Package ‘latenetwork’

August 8, 2023

Type Package

Title Inference on LATEs under Network Interference of Unknown Form

Version 1.0.1

Description

Estimating causal parameters in the presence of treatment spillover is of great interest in statistics. This package provides tools for instrumental variables estimation of average causal effects under network interference of unknown form.
The target parameters are the local average direct effect, the local average indirect effect, the local average overall effect, and the local average spillover effect.
The methods are developed by Hoshino and Yanagi (2023) <doi:10.48550/arXiv.2108.07455>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports igraph, methods, statip, stats

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

URL `https://tkhdyanagi.github.io/latenetwork/`

NeedsCompilation no

Author Tadao Hoshino [aut, cph] (<`https://orcid.org/0000-0001-9484-2943`>),
Takahide Yanagi [aut, cre, cph]
(<`https://orcid.org/0000-0003-3975-5518`>)

Maintainer Takahide Yanagi <`yanagi@econ.kyoto-u.ac.jp`>

Repository CRAN

Date/Publication 2023-08-08 16:10:02 UTC

R topics documented:

`datageneration` *Generate Artificial Data by Simulation*

Description

The `datageneration()` function generates artificial ring-network data by simulation. The function is used in the package vignette.

Usage

```
datageneration(n)
```

Arguments

`n`	The sample size

Value

A list containing the outcome vector, the treatment vector, the instrumental vector, and the true instrumental exposure vector, and the symmetric binary adjacency matrix.

Examples

```
latenetwork::datageneration(n = 2000)
```

`direct` *Inference on Average Direct Effect Parameters*

Description

Inference on the average direct effect of the IV on the outcome, that on the treatment receipt, and the local average direct effect in the presence of network spillover of unknown form

Usage

```
direct(
  Y,
  D,
  Z,
  IEM = NULL,
  S,
  A,
  K = 1,
  t = NULL,
  bw = NULL,
  B = NULL,
  alp = 0.05
)
```

Arguments

`Y`	An n-dimensional outcome vector
`D`	An n-dimensional binary treatment vector
`Z`	An n-dimensional binary instrumental vector
`IEM`	An n-dimensional instrumental exposure vector. If `IEM = NULL` or `t = NULL`, the constant IEM is used. Default is NULL.
`S`	An n-dimensional logical vector to indicate whether each unit belongs to the sub-population S
`A`	An n times n symmetric binary adjacency matrix
`K`	A scalar to indicate the range of neighborhood used for constructing the interference set. Default is 1. In the `direct()` function, `K` is used only for computing the bandwidth.
`t`	A scalar of the evaluation point of IEM. Default is NULL.
`bw`	A scalar of the bandwidth used for the HAC estimation and the wild bootstrap. If `bw = NULL`, the rule-of-thumb bandwidth proposed by Leung (2022) is used. Default is NULL.
`B`	The number of bootstrap repetitions. If `B = NULL`, the wild bootstrap is skipped. Default is NULL.
`alp`	The significance level. Default is 0.05.

Details

The `direct()` function estimates the average direct effect of the IV on the outcome, that on the treatment receipt, and the local average direct effect via inverse probability weighting in the approximate neighborhood interference framework. The function also computes the standard errors and the confidence intervals for the target parameters based on the network HAC estimation and the wild bootstrap. For more details, see Hoshino and Yanagi (2023). The lengths of `Y`, `D`, `Z`, `S` and of the row and column of `A` must be the same. `IEM` must be `NULL` or a vector of the same length as `Y`. `t` must be `NULL` or a value in the support of `IEM`. `K` must be a positive integer. `bw` must be `NULL` or a non-negative integer. `B` must be `NULL` or a positive number. `alp` must be a positive number between 0 and 0.5.

Value

A data.frame containing the following elements:

est	The parameter estimate
HAC_SE	The standard error computed by the network HAC estimation
HAC_CI_L	The lower bound of the confidence interval computed by the network HAC estimation
HAC_CI_U	The upper bound of the confidence interval computed by the network HAC estimation
wild_SE	The standard error computed by the wild bootstrap
wild_CI_L	The lower bound of the confidence interval computed by the wild bootstrap
wild_CI_U	The upper bound of the confidence interval computed by the wild bootstrap
bw	The bandwidth used for the HAC estimation and the wild bootstrap
size	The size of the subpopulation S

References

Hoshino, T., & Yanagi, T. (2023). Causal inference with noncompliance and unknown interference. arXiv preprint arXiv:2108.07455.

Leung, M.P. (2022). Causal inference under approximate neighborhood interference. Econometrica, 90(1), pp.267-293.

Examples

```
# Generate artificial data
set.seed(1)
n <- 2000
data <- latenetwork::datageneration(n = n)

# Arguments
Y   <- data$Y
D   <- data$D
Z   <- data$Z
IEM <- data$IEM
S   <- rep(TRUE, n)
A   <- data$A
K   <- 1
t   <- 0
bw  <- NULL
B   <- NULL
alp <- 0.05

# Estimation
latenetwork::direct(Y = Y,
                    D = D,
                    Z = Z,
                    IEM = IEM,
                    S = S,
```

```
A = A,
K = K,
t = t,
bw = bw,
B = B,
alp = alp)
```

`indirect` *Inference on Average Indirect Effect Parameters*

Description

Inference on the average indirect effect of the IV on the outcome, that on the treatment receipt, and the local average indirect effect in the presence of network spillover of unknown form

Usage

```
indirect(Y, D, Z, S, A, K = 1, bw = NULL, B = NULL, alp = 0.05)
```

Arguments

Y	An n-dimensional outcome vector
D	An n-dimensional binary treatment vector
Z	An n-dimensional binary instrumental vector
S	An n-dimensional logical vector to indicate whether each unit belongs to the sub-population S
A	An n times n symmetric binary adjacency matrix
K	A scalar to indicate the range of neighborhood used for constructing the interference set. Default is 1.
bw	A scalar of the bandwidth used for the HAC estimation and the wild bootstrap. If `bw = NULL`, the rule-of-thumb bandwidth proposed by Leung (2022) is used. Default is NULL.
B	The number of bootstrap repetitions. If `B = NULL`, the wild bootstrap is skipped. Default is NULL.
alp	The significance level. Default is 0.05.

Details

The `indirect()` function estimates the average indirect effect of the IV on the outcome, that on the treatment receipt, and the local average indirect effect via inverse probability weighting in the approximate neighborhood interference framework. The function also computes the standard errors and the confidence intervals for the target parameters based on the network HAC estimation and the wild bootstrap. For more details, see Hoshino and Yanagi (2023). The lengths of `Y`, `D`, `Z`, `S` and of the row and column of `A` must be the same. `K` must be a positive integer. `bw` must be `NULL` or a non-negative number. `B` must be `NULL` or a positive number. `alp` must be a positive number between 0 and 0.5.

Value

A data.frame containing the following elements:

est	The parameter estimate
HAC_SE	The standard error computed by the network HAC estimation
HAC_CI_L	The lower bound of the confidence interval computed by the network HAC estimation
HAC_CI_U	The upper bound of the confidence interval computed by the network HAC estimation
wild_SE	The standard error computed by the wild bootstrap
wild_CI_L	The lower bound of the confidence interval computed by the wild bootstrap
wild_CI_U	The upper bound of the confidence interval computed by the wild bootstrap
bw	The bandwidth used for the HAC estimation and the wild bootstrap
size	The size of the subpopulation S

References

Hoshino, T., & Yanagi, T. (2023). Causal inference with noncompliance and unknown interference. arXiv preprint arXiv:2108.07455.

Leung, M.P. (2022). Causal inference under approximate neighborhood interference. Econometrica, 90(1), pp.267-293.

Examples

```
# Generate artificial data
set.seed(1)
n <- 2000
data <- latenetwork::datageneration(n = n)

# Arguments
Y   <- data$Y
D   <- data$D
Z   <- data$Z
S   <- rep(TRUE, n)
A   <- data$A
K   <- 1
bw  <- NULL
B   <- NULL
alp <- 0.05

# Estimation
latenetwork::indirect(Y = Y,
                      D = D,
                      Z = Z,
                      S = S,
                      A = A,
                      K = K,
                      bw = bw,
```

```
B = B,
alp = alp)
```

overall *Inference on Average Overall Effect Parameters*

Description

Inference on the average overall effect of the IV on the outcome, that on the treatment receipt, and the local average overall effect in the presence of network spillover of unknown form

Usage

```
overall(Y, D, Z, S, A, K = 1, bw = NULL, B = NULL, alp = 0.05)
```

Arguments

Y	An n-dimensional outcome vector
D	An n-dimensional binary treatment vector
Z	An n-dimensional binary instrumental vector
S	An n-dimensional logical vector to indicate whether each unit belongs to the sub-population S
A	An n times n symmetric binary adjacency matrix
K	A scalar to indicate the range of neighborhood used for constructing the interference set. Default is 1.
bw	A scalar of the bandwidth used for the HAC estimation and the wild bootstrap. If `bw = NULL`, the rule-of-thumb bandwidth proposed by Leung (2022) is used. Default is NULL.
B	The number of bootstrap repetitions. If `B = NULL`, the wild bootstrap is skipped. Default is NULL.
alp	The significance level. Default is 0.05.

Details

The `overall()` function estimates the average overall effect of the IV on the outcome, that on the treatment receipt, and the local average overall effect via inverse probability weighting in the approximate neighborhood interference framework. The function also computes the standard errors and the confidence intervals for the target parameters based on the network HAC estimation and the wild bootstrap. For more details, see Hoshino and Yanagi (2023). The lengths of `Y`, `D`, `Z`, `S` and of the row and column of `A` must be the same. `K` must be a positive integer. `bw` must be `NULL` or a non-negative number. `B` must be `NULL` or a positive number. `alp` must be a positive number between 0 and 0.5.

Value

A data.frame containing the following elements:

est	The parameter estimate
HAC_SE	The standard error computed by the network HAC estimation
HAC_CI_L	The lower bound of the confidence interval computed by the network HAC estimation
HAC_CI_U	The upper bound of the confidence interval computed by the network HAC estimation
wild_SE	The standard error computed by the wild bootstrap
wild_CI_L	The lower bound of the confidence interval computed by the wild bootstrap
wild_CI_U	The upper bound of the confidence interval computed by the wild bootstrap
bw	The bandwidth used for the HAC estimation and the wild bootstrap
size	The size of the subpopulation S

References

Hoshino, T., & Yanagi, T. (2023). Causal inference with noncompliance and unknown interference. arXiv preprint arXiv:2108.07455.

Leung, M.P. (2022). Causal inference under approximate neighborhood interference. Econometrica, 90(1), pp.267-293.

Examples

```
# Generate artificial data
set.seed(1)
n <- 2000
data <- latenetwork::datageneration(n = n)

# Arguments
Y   <- data$Y
D   <- data$D
Z   <- data$Z
S   <- rep(TRUE, n)
A   <- data$A
K   <- 1
bw  <- NULL
B   <- NULL
alp <- 0.05

# Estimation
latenetwork::overall(Y = Y,
                     D = D,
                     Z = Z,
                     S = S,
                     A = A,
                     K = K,
                     bw = bw,
```

```
B = B,
alp = alp)
```

spillover *Inference on Average Spillover Effect Parameters*

Description

Inference on the average spillover effect of the IV on the outcome, that on the treatment receipt, and the local average spillover effect in the presence of network spillover of unknown form

Usage

```
spillover(
  Y,
  D,
  Z,
  IEM,
  S,
  A,
  K = 1,
  z,
  t0,
  t1,
  bw = NULL,
  B = NULL,
  alp = 0.05
)
```

Arguments

`Y`	An n-dimensional outcome vector
`D`	An n-dimensional binary treatment vector
`Z`	An n-dimensional binary instrumental vector
`IEM`	An n-dimensional instrumental exposure vector
`S`	An n-dimensional logical vector to indicate whether each unit belongs to the sub-population S
`A`	An n times n symmetric binary adjacency matrix
`K`	A scalar to indicate the range of neighborhood used for constructing the interference set. Default is 1. In the `spillover()` function, `K` is used only for computing the bandwidth.
`z`	A scalar of the evaluation point of Z
`t0`	A scalar of the evaluation point of instrumental exposure (from)
`t1`	A scalar of the evaluation point of instrumental exposure (to)

bw	A scalar of the bandwidth used for the HAC estimation and the wild bootstrap. If bw = NULL, the rule-of-thumb bandwidth proposed by Leung (2022) is used. Default is NULL.
B	The number of bootstrap repetitions. If B = NULL, wild bootstrap is skipped. Default is NULL.
alp	The significance level. Default is 0.05.

Details

The spillover() function estimates the average spillover effect of the IV on the outcome, that on the treatment receipt, and the local average spillover effect via inverse probability weighting in the approximate neighborhood interference framework. The function also computes the standard errors and the confidence intervals for the target parameters based on the network HAC estimation and the wild bootstrap. For more details, see Hoshino and Yanagi (2023). The lengths of Y, D, Z, IEM, S and of the row and column of A must be the same. z must be 0 or 1. t0 and t1 must be values in the support of IEM. bw must be NULL or a non-negative number. B must be NULL or a positive integer. alp must be a positive number between 0 and 0.5.

Value

A data frame containing the following elements:

est	The parameter estimate
HAC_SE	The standard error computed by the network HAC estimation
HAC_CI_L	The lower bound of the confidence interval computed by the network HAC estimation
HAC_CI_U	The upper bound of the confidence interval computed by the network HAC estimation
wild_SE	The standard error computed by the wild bootstrap
wild_CI_L	The lower bound of the confidence interval computed by the wild bootstrap
wild_CI_U	The upper bound of the confidence interval computed by the wild bootstrap
bw	The bandwidth used for the HAC estimation and the wild bootstrap
size	The size of the subpopulation S

References

Hoshino, T., & Yanagi, T. (2023). Causal inference with noncompliance and unknown interference. arXiv preprint arXiv:2108.07455.

Leung, M.P. (2022). Causal inference under approximate neighborhood interference. Econometrica, 90(1), pp.267-293.

Examples

```
# Generate artificial data
set.seed(1)
n <- 2000
data <- latenetwork::datageneration(n = n)
```

```
# Arguments
Y   <- data$Y
D   <- data$D
Z   <- data$Z
S   <- rep(TRUE, n)
A   <- data$A
K   <- 1
IEM <- ifelse(A %*% Z > 0, 1, 0)
z   <- 1
t0  <- 0
t1  <- 1
bw  <- NULL
B   <- NULL
alp <- 0.05

# Estimation
latenetwork::spillover(Y = Y,
                       D = D,
                       Z = Z,
                       IEM = IEM,
                       S = S,
                       A = A,
                       K = K,
                       z = z,
                       t0 = t0,
                       t1 = t1,
                       bw = bw,
                       B = B,
                       alp = alp)
```

Index