

Package ‘HSROC’

February 15, 2013

Type Package

Title Joint meta-analysis of diagnostic test sensitivity and specificity with or without a gold standard reference test

Version 2.1.6

Date 2013-01-07

Author Ian Schiller and Nandini Dendukuri

Maintainer Ian Schiller <ian.schiller@clinepi.mcgill.ca>

Depends R (>= 2.15.2), lattice, coda, MASS, MCMCpack

Description This package implements a model for joint meta-analysis of sensitivity and specificity of the diagnostic test under evaluation, while taking into account the possibly imperfect sensitivity and specificity of the reference test. This hierarchical model accounts for both within and between study variability. Estimation is carried out using a Bayesian approach, implemented via a Gibbs sampler. The model can be applied in situations where more than one reference test is used in the selected studies.

License GPL (>= 2)

URL <http://www.nandinidendukuri.com/>

Encoding latin1

Repository CRAN

Date/Publication 2013-01-07 17:57:40

NeedsCompilation yes

R topics documented:

beta.parameter	2
HSROC	3
HSROCSummary	8
In.house	11
MRI	12
simdata	13
Index	17

beta.parameter	<i>A function that returns the shape parameters of the beta distribution</i>
----------------	--

Description

When provided the minimum and maximum values corresponding to the prior information on a parameter (within [0, 1]), the function returns the shape parameters of the beta distribution, namely α and β , that covers this range.

Usage

beta.parameter(low, up)

Arguments

- low numeric argument. Lower bound of the range
- up numeric argument. Upper bound of the range

Details

The low argument must always be less than the up argument. Also, both low and up arguments must remain within [0, 1].

Value

It returns a list object consisting of the shape parameters alpha and beta of the Beta distribution such that

$$\frac{\alpha}{\alpha + \beta} = \frac{low + up}{2}$$

and

$$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{up - low}{4}$$

The function has been built such that the particular case where $low = up = 1$ will return $\alpha = \beta = 1$.

Examples

```
## Not run: beta.parameter(-1, 0.5) #Returns error!
## Not run: beta.parameter(0, 0) #Not allowed. Returns error!
## Not run: beta.parameter(0.75, 0.25) #Returns error!

beta.parameter(0, 1)
beta.parameter(0.5, 1)
beta.parameter(0.1, 0.7)
```

HSROC	<i>A function for joint meta-analysis of sensitivity and specificity of a diagnostic test.</i>
-------	--

Description

This function is used to estimate the parameters of a hierarchical summary receiver operating characteristic (HSROC) model allowing for the reference standard to be possibly imperfect, and assuming it is conditionally independent from the test under evaluation. The estimation is carried out using a Gibbs sampler.

Usage

```
HSROC(data, iter.num, init = NULL, sub_rs=NULL,
      first.run = TRUE, path=getwd(), refresh = 100,
      prior.Seref=NULL, prior.SPref=NULL, prior.PI=c(0,1),
      prior.LAMBDA = c(-3,3), prior.THETA = c(-1.5,1.5),
      prior_sd_alpha = list(0,2,"sd"), prior_sd_theta = list(0,2,"sd"),
      prior_beta = c(-0.75,0.75))
```

Arguments

<code>data</code>	a matrix with the number of rows equal to the number of studies and 4 columns. Each row consists of the entries of the 2x2 table of the index test (i.e. test under evaluation) vs. the reference test reported in each study. The ordering of the columns is ++, +-, -+, -, where the first entry refers to the result of the test under evaluation and the second entry refers to the result of the reference test.
<code>iter.num</code>	the number of iterations of the Gibbs sampler.
<code>init</code>	a list of initial values. See details for further explanation.
<code>sub_rs</code>	a list that describes the partition of the reference standards among the studies, if any. The first element of the list is the number of different reference standards used. Element 2 to <code>sub_rs[1]</code> are vectors of integers specifying the study numbers that used each reference test. See details.

prior.SEref	a vector of values specifying the lower and upper limits of the sensitivity of the reference test(s) based on prior information. If there are multiple reference standards, the lower and upper limits are each specified as a vector of length <code>sub_rs[1]</code> . The default value is NULL, implying a single reference standard with perfect sensitivity.
prior.SPref	a vector of values specifying the lower and upper limits of the specificity of the reference test(s) based on prior information. If there are multiple reference standards, the lower and upper limits are each specified as a vector of length <code>sub_rs[1]</code> . The default value is NULL, implying a single reference standard with perfect specificity.
first.run	logical. If TRUE (default), the <code>gibbs.sampler</code> function is run for the first time. If FALSE, the function assumes k iterations where already run and it continues from where it left off, that is from iteration k+1.
path	a character string pointing to the directory where the sample drawn at each iteration of the Gibbs sampler are to be stored.
refresh	A positive integer defining the frequency at which the Gibbs sampler's progress will be displayed on the R GUI. The default is 100.
prior.PI	a vector with 2 components specifying the minimum and maximum values for the prevalence in each study based on prior information. If unspecified, $Beta(1, 1)$ priors are used.
prior.LAMBDA	a vector with 2 components specifying the minimum and maximum values for the difference in mean values (diagnostic accuracy) among disease positive and negative groups based on prior information. The default value is $c(-3, 3)$ implying a $U(-3, 3)$.
prior.THETA	a vector with 2 components specifying the minimum and maximum values for the overall mean cut-off value for defining a positive test based on prior information. The default value is $c(-1.5, 1.5)$ implying a $U(-1.5, 1.5)$.
prior_sd_alpha	a list with 3 components. The first 2 components are specifying the minimum and maximum values for the between study standard deviation in the difference in mean values of the disease positive and negative groups in the i^{th} study, α_i , based on prior information. The third component determine whether we want the prior to be on the standard deviation (sd) or the variance (v). The default value is <code>list(0, 2, "sd")</code> implying a $U(0, 2)$ prior.
prior_sd_theta	a list with 3 components. The first 2 components are specifying the minimum and maximum values for the between study standard deviation in the cut-off, θ_i , in the i^{th} study based on prior information. The third component determine whether we want the prior to be on the standard deviation (s) or the variance (v). The default value is <code>list(0, 2, "sd")</code> implying a $U(0, 2)$ prior.
prior_beta	a vector with 2 components specifying the minimum and maximum values for the logarithm of the ratio of the standard deviation of test results among patients with and without the disease, based on prior belief. This parameter is assumed to be constant across studies. The default value is $c(-0.75, 0.75)$ implying a $U(-0.75, 0.75)$. If the argument is (NULL) the function assumes a range of $(-\log((LAMBDA.up/3) + 1), \log((LAMBDA.up/3) + 1))$, where LAMBDA.up is the upper limit of prior.LAMBDA.

Details

Our HSROC model uses a probit link function and not the logit link function used in the HSROC model developed by Rutter and Gatsonis (2001).

The probability of a positive result on the index test for the j^{th} individual in the i^{th} study is given by :

$$1 - \Phi((\theta_i - \alpha_i D_{ij}) \exp(-\beta D_{ij})),$$

while the probability of a negative result on the index test for the j^{th} individual in the i^{th} study is given by :

$$\Phi((\theta_i - \alpha_i D_{ij}) \exp(-\beta D_{ij})),$$

where $D_{ij} = 0.5$ if the j^{th} individual in the i^{th} study is disease positive and -0.5 if the j^{th} individual in the i^{th} study is not disease negative.

When ranges are provided for `prior.SEref`, `prior.SPref` and `prior.PI` they are transformed into Beta prior distributions using the method described in `beta.parameter`.

If the argument `init=NULL`, the function will randomly generate initial values based on the prior information. Otherwise, the user can provide his own initial values for the within-study and between-study parameters and for the reference standards through the `init` argument, with the within-study, between study and reference standard initial values as the first, second and third element of `init`, respectively. Furthermore, the within-study parameters must be a matrix-like object with each column being initial values for α_i , θ_i , sensitivity of the test under evaluation S_{1i} , specificity of the test under evaluation C_{1i} and prevalence π_i . The between-study parameters must be a vector of the following initial values : Θ , σ_θ , Λ , σ_α and β . The initial values for the reference standard must be a 2 X `sub_rs[[1]]` matrix-like object. The first row must be the initial values of the sensitivity of the reference standard, while the second row must correspond to initial values of the specificity of the reference standard. The ordering described above in the within-study, between-study and reference standard parameters must be preserved.

The first element of the list-object `sub_rs` corresponds to the number of different reference standards. The default value is 1. The number of additional elements will depend on the value of the first element. There must be as many additional elements in `sub_rs` as there are different reference standards. Assuming the studies are labelled 1, ..., N, each of these additional elements must be a vector (possibly of length one) taking as their values the labelling of the corresponding studies sharing the same reference standard. For example, if we have 2 reference tests, the first one applied over study 1-10 and the second one applied over study 11-15 then the `sub_rs` list-argument should be of length 3 with the following elements : 3, 1:10, 11:15

The `path` argument points to the working directory where files written at each iteration of the Gibbs sampler will be saved. If no path is provided, the current working directory will be used

For more help on this function, see the tutorial pdf file available at <http://www.nandinidendukuri.com/files/onjoomla/site/HSROC>

Value

Text files with samples from the joint posterior distribution of the between-study parameters, within-study parameters and performance parameters of the reference standard(s) are created in the `path` directory. These results can be summarized using the `HSROCSummary` function.

The following files are also created and saved in the path directory :

“Prior.information.txt”, lists the prior distributions.

“Initial values.txt”, lists the initial values used. If the argument `init = NULL`, the initial value file is called “Random Initial values.txt”.

A series of files listing the values of various parameters sampled in the last iteration of the Gibbs sampler as follows :

“REstarting values.txt”, contains values of the within-study parameters (α_i , θ_i , sensitivity of test under evaluation (S_{1i}), specificity of test under evaluation (C_{1i}) and prevalence (π_i)).

“REstarting values 2.txt”, contains values of the between-study parameters (Λ , standard deviation of α_i , (σ_α), Θ , standard deviation of θ_i (σ_θ) and β).

“REstarting REFSTD.txt”, contains values of the sensitivity and specificity of the reference test.

“REstart values index.txt”, lists the variable names in the 3 files described above.

References

Dendukuri, N., Schiller, I., Joseph, L., and Pai, M. (2012) *Bayesian meta-analysis of the accuracy of a test for tuberculosis pleuritis in the absence of a gold-standard reference*. Biometrics. doi:10.1111/j. 1541-0420.2012.01773.x

Rutter, C. M., and Gatsonis, C. A. (2001) *A hierarchical regression approach to meta-analysis of diagnostic accuracy evaluations*. Statistics in Medicine, 20(19):2865-2884.

Examples

```
#=====
#TO SET UP THE REFERENCE STANDARD
#=====

#There were three different reference standards for the In.house dataset.
#The first reference standard was used in study 1 and 2,
#the second was used in studies 3 and 4 and the third in study 5 to 12.
REFSTD = list(3, 1:2, 3:4, 5:11)

#=====
#TO SET UP DATA AND INITIAL VALUES
#=====

data(In.house)
M = length(In.house[,1])

#Initial values for the within-study parameters
init.alpha = rep(2.5, M) ; init.theta = rep(1, M) ;
init.s1 = rep(0.5, M) ; init.c1 = rep(0.5, M) ;
init.pi = rep(0.5, M)

#Initial values for the between-study parameters
init.THETA = 1 ; init.sd.theta = 0.5 ;
```

```

init.LAMBDA = 2.5 ; init.sd.alpha = 0.5 ;
init.beta = 0 ;

#Initial values for the reference standard sensitivities and specificities
init.s2 = rep(0.5, REFSTD[[1]]) ; init.c2 = rep(0.5, REFSTD[[1]])

#The ordering of the initial values is important!
init1 = cbind(init.alpha, init.theta, init.s1, init.c1, init.pi)
init2 = c(init.THETA, init.sd.theta, init.LAMBDA, init.sd.alpha, init.beta)
init3 = rbind(init.s2, init.c2)

init = list(init1, init2, init3)
#=====
#TO PROVIDE PRIOR INFORMATION
#=====

S2.a = c(0.2, 0.2, 0.7) ; S2.b = c(0.6, 0.7, 0.9)
C2.a = rep(0.9, 3) ; C2.b = rep(1, 3)

#=====
#TO RUN GIBBS SAMPLER
#=====

## Not run:

set.seed(10)
HSROC(data=In.house, init=init, iter.num=50000,
      prior.SEref=c(S2.a,S2.b), prior.SPref=c (C2.a,C2.b), sub_rs=REFSTD)

#Putting prior information on sigma_alpha^2 (sigma_alphs squared)
#instead of sigma_alpha
set.seed(10)
HSROC(data=In.house, init=init, iter.num=50000,
      prior.SEref=c(S2.a,S2.b), prior.SPref=c (C2.a,C2.b),
      sub_rs=REFSTD, prior_sd_alpha = list(0,2,"v"))

#Letting the function select randomly its own initial values
set.seed(10)
HSROC(data=In.house, iter.num=50000,
      prior.SEref=c(S2.a,S2.b), prior.SPref=c (C2.a,C2.b),
      sub_rs=REFSTD)

#For more help on this function, see the tutorial pdf file availalbe
#at http://www.nandinidendukuri.com/filesonjoomlasite/HSROC\_R\_Tutorial.pdf

## End(Not run)

```

Description

This function produces summaries for the HSROC model parameters

Usage

```
HSROCSummary(data, burn_in = 0, iter.keep = NULL, Thin = 1, sub_rs=NULL,
  point_estimate = c("median", "mean"), summary.path = getwd(),
  chain = getwd(), tv = NULL, digit = 6, print_plot = FALSE,
  plot.ind.studies = TRUE, cred_region = TRUE, predict_region = TRUE,
  col.pooled.estimate = "red", col.predict.region = "blue",
  lty.cred.region = "dotted", lty.predict.region = "dotted",
  region_level = 0.95, trunc_low = 0.025, trunc_up = 0.025 )
```

Arguments

data	a matrix with the number of rows equal to the number of studies and 4 columns. Each row consists of the entries of the 2x2 table of the index test (i.e. test under evaluation) vs. the reference test reported in each study. The ordering of the columns is ++, +-, -+, --, where the first entry refers to the result of the test under evaluation and the second entry refers to the result of the reference test.
burn_in	The number of early iterations that are to be dropped. The default value is 0.
iter.keep	Maximum number of iteration we want to keep.
Thin	a single numeric value. It sets the numerical field used to select every thin-th iteration to contribute to the estimates being calculated. The default value is 1.
sub_rs	a list that describes the partition of the reference standard among the studies, if any. See details for further explanations.
point_estimate	a character string indicating which method is to be used to calculate the estimates. One of "median" (default) or "mean", can be used
summary.path	a character string pointing to the directory where the SUMMARY files are to be stored.
chain	A list of character strings pointing to each directory, one for each chain, where the files created during the Gibbs sampler process are stored.
tv	a list of true parameter values. See details for further explanations
digit	integer indicating the number of decimal places to be used. The default value is 6.
print_plot	logical. If TRUE, pdf files of trace, density and summary receiver operating characteristic (SROC) curve plots are saved in the summary.path working directory to help assess convergence of the Gibbs sampler.

<code>plot.ind.studies</code>	logical. If TRUE (default), empty circles representing individual studies are drawn on the SROC plot. The radius of the circles are proportional to the sample size of the studies.
<code>cred_region</code>	logical. If TRUE (default), a credible region curve is drawn on the SROC plot.
<code>predict_region</code>	logical. If TRUE (default), a prediction region curve is drawn on the SROC plot.
<code>col.pooled.estimate</code>	A specification for the default SROC plotting color of the pooled estimate point and credible region curve. Defaults to "red".
<code>col.predict.region</code>	A specification for the default SROC plotting color of the prediction region curve. Defaults to "blue".
<code>lty.cred.region</code>	The credible region line type. Default to "dotdash"
<code>lty.predict.region</code>	The prediction region line type. Default to "dotted"
<code>region_level</code>	The credible (prediction) level required for the credible (prediction) region.
<code>trunc_low</code>	Lower limit truncation entering the creation of the SROC curve. Default to 0.025. See details for further explanations.
<code>trunc_up</code>	Upper limit truncation entering the creation of the SROC curve. Default to 0.025. See details for further explanations.

Details

The first element of the list-object `sub_rs` corresponds to the number of different reference standards. The default value is 1. The number of additional elements will depend on the value of the first element. There must be as many additional element in `sub_rs` as there are different reference standards. Assuming the studies are labelled 1, ..., N, each of these additional elements must be a vector (possibly of length one) taking as their values the labelling of the corresponding studies sharing the same reference standard. For example, if we have 2 reference tests, the first one applied over study 1-10 and the second one applied over study 11-15 then the `sub_rs` list-argument should be of length 3 with the following elements : 3, 1:10, 11:15

If the argument `tv` is equal to NULL, the function assumes the data are coming from a real life example. Otherwise, it assumes the data are coming from a simulated dataset in which case the user must provide the "true parameters" that were used to simulate the data for the within-study and between-study parameters and for the reference standards through the `tv` argument. The within-study parameters must be a matrix-like object with each column being true values for α_i , θ_i , sensitivity of test under evaluation S_{1i} , specificity of test under evaluation C_{1i} and prevalence π_i . The between-study parameters must be a vector of the following true values : Θ , σ_θ , Λ , σ_α and β . The reference standard initial values must be a 2 X `sub_rs[[1]]` matrix-like object. The first row must be the true values of the sensitivity of the reference standard, while the second row must correspond to true values of the specificity of the reference standard. The ordering described above in the within-study, between-study and reference standard true parameters must be preserved otherwise an error message will be displayed.

If the argument `print_plot` is equal to TRUE, the function will create and save in the `summary.path` working directory 3 type of plots to help the user judge if the descriptive statistics are reliable.

First, a trace plot for each parameter will be created to help evaluate whether the Gibbs sampler has converged. Each trace plot is a scatter plot of the posterior sample of a single parameter vs the iteration number of the Gibbs sampler. Second, a density plot for each parameter will also be created. It plots a smoothed posterior kernel density estimate for each parameter. Finally, a SROC curve plot will also be created by the function. It summarizes the performance of diagnostic tests by plotting the relationship between the true positive rate and the false positive rate of the tests, as the threshold used to distinguish disease cases from noncases varies.

The SROC curve plots Sensitivity (λ) vs Specificity (β) at the posterior mean values of λ and β as the value of T varies over the posterior sample of Θ . The `trunc_low` and `trunc_up` arguments truncate the posterior sample of Θ by $(\text{trunc_low}) \times 100$

For more help on this function, see the tutorial pdf file available at <http://www.nandinidendukuri.com/files/onjoomla/site/HSROC>

Value

A list of : Point estimates and 95% highest posterior density (HPD) intervals of the between-study parameters, within-study parameters and reference standard parameters. All estimates are obtained via a Gibbs sampler process.

It also prints in the `summary.path` working directory a text file with a more complete summary of the results listed above (i.e. the results also include standard errors and MC errors plus the posterior predictive value for the sensitivity and specificity of a new study that has not yet taken place).

References

- Dendukuri, N., Schiller, I., Joseph, L., and Pai, M. (2012) *Bayesian meta-analysis of the accuracy of a test for tuberculosis pleuritis in the absence of a gold-standard reference*. Biometrics. doi:10.1111/j. 1541-0420.2012.01773.x
- Rutter, C. M., and Gatsonis, C. A. (2001) *A hierarchical regression approach to meta-analysis of diagnostic accuracy evaluations*. Statistics in Medicine, 20(19):2865-2884.
- Scheidler J, Hricak H, Yu KK, Subak L, Segal MR. *Radiological evaluation of lymph node metastases in patients with cervical cancer : a meta-analysis*. Journal of the American Medical Association 1997 ; 278(13):1096-1101.
- Pai, M. et al. (2004) *Nucleic acid amplification in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis*. BMC Infect Dis 2004, 4:6.
- Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., and Zwinderman, A. H. (2005) *Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews*. Journal of Clinical Epidemiology ; 58(10), 982-990.
- Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P., and Sterne, J. A. (2006) *Unification of models for meta-analysis of diagnostic accuracy studies*. Biostatistics ; 8(2), 239-251.

Examples

```
#REAL-LIFE EXAMPLES
#
#PLEASE NOTE THAT BOTH EXAMPLES BELOW ASSUME THE EXISTANCE OF POSTERIOR SAMPLES OBTAINED FROM THE 'HSROC' FUNCTION
#IN OTHER WORDS' ONE NEEDS TO RUN THE 'HSROC' FUNCTION BEFORE USING THE 'HSROCSummary' FUNCTION.
#
```

```

#Example 1
#To get descriptive statistics and graphical summaries for the MRI data
#(Scheidler et al. 1997) after dropping the first 5,000 iterations.

data(MRI) #load the data
## Not run:
HSROCSummary(data = MRI, burn_in=5000, print_plot=TRUE )

## End(Not run)

#Example 2
#To get descriptive statistics and graphical summaries for the In.house
#data (Pai et al. 2004) coming from 2 different chains.
#We provide the path to each chain's directory, i.e. the directory where
#all files created during the Gibbs sampler process are stored for
#each chain. Let's assume there are two fictional directoies
#chain_path = list("C:/path_to_chain_1", "C:/path_to_chain_2").
#Let's assume we drop the first 5,000 iterations and we use a thinning
#interval of 10.

data(In.house) #load the data
## Not run:
HSROCSum1<- HSROCSummary(data = In.house, burn_in=5000, Thin=10,
chain=chain_path, print_plot=TRUE,
sub_rs=REFSTD )

#For more help on this function, see the tutorial pdf file availalbe
#at http://www.nandinidendukuri.com/filesonjoomlasite/HSROC\_R\_Tutorial.pdf

## End(Not run)

```

In.house

*IN-HOUSE NUCLEIC ACID AMPLIFICATION TESTS (INH) FOR
TB PLEURITIS*

Description

This dataset gives the observed cross-tabulation of the test under evaluation (nucleic acid amplification test) and the reference test. In studies 1 and 2, the reference test was a culture test, in studies 3 and 4 it was a composite reference test of culture and clinical data (including signs, symptoms and clinical response to empiric TB therapy) and in studies 5 to 11 it was a composite reference test of culture and biopsy.

Usage

```
data(In.house)
```

Format

A matrix with 11 observations on the following 4 variables.

- ++ Observed individuals who tested positive on both tests
- +− Observed individuals who tested positive on the test under evaluation and negative on the reference test
- −+ Observed individuals who tested negative on the test under evaluation and positive on the reference test
- Observed individuals who tested negative on both tests

References

Pai, M. et al. (2004) *Nucleic acid amplification in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis*. BMC Infect Dis 2004, 4:6.

Examples

```
data(In.house)
In.house
```

MRI

MAGNETIC RESONANCE IMAGING TEST (MRI) for evaluation of lymph node metastasis in women with invasive cervical cancer

Description

This dataset gives the observed cross-tabulation of the MRI (test under evaluation) and histologic/cytologic specimens obtained by surgery or lymph node biopsy (reference test).

Usage

```
data(MRI)
```

Format

A matrix with 10 observations on the following 4 variables.

- ++ Observed individuals who tested positive on both tests
- +− Observed individuals who tested positive on the test under evaluation and negative on the reference test
- −+ Observed individuals who tested negative on the test under evaluation and positive on the reference test
- Observed individuals who tested negative on both tests

References

Scheidler J, Hricak H, Yu KK, Subak L, Segal MR. *Radiological evaluation of lymph node metastases in patients with cervical cancer : a meta-analysis*. Journal of the American Medical Association 1997 ; 278(13):1096-1101.

C. M. Rutter and C. A. Gatsonis. *A hierarchical regression approach to meta-analysis of diagnostic accuracy evaluations*. Statistics in Medicine 2001 ; 20(19):2865-2884.

Examples

```
data(MRI)
MRI
```

simdata	<i>Simulate a dataset based on a HSROC model</i>
---------	--

Description

This function simulates a dataset based on the HSROC diagnostic meta-analysis model. It allows for the reference standard to be imperfect or perfect.

Usage

```
simdata(N, n, n.random = "FALSE", sub_rs = NULL, prev, se_ref = NULL,
        sp_ref = NULL, T, range.T = c(-Inf, Inf), L, range.L = c(-Inf, Inf),
        sd_t, sd_a, b, path = getwd() )
```

Arguments

N	the number of studies to be included in the meta-analysis.
n	numerical vector, possibly a single value, specifying the number of individuals within each study. See details for further explanations.
n.random	if TRUE, the number of individuals within each study is drawn from n with replacement.
sub_rs	a list that specifies the reference standard used by each study. See details for further explanations.
prev	a vector of length N giving the prevalence in each study.
se_ref	a vector of length equal to the number of reference standards giving the sensitivity for each reference test.
sp_ref	a vector of length equal to the number of reference standards giving the specificity for each reference test.
T	single numeric value, the overall mean cut-off value to define a positive test.
range.T	a vector of length 2 specifying a range of values for the individual cut-off θ_i . See details for further explanations.

L	single numeric value, the overall difference in mean values (diagnostic accuracy) on the continuous index test result comparing the diseased group and the non-diseased group.
range.L	a vector of length 2 specifying a range of values for the individual difference in mean values (diagnostic accuracy) on the continuous index test result comparing the diseased group and the non-diseased group α_i . See details for further explanations.
sd_t	single numeric value, the between study standard deviation in the cut-off θ_i .
sd_a	single numeric value, the between study standard deviation in the mean value of the index test disease group α_i
b	single numeric value, the ratio of the continuous standard deviation of the index test results on patients with the disease compared to patients without the disease.
path	a character string pointing to the directory where the simulated data will be saved to.

Details

The HSROC model uses the following parametrization : $S_i = \Phi(\frac{-\theta_i + \alpha_i/2}{\exp(\beta/2)})$ and $C_i = \Phi(\frac{\theta_i + \alpha_i/2}{\exp(-\beta/2)})$

If n.random is FALSE, the number of components in n must match the value of N, unless n is equal to a single value. For the latter case, all studies would be assumed to have the same number of individuals, that is n. If n.random is TRUE, the number of elements may not necessarily be equal to the value of N.

The first element of the list-object sub_rs corresponds to the number of different reference standards. The default value is 1. The number of additional elements will depend on the value of the first element. There must be as many additional elements in sub_rs as there are different reference standards. Assuming the studies are labelled 1, ..., N, each of these additional elements must be a vector (possibly of length one) taking as their values the labels of the studies sharing the same reference standard. For example, if we have 2 reference tests, the first one applied over studies 1-10 and the second one applied over studies 11-15 then the sub_rs list-argument should be of length 3 with the following elements : 2, 1:10, 11:15

The range.T argument ensures the individual θ_i will be generated within the range provided. If no range is provided by the user (default) the function assumes no restrictions are made on the possible values of θ_i . The range.L argument ensures the individual α_i will be generated within the range provided. If no range is provided by the user (default) the function assumes no restrictions are made on the possible values of θ_i .

For more help on this function, see the tutorial pdf file available on <http://www.nandinidendukuri.com/filesonjoomlasite/HSROC>

Value

A list of the 2x2 tables for each study, the between-study parameters, the within-study parameters and the reference standard.

Text files are created in the path directory. These files are :

“True_values.txt”, reports the within-study parameters α_i , θ_i , sensitivity of test under evaluation (S_{1i}), specificity of test under evaluation (C_{1i}) and prevalence (π_i) used in the simulation.

“True_values2.txt”, reports the values of the between-study parameters Λ , standard deviation of α_i (σ_α), Θ , standard deviation of θ_i (σ_θ) and β used to simulate the data.

“True_REFSTD.txt”, reports the values of the reference standard used to simulate the data.

“True_values_index.txt”, reports the variable names of the 3 files described above.

.

References

Dendukuri, N., Schiller, I., Joseph, L., and Pai, M. (2012) *Bayesian meta-analysis of the accuracy of a test for tuberculosis pleuritis in the absence of a gold-standard reference*. Biometrics. doi:10.1111/j. 1541-0420.2012.01773.x

Rutter, C. M., and Gatsonis, C. A. (2001) *A hierarchical regression approach to meta-analysis of diagnostic accuracy evaluations*. Statistics in Medicine, 20(19):2865-2884.

Examples

```
#EXAMPLE 1
#We want to simulate data for 10 studies based on an HSROC model assuming
#each study uses the same imperfect reference standard.
```

```
N = 10
LAMBDA = 2
sd_alpha = 0.75
THETA = 1.5
sd_theta = 0.5
beta = 0
pi = runif(10,0,1)
```

```
REFSTD = list(1, 1:10) #Only 1 reference standard ...
s2 = c(0.5) #Sensitivity of the reference test
c2 = c(0.85) #Specificity of the reference test
```

```
sim.data = simdata(N=N, n = c(50,50,60,60,70,70,80,80,90,90),
  sub_rs = REFSTD, prev=pi, se_ref=s2, sp_ref=c2, T=THETA,
  L=LAMBDA, sd_t=sd_theta, sd_a=sd_alpha, b=beta)
```

```
#EXAMPLE 2
#We want to simulate data for 15 studies based on an HSROC model such that
#the first 5 studies share a common reference standard and the remaining
#10 studies also share a common reference standard.
```

```
N = 15
LAMBDA = 3.6
sd_alpha = 1.15
THETA = 2.3
sd_theta = 0.75
beta = 0.15
pi = runif(15,0.1,0.5)
```

```

REFSTD = list(2, 1:5, 6:15) #Two different reference standards ...
s2 = c(0.40, 0.6) #Sensitivity of the reference tests
c2 = c(0.75,0.95) #Specificity of the reference tests

#Thus, for the first 5 studies, S2 = 0.40 and C2 = 0.75 while for the last
#10 studies s2 = 0.6 and c2 = 0.95

sim.data = simdata(N=N, n=seq(30,120,1), n.random=TRUE, sub_rs = REFSTD,
  prev=pi, se_ref=s2, sp_ref=c2, T=THETA, L=LAMBDA, sd_t=sd_theta,
  sd_a=sd_alpha, b=beta)

#EXAMPLE 3
#Assume the same context as the one in EXAMPLE 2 and let's suppose
#that each individual cut-off theta_i should lie between [-5,5]

N = 15
LAMBDA = 3.6
sd_alpha = 1.15
THETA = 2.3
sd_theta = 0.75
beta = 0.15
pi = runif(15,0.1,0.5)

REFSTD = list(2, 1:5, 6:15) #Two different reference standards ...
s2 = c(0.40, 0.6) #Sensitivity of the reference tests
c2 = c(0.75,0.95) #Specificity of the reference tests

#Thus, for the first 5 studies, S2 = 0.40 and C2 = 0.75 while for the last
#10 studies s2 = 0.6 and c2 = 0.95

sim.data = simdata(N=N, n=seq(30,120,1), n.random=TRUE, sub_rs = REFSTD,
  prev=pi, se_ref=s2, sp_ref=c2, T=THETA, range.T=c(-5,5),L=LAMBDA,
  sd_t=sd_theta,sd_a=sd_alpha, b=beta)

```


Index

- *Topic **datagen**
 - simdata, [13](#)
- *Topic **datasets**
 - In.house, [11](#)
 - MRI, [12](#)
- *Topic **methods**
 - beta.parameter, [2](#)
- *Topic **models**
 - HSROC, [3](#)
 - HSROCSummary, [8](#)
- beta.parameter, [2](#)
- HSROC, [3](#)
- HSROCSummary, [8](#)
- In.house, [11](#)
- MRI, [12](#)
- simdata, [13](#)