Rasch and IRT Models: A Case Study

Item response models may be used to model the responses of subjects to a number of questions or test items. An item response model with one parameter for item difficulty is known as a Rasch model. Georg Rasch (1901-1980), a Danish statistician, gave an axiomatic derivation of the model in the 1960s. We will be using a conditional (fixed-effects) logit model to illustrate the model, however, Rasch's derivation used a different approach, but one that turns out to be equivalent to the fixed-effects logit. Rasch models are one of the dominant models for binary items (e.g., success/failure on test items) in psychometrics.

Suppose n students are taking a test consisting of k true/false questions. In the Rasch model the log odds of subject i giving a correct response to item j may be modeled using a one-parameter logistic model

$$\begin{aligned} y_{ij} &\sim \mathcal{B}er(p_{ij}), \\ \text{logit}(p_{ij}) &= \alpha_i - \delta_j,\ 1 \le i \le n, 1 \le j \le k, \end{aligned} \tag{1}$$

where p_{ij} is the probability of subject i answering the item j correctly, α_i represent the *ability* of the subject i, and δ_j represent the *difficulty* of the item j. Observed y_{ij} is a scoring – has value 1 if the answer is correct, and 0 if the answer is wrong.

In terms of probability

$$P(y_{ij} = 1) = \text{logit}^{-1}(\alpha_i - \delta_j) = \frac{\exp\{\alpha_i - \delta_j\}}{1 + \exp\{\alpha_i - \delta_j\}}.$$

Figure 1: Georg William Rasch (1901-1980), Danish Statistician.

The failure of an item to fit the model can be traced to two main sources. One is that the model is too simple. It takes account of only one item characteristic – item easiness. This model assumes that all items have the same discrimination, and that the effect of guessing is negligible. Parameters for discrimination and guessing can be included in a more general model. However, their inclusion makes the application of the model to actual measurement involved.

The model (1) is not identifiable because a common constant can be added to all the abilities α_i and all the difficulties δ_j, and the predictions of the model will remain the same,

$$\alpha_i - \delta_j = \alpha_i + c - (\delta_j + c) = \alpha'_i - \delta'_j.$$

The probabilities depend only on the difference of the ability and difficulty parameters, but not on their individual locations.

Figure 2: Test

From the standpoint of classical logistic regression, this nonidentifiability is a simple case of collinearity and can be resolved by several ways. For example, by constraining the parameters (i) setting $\alpha_1 = 0$ (that is, using the person #1 as a baseline), (ii) by setting $\delta_1 = 0$ (so that the first item is the comparison point), (iii) constraining $\sum_i \alpha_i$ to 0, or (iv) constraining $\sum_j \delta_j$ to 0.

A common Bayesian model for (1) assigns normal priors to the ability and difficulty parameters:

$$\begin{aligned} \alpha_i &\sim \mathcal{N}(\mu_\alpha, \tau_\alpha),\ i = 1, \ldots, n \\ \delta_j &\sim \mathcal{N}(\mu_\delta, \tau_\delta),\ j = 1, \ldots, k. \end{aligned}$$

The priors for these parameters are assigned hyperpriors and estimated conditional on the data. This is also referred to as a partial pooling or hierarchical approach (remember the concept of "borrowing strength" in hierarchical models). The model is nonidentifiable for the reasons mentioned above: this time, it is μ_α and μ_δ that are not identifiable, because a constant can be added to each without changing the predictions. Bayesian analysis can proceed with or without the model parameters being identified, since identification is a property of a likelihood. Priors do not really "solve" identification problems, except in the degenerate case of a point-mass spike priors (i.e., parameter restrictions by any other means), however, they could place the estimators in the range where classical restricted estimators would fall. The simplest way to identify (in this Bayesian-vague way) this hierarchical model is set μ_α to 0 (or to set μ_δ to 0, but probably not both due to their relationship in the likelihood).

```
model Rasch
 {
for (i in 1:n)
    {      for (j in 1:k)
           {         y[i,j] ~ dbern(p[i,j])
                     logit(p[i,j]) <- alpha[i] - delta[j]
        #     or  p[i,j] <- (exp(alpha[i] - delta[j])/(1+exp(alpha[i] - delta[j])))
```

```
          }
    alpha[i] ~ dnorm(0,tau.alpha)
    }
 for (j in 1:k)
   {
  delta[j] ~ dnorm(mu.delta, tau.delta)
   }

tau.alpha ~ dnorm(0, 0.001) I(0,)
mu.delta ~ dnorm(0, 0.001)
tau.delta ~ dnorm(0, 0.001) I(0,)
var.alpha <- 1/tau.alpha
var.delta <- 1/tau.delta
}

#data
list(n=162,k=33, y=structure(.Data=
c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
0,0,1,1,1,1,1,1,0,0,1,1,1,1,0,0,1,0,1,1,0,1,1,1,1,1,0,1,1,1,1,1,1,
0,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,1,0,0,1,0,1,1,1,1,1,1,1,1,1,0,1,0,
... 156 lines deleted ....
0,1,1,1,1,0,0,1,1,0,1,1,0,1,1,0,0,1,1,1,0,0,0,1,1,1,0,0,0,0,1,1,0,
0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,1,1,0,0,0,0,0,0,0,
1,0,0,0,1,0,1,1,1,1,1,1,1,1,1,1,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0
), .Dim=c(162,33)))

#initial values

list(delta=c(2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2),
tau.alpha = 0.5, tau.delta=0.5, mu.alpha=0)
```

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
alpha[1]	3.907	0.774	0.01728	2.548	3.843	5.575	501	2500
alpha[2]	1.45	0.4141	0.01097	0.6703	1.447	2.305	501	2500
alpha[3]	0.1096	0.385	0.01109	-0.6411	0.1078	0.8749	501	2500
alpha[4]	2.006	0.4738	0.01146	1.09	1.997	2.979	501	2500
...								
alpha[159]	-0.6511	0.4173	0.01011	-1.523	-0.6473	0.1385	501	2500
alpha[160]	0.543	0.3886	0.0105	-0.2339	0.5378	1.313	501	2500
alpha[161]	-1.599	0.4832	0.01135	-2.586	-1.583	-0.680	501	2500
alpha[162]	-0.0371	0.4012	0.01153	-0.8194	-0.0364	0.7583	501	2500
delta[1]	0.9355	0.2273	0.008836	0.4836	0.9393	1.385	501	2500
delta[2]	0.5556	0.2165	0.008035	0.1311	0.556	0.9749	501	2500
delta[3]	0.2443	0.2224	0.008775	-0.2008	0.2396	0.684	501	2500
delta[4]	1.428	0.2378	0.009084	0.971	1.427	1.89	501	2500

...

```
delta[30]   -0.4043 0.2141  0.008461 -0.8185  -0.4013 0.01447 501 2500
delta[31]   0.3747  0.215   0.008428 -0.03354 0.3738  0.7909  501 2500
delta[32]   -0.8829 0.2142  0.008372 -1.32    -0.8804 -0.4746 501 2500
delta[33]   0.7065  0.2145  0.007964 0.3084   0.702   1.142   501 2500
```

2 Parameter IRT Model (2PLM)

The Rasch model can be generalized to 2-parameter item-response model, known as 2PLM, by allowing the slope of the logistic regression to vary by item:

$$\begin{aligned} y_{ij} &\sim \mathcal{B}er(p_{ij}), \\ \text{logit}(p_{ij}) &= \gamma_j(\alpha_i - \delta_j),\ 1 \le i \le n, 1 \le j \le k, \end{aligned} \tag{2}$$

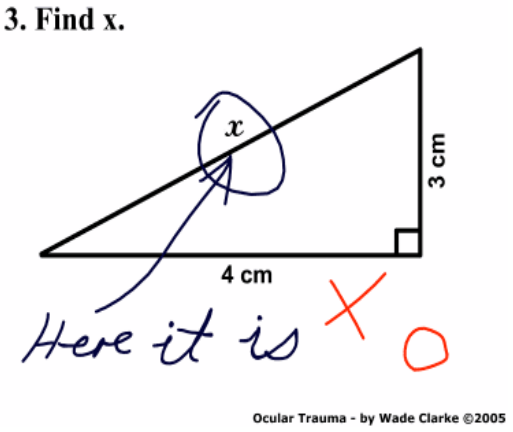

Figure 3: Example of an y_{ij}

In the model (3), parameter γ_j is called the *discrimination* of item j: if $\gamma_j = 0$, then the item does not discriminate at all, and $P(y_{ij} = 1) = 0.5$ for any person, whereas high values of γ_j correspond to a strong relation between ability and the probability of getting a correct response. Negative values of γ_j correspond to items where low-ability persons do better. Such items typically represent mistakes in the construction of the test since test designers generally try to create questions with a high positive discrimination value.

The addition of the discrimination parameter induces a new invariance problem. Model (3) has a so called multiplicative aliasing. This arises when multiplying the γ_j by a constant and dividing the $\alpha_i - \delta_j$ by that same constant.

This indeterminacy can be resolved constraining the α_i's to have mean 0 and standard deviation 1 or, in Bayesian context, by giving the α_i's a fixed prior distribution, for example $\mathcal{N}(0, 1)$.

```
model 2-parameter IR
 {
for (i in 1:n)
    {      for (j in 1:k)
           {         y[i,j] ~ dbern(p[i,j])
                     logit(p[i,j]) <- gamma[j] * ( alpha[i] - delta[j] )
```

```
          }
    alpha[i] ~ dnorm(0,1)
    }
 for (j in 1:k)
   {
  delta[j]     ~ dnorm(mu.delta, tau.delta)
  gamma[j] ~ dnorm(mu.gamma, tau.gamma)
   }
mu.gamma ~ dnorm(0, 0.001)
tau.gamma ~ dnorm(0, 0.001) I(0,)
mu.delta ~ dnorm(0, 0.001)
tau.delta ~ dnorm(0, 0.001) I(0,)
var.gamma <- 1/tau.gamma
var.delta <- 1/tau.delta
}
```

3 Parameter IRT Model (3PLM)

The 3LPM model is given by

$$\begin{aligned} y_{ij} &\sim \mathcal{B}er(p_{ij}), \\ p_{ij} &= c_j + (1-c_j)\frac{\exp\{\gamma_j(\alpha_i - \delta_j)\}}{1+\exp\{\gamma_j(\alpha_i - \delta_j)\}},\ 1 \le i \le n, 1 \le j \le k. \end{aligned} \tag{3}$$

The c parameter is commonly called the *guessing* parameter or *lower asymptote* parameter, because it indicates the probability of responding positively for examinees having very low α.

The more general IRT model is a diffuse polytomous model where each question has m categories of response, with $m > 2$. In this case the probability related to each category depends on item and person parameter as in the other case and also on the category thresholds.

Further Reading

There are many papers in fields different than education where Rasch-type models are appropriate: manufacturing and industry (conforming/non-conforming products), medicine and health systems (healthy/non-healthy), psychology (conscientiousness or cognitive ability/oposite), genetics, etc.

Some references are:

Hahn et al. (2005). Cross-Cultural Evaluation of Health Status Using Item Response Theory: FACT-B Comparisons. *Eval Health Prof.*,**28**, 233–259.

Strong, D. R., Breen, R., & Lejuez, C. W. (2004). Using item response theory to examine gambling affinity as an underlying vulnerability across a continuum of gambling involvement. *Personality and Individual Differences*, **36**, 1515–1529.

Leplege A, Ecosse E. and the WHOQOL Rasch Project Scientific Committee, (2000) Methodological issues in using the Rasch model to develop a quality of life index: the analysis of four WHOQOL-100 data sets (Argentina, France, Hong Kong,, UK). *Journal of Applied Measurement*,**1** (4) 389–418.

Li H and Hong F. (2001) Cluster-Rasch models for microarray gene expression data. *Genome Biol.*,**2(8).**