

Package ‘ebdm’

October 16, 2025

Type Package

Title Estimating Bivariate Dependency from Marginal Data

Version 3.0.0

Description Provides statistical methods for estimating bivariate dependency (correlation) from marginal summary statistics across multiple studies.

The package supports three modules: (1) bivariate correlation estimation for binary outcomes, (2) bivariate correlation estimation for continuous outcomes, and (3) estimation of component-wise means and variances under a conditional two-component Gaussian mixture model for a continuous variable stratified by a binary class label. These methods enable privacy-preserving joint estimation when individual-level data are unavailable.

The approaches are de-

tailed in Shang, Tsao, and Zhang (2025a) <[doi:10.48550/arXiv.2505.03995](https://doi.org/10.48550/arXiv.2505.03995)> and Shang, Tsao, and Zhang (2025b) <[doi:10.48550/arXiv.2505.03995](https://doi.org/10.48550/arXiv.2505.03995)>

License GPL (>= 3)

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports stats

RoxygenNote 7.3.2

NeedsCompilation no

Author Longwen Shang [aut, cre],
Min Tsao [aut],
Xuekui Zhang [aut]

Maintainer Longwen Shang <shanglongwen0918@gmail.com>

Repository CRAN

Date/Publication 2025-10-16 20:20:17 UTC

Contents

bin_example	2
cont_example	2

cor_bin	3
cor_cont	4
est_mixture	5
mixture_example	6

Index	7
--------------	----------

bin_example	<i>Example Data: Binary Variables</i>
-------------	---------------------------------------

Description

Simulated dataset for testing the cor_bin() function.

Usage

data(bin_example)

Format

- A data frame with 3 columns:
- ni** Sample size per study
 - xi** Count of first binary variable
 - yi** Count of second binary variable

cont_example	<i>Example Data: Continuous Variables</i>
--------------	---

Description

Simulated dataset for testing the cor_cont() function.

Usage

data(cont_example)

Format

- A data frame with 5 columns:
- Sample_Size** Sample size for each study.
 - Mean_X** Sample mean of variable X.
 - Mean_Y** Sample mean of variable Y.
 - Variance_X** Sample variance of variable X.
 - Variance_Y** Sample variance of variable Y.

cor_bin	<i>Estimate the Joint Distribution of Two Binary Variables from Marginal Summaries</i>
---------	--

Description

Performs maximum likelihood estimation (MLE) of the joint distribution of two binary variables using only marginal summary data from multiple studies.

Usage

```
cor_bin(ni, xi, yi, ci_method = c("none", "normal", "lr"))
```

Arguments

ni	Numeric vector. Sample sizes for each dataset.
xi	Numeric vector. Count of observations where variable 1 equals 1.
yi	Numeric vector. Count of observations where variable 2 equals 1.
ci_method	Character string. Method for confidence interval computation. Options are "none" (default), "normal", or "lr" (likelihood ratio).

Value

A named list with point estimates, variance, standard error, and confidence interval (if requested).

p1_hat Estimated marginal probability for variable 1.

p2_hat Estimated marginal probability for variable 2.

p11_hat Estimated joint probability.

var_hat Estimated variance of p11_hat.

sd_hat Standard error of p11_hat.

ci Confidence interval for p11_hat, if requested.

Examples

```
data(bin_example)
cor_bin(bin_example$ni, bin_example$xi, bin_example$yi, ci_method = "lr")
```

cor_cont

*Estimate the Bivariate Normal Distribution from Marginal Summaries***Description**

Estimate the correlation coefficient ρ (and marginal means / SDs) of two normally-distributed variables using summary-level data from multiple independent studies.

Usage

```
cor_cont(
  n,
  xbar,
  ybar,
  s2x = NULL,
  s2y = NULL,
  method = c("proposed", "weighted"),
  ci_method = c("none", "normal", "lr")
)
```

Arguments

n	Numeric vector. Sample size of each study.
xbar, ybar	Numeric vectors. Sample means of the two variables.
s2x, s2y	Numeric vectors. Sample variances; required for method = "proposed".
method	Character. "proposed" uses the proposed MLE method in the paper; "weighted" replicates the weighted mean based method (Baseline) when no variances are available.
ci_method	Confidence interval type: "none", "normal", or "lr" (likelihood ratio). Only implemented when method = "proposed".

Value

A list with elements

- mu_x, mu_y : estimated marginal means
- sigma_x, sigma_y : estimated SDs
- rho : estimated correlation
- se : standard error of rho (proposed only)
- ci : confidence interval for rho (if requested)

Examples

```
data(cont_example)
# Example with full summaries
cor_cont(cont_example$Sample_Size, cont_example$Mean_X, cont_example$Mean_Y,
  cont_example$Variance_X, cont_example$Variance_Y, method = "proposed", ci_method = "lr")

# Only means + n, weighted mean method
cor_cont(cont_example$Sample_Size, cont_example$Mean_X, cont_example$Mean_Y, method = "weighted")
```

est_mixture	<i>Estimate Parameters in a Two-Component Gaussian Mixture Using Study-Level Summaries</i>
-------------	--

Description

Estimates group-specific means and standard deviations ($\mu_1, \mu_0, \sigma_1, \sigma_0$) in a two-component normal mixture model based on aggregate data across multiple studies. The continuous variable X is assumed to follow a Gaussian mixture conditional on a binary group indicator $Y \in \{0, 1\}$, with each study reporting only summary-level statistics.

Usage

```
est_mixture(ni, xbar, mi, s2 = NULL, method = c("gmm", "naive"))
```

Arguments

ni	Integer vector of sample sizes per study.
xbar	Numeric vector of sample means per study.
mi	Integer vector of group 1 counts per study.
s2	Numeric vector of sample variances per study. Required if method = "gmm".
method	Estimation method to use. One of "naive" or "gmm". Default is "gmm".

Details

#' Two estimation methods are available:

- **"naive"**: Likelihood-based estimator using only sample means.
- **"gmm"**: Generalized method of moments (GMM) estimator using sample means and variances.

Value

A named list containing:

mu1_hat, mu0_hat	Estimated means of the two groups.
sigma1_hat, sigma0_hat	Estimated standard deviations.
se	Standard errors of the parameter estimates (NA if method = "naive").
ci	List of 95% confidence intervals for each parameter (NULL if method = "naive").
method	A character string indicating the method used.

Examples

```
# Load example dataset included in the package
data(mixture_example)

# Estimate using GMM (recommended) with full summary statistics
est_mixture(
  ni = mixture_example$ni,
  xbar = mixture_example$xbar,
  s2 = mixture_example$s2,
  mi = mixture_example$mi,
  method = "gmm"
)

# Estimate using naive likelihood method (only means used)
est_mixture(
  ni = mixture_example$ni,
  xbar = mixture_example$xbar,
  mi = mixture_example$mi,
  method = "naive"
)
```

mixture_example

Example Data: Mixture Model Summaries

Description

Simulated dataset for testing the `est_mixture()` function. Each row corresponds to one study providing summary-level data from a two-component normal mixture.

Usage

```
data(mixture_example)
```

Format

A data frame with 4 columns:

- ni** Sample size for each study.
- mi** Count of group 1 individuals in each study.
- xbar** Sample mean of the outcome variable.
- s2** Sample variance of the outcome variable.

Index

* **datasets**

bin_example, [2](#)

cont_example, [2](#)

mixture_example, [6](#)

bin_example, [2](#)

cont_example, [2](#)

cor_bin, [3](#)

cor_cont, [4](#)

est_mixture, [5](#)

mixture_example, [6](#)