Package 'ebdm'

June 5, 2025

Type Package

Title Implementation of Estimating Binary Dependency from Marginal Data

Version 1.0.0

Description Provides a maximum likelihood estimation method to recover the joint distribution of two binary variables using only marginal summary data from multiple studies. This approach allows for privacy-preserving estimation in settings where individual-level data are unavailable.
The method is fully described in the manuscript by Shang, Tsao and Zhang (2025) <doi:10.48550/arXiv.2505.03995>: ``Estimating the Joint Distribution of Two Binary Variables from Their Marginal Summaries".

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 2025-06-05 11:10:10 UTC

Contents

ebdm_estimate *Estimate the Joint Distribution of Two Binary Variables*

Description

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

Usage

```
ebdm_estimate(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(eg_data)
ebdm_estimate(eg_data$ni, eg_data$xi, eg_data$yi, ci_method = "lr")
```

`eg_data` *Example Dataset*

Description

Simulated dataset for testing the `ebdm_estimate()` function.

Usage

```
data(eg_data)
```

Format

A data frame with 3 columns:

ni Sample size per study

xi Count of first binary variable

yi Count of second binary variable

Index