[1]+m.probit$coef[2]*x)
pc <- 1-exp(-exp((m.clog$coef[1]+m.clog$coef[2]*x)))
```


Logistic vs. gaussian model (Con't)



Application of the Rasch model: Toy example

```
data(LSAT,package="ltm")
ftable(LSAT)
library(ltm)
model.rasch <- rasch(LSAT,constraint=rbind(c(dim(LSAT)[2],1)))
summary(model.rasch)
model.rasch$coefficients
```

The Rasch model as a Mixed-effects GLLM

The case of small sample: This is where Bayesian estimation (deterministic or stochastic) can be VERY helpful. You can fit a model that's a compromise between the Rasch and 2PL by using a hyper-parameter on the slopes, for instance, to shrink things towards a common mean value. Make this prior

very informative and you have a Rasch model. Make it very uninformative and you have a 2PL model. There is two a frequentist appoach which is the ONE Parameter Logistic Model (OPLM) which is different of the 1-PLM(=Rasch model). The OPLM allows defining a value of slope (discriminating power) different for each item. The difference with the 2PLM is that this slopes are a priori fixed by the user. The properties of the OPLM are very close of the Rasch model (objective measure, exhaustivity of the score), with a besser flexibility compare to the Rasch model. This is possible to implement this model with gllamm The 2PL model is the Spearman factor model analog for logistic regression. NLMIXED has one really big advantage over -gllamm- in many circumstances. It uses analytic derivatives via automatic differentiation rather than numerical derivatives. This is a potentially huge speedup because it cuts the number of fevals down a lot. I have gotten quite complex models to run in NLMIXED quite rapidly (seconds to minutes) given good starting values. Sure, it's not as quick as, say, BILOG, but it's a whole lot more flexible. It is imperative that it have good starting values, however.

Anderson, C. J., Li, Z., & Vermunt, J. K. (2007). Estimation of models in a Rasch family for polytomous items and multiple latent variables. Journal of Statistical Software, 20(6), 1-36.

Other formulation for the RM: GLMM

Agresti [1, p. 495] presents the RM as an extension of the logit matched-pairs model, which is basically an extension of the GLMM itself, where

$$g(\mu_{it}) = \boldsymbol{x'}_{it}\boldsymbol{\beta} + \boldsymbol{z'}_{it}\boldsymbol{u}_i, \quad \text{with } \mu_i = \mathbb{E}(Y_{it}|\boldsymbol{u}_i). \tag{2}$$

The random effect vector $\boldsymbol{u}_i$ is assumed to follow an $\mathcal{N}(0,\Sigma)$ distribution, with the covariance matrix Σ depending on unknown variance components (and possibly correlation parameters). Here, the random-intercept model takes the form

$$\text{logit}\big[\Pr(Y_{it} = 1|u_i)\big] = u_i + \beta_i, \tag{3}$$

where $u_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2)$.

Other formulation for the RM: Log-linear

In [**?**] the RM is formulated as a log-linear model of quasi-symmetry, but see also [1, pp. 425-426].

```
lsat.table <- data.frame(n=counts,Y=X,Yplus=as.factor(apply(X,1,sur
rasch.glm <- glm(n ~ .,data=lsat.table,family=poisson)
summary(rasch.glm)
```


References

[1] Alan Agresti. *Categorical Data Analysis*. Wiley-Interscience, 1990.

[2] D Andrich. A rating formulation for ordered response categories. *Psychometrika*, 43:561–573, 1978.

[3] A Birnbaum. Some latent trait models and their use in inferring an examinees ability. In F M Lord and M R Novick, editors, *Statistical Theories of Mental Test Scores*, pages 397–472. Reading, MA: Addison-Wesley, 1968.

[4] R D Bock. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37:29–51, 1972.

[5] A Boomsma, M A J van Duijn, and T A B Snijders. *Essays on Item Response Theory*. Springer, 2001.

[6] J Faraway. *Extending the Linear Model with R*. Chapman & Hall/CRC Press, 2005.

[7] D C Haley. Estimation of the dosage mortality relationship when the dose is subject to error. Technical Report 15, Stanford, CA: Stanford University, Applied Mathematics and Statistics Laboratory, 1952.

[8] G N Masters. A rasch model for partial credit scoring. *Psychometrika*, 47:149–174, 1982.

[9] C R Rao and S Sinharay, editors. *Handbook of Statistics, Vol. 26: Psychometrics*. Elsevier Science B.V.: The Netherlands, 2007.

[10] G Rasch. *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press, 1960. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright.

[11] F Samejina. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, 17(34), 1969.

[12] D Thissen and L Steinberg. A taxonomy of item response models. *Psychometrika*, 51(4):567–577, 1986.

[13] W J van der Linden and R K Hambleton. *Handbook of Modern Item Response Theory*. Springer, 1996.

[14] B D Wright and G N Masters. *Rating Scale Analysis*. Chicago: MESA Press, 1982.