

Package ‘EvidenceSynthesis’

July 31, 2020

Type Package

Title An R Package for Combining Evidence from Multiple Sources

Version 0.1.0

Date 2020-07-21

Maintainer Martijn Schuemie <schuemie@ohdsi.org>

Description Routines for combining evidence and diagnostics across multiple sources, such as multiple data sites in a distributed study. This includes functions for performing meta-analysis and forest plots.

Depends survival

Imports ggplot2,
gridExtra,
meta,
EmpiricalCalibration,
rJava,
BeastJar,
Cyclops,
HDInterval,
rlang,
methods

Suggests knitr,
testthat

Remotes beast-dev/BeastJar

License Apache License 2.0

URL <https://github.com/OHDSI/EvidenceSynthesis>

BugReports <https://github.com/OHDSI/EvidenceSynthesis/issues>

VignetteBuilder knitr

RoxygenNote 7.1.0

Roxygen list(markdown = TRUE)

Encoding UTF-8

R topics documented:

approximateLikelihood	2
computeBayesianMetaAnalysis	3

computeFixedEffectMetaAnalysis	4
createSimulationSettings	5
customFunction	6
plotCovariateBalances	6
plotEmpiricalNulls	7
plotLikelihoodFit	8
plotMcmcTrace	9
plotMetaAnalysisForest	10
plotPerDbMcmcTrace	11
plotPerDbPosterior	12
plotPosterior	12
plotPreparedPs	13
preparePsPlot	14
simulatePopulations	15
skewNormal	15
Index	17

approximateLikelihood	<i>Approximate a likelihood function</i>
-----------------------	--

Description

Approximate the likelihood function using a parametric (normal, skew-normal, or custom parametric), or grid approximation. The approximation does not reveal person-level information, and can therefore be shared amongst data sites. When counts are low, a normal approximation might not be appropriate.

Usage

```
approximateLikelihood(
  cyclopsFit,
  parameter = 1,
  approximation = "custom",
  bounds = c(log(0.1), log(10))
)
```

Arguments

cyclopsFit	A model fitted using the <code>Cyclops::fitCyclopsModel()</code> function.
parameter	The parameter in the cyclopsFit object to profile.
approximation	The type of approximation. Valid options are 'normal', 'skew normal', 'custom', or 'grid'.
bounds	The bounds on the effect size used to fit the approximation.

Value

A vector of parameters of the likelihood approximation.

Examples

```
populations <- simulatePopulations()
cyclopsData <- Cyclops::createCyclopsData(Surv(time, y) ~ x + strata(stratumId),
                                         data = populations[[1]],
                                         modelType = "cox")
cyclopsFit <- Cyclops::fitCyclopsModel(cyclopsData)
approximation <- approximateLikelihood(cyclopsFit, "x")
```

computeBayesianMetaAnalysis

Compute a Bayesian meta-analysis

Description

Compute a Bayesian meta-analysis using the Markov chain Monte Carlo (MCMC) engine BEAST.

A normal and half-normal prior are used for the mu and tau parameters, respectively, with standard deviations as defined by the priorSd argument.

Usage

```
computeBayesianMetaAnalysis(
  data,
  chainLength = 1100000,
  burnIn = 1e+05,
  subSampleFrequency = 100,
  priorSd = c(2, 0.5),
  alpha = 0.05
)
```

Arguments

data	A data frame containing either normal, skew-normal, custom parametric, or grid likelihood data, with one row per database.
chainLength	Number of MCMC iterations.
burnIn	Number of MCMC iterations to consider as burn in.
subSampleFrequency	Subsample frequency for the MCMC.
priorSd	A two-dimensional vector with the standard deviation of the prior for mu and tau, respectively.
alpha	The alpha (expected type I error) used for the credible intervals.

Value

A data frame with the point estimates and 95% credible intervals for the mu and tau parameters (the mean and standard deviation of the distribution from which the per-site effect sizes are drawn). Attributes of the data frame contain the MCMC trace and the detected approximation type.

Examples

```

populations <- simulatePopulations()

fitModelInDatabase <- function(population) {
  cyclopsData <- Cyclops::createCyclopsData(Surv(time, y) ~ x + strata(stratumId),
                                             data = population,
                                             modelType = "cox")

  cyclopsFit <- Cyclops::fitCyclopsModel(cyclopsData)
  approximation <- approximateLikelihood(cyclopsFit, "x")
  return(approximation)
}
approximations <- lapply(populations, fitModelInDatabase)
approximations <- do.call("rbind", approximations)
estimate <- computeBayesianMetaAnalysis(approximations)
estimate
# mu      mu95Lb    mu95Ub      muSe      tau      tau95Lb    tau95Ub
# 0.0003129562 -0.1747429 0.1723472 0.089661 0.07759992 0.0002024991 0.301007

```

```
computeFixedEffectMetaAnalysis
```

Compute a fixed-effect meta-analysis

Description

Compute a fixed-effect meta-analysis.

Usage

```
computeFixedEffectMetaAnalysis(data, alpha = 0.05)
```

Arguments

data	A data frame containing either normal, skew-normal, custom parametric, or grid likelihood data. One row per database.
alpha	The alpha (expected type I error) used for the confidence intervals.

Examples

```

populations <- simulatePopulations()

fitModelInDatabase <- function(population) {
  cyclopsData <- Cyclops::createCyclopsData(Surv(time, y) ~ x + strata(stratumId),
                                             data = population, modelType = "cox")

  cyclopsFit <- Cyclops::fitCyclopsModel(cyclopsData)
  approximation <- approximateLikelihood(cyclopsFit, "x")
  return(approximation)
}
data <- lapply(populations, fitModelInDatabase)
data <- do.call("rbind", data)
computeFixedEffectMetaAnalysis(data)

```

`createSimulationSettings`*Create simulation settings*

Description

Create an object specifying a simulation. Currently only Cox proportional hazard models are supported.

Usage

```
createSimulationSettings(  
  nSites = 5,  
  n = 10000,  
  treatedFraction = 0.2,  
  nStrata = 10,  
  minBackgroundHazard = 2e-07,  
  maxBackgroundHazard = 2e-05,  
  hazardRatio = 2,  
  randomEffectSd = 0  
)
```

Arguments

<code>nSites</code>	Number of database sites to simulate.
<code>n</code>	Number of subjects per site. Either a single number, or a vector of length <code>nSites</code> .
<code>treatedFraction</code>	Fraction of subjects that is treated. Either a single number, or a vector of length <code>nSites</code> .
<code>nStrata</code>	Number of strata per site. Either a single number, or a vector of length <code>nSites</code> .
<code>minBackgroundHazard</code>	Minimum background hazard. Either a single number, or a vector of length <code>nSites</code> .
<code>maxBackgroundHazard</code>	Maximum background hazard. Either a single number, or a vector of length <code>nSites</code> .
<code>hazardRatio</code>	Hazard ratio.
<code>randomEffectSd</code>	Standard deviation of the $\log(\text{hazardRatio})$. Fixed effect if equal to 0.

Value

An object of type `simulationSettings`.

customFunction	<i>A custom function to approximate a log likelihood function</i>
----------------	---

Description

A custom function to approximate a log likelihood function

Usage

```
customFunction(x, mu, sigma, gamma)
```

Arguments

x	The log(hazard ratio) for which to approximate the log likelihood.
mu	The position parameter.
sigma	The scale parameter.
gamma	The skew parameter.

Value

The approximate log likelihood for the given x.

plotCovariateBalances	<i>Plot covariate balances</i>
-----------------------	--------------------------------

Description

Plots the covariate balance before and after matching for multiple data sources.

Usage

```
plotCovariateBalances(
  balances,
  labels,
  threshold = 0,
  beforeLabel = "Before matching",
  afterLabel = "After matching",
  fileName = NULL
)
```

Arguments

balances	A list of covariate balance objects as created using the computeCovariateBalance() function in the CohortMethod package. Each balance object is expected to be a data frame with at least these two columns: beforeMatchingStdDiff and afterMatchingStdDiff.
labels	A vector containing the labels for the various sources.
threshold	Show a threshold value for the standardized difference.

beforeLabel	Label for before matching / stratification / trimming.
afterLabel	Label for after matching / stratification / trimming.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave for supported file formats.

Details

Creates a plot showing the covariate balance before and after matching. Balance distributions are displayed as box plots combined with scatterplots.

Value

A Ggplot object. Use the [ggplot2::ggsave](#).

Examples

```
# Some example data:
balance1 <- data.frame(beforeMatchingStdDiff = rnorm(1000, 0.1, 0.1),
                        afterMatchingStdDiff = rnorm(1000, 0.0, 0.01))
balance2 <- data.frame(beforeMatchingStdDiff = rnorm(1000, 0.2, 0.1),
                        afterMatchingStdDiff = rnorm(1000, 0.0, 0.05))
balance3 <- data.frame(beforeMatchingStdDiff = rnorm(1000, 0.0, 0.1),
                        afterMatchingStdDiff = rnorm(1000, 0.0, 0.03))
plotCovariateBalances(balances = list(balance1, balance2, balance3),
                      labels = c("Site A", "Site B", "Site C"))
```

plotEmpiricalNulls	<i>Plot empirical null distributions</i>
--------------------	--

Description

Plot the empirical null distribution for multiple data sources.

Usage

```
plotEmpiricalNulls(
  logRr,
  seLogRr,
  labels,
  xLabel = "Relative risk",
  limits = c(0.1, 10),
  showCis = TRUE,
  fileName = NULL
)
```

Arguments

logRr	A numeric vector of effect estimates for the negative controls on the log scale.
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = $(\log(\text{lower bound 95 percent confidence interval}) - \log(\text{effect estimate}))/qnorm(0.025)$.
labels	A vector containing the labels for the various sources. Should be of equal length as logRr and seLogRr.
xLabel	The label on the x-axis: the name of the effect estimate.
limits	The limits of the effect size axis.
showCis	Show the 95 percent confidence intervals on the null distribution and distribution parameter estimates?
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function <code>ggplot2::ggsave()</code> for supported file formats.

Details

Creates a plot showing the empirical null distributions. Distributions are shown as mean plus minus one standard deviation, as well as a distribution plot.

Value

A Ggplot object. Use the `ggplot2::ggsave()` function to save to file.

Examples

```
# Some example data:
site1 <- EmpiricalCalibration::simulateControls(n = 50, mean = 0, sd = 0.1, trueLogRr = 0)
site1$label <- "Site 1"
site2 <- EmpiricalCalibration::simulateControls(n = 50, mean = 0.1, sd = 0.2, trueLogRr = 0)
site2$label <- "Site 2"
site3 <- EmpiricalCalibration::simulateControls(n = 50, mean = 0.15, sd = 0.25, trueLogRr = 0)
site3$label <- "Site 3"
sites <- rbind(site1, site2, site3)

plotEmpiricalNulls(logRr = sites$logRr,
                   seLogRr = sites$seLogRr,
                   labels = sites$label)
```

plotLikelihoodFit

Plot the likelihood approximation

Description

Plot the likelihood approximation

Usage

```
plotLikelihoodFit(
  approximation,
  cyclopsFit,
  parameter = "x",
  logScale = TRUE,
  xLabel = "Hazard Ratio",
  limits = c(0.1, 10),
  fileName = NULL
)
```

Arguments

approximation	An approximation of the likelihood function as fitted using the approximateLikelihood() function.
cyclopsFit	A model fitted using the Cyclops::fitCyclopsModel() function.
parameter	The parameter in the cyclopsFit object to profile.
logScale	Show the y-axis on the log scale?
xLabel	The title of the x-axis.
limits	The limits on the x-axis.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave in the ggplot2 package for supported file formats.

Value

A Ggplot object. Use the [ggplot2::ggsave](#) function to save to file.

Examples

```
# Simulate a single database population:
population <- simulatePopulations(createSimulationSettings(nSites = 1))[[1]]

cyclopsData <- Cyclops::createCyclopsData(Surv(time, y) ~ x + strata(stratumId),
                                          data = population,
                                          modelType = "cox")

cyclopsFit <- Cyclops::fitCyclopsModel(cyclopsData)
approximation <- approximateLikelihood(cyclopsFit, parameter = "x")

plotLikelihoodFit(approximation, cyclopsFit, parameter = "x")
```

plotMcmcTrace

Plot MCMC trace

Description

Plot MCMC trace

Usage

```
plotMcmcTrace(
  estimate,
  showEstimate = TRUE,
  dataCutoff = 0.01,
  fileName = NULL
)
```

Arguments

estimate	An object as generated using the computeBayesianMetaAnalysis() function.
showEstimate	Show the parameter estimates (mode) and 95 percent confidence intervals?
dataCutoff	This fraction of the data at both tails will be removed.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave in the ggplot2 package for supported file formats.

Value

A Ggplot object. Use the [ggplot2::ggsave](#) function to save to file.

plotMetaAnalysisForest

Perform a meta-analysis and create a forest plot

Description

Perform a meta-analysis and creates a forest plot of effect size estimates.

Usage

```
plotMetaAnalysisForest(
  logRr,
  logLb95Ci,
  logUb95Ci,
  labels,
  xLabel = "Relative risk",
  limits = c(0.1, 10),
  hakn = FALSE,
  fileName = NULL
)
```

Arguments

logRr	A numeric vector of effect estimates on the log scale.
logLb95Ci	The lower bound of the 95 percent confidence interval on the log scale.
logUb95Ci	The upper bound of the 95 percent confidence interval on the log scale.
labels	A vector containing the labels for the various estimates.
xLabel	The label on the x-axis: the name of the effect estimate.

limits	The limits of the effect size axis.
hakn	A logical indicating whether method by Hartung and Knapp should be used to adjust test statistics and confidence intervals.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave ifor supported file formats.

Details

Creates a forest plot of effect size estimates, and includes a meta-analysis estimate using a random effects model. The DerSimonian-Laird estimate (1986) is used.

Value

A Ggplot object. Use the [ggplot2::ggsave](#) function to save to file.

References

DerSimonian R, Laird N (1986), Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-188.

Examples

```
plotMetaAnalysisForest(logRr = c(0, 0.2, -0.2, 0, 0.2, -0.2),
  logLb95Ci = c(-0.2, -0.2, -0.6, -0.2, -0.2, -0.6),
  logUb95Ci = c(0.2, 0.6, 0.2, 0.2, 0.6, 0.2),
  labels = c("Site A", "Site B", "Site C", "Site D", "Site E", "Site F"))
```

plotPerDbMcmcTrace	<i>Plot MCMC trace for individual databases</i>
--------------------	---

Description

Plot MCMC trace for individual databases

Usage

```
plotPerDbMcmcTrace(
  estimate,
  showEstimate = TRUE,
  dataCutoff = 0.01,
  fileName = NULL
)
```

Arguments

estimate	An object as generated using the computeBayesianMetaAnalysis() function.
showEstimate	Show the parameter estimates (mode) and 95 percent confidence intervals?
dataCutoff	This fraction of the data at both tails will be removed.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave in the ggplot2 package for supported file formats.

Value

A Ggplot object. Use the [ggplot2::ggsave](#) function to save to file.

plotPerDbPosterior	<i>Plot posterior density per database</i>
--------------------	--

Description

Plot posterior density per database

Usage

```
plotPerDbPosterior(
  estimate,
  showEstimate = TRUE,
  dataCutoff = 0.01,
  fileName = NULL
)
```

Arguments

estimate	An object as generated using the computeBayesianMetaAnalysis() function.
showEstimate	Show the parameter estimates (mode) and 95 percent confidence intervals?
dataCutoff	This fraction of the data at both tails will be removed.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave in the ggplot2 package for supported file formats.

Value

A Ggplot object. Use the [ggplot2::ggsave](#) function to save to file.

plotPosterior	<i>Plot posterior density</i>
---------------	-------------------------------

Description

Plot posterior density

Usage

```
plotPosterior(
  estimate,
  showEstimate = TRUE,
  dataCutoff = 0.01,
  fileName = NULL
)
```

Arguments

estimate	An object as generated using the computeBayesianMetaAnalysis() function.
showEstimate	Show the parameter estimates (mode) and 95 percent confidence intervals?
dataCutoff	This fraction of the data at both tails will be removed.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave in the ggplot2 package for supported file formats.

Value

A Ggplot object. Use the [ggplot2::ggsave](#) function to save to file.

plotPreparedPs	<i>Plot the propensity score distribution</i>
----------------	---

Description

Plot the propensity score distribution

Usage

```
plotPreparedPs(
  preparedPsPlots,
  labels,
  treatmentLabel = "Target",
  comparatorLabel = "Comparator",
  fileName = NULL
)
```

Arguments

preparedPsPlots	list of prepared propensity score data as created by the preparePsPlot() function.
labels	A vector containing the labels for the various sources.
treatmentLabel	A label to us for the treated cohort.
comparatorLabel	A label to us for the comparator cohort.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggplot2::ggsave for supported file formats.

Value

A ggplot object. Use the [ggplot2::ggsave](#) function to save to file in a different format.

Examples

```
treatment <- rep(0:1, each = 100)
propensityScore <- c(rnorm(100, mean = 0.4, sd = 0.25), rnorm(100, mean = 0.6, sd = 0.25))
data <- data.frame(treatment = treatment, propensityScore = propensityScore)
data <- data[data$propensityScore > 0 & data$propensityScore < 1, ]
preparedPlot <- preparePsPlot(data)
# Just reusing the same data three times for demonstration purposes:
preparedPsPlots <- list(preparedPlot, preparedPlot, preparedPlot)
labels <- c("Data site A", "Data site B", "Data site C")
```

```
preparePsPlot
```

```
Prepare to plot the propensity score distribution
```

Description

Prepare to plot the propensity (or preference) score distribution. It computes the distribution, so the output does not contain person-level data.

Usage

```
preparePsPlot(data, unfilteredData = NULL, scale = "preference")
```

Arguments

<code>data</code>	A data frame with at least the two columns described below
<code>unfilteredData</code>	To be used when computing preference scores on data from which subjects have already been removed, e.g. through trimming and/or matching. This data frame should have the same structure as <code>data</code> .
<code>scale</code>	The scale of the graph. Two scales are supported: <code>scale = 'propensity'</code> or <code>scale = 'preference'</code> . The preference score scale is defined by Walker et al (2013).

Details

The data frame should have at least the following two columns:

- **treatment** (integer): Column indicating whether the person is in the treated (1) or comparator (0) group.
- **propensityScore** (numeric): Propensity score.

Value

A data frame describing the propensity score (or preference score) distribution at 100 equally-spaced points.

References

Walker AM, Patrick AR, Lauer MS, Hornbrook MC, Marin MG, Platt R, Roger VL, Stang P, and Schneeweiss S. (2013) A tool for assessing the feasibility of comparative effectiveness research, *Comparative Effective Research*, 3, 11-20

Examples

```
treatment <- rep(0:1, each = 100)
propensityScore <- c(rnorm(100, mean = 0.4, sd = 0.25), rnorm(100, mean = 0.6, sd = 0.25))
data <- data.frame(treatment = treatment, propensityScore = propensityScore)
data <- data[data$propensityScore > 0 & data$propensityScore < 1, ]
preparedPlot <- preparePsPlot(data)
```

simulatePopulations	<i>Simulate survival data for multiple databases</i>
---------------------	--

Description

Simulate survival data for multiple databases

Usage

```
simulatePopulations(settings = createSimulationSettings())
```

Arguments

settings	An object of type simulationSettings, created by the createSimulationSettings() function.
----------	---

Value

A object of class simulation, which is a list of populations, each a data frame with columns rowId, stratumId, x, time, and y.

Examples

```
simulation <- simulatePopulations()
```

skewNormal	<i>The skew normal function to approximate a log likelihood function</i>
------------	--

Description

The skew normal function to approximate a log likelihood function

Usage

```
skewNormal(x, mu, sigma, alpha)
```

Arguments

x	The log(hazard ratio) for which to approximate the log likelihood.
mu	The position parameter.
sigma	The scale parameter.
alpha	The skew parameter.

Value

The approximate log likelihood for the given x.

Index

`approximateLikelihood`, [2](#)
`approximateLikelihood()`, [9](#)

`computeBayesianMetaAnalysis`, [3](#)
`computeBayesianMetaAnalysis()`, [10–13](#)
`computeFixedEffectMetaAnalysis`, [4](#)
`createSimulationSettings`, [5](#)
`createSimulationSettings()`, [15](#)
`customFunction`, [6](#)
`Cyclops::fitCyclopsModel()`, [2, 9](#)

`ggplot2::ggsave`, [7, 9–13](#)
`ggplot2::ggsave()`, [8](#)

`plotCovariateBalances`, [6](#)
`plotEmpiricalNulls`, [7](#)
`plotLikelihoodFit`, [8](#)
`plotMcmcTrace`, [9](#)
`plotMetaAnalysisForest`, [10](#)
`plotPerDbMcmcTrace`, [11](#)
`plotPerDbPosterior`, [12](#)
`plotPosterior`, [12](#)
`plotPreparedPs`, [13](#)
`preparePsPlot`, [14](#)
`preparePsPlot()`, [13](#)

`simulatePopulations`, [15](#)
`skewNormal`, [15](#)