# Standard approaches to modeling latent abilities (Export)

# Basic setup: the Bayesian Rasch model

In the context of educational assessment, typical data are the responses of a sample of test takers on a set of test items. While a broad set of IRT models is available to analyze such data, this study focuses on the Rasch model (Rasch, 1960) under a Bayesian framework (Fox, 2010; Levy & Mislevy, 2016; Debelak et al., 2022; more citations?). Let denotes person ’s response to item for and . indicates the correct response for person at the th item, whereas represents the incorrect answer.

Under a Bayesian hierarchical framework, the first stage of the Rasch model can be written as

where denotes the probability that person answers item correctly given the model parameters and , which respectively represent the latent trait of the -th individual and the item difficulty for the -th item. Since the model parameters are both unobservable, researcher only have access to their estimates, and , and the corresponding squared standard errors, and . In particular, the and can be obtained by maximum likelihood (ML) estimation using only the *within-person* observed response data from person for the Rasch model. is estimated as an inverse of the following Fisher information function (Fischer & Molenaar, 1995):

The test information function describes the amount of information provided by the item responses at a specific point on the person parameter continuum (Debelak et al., 2022). Accordingly, larger means smaller sampling error variance of each around its true person-specific latent trait , that is, smaller .

The second stage of the Bayesian Rasch model assumes that are independent and identically distributed with a certain prior distribution ,

The prior distribution is unknown in general, but for the Rasch model, we typically specify as a Gaussian distribution with two hyperparameters: and , the mean and variance of true person-specific latent traits ’s, respectively, both defined at the population level. For identification, is constrained to be zero to fix the location of the latent scale. can be assumed to have a prior distribution such as an inverse-gamma distribution (Fox, 2010; Paulon et al., 2018; Paganin et al., 2021) or an inverse- distribution (Debelak et al., 2022).

The second stage model also assumes that the item difficulty parameter ’s are independent draws from a Gaussian distribution with a mean of and a variance of :

A non-informative Gaussian prior with a large variance or an improper uniform prior density can be used for the hyperprior distribution of . Again, can have either an inverse-gamma or an inverse- prior distribution.

# Estimation of person-specific latent traits

The estimation of hyperparameters such as and using Gaussian hierarchical models is found to be robust to misspecification of the prior distribution (McCulloch & Neuhaus, 2011). The focus of this study is on inferences for , however, which can be sensitive to misspecification of (citations needed). Under the model’s normality assumptions, the conditional posterior distribution of given the person’s observed responses, the item parameters, and the hyperparameters is normal (citation needed):

where and are the conditional posterior mean and variance, respectively. When ML estimates for the item parameters are plugged in, becomes an expected a posteriori (EAP) estimate (citation needed). The conditional posterior mean is a weighted average of the prior mean (i.e., fixed at zero) and the ML estimate obtained using the person ’s observed response data. The weights are given by the so-called *precisions*, the inverse of the and . The posterior variance is proportional to the sum of the two precisions. If , the ML estimator of is perfectly precise and reliable, and thus the posterior mean and the ML estimator are identical (). A large indicates relatively less informative data about the than its prior distribution, which results in the posterior mean shrunken more toward the fixed prior mean zero.

The key insight of the Bayesian hierarchical Rasch model is that the observed variation in estiamted person-specific latent traits, , reflects two sources of variation: (1) the between-person heterogeneity in true latent traits (), and (2) the within-person sampling variation of each around its (). The *shrinkage factor*, , is defined as the proportion of the variance of the ML estimator that is due to the underlying heterogeneity across persons. If the shrinkage factor is larger than 0.5 and close to 1.0 for person , it indicates that is smaller than , suggesting more weights on the ML estimate obtained using the person ’s observed response data. To summarize how informative the ’s are *on average*, we calculate the mean of the squared standard errors over all respondents and plug it into the shrinkage factor formula: . The average shrinkage factor is equivalent to the estimate of a test’s reliability in Rasch measurement (Wright & Masters, 1982). A large value indicates less noisy, more reliable sample-based estimates ’s. In Rasch measurement context, therefore, large values indicate that a test is more suited to distinguish between respondents in the sample with different levels.