

Posterior predictive model checking and cross-validation for explanatory item response models

Daniel Furr

October 16, 2015

Contents

1	Introduction	2
2	A doubly explanatory item response model	2
2.1	General formulation	2
2.2	Hierarchical Bayes modeling approach	3
2.3	Frequentist modeling approach	5
2.4	Special cases	6
3	Estimated and predicted quantities	7
3.1	Inferences for within-sample persons and items	8
3.2	Inferences for out-of-sample persons and items	8
3.3	Inferences for mixed within-sample and out-of-sample responses	9
3.4	Inferences for special cases	9
3.5	Frequentist prediction methods	10
4	Posterior predictive model checking	10
5	Cross-validation	10
6	Dissertation chapters	10
	References	10

1 Introduction

2 A doubly explanatory item response model

2.1 General formulation

A popular model for dichotomous item response data is the Rasch model (Rasch, 1960):

$$\Pr(y_{ip}|\theta_p, \delta_i) = \frac{\exp(\theta_p - \delta_i)^{y_{ip}}}{1 + \exp(\theta_p - \delta_i)}, \quad (1)$$

where $y_{ip} = 1$ if person p ($p = 1, \dots, P$) responded to item i ($i = 1, \dots, I$) correctly and $y_{ip} = 0$ otherwise, θ_p is the ability parameter for person p , and δ_i is the difficulty parameter for item i . The individual instances of θ_p and δ_i may be collected into vectors $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$, respectively. This is a “descriptive” item response model (Wilson & De Boeck, 2004); it fully accounts for abilities and difficulties (assuming the appropriateness of the model), but does not offer insight into the the person or item predictors associated with abilities and difficulties.

The model in Equation 1 may be expanded to a “person explanatory” model by decomposing θ_p as

$$\theta_p = \mathbf{w}_p' \boldsymbol{\gamma} + \zeta_p, \quad (2)$$

where \mathbf{w}_p is a row from the design matrix of person-related covariates \mathbf{W} , $\boldsymbol{\gamma}$ is a vector of parameters, and ζ_p is the residual person abilities. The above may be interpreted as a latent regression of abilities on \mathbf{W} . Henceforth, θ_p will be referred to as the composite ability, $\mathbf{w}_p' \boldsymbol{\gamma}$ the structured part of ability, and ζ_p the residual part.

Similarly, decomposing δ_i as

$$\delta_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i \quad (3)$$

results in an “item explanatory” model, in which \mathbf{x}_i is a row from the design matrix of item-related covariates \mathbf{X} , $\boldsymbol{\beta}$ is a vector of parameters, and ϵ_i is the residual item difficulty. The above is then a latent regression of item difficulties on \mathbf{X} . In parallel with the preceding terminology for ability, δ_i will be referred to as the composite difficulty, $\mathbf{x}_i' \boldsymbol{\beta}$ the structured part of difficulty, and ϵ_i the residual part.

Equations 1, 2, and 3 together form a “doubly explanatory” item response model, given that it incorporates covariates associated with both the person and item “sides.” Note that the model may still serve descriptive purpose as the composite abilities and difficulties remain a part of the model. The final step in formulating the model is to specify distributions for the residuals. Henceforth, normal distributions are assumed,

$$\zeta_p \sim N(0, \sigma^2) \quad (4)$$

and

$$\epsilon_i \sim N(0, \tau^2), \quad (5)$$

though other choices could be considered. The sides of the model are specified in directly parallel ways, and much of the discussion that follows will make use of this point.

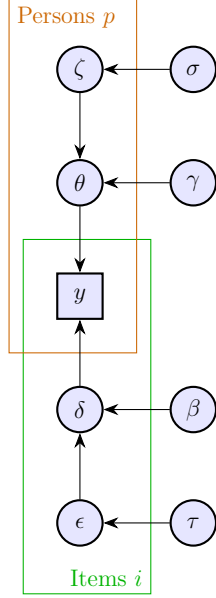


Figure 1: The full model. Circles represent parameters and squares represent data. Person covariates \mathbf{W} and item covariates \mathbf{X} are omitted.

The model is presented diagrammatically in Figure 1. In the diagram, the dependence on covariates \mathbf{W} and \mathbf{X} is suppressed. Parameters (loosely defined for the moment) are indicated by circles, and data are indicated by squares. The boxed regions indicate whether the parameters vary over persons, items or neither. The response data \mathbf{y} of course varies over both.

2.2 Hierarchical Bayes modeling approach

In Bayesian methodology, the joint posterior distribution for the parameters is factorized by way of Bayes theorem:

$$p(\boldsymbol{\omega}|\mathbf{y}) \propto p(\boldsymbol{\omega})p(\mathbf{y}|\boldsymbol{\omega}), \quad (6)$$

which indicates that the joint posterior distribution is proportional to the product of the joint prior distribution and the likelihood. In the above, $\boldsymbol{\omega}$ is the set of model parameters. A taxonomy of parameters types is given below.

1. *Basic parameters* are the foundational parameters. They are plugged into the likelihood directly or affect it indirectly through intermediate parameters. Priors are specified for them.
 - (a) *Exchangeable basic parameters* have hierarchical prior distributions. They are exchangeable draws from a distribution, the characteristics of which are determined by hyperparameters.
 - (b) *Non-exchangeable basic parameters* have non-hierarchical (often non-informative) priors. They are not thought of as exchangeable or drawn from a distribution (except in the loose sense that there is a prior distribution).

2. *Intermediate parameters* are composites built from basic parameters. They are plugged into the likelihood in place of basic parameters.
3. *Hyperparameters* are parameters for the estimated distributions for exchangeable basic parameters. These are not plugged into the likelihood but rather determine the prior distributions for the exchangeable parameters.

The doubly descriptive model has a likelihood (Equation 1) based on intermediate parameters $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$, which are in turn built from basic parameters $\boldsymbol{\gamma}$, $\boldsymbol{\zeta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ (Equations 2 and 3). $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are non-exchangeable basic parameters, while $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$ are exchangeable basic parameters whose priors depend on hyperparameters σ and τ , respectively.

The prior distribution in Equation 6 may be rewritten as

$$p(\boldsymbol{\omega}) = p(\boldsymbol{\gamma})p(\sigma) \left[\prod_{p=1}^P p(\zeta_p|\sigma) \right] p(\boldsymbol{\beta})p(\tau) \left[\prod_{i=1}^I p(\epsilon_i|\tau) \right], \quad (7)$$

assuming independent priors are specified, which is the usual case. No prior is included for $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ as they are wholly determined from the basic parameters. The likelihood part of Equation 6 may be rewritten in terms of basic parameters as

$$p(\mathbf{y}|\boldsymbol{\omega}) = p(\mathbf{y}|\mathbf{W}, \mathbf{X}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}) = \prod_{p=1}^P \prod_{i=1}^I \Pr(y_{ip}|\mathbf{w}_p, \mathbf{x}_i, \zeta_p, \boldsymbol{\gamma}, \epsilon_i, \boldsymbol{\beta}) \quad (8)$$

or in terms of intermediate parameters as

$$p(\mathbf{y}|\boldsymbol{\omega}) = p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\delta}) = \prod_{p=1}^P \prod_{i=1}^I \Pr(y_{ip}|\theta_p, \delta_i). \quad (9)$$

Given that both the prior and the likelihood may be specified ignoring the intermediate parameters $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$, it is clear that they are redundant. For many applications, however, they have useful interpretations. Further, the posterior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ can be estimated easily using the posterior draws of the basic parameters.

The posterior for a single parameter may be isolated by integrating the joint posterior over the other parameters. Let $\mathbf{D} = \{\mathbf{y}, \mathbf{W}, \mathbf{X}\}$ represent the full data. Then,

$$p(\sigma|\mathbf{D}) = \iiint p(\boldsymbol{\zeta}, \boldsymbol{\gamma}, \sigma, \boldsymbol{\epsilon}, \boldsymbol{\beta}, \tau|\mathbf{D}) d\boldsymbol{\zeta} d\boldsymbol{\gamma} d\boldsymbol{\epsilon} d\boldsymbol{\beta} d\tau \quad (10)$$

is the posterior for the standard deviation of the ability residuals. The mean and standard deviation of the marginal posterior for a parameter may be taken to represent a point estimate and standard error. Further, the joint posterior of a subset of parameters, $p(\boldsymbol{\beta}, \zeta_p, \boldsymbol{\gamma}, \epsilon_i|\mathbf{D})$ for example, likewise may be obtained by integrating out the other parameters. Despite the high-dimensional integral involved, these quantities are readily available from Monte Carlo simulation by simply ignoring the draws for the parameters to be integrated out, and so no special effort is required to obtain them.

Some alternative approaches to specifying the model bear mentioning. First, the model could equivalently be specified using hierarchical centering (Gelfand, Sahu, & Carlin, 1995) by replacing the prior with

$$p(\boldsymbol{\omega}) = p(\boldsymbol{\gamma})p(\sigma) \left[\prod_{p=1}^P p(\theta_p | \mathbf{w}_p, \boldsymbol{\gamma}, \sigma) \right] p(\boldsymbol{\beta})p(\tau) \left[\prod_{i=1}^I p(\delta_i | \mathbf{x}_i, \boldsymbol{\beta}, \tau) \right] \quad (11)$$

where $p(\theta_p | \mathbf{w}_p, \boldsymbol{\gamma}, \sigma) = N(\mathbf{w}_p \boldsymbol{\gamma}, \sigma^2)$ and $p(\delta_i | \mathbf{x}_i, \boldsymbol{\beta}, \tau) = N(\mathbf{x}_i \boldsymbol{\beta}, \tau^2)$, respectively. The likelihood is specified as in Equation 9, and $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$ are omitted altogether. Depending on the algorithm used for Monte Carlo simulation, this may improve the efficiency. Second, a likelihood may be chosen that is marginal in regards to $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$:

$$p(\mathbf{y} | \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma, \tau) = \iint \prod_{p=1}^P \prod_{i=1}^I \Pr(y_{ip} | \zeta_p, \boldsymbol{\gamma}, \epsilon_i, \boldsymbol{\beta}) p(\zeta_p | \sigma) p(\epsilon_i | \tau) d\zeta_p d\epsilon_i. \quad (12)$$

The residuals ζ_p and ϵ_i are integrated out, and they are not included in the prior, which is simply $p(\boldsymbol{\gamma})p(\sigma)p(\boldsymbol{\beta})p(\tau)$.

2.3 Frequentist modeling approach

In the “frequentist” approach, only the non-exchangeable basic parameters ($\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$) and the hyperparameters (σ and τ) are treated as parameters to be estimated. The likelihood is specified as in Equation 12, in which the probability of a response is marginal over the distributions for person and item residuals. Point estimates $\hat{\boldsymbol{\gamma}}$, $\hat{\sigma}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\tau}$ are obtained by a process that maximizes this likelihood.

Within this framework, the exchangeable parameters $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$ are called latent variables or random effects because parameters cannot have distributions. Rather than obtain direct estimates for random effects, marginal maximum likelihood estimation obtains estimates for the parameters of their distributions, in this case, $\hat{\sigma}$ and $\hat{\tau}$. The non-exchangeable basic parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are sometimes referred to as “fixed-effects.” [*Can choose whether to treat ζ_p or ϵ_i as latent variables (random effects) or just consider their marginal likelihood, for example, to just get a covariance structure.* Ref: Molenberghs and Verbeke. Hard to find because they wrote a million articles together.]

A model of this kind may be formulated in the generalized linear mixed model framework. The response variable, conditional on covariates and so-called random effects, is specified as arising from a Bernoulli distribution:

$$y_{ip} | \mathbf{w}_p, \mathbf{x}_i, \zeta_p, \epsilon_i \sim \text{Bernoulli}(\pi_{ip}). \quad (13)$$

Then the model may be written in terms of an inverse link function

$$\pi_{ip} = \Pr(y_{ip} = 1 | \mathbf{w}_p, \mathbf{x}_i, \zeta_p, \epsilon_i) = \text{logit}^{-1}[\eta_{ip}] \quad (14)$$

and a linear predictor

$$\eta_{ip} = (\mathbf{w}_p' \boldsymbol{\gamma} + \zeta_p) - (\mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i). \quad (15)$$

Because the random-effects ζ_p and ϵ_i are not nested, the model may be described as a crossed-mixed effects model. Such a model is difficult to estimate efficiently via marginal maximum likelihood because the integrals in Equation ?? do not factorize as they do with nested random effects. (). The result is an $I \times P$ dimensional integral, though Rasbash and Goldstein (1994) describe a means of reducing this to an $I + 1$ dimensional integral.

2.4 Special cases

Many dichotomous item response models are special cases of the doubly explanatory model that arise from restrictions placed on the composite abilities and difficulties. For example, the variant of the Rasch model (Rasch, 1960) associated with marginal maximum likelihood estimation (Bock & Aitkin, 1981) can be written as

$$\Pr(y_{ip}|\theta_p, \delta_i) = \frac{\exp(\theta_p - \delta_i)^{y_{ip}}}{1 + \exp(\theta_p - \delta_i)} \quad (16)$$

$$\theta_p = \zeta_p \quad (17)$$

$$\delta_i = \mathbf{x}'_i \boldsymbol{\beta}, \quad (18)$$

where \mathbf{X} is an $I \times I$ identity matrix (\mathbf{I}_I) and $\boldsymbol{\beta}$ is a vector of length I , such that $\delta_i = \beta_i$. In other words, θ_p is set equal to the ability residuals, and δ_i is set equal to the (unstructured) structural part.

In the Bayesian approach, the posterior for this Rasch model variant is given by

$$p(\boldsymbol{\theta}, \sigma, \boldsymbol{\delta}|\mathbf{y}) \propto \left[p(\boldsymbol{\delta})p(\sigma) \prod_{p=1}^P p(\theta_p|\sigma) \right] \left[\prod_{p=1}^P \prod_{i=1}^I \Pr(y_{ip}|\theta_p, \delta_i) \right], \quad (19)$$

in which the left hand bracketed quantity is the prior and the right hand quantity is the likelihood. The marginal likelihood for the frequentist approach is

$$p(\mathbf{y}|\sigma, \boldsymbol{\delta}) = \prod_{p=1}^P \int \prod_{i=1}^I \Pr(y_{ip}|\theta_p, \delta_i) p(\theta_p|\sigma) d\theta_p. \quad (20)$$

The single dimensional integration is simpler than the $I \times P$ dimensional integral in Equation 12 and is normally approximated using adaptive quadrature.

Other special cases arise from different choices of restrictions placed on the composite abilities and difficulties, and these are summarized in Table 1. The variant of the Rasch model associated with joint maximum likelihood estimation (e.g., Embretson & Reise, 2000) includes only the structured parts of ability and difficulty with identity matrices for \mathbf{W} and \mathbf{X} (one difficulty or ability parameter must be constrained for identifiability). In contrast, the random item Rasch model (e.g., De Boeck, 2008) has only the residual parts for both sides (a model intercept must be added). The latent regression item response model (Mislevy, 1985; Adams, Wilson, & Wu, 1997) includes the both parts of the composite ability and the structured part of item difficulty, where \mathbf{X} is an identity matrix. The linear logistic test model (LLTM) (Fischer, 1973), has the residual part for ability and the structured part for difficulty. Its extension, the linear logistic test model with error (LLTM-E) (e.g., Mislevy, 1988; Janssen, Schepers, & Peres, 2004), adds an item difficulty residual.

Model	θ_p	δ_i	Notes
MML Rasch	ζ_p	$\mathbf{x}'_i \boldsymbol{\beta}$	$\mathbf{X} = \mathbf{I}_I$
JML Rasch	$\mathbf{w}'_p \boldsymbol{\gamma}$	$\mathbf{x}'_i \boldsymbol{\beta}$	$\mathbf{W} = \mathbf{I}_{P-1}, \mathbf{X} = \mathbf{I}_I$
Random item Rasch	ζ_p	ϵ_i	
Latent regression	$\mathbf{w}'_p \boldsymbol{\gamma} + \zeta_p$	$\mathbf{x}'_i \boldsymbol{\beta}$	$\mathbf{X} = \mathbf{I}_I$
Linear logistic test	ζ_p	$\mathbf{x}'_i \boldsymbol{\beta}$	
Linear logistic test with error	ζ_p	$\mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i$	
Doubly explanatory	$\mathbf{w}'_p \boldsymbol{\gamma} + \zeta_p$	$\mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i$	

Table 1: Specification of several special cases.

3 Estimated and predicted quantities

Several quantities from the fitted model may be of interest. At the macro-level, $\boldsymbol{\gamma}$ represents the effects of the person covariates, and $\mathbf{W}\boldsymbol{\gamma}$ together with σ describes the conditional distribution for person abilities. Likewise, $\boldsymbol{\beta}$ represents the effects of the item covariates, and $\mathbf{X}\boldsymbol{\beta}$ together with τ describes the conditional distribution for item difficulties. Depending on the choice of either a frequentist and Bayesian framework, the maximum likelihood estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$ or posterior distributions $p(\boldsymbol{\gamma}|\mathbf{D})$ and $p(\boldsymbol{\beta}|\mathbf{D})$ will be obtained for these parameters.

For some applications, such as measurement “per se” and model assessment, the individual persons and items may be of interest. In this case, focus will be placed on $\boldsymbol{\theta}$, $\boldsymbol{\delta}$, $\boldsymbol{\zeta}$, and $\boldsymbol{\epsilon}$. The intermediate parameters $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ will be of interest for measurement per se, while the residuals $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$ will be useful for model assessment. These are within-sample quantities; that is, the estimation sample contains a person p who is associated with ζ_p and θ_p and also an item i that is associated with ϵ_i and δ_i .

There may be a (real or hypothetical) person p' who is not represented in the estimation data. This out-of-sample person p' has a covariate vector $\mathbf{w}_{p'}$ and is associated with parameters $\tilde{\zeta}_{p'}$ and $\tilde{\theta}_{p'}$, none of which play a role in fitting the model. Likewise, an out-of-sample item i' associated with $\mathbf{w}_{p'}$, $\tilde{\epsilon}_{i'}$, and $\tilde{\delta}_{i'}$ may be envisioned.

Inferences for the within-sample quantities θ_p , δ_i , ζ_p , and ϵ_i are called predictions in the frequentist framework because they are random quantities. Such inferences are called estimates in a Bayesian setting because they are either a part of the posterior or obtainable from the posterior. [Not sure the “because” is correct, but I think some explanation is needed. For example, $\boldsymbol{\zeta}$ is not directly available from the posterior if a marginal likelihood approach is used. Will have to read up on the specifics.] Inferences for the out-of-sample quantities $\tilde{\theta}_{p'}$, $\tilde{\delta}_{i'}$, $\tilde{\zeta}_{p'}$, and $\tilde{\epsilon}_{i'}$ and are considered predictions in either case.

Lastly, inferences may be made regarding new responses, which are always considered predictions. A new response may be conceived as arising from a within-sample person-item pair, indicated by \tilde{y}_{ip} . This is, in other words, simply the model-predicted response for a within-sample observation. Several possibilities exist for out-of-sample responses: $\tilde{y}_{i'p}$ represents a new response from a within-sample person to an out-of-sample item, $\tilde{y}_{ip'}$ represents a new response from an out-of-sample person to a within-sample item, and $\tilde{y}_{i'p'}$ represents a totally out-of-sample response. Predictions for out-of-sample responses are termed forecasting in a longitudinal setting.

3.1 Inferences for within-sample persons and items

If the doubly explanatory model is estimated using the Bayesian estimation approach outlined above, posterior distributions for both the composite abilities, $p(\theta_p|\mathbf{D})$, and residual abilities, $p(\zeta_p|\mathbf{D})$, will be available. Likewise, posteriors for the composite item difficulties, $p(\delta_i|\mathbf{D})$, and residual item difficulties, $p(\epsilon_i|\mathbf{D})$, will be available.

The posterior predictive distribution (Rubin, 1984) for new a response \tilde{y}_{ip} from a within-sample person-item pair, in terms of intermediate parameters, is

$$p(\tilde{y}_{ip}|\mathbf{D}) = \iint \Pr(\tilde{y}_{ip}|\theta_p, \delta_i)p(\theta_p, \delta_i|\mathbf{D}) d\theta_p d\delta_i, \quad (21)$$

or equivalently, in terms of basic and hyperparameters is

$$p(\tilde{y}_{ip}|\mathbf{D}; \zeta_p, \epsilon_i) = \iiint \Pr(\tilde{y}_{ip}|\gamma, \zeta_p, \beta, \delta_i)p(\gamma, \zeta_p, \beta, \delta_i|\mathbf{D}) d\gamma d\zeta_p d\beta d\delta_i. \quad (22)$$

This is depicted in Figure 2a. On the left side is a graphical representation of the model, similar to the one shown earlier, and on the right side is a shaded region for out-of-sample predictions. The diagram makes clear that \tilde{y}_{ip} will closely resemble y_{ip} because both arise from the same θ_p and δ_i , assuming the correctness of the model.

3.2 Inferences for out-of-sample persons and items

The predictive distribution for the residual ability for an out-of-sample person is

$$p(\tilde{\zeta}_{p'}|\mathbf{D}) = \int p(\tilde{\zeta}_{p'}|\sigma)p(\sigma|\mathbf{D}) d\sigma, \quad (23)$$

where $\tilde{\zeta}_{p'}$ indicates the new residual. This is referred to as mixed predictive distribution (Gelman, Meng, & Stern, 1996), as $\tilde{\zeta}_{p'}$ is drawn from its conditional prior distribution given σ and σ is drawn from its posterior. The mixed predictive distribution for a new composite ability is denoted by $p(\tilde{\theta}_{p'}|\mathbf{D})$, which is a function of $p(\gamma, \tilde{\zeta}_{p'}|\mathbf{D}) = \int p(\tilde{\zeta}_{p'}|\sigma)p(\gamma, \sigma|\mathbf{D}) d\sigma$ and the covariate vector $\mathbf{w}_{p'}$. [*This is vague but at least correct, I think.*]

Predictive distributions for out-of-sample items may be discussed in a manner parallel to out-of-sample persons. The mixed predictive distribution for a new item residual, $p(\tilde{\epsilon}_{i'}|\mathbf{D})$, is derived in a similar way as in Equation 23. The mixed predictive distribution for the composite difficulty for a new item is $p(\tilde{\delta}_{i'}|\mathbf{D})$, and this is a function of $p(\beta, \tilde{\epsilon}_{i'}|\mathbf{D})$ and \mathbf{x}_i .

The predictive distribution for a new response arising from an out-of-sample person and out-of-sample item may be specified in terms of the composite abilities and difficulties:

$$p(\tilde{y}_{i'p'}|\mathbf{D}) = \iint \Pr(\tilde{y}_{i'p'}|\tilde{\theta}_{p'}, \tilde{\delta}_{i'})p(\tilde{\theta}_{p'}, \tilde{\delta}_{i'}|\mathbf{D}) d\tilde{\theta}_{p'} d\tilde{\delta}_{i'}. \quad (24)$$

Alternatively, it may be specified in terms of the residual parts, structural parts, and covariates:

$$p(\tilde{y}_{i'p'}|\mathbf{D}) = \iiint \Pr(\tilde{y}_{i'p'}|\gamma, \tilde{\zeta}_{p'}, \beta, \tilde{\epsilon}_{i'}, \mathbf{w}_{p'}, \mathbf{x}_i)p(\gamma, \tilde{\zeta}_{p'}, \beta, \tilde{\epsilon}_{i'}|\mathbf{D}) d\gamma d\tilde{\zeta}_{p'} d\beta d\tilde{\epsilon}_{i'}, \quad (25)$$

where

$$p(\boldsymbol{\gamma}, \tilde{\zeta}_{p'}, \boldsymbol{\beta}, \tilde{\epsilon}_{i'} | \mathbf{D}) = \iint p(\tilde{\zeta}_{p'} | \sigma) p(\tilde{\epsilon}_{i'} | \tau) p(\boldsymbol{\gamma}, \sigma, \boldsymbol{\beta}, \tau | \mathbf{D}) d\sigma d\tau. \quad (26)$$

Figure 2b depicts how the predictive distribution for $\tilde{y}_{i'p'}$ is formed. Because $\tilde{y}_{i'p'}$ arises from $\tilde{\theta}_p$ and $\tilde{\delta}_{i'}$, it will differ substantially from y_{ip} unless $\mathbf{W}\boldsymbol{\gamma}$ and $\mathbf{X}\boldsymbol{\beta}$ are such that the estimated σ and τ are near zero.

3.3 Inferences for mixed within-sample and out-of-sample responses

A new response could, for example, come from an within-sample person but an out-of-sample item. The predictive distribution for such a response,

$$p(\tilde{y}_{i'p} | \mathbf{D}) = \iint \Pr(\tilde{y}_{i'p} | \theta_p, \tilde{\delta}_{i'}) p(\theta_p, \tilde{\delta}_{i'} | \mathbf{D}) d\theta_p d\tilde{\delta}_{i'}, \quad (27)$$

relies on the joint distribution comprised of the posterior for θ_p and the mixed predictive distribution for $\tilde{\delta}_{i'}$. Alternatively, it may be specified in terms of the residual parts, structural parts, and covariates:

$$p(\tilde{y}_{i'p} | \mathbf{D}) = \iiint \Pr(\tilde{y}_{i'p} | \boldsymbol{\gamma}, \zeta_p, \boldsymbol{\beta}, \tilde{\epsilon}_{i'}, \mathbf{w}_p, \mathbf{x}_i) p(\boldsymbol{\gamma}, \zeta_p, \boldsymbol{\beta}, \tilde{\epsilon}_{i'} | \mathbf{D}) d\boldsymbol{\gamma} d\zeta_p d\boldsymbol{\beta} d\tilde{\epsilon}_{i'}, \quad (28)$$

where

$$p(\boldsymbol{\gamma}, \zeta_p, \boldsymbol{\beta}, \tilde{\epsilon}_{i'} | \mathbf{D}) = \int p(\tilde{\epsilon}_{i'} | \tau) p(\boldsymbol{\gamma}, \zeta_p, \boldsymbol{\beta}, \tau | \mathbf{D}) d\tau. \quad (29)$$

The prediction for $\tilde{y}_{i'p}$ is depicted in Figure 2d.

If instead the predictive distribution for a new response from a out-of-sample person but within-sample item is needed, then $p(\tilde{y}_{ip'} | \mathbf{D})$ may be derived based on a joint distribution comprised of the posterior for δ_i and the mixed predictive distribution for $\tilde{\theta}_{p'}$. This is depicted in Figure 2c.

3.4 Inferences for special cases

If a special case of the full model is fitted, such as any described in Section 2.4, some predictive inferences may not be available. Specifically, a parameter must be exchangeable (that is, drawn from a distribution) in order to specify a predictive distribution for it. For example, with the Rasch model (either of the marginal or joint maximum likelihood formulations) predictive distributions for $\tilde{\delta}_{i'}$ and $\tilde{y}_{i'p}$ are unavailable due the restriction $\boldsymbol{\delta} = \mathbf{X}\boldsymbol{\beta}$. *[I may decide this section isn't important.]*

3.5 Frequentist prediction methods

4 Posterior predictive model checking

5 Cross-validation

6 Dissertation chapters

Proposed chapters:

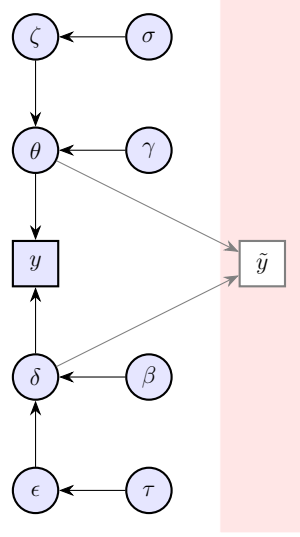
1. *This introduction.*
2. *Cross-validation using marginal likelihood to select item covariates.* A frequentist-leaning approach. Simulate dataset from the full model. Fit three models without the “random” item aspects: (1) one missing a correct item covariate interaction, (2) one containing the correct interaction, and (3) one containing the correct interaction along with an incorrect interaction. Compare the three on the basis of (1) within-sample log-likelihood, (2) out-of-sample log-likelihood, and (3) AIC.
3. *Cross-validation by leaving one item out.* Specific approaches: (1) actually do with marginal model, (2) actually do with monte carlo simulation, (3) approximate using LOO/WAIC, (4) approximate by performing OLS on Rasch-style difficulties.
4. *Predictive model checking.* Already wrote about this a bit.

Parameters for simulating datasets for proposed chapter 2:

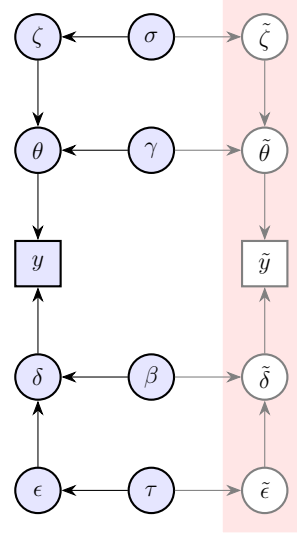
- Simulate according to the full model.
- 300 persons. Two binary covariates in a 2×2 factorial design. (Number of persons and person covariates may not matter much when focus is on item prediction.)
- 24 items. Three binary covariates in a $2 \times 2 \times 2$ factorial design with 3 items per cell. Include one interaction in data generation.
- Item variance τ may be manipulated to control quality of item predictors.

References

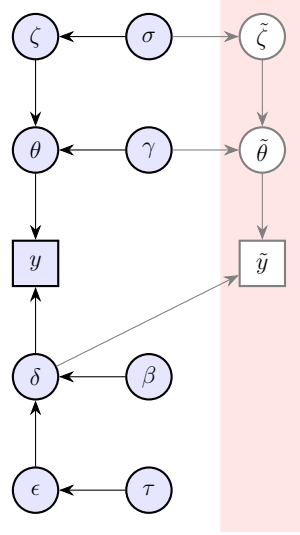
- Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, 22, 47–76.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443–459.
- De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73(4), 533–559.



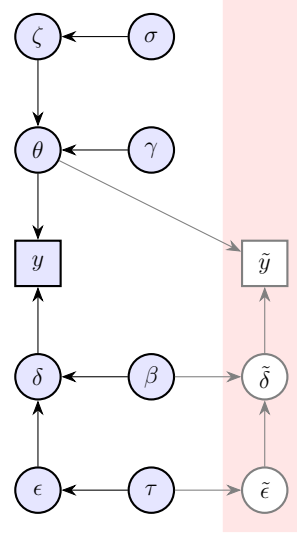
(a) Within-sample persons and items



(b) Out-of-sample persons and items



(c) Out-of-sample persons, within-sample items



(d) Within-sample persons, out-of-sample items

Figure 2: Diagrams for generating predictive distributions of various forms. Circles represent parameters and squares represent data. The shaded region indicates predictive quantities that are not involved in the estimation. Covariates \mathbf{W} and \mathbf{X} are omitted.

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory*. Mahwah, New Jersey: Psychology Press.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359–374.
- Gelfand, A. E., Sahu, S. K., & Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. *Biometrika*, 82(3), 479–488.
- Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733–807.
- Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. D. Boeck & M. Wilson (Eds.), *Explanatory item response models* (pp. 189–212). New York: Springer.
- Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, 80, 993–997.
- Mislevy, R. J. (1988). Exploiting auxiliary information about items in the estimation of rasch item difficulty parameters. *Applied Psychological Measurement*, 12(3), 281–296.
- Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral statistics*, 19(4), 337–350.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12(4), 1151–1172.
- Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In M. Wilson & P. De Boeck (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 43–74). New York: Springer.