0.1 `irtkd`: k-Dimensional Item Response Theory Model

Given several observed dependent variables and an unobserved explanatory variable, item response theory estimates the latent variable (ideal points). The model is estimated using the Markov Chain Monte Carlo algorithm, via a combination of Gibbs sampling and data augmentation. Use this model if you believe that the ideal points lie in k dimensions. See the unidimensional item response model (Section **??**) for a single hypothesized latent variable.

Syntax

```
> z.out <- zelig(cbind(Y1, Y2, Y3) ~ NULL, dimensions = 1,
                 model = "irtkd", data = mydata)
```

Inputs

`irtkd` accepts the following arguments:

- `Y1`, `Y2`, and `Y3`: `Y1` contains the items for subject "Y1", `Y2` contains the items for subject "Y2", and so on.
- `dimensions`: The number of dimensions in the latent space. The default is 1.

Additional arguments

`irtkd` accepts the following additional arguments for model specification:

- `item.constraints`: a list of lists specifying possible simple equality or inequality constraints on the item parameters. A typical entry has one of the following forms:
  - `varname = list()`: by default, no constraints are imposed.
  - `varname = list(d, c)`: constrains the dth item parameter for the item named `varname` to be equal to `c`.
  - `varname = list(d, "+")`: constrains the dth item parameter for the item named `varname` to be positive;
  - `varname = list(d, "-")`: constrains the dth item parameter for the item named `varname` to be negative.

  In a k dimensional model, the first item parameter for item i is the difficulty parameter α_i, the second item parameter is the discrimination parameter on dimension 1, $(\beta_{i,1})$, the third item parameter is the discrimination parameter on dimension 2, $(\beta_{i,2}), \ldots$, and $(k+1)$th item parameter is the discrimination parameter on dimension k, $(\beta_{i,k})$. The item difficulty parameter(α) should not be constrained in general.

`irtkd` accepts the following additional arguments to monitor the sampling scheme for the Markov chain:

- `burnin`: number of the initial MCMC iterations to be discarded (defaults to 1,000).

- `mcmc`: number of the MCMC iterations after burnin (defaults to 20,000).

- `thin`: thinning interval for the Markov chain. Only every `thin`-th draw from the Markov chain is kept. The value of `mcmc` must be divisible by this value. The default value is 1.

- `verbose`: defaults to `FALSE`. If `TRUE`, the progress of the sampler (every 10%) is printed to the screen. The default is `FALSE`.

- `zelig.data`: the input data frame if `save.data = TRUE`.

- `seed`: seed for the random number generator. The default is `NA` which corresponds to a random seed 12345.

- `alphabeta.start`: starting values for the item parameters α and β, either a scalar or a $(k+1) \times items$ matrix. If it is a scalar, then that value will be the starting value for all the elements of `alphabeta.start`. The default is `NA` which sets the starting values for the unconstrained elements based on a series of proportional odds logistic regressions. The starting values for the inequality constrained elements are set to be either 1.0 or -1.0 depending on the nature of the constraints.

- `store.item`: defaults to `FALSE`. If `TRUE` stores the posterior draws of the item parameters. (For a large number of draws or a large number observations, this may take a lot of memory.)

- `store.ability`: defaults to `TRUE`, storing the posterior draws of the subject abilities. (For a large number of draws or a large number observations, this may take a lot of memory.)

- `drop.constant.items`: defaults to `TRUE`, dropping items with no variation before fitting the model.

`irtkd` accepts the following additional arguments to specify prior parameters used in the model:

- `b0`: prior mean of (α, β), either as a scalar or a vector of compatible length. If a scalar value, then the prior means for both α and β will be set to that value. The default is 0.

- `B0`: prior precision for (α, β), either a scalar or a $(k+1) \times items$ matrix. If a scalar value, the prior precision will be a blocked diagonal matrix with elements `diag(B0,items)`. The prior precision is assumed to be same for all the items. The default is 0.25.

Zelig users may wish to refer to `help(MCMCirtKd)` for more information.

Convergence

Users should verify that the Markov Chain converges to its stationary distribution. After running the `zelig()` function but before performing `setx()`, users may conduct the following convergence diagnostics tests:

- `geweke.diag(z.out$coefficients)`: The Geweke diagnostic tests the null hypothesis that the Markov chain is in the stationary distribution and produces z-statistics for each estimated parameter.

- `heidel.diag(z.out$coefficients)`: The Heidelberger-Welch diagnostic first tests the null hypothesis that the Markov Chain is in the stationary distribution and produces p-values for each estimated parameter. Calling `heidel.diag()` also produces output that indicates whether the mean of a marginal posterior distribution can be estimated with sufficient precision, assuming that the Markov Chain is in the stationary distribution.

- `raftery.diag(z.out$coefficients)`: The Raftery diagnostic indicates how long the Markov Chain should run before considering draws from the marginal posterior distributions sufficiently representative of the stationary distribution.

If there is evidence of non-convergence, adjust the values for `burnin` and `mcmc` and rerun `zelig()`.

Advanced users may wish to refer to `help(geweke.diag)`, `help(heidel.diag)`, and `help(raftery.diag)` for more information about these diagnostics.

Examples

1. Basic Example
   Attaching the sample dataset:

   ```
   > data(SupremeCourt)
   > names(SupremeCourt) <- c("Rehnquist", "Stevens", "OConnor", "Scalia",
   +     "Kennedy", "Souter", "Thomas", "Ginsburg", "Breyer")
   ```

   Fitting a one-dimensional item response theory model using `irtkd`:

   ```
   > z.out <- zelig(cbind(Rehnquist, Stevens, OConnor, Scalia, Kennedy,
   +     Souter, Thomas, Ginsburg, Breyer) ~ NULL, dimensions = 1,
   +     data = SupremeCourt, model = "irtkd", B0 = 0.25, burnin = 5000,
   +     mcmc = 50000, thin = 10, verbose = TRUE)
   ```

   Checking for convergence before summarizing the estimates:

   ```
   > geweke.diag(z.out$coefficients)
   ```

```
> heidel.diag(z.out$coefficients)

> raftery.diag(z.out$coefficients)

> summary(z.out)
```

Model

Let Y_i be a vector of choices on J items made by subject i for $i = 1, \dots, n$. The choice Y_{ij} is assumed to be determined by unobserved utility Z_{ij}, which is a function of subject abilities (ideal points) θ_i and item parameters α_j and β_j,

$$Z_{ij} = -\alpha_j + \beta_j' \theta_i + \epsilon_{ij}.$$

In the k-dimensional item response theory model, each subject's ability is represented by a k-vector, θ_i. Each item has a difficulty parameter α_j and a k-dimensional discrimination parameter β_j. In one-dimensional item response theory model, $k = 1$.

- The *stochastic component* is given by

  $$\begin{aligned} Y_{ij} &\sim \text{Bernoulli}(\pi_{ij}) \\ &= \pi_{ij}^{Y_{ij}} (1 - \pi_{ij})^{1 - Y_{ij}}, \end{aligned}$$

  where $\pi_{ij} = \Pr(Y_{ij} = 1) = E(Z_{ij})$.

  The error term in the unobserved utility equation has a standard normal distribution,

  $$\epsilon_{ij} \sim \text{Normal}(0, 1).$$

- The *systematic component* is given by

  $$\pi_{ij} = \Phi(-\alpha_j + \beta_j \theta_i),$$

  where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution with mean 0 and variance 1, while θ_i contains the k-dimensional subject abilities(ideal points), and α_j and β_j are the item parameters. Both subject abilities and item parameters need to estimated from the model. The model is identified by placing constraints on the item parameters.

- The *prior* for θ_i is given by

  $$\theta_i \sim \text{Normal}_k(0, I_k)$$

- The joint *prior* for α_j and β_j is given by

  $$(\alpha_j, \beta_j)' \sim \text{Normal}_{k+1}\left(b_{0_j}, B_{0_j}^{-1}\right)$$

  where b_{0_j} is a $(k+1)$-vector of prior mean and B_{0_j} is a $(k+1) \times (k+1)$ prior precision matrix which is assumed to be diagonal.

Output Values

The output of each Zelig command contains useful information which you may view. For example, if you run:

```
z.out <- zelig(cbind(Y1, Y2, Y3) ~ NULL, model = "irtkd", data)
```

then you may examine the available information in `z.out` by using `names(z.out)`, see the draws from the posterior distribution of the `coefficients` by using `z.out$coefficients`, and view a default summary of information through `summary(z.out)`. Other elements available through the `$` operator are listed below.

- From the `zelig()` output object `z.out`, you may extract:
  - `coefficients`: draws from the posterior distributions of the estimated subject abilities(ideal points). If `store.item = TRUE`, the estimated item parameters α and β are also contained in `coefficients`.
  - `data`: the name of the input data frame.
  - `seed`: the random seed used in the model.
- Since there are no explanatory variables, the `sim()` procedure is not applicable for item response models.

Contributors

The k dimensional item response function is part of the MCMCpack library by Andrew D. Martin and Kevin M. Quinn. If you use this model, please cite:

.

The convergence diagnostics are part of the CODA library by Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. These diagnostics should be cited as:

.

Sample data are adapted from

.

Ben Goodrich and Ying Lu enabled `irtkd` to work with Zelig.