

# Package ‘BayesTwin’

June 26, 2017

**Version** 1.0

**Title** Bayesian Analysis of Item-Level Twin Data

**Date** 2017-06-23

**Author** Inga Schwabe

**Maintainer** Inga Schwabe <bayestwin@gmail.com>

**Depends** R (>= 3.4.0), foreign, coda, matrixStats, rjags

## Description

Bayesian analysis of item-level hierarchical twin data using an integrated item response theory model. Analyses are based on Schwabe & van den Berg (2014) <doi:10.1007/s10519-014-9649-7>, Molenaar & Dolan (2014) <doi:10.1007/s10519-014-9647-9>, Schwabe, Jonker & van den Berg (2016) <doi:10.1007/s10519-015-9768-9> and Schwabe, Boomsma & van den Berg (2016) <doi:10.1016/j.lindif.2017.01.018>.

**License** GPL (>= 2)

**URL** <http://www.ingaschwabe.com>

**BugReports** <https://github.com/ingaschwabe/BayesTwin>

**RoxygenNote** 6.0.1

**RemoteType** github

**RemoteHost** <https://api.github.com>

**RemoteRepo** BayesTwin

**RemoteUsername** ingaschwabe

**RemoteRef** master

**RemoteSha** 1bc0b5c9c3bae6b17e71231f318baa00a8b1b86a

**GithubRepo** BayesTwin

**GithubUsername** ingaschwabe

**GithubRef** master

**GithubSHA1** 1bc0b5c9c3bae6b17e71231f318baa00a8b1b86a

## R topics documented:

BayesTwin	2
geplot	4
HPD	5

IRTtwin . . . . .	6
plotbayestwin . . . . .	8
simulatetwin . . . . .	9
summary.bayestwin . . . . .	11

<b>Index</b>	<b>12</b>
--------------	-----------

---

BayesTwin	<i>Bayesian analysis of item-level twin data</i>
-----------	--

---

## Description

This package can be used to perform bayesian analysis of item-level twin data. Simultaneously with the biometric model (ACE/ADE or AE), an item response theory (IRT) model is estimated to take into account properties of the measurement scale. Functions are included in the package that help plot relevant information in figures and compute posterior statistics such as HPD intervals.

Caution! The subroutines of this package rely on the program JAGS, which can be freely obtained from <http://mcmc-jags.sourceforge.net>.

## Details

Package: BayesTwin  
 Type: Package  
 Version: 0.1.0  
 Date: 2017-01-06  
 License: GPL-2

The main function `IRT_twin` can be used to analyse item-level twin data under the ACE, ADE or AE model. Simultaneously with the biometric model, an item response theory (IRT) measurement model is estimated. For dichotomous item data, the 1 parameter model (1PL) or the 2 parameter model (2PL) can be used and for polytomous item data, the partial credit model (PCM) or the generalized partial credit model (GPCM). Optionally, genotype by environment interaction (GxE) can be estimated and/or covariates can be included in the analysis. GxE is assessed in the case that the unique environment features as latent (i.e., unmeasured) variable, using the parametrization as described in Schwabe & van den Berg (2014, 2016) and Schwabe, Boomsma & van den Berg (2017). The function returns MCMC samples as well as posterior means, standard deviations and 95% HPD intervals for all variance components. Objects returned from the function are assigned the class "Bayestwin" which has its own summary method that can be used to summarize the output. The function `plotbayestwin` can be used to create sampling plots or plot posterior distributions and the function `geplot` to plot the 95% credibility region of the GxE interaction effect.

The function `simulate_twin_data` can be used to simulate item-level twin data under all common biometric models (ACE/ADE or AE). For the simulation of the item data, a 1PL, 2PL, PCM or GPCM can be used. Optionally, the data can be simulated with GxE, using the parametrization as described in Schwabe & van den Berg (2014, 2016) and Schwabe, Boomsma & van den Berg (2017).

## Note

To work, this packages requires the program JAGS to be in the PATH variable. JAGS can be freely obtained from <http://mcmc-jags.sourceforge.net>.

**Author(s)**

Inga Schwabe <bayestwin@gmail.com>

**References**

- Schwabe, I. & van den Berg, S.M. (2014). Assessing Genotype by Environment Interaction in Case of Heterogeneous Measurement Error, *Behavior Genetics*, 44 (4), 394-406.
- Molenaar, D. & Dolan, C.V. (2014). Testing Systematic Genotype by Environment Interactions Using Item Level Data, *Behavior Genetics*, 44(3), 212-231.
- Schwabe, I., Jonker, W. & van den Berg, S.M. (2016). Genes, Culture and Conservatism - A Psychometric-Genetic Approach, *Behavior Genetics*, 46 (4), 516-52.
- Schwabe, I., Boomsma, D.I. & van den Berg, S.M. (2017). Increased Environmental Sensitivity in High Mathematics Performance. *Learning and Individual Differences*, 54, 196-201.

**Examples**

```
data(results)
summary(results)

#Using the output to obtain the 95% HPD for additive genetic variance
HPD(results$samples_var_a)

#Using the output to obtain the 90% HPD for all item difficulty parameters
apply(results$samples_item_b, 1, function (x) HPD(x, 0.90))

## Not run:
##Simulate Item-level twin data under the 1PL Rasch model
data = simulatetwin(irt_model = "1PL", var_a = 0.5, var_c = 0.3, ge_beta0 = log(0.2),
                  ge = TRUE)

data_mz = data$y_mz
data_dz = data$y_dz

##Analyse the simulated data under an 1PL model with GxE
results = IRTtwin(data_mz, data_dz, 1:20, 21:40, ge = TRUE)

##Summarize results:
summary(results)

#Using the output to obtain the 95% HPD for additive genetic variance
HPD(results$samples_var_a)

#Using the output to obtain the 90% HPD for all item difficulty parameters
apply(results$samples_item_b, 1, function (x) HPD(x, 0.90))

##Plot trace lines for var(A)
plotbayestwin(results$samples_var_a, type = "trace")

##Plot posterior distribution of var(A)
plotbayestwin(results$samples_var_a)

##Plot 95% credibility region of GxE interaction effect
geplot(results$var_a, results$samples_beta0, results$samples_beta1)

## End(Not run)
```

---

geplot

---

*Plots 95% credibility interval of GxE interaction*


---

### Description

This function plots the 95% credibility region of a GxE interaction effect.

### Usage

```
geplot(var_a, samples_beta0, samples_beta1, main, xlab, ylab, col, ...)
```

### Arguments

var_a	A single value representing the posterior point estimate for genetic variance
samples_beta0	A vector representing draws from the posterior distribution of beta0, as produced by the main function IRT_twin of this package
samples_beta1	A vector representing draws from the posterior distribution of beta1, as produced by the main function IRT_twin of this package
main	An overall title for the plot
xlab	A title for the x axis
ylab	A title for the y axis
col	Color of the lines
...	Further arguments for the default S3 plot method

### Details

The 95 % credibility region for the GxE interaction effect is displayed for the entire range of estimated genotypic values in the interval of -2 and 2 standard deviations of the posterior point estimate of genetic variance.

The function expects as input the posterior point estimate for genetic variance and two vectors representing draws from the posterior distributions of beta0 and beta1 respectively, as produced by the main function IRT\_twin of this package.

### Value

Plot of the 95% credibility region of the GxE interaction.

### Author(s)

Inga Schwabe

### Examples

```
#Using the output to plot the 95% credibility region
data(results)
geplot(results$var_a, results$samples_beta0, results$samples_beta1)
```

HPD

*Calculate highest posterior density interval***Description**

This function calculates the Bayesian highest posterior density interval (HPD) based on a parameters' posterior sample.

**Usage**

```
HPD(sample, cred_int = 0.95)
```

**Arguments**

sample	A vector representing draws from the target distribution of the parameter of interest, as produced by the main function <code>IRT_twin</code> of this package.
cred_int	The desired accuracy of the HPD. Default value is 0.95 for 95%.

**Details**

The highest posterior density interval (HPD, see e.g. Box & Tia, 1992) contains the required mass such that all points within the interval have a higher probability density than points outside of the interval.

The function expects as input a vector representing draws from the target distribution of the parameter of interest, such as produced by the main function `IRT_twin` of this package.

The result is a vector consisting of two values, the first value representing the lower bound of the HPD and the second value representing the upper bound.

**Value**

A vector of length 2 with the lower (first value) and upper (second value) bound of the HPD.

**Author(s)**

Inga Schwabe

**References**

Box, G., & Tiao, G. (1992). Bayesian inference in statistical analysis. New York: John Wiley & Sons.

**Examples**

```
data(results)

#Obtain the 95% HPD for additive genetic variance
HPD(results$samples_var_a)

#Obtain the 90% HPD for all item difficulty parameters
apply(results$samples_item_b, 1, function(x) HPD(x, 0.90))
```

## Description

Bayesian analysis of item-level twin data under the AE, ACE or ADE model. Covariates can be included and a GxE interaction effect can be estimated.

## Usage

```
IRTtwin(data_mz, data_dz,
        twin1_datacols_p, twin2_datacols_p,
        twin1_datacols_cov = NA, twin2_datacols_cov = NA,
        decomp_model = "ACE", irt_model = "1PL", ge = FALSE,
        n_iter = 5000, n_burnin = 7000, n_chains = 1,
        fit_stats = FALSE, var_prior = "INV_GAMMA",
        N_cov = 0, inits = NA, Nk = 0)
```

## Arguments

<code>data_mz</code>	Phenotypic as well as covariate data of MZ twins in matrix form.
<code>data_dz</code>	Phenotypic as well as covariate data of DZ twins in matrix form.
<code>twin1_datacols_p</code>	Columns of <code>data_mz</code> or respectively <code>data_dz</code> in which phenotypic data is stored for the first twin of all MZ and DZ twins, e.g. <code>c(1,2,3)</code> if they are stored in the first three columns or <code>c(20:40)</code> if stored in column 20 through 40.
<code>twin2_datacols_p</code>	Columns of <code>data_mz</code> or respectively <code>data_dz</code> in which phenotypic data is stored for the second twin of all MZ and DZ twins.
<code>twin1_datacols_cov</code>	Columns of <code>data_mz</code> or respectively <code>data_dz</code> in which covariate data is stored for the first twin of all MZ and DZ twins.
<code>twin2_datacols_cov</code>	Columns of <code>data_mz</code> or respectively <code>data_dz</code> in which covariate data is stored for the second twin of all MZ and DZ twins.
<code>decomp_model</code>	Choice of the biometric model under which the data is analyzed, either "ACE", "ADE" or "AE". Default model is "ACE" model.
<code>irt_model</code>	Choice of the measurement (IRT) model under which the item data is simulated. "1PL" = 1 PL model, "2PL" = 2PL model, "PCM" = Partial credit model, "GPCM" = Generalized partial credit model..
<code>ge</code>	Indication whether data is analyzed with ( <code>ge = TRUE</code> ) or without ( <code>ge = FALSE</code> ) GxE. Default value is FALSE.
<code>n_iter</code>	Total number of MCMC iterations (without burn-in period). Default number is 5.000
<code>n_burnin</code>	Total number of burn-in iterations for the MCMC analysis. Default number is 7.000
<code>n_chains</code>	Number of MCMC chains. Default number is 1.

<code>fit_stats</code>	Should be set to TRUE when fit statistics are needed (e.g., DIC, BIC). In this case, multiple chains need to be used (the parameter <code>n_chains</code> should be set to at least 2)
<code>var_prior</code>	Choice of prior distributions used for the variance components, either "INV_GAMMA" for an inverse gamma distribution or "UNIF" for uniform distribution. For a low total phenotypic variance, uniform prior distributions are recommended. Default value is "INV_GAMMA".
<code>N_cov</code>	Number of covariates. Default value is 0.
<code>inits</code>	Initial values for the MCMC analysis. Default is NULL.
<code>Nk</code>	Number of categories in case of polytomous item data. As the default IRT model is a 1PL, the default value is set 0.

## Details

Bayesian analysis of item-level twin data under all common biometric models (i.e., A(C)E, ADE) with or without genotype by environment interaction (GxE). For the simulation of dichotomous item data, the one parameter logistic model (1PL) or the 2 parameter logistic model (2PL) can be used. For polytomous item data, the partial credit model (PCM) or the generalized partial credit model (GPCM) can be used. Optionally, covariates can be included in the analysis.

GxE is defined modelled as heterogeneous unique-environmental variance and is defined as  $\sigma^2 E_{ij} = \exp(\beta_0 + \beta_1 A_{ij})$  under the ACE model and as  $\sigma^2 E_{ij} = \exp(\beta_0 + \beta_1 G_{ij})$  under the ADE model where  $\beta_0$  refers to an intercept (unique-environmental variance when  $A_{ij}$  or respectively  $G_{ij} = 0$ ) and  $\beta_1$  represents the GxE interaction effect.  $A_{ij}$  represents the genotypic value for twin  $j$  from family  $i$ , likewise  $G$  represents the total genotypic value (both additive and non-additive genetic effects) for twin  $j$  from family  $i$ . For more detail see Schwabe & van den Berg (2014), Behavior Genetics, 44(4), 394-406 and Schwabe, Boomsma & van den Berg (2017), Learning and Individual Differences, 54, 196-201.

## Value

Output of the class "bayestwin" including samples, posterior point estimates, standard deviations and the 95% HPD interval for variance components and, if applicable, regression parameters. Call `str(results)` if the output of IRT\_twin is stored in an object called results to see a list of all stored variables.

Results are printed on the fly for variance components and, if applicable, regression parameters.

## Author(s)

Inga Schwabe

## Examples

```
data(results)

#Summarize results
summary(results)

##Plot posterior distribution of var(A)
plotbayestwin(results$samples_var_a)

##Plot 95% credibility region of GxE interaction effect
geplot(results$var_a, results$samples_beta0, results$samples_beta1)
```

```
## Not run:
##Simulate data
data = simulate(irt_model = "1PL", var_a = 0.5, var_c = 0.3, ge_beta0 = log(0.2),
               ge = TRUE)
data_mz = data$y_mz
data_dz = data$y_dz

##Run analysis
results = IRTtwin(data_mz, data_dz, 1:20, 21:40, ge = TRUE)

##Summarize results:
summary(results)

##Plot trace lines for var(A)
plotbayestwin(results$samples_var_a, type = "trace")

##Plot posterior distribution of var(A)
plotbayestwin(results$samples_var_a)

##Plot 95% credibility region of GxE interaction effect
geplot(results$var_a, results$samples_beta0, results$samples_beta1)

## End(Not run)
```

---

plotbayestwin

*Summary plots of BayesTwin MCMC-objects*


---

## Description

plotbayestwin returns either a density or trace plot for the target parameter.

## Usage

```
plotbayestwin(sample, t = "density", main, xlab, ylab, legend = TRUE, lines = TRUE, ...)
```

## Arguments

sample	A vector representing draws from the posterior distribution of the parameter of interest, as produced by the main function IRT_twin of this package
t	Type of plot that is produced, either "density" to get a histogram of the posterior distribution or "trace" to get a trace plot. Default value is "density"
main	An overall title for the plot
xlab	A title for the x axis
ylab	A title for the y axis
legend	Indicating if a legend should be added. Default value is TRUE.
lines	Indicating if lines for mean, median and lower and upper limit of the 95% HPD should be added. Default value is TRUE
...	Further arguments for the generic S3 method histogram (for the density plot) or plot (for the trace plot) method



## Details

Creates either a density or trace plot for the target parameter.

The function expects as input a vector representing draws from the posterior distributions of the parameter of interest, as produced by the main function `IRT_twin` of this package.

## Author(s)

Inga Schwabe

## Examples

```
data(results)

##Plot posterior distribution of var(A)
plotbayestwin(results$samples_var_a)
```

---

simulatetwin	<i>Simulate item-level twin data</i>
--------------	--------------------------------------

---

## Description

Simulation of item-level twin data under the AE, ACE or ADE model with or without genotype by environment interaction (GxE).

## Usage

```
simulatetwin(n_mz = 140, n_dz = 360, var_a = 0.5, var_c = 0.3,
var_e = 0.2, var_d = 0, model = "ACE", n_items = 20, n_cat = 0,
ge = FALSE, ge_beta0 = 0, ge_beta1 = 0, irt_model = "1PL")
```

## Arguments

<code>n_mz</code>	Number of monozygotic (MZ) twin pairs. Default value is 140.
<code>n_dz</code>	Number of dizygotic (DZ) twin pairs. Default value is 360.
<code>var_a</code>	Variance explained by additive genetic influences. Default value is 0.5.
<code>var_c</code>	Variance explained by shared environmental influences. Default value is 0.3.
<code>var_e</code>	Variance explained by unique environmental influences. Default value is 0.2.
<code>var_d</code>	Variance explained by dominance genetic influences. As the default model is an "ACE" model, the default value is set to 0.
<code>model</code>	Biometric model used for the simulation, either "ACE", "ADE" or "AE". Default model is "ACE" model.
<code>n_items</code>	Number of test items. Default value is 20.
<code>n_cat</code>	Number of categories for the simulation of polytomous item data. As the default IRT model is a 1PL, the default is set 0.
<code>ge</code>	Indication whether data is simulated with ( <code>ge = TRUE</code> ) or without ( <code>ge = FALSE</code> ) GxE. Default value is FALSE

<code>ge_beta0</code>	When data is simulated with <code>ge = TRUE</code> , average environmental variance (when $A = 0$ ) has to be defined. Default value is 0.
<code>ge_beta1</code>	When data is simulated with <code>ge = TRUE</code> , the GxE interaction parameter has to be defined. Default value is 0.
<code>irt_model</code>	Choice of the measurement (IRT) model under which the item data is simulated. "1PL" = 1 PL model, "2PL" = 2PL model, "PCM" = Partial credit model, "GPCM" = Generalized partial credit model.

## Details

Item twin data is simulated under all common biometric models (i.e., A(C)E, ADE) with or without genotype by environment interaction (GxE). Data is simulated with an average phenotypic value of 0. For the simulation of dichotomous item data, the 1 PL or 2 PL can be used. For polytomous item data with three or more categories, the partial credit model (PCM) or the generalized partial credit model (GPCM) can be used.

Genotype by environment interaction is defined modelled as heterogeneous unique-environmental variance and is defined as  $\sigma^2 Eij = \exp(\beta_0 + \beta_1 Aij)$  under the ACE model and as  $\sigma^2 Eij = \exp(\beta_0 + \beta_1 Gij)$  under the ADE model where  $\beta_0$  refers to an intercept (unique-environmental variance when  $Aij$  or respectively  $Gij = 0$ ) and  $\beta_1$  represents the GxE interaction effect.  $Aij$  represents the genotypic value for twin  $j$  from family  $i$ , likewise  $G$  represents the total genotypic value (both additive and non-additive genetic effects) for twin  $j$  from family  $i$ . For more detail see Schwabe & van den Berg (2014, 2016) or Schwabe, Boomsma & van den Berg (2017).

## Value

Two matrices are returned, `y_mz` for MZ twin pairs and `y_dz` for DZ twin pairs. For a total of `n_items`, the result is a matrix of `n_mz` rows for the  $i$ -th MZ family and  $2 * n_items$  columns with the item answers of the first (columns 1:`n_items`) and second twin (columns `n_items + 1`: $2 * n_items$ ) of a family. For example, `y_mz[1, 22]` is the response of the first twin from family 1 to item 22 and `y_mz[1, 23]` is the response of the second twin from family 1 to item 1 if `n_items = 22`. The same logic applies to the data of DZ families.

## Author(s)

Inga Schwabe

## References

- Schwabe, I. & van den Berg, S.M. (2014). Assessing Genotype by Environment Interaction in Case of Heterogeneous Measurement Error, *Behavior Genetics*, 44 (4), 394-406.
- Schwabe, I., Jonker, W. & van den Berg, S.M. (2016). Genes, Culture and Conservatism - A Psychometric-Genetic Approach, *Behavior Genetics*, 46 (4), 516-52.
- Schwabe, I., Boomsma, D.I. & van den Berg, S.M. (2017). Increased Environmental Sensitivity in High Mathematics Performance. *Learning and Individual Differences*, 54, 196-201.

## Examples

```
#100 MZ twins, 200 DZ twins, ACE model, no GxE, 1 PL:
simulatetwin(n_mz = 100, n_dz = 200)

## Not run:
#500 MZ twins, 800 DZ twins, ACE model, no GxE, 1 PL:
simulatetwin(n_mz = 500, n_dz = 800)
```

```
#140 MZ twins, 360 DZ twins, ADE model, GPCM:
simulatetwin(var_a = 0.4, var_d = 0.2, var_e = 0.4,
             model = "ADE", irt_model = "GPCM")

## End(Not run)
```

---

`summary.bayestwin`*Summary statistics for BayesTwin analysis*

---

## Description

`summary.bayestwin` produces summary statistics for variance components and, if applicable, regression parameters: The posterior point estimate and its standard deviation and the 95% HPD.

## Usage

```
## S3 method for class 'bayestwin'
summary(object, ...)
```

## Arguments

<code>object</code>	An object storing output from the main function <code>IRT_twin</code>
<code>...</code>	Further arguments for the default S3 summary method

## Author(s)

Inga Schwabe

## Examples

```
data(results)
summary(results)
```

# Index

## \*Topic **IRTtwin**

IRTtwin, [6](#)

summary.bayestwin, [11](#)

## \*Topic **utilities**

geplot, [4](#)

HPD, [5](#)

plotbayestwin, [8](#)

simulatetwin, [9](#)

BayesTwin, [2](#)

ge\_plot (geplot), [4](#)

GEplot (geplot), [4](#)

geplot, [4](#)

HPD, [5](#)

hpd (HPD), [5](#)

IRTtwin, [6](#)

plotBayestwin (plotbayestwin), [8](#)

plotbayestwin, [8](#)

Simulatedata (simulatetwin), [9](#)

simulatetwin, [9](#)

simulatetwindata (simulatetwin), [9](#)

summary.bayestwin, [11](#)

summaryBayestwin (summary.bayestwin), [11](#)

summarybayestwin (summary.bayestwin), [11](#)