Package ‘fasttreeid’

November 7, 2025

Type Package

Title Identifies Parameters in a Tree-Shaped SCM

Version 1.0.0

Description Implements the algorithm by Briefs and Bläser (2025) <https://openreview.net/forum?id=8PHOPPH35D>, based on the approach of Gupta and Bläser (2024) <doi:10.1609/aaai.v38i18.30023>. It determines, for a structural causal model (SCM) whose directed edges form a tree, whether each parameter is unidentifiable, 1-identifiable or 2-identifiable (other cases cannot occur), using a randomized algorithm with provable running time O(n^3 log^2 n).

License MIT + file LICENSE

Encoding UTF-8

Imports Rcpp

Suggests gmp, testthat

LinkingTo Rcpp

URL https://github.com/yasminebriefs/FastTreeID

BugReports https://github.com/yasminebriefs/FastTreeID/issues

NeedsCompilation yes

Author Yasmine Briefs [aut, cre],
Markus Bläser [aut]

Maintainer Yasmine Briefs <ybriefs@mpi-inf.mpg.de>

Repository CRAN

Date/Publication 2025-11-07 13:50:07 UTC

Contents

fasttreeid_identify *Identifies parameters in a tree-shaped SCM*

Description

Implements the algorithm from the paper 'Faster Generic Identification in Tree-Shaped Structural Causal Models' (2025) by Briefs and Bläser, based on the approach of Gupta and Bläser in 'Identification for Tree-Shaped Structural Causal Models in Polynomial Time' (2024). It determines, for a structural causal model (SCM) whose directed edges form a tree, whether each parameter is unidentifiable, 1-identifiable or 2-identifiable (other cases cannot occur), using a randomized algorithm with provable running time O(n^3 log^2 n).

Usage

```
fasttreeid_identify(bidirected, directed, seed = NULL, prime = NULL)
```

Arguments

`bidirected`	A 2-column matrix in which each row specifies a bidirected edge (1-indexed)
`directed`	The directed parents p_i of nodes i from 2 to n (1 <= p_i < i)
`seed`	The seed to use as a decimal string. Picks a random seed by default. Must fit into an unsigned 64-bit integer
`prime`	The prime to use as a decimal string. Picks a random prime by default. If no prime is given, the package gmp must be installed. The prime must be between 2^58 and 2^59

Value

Returns a nested named list describing the identification result:

identification A list with n entries for all nodes i from 1 to n. For each node, it contains a list describing the result for the node. For all nodes, this list contains an entry `identifiability` that can take the values 0 (unidentifiable), 1 (1-identifiable) and 2 (2-identifiable). For all nodes that are not unidentifiable, this list contains an entry `type` that can take the following values:

path This means that this node can be identified by missing bidirected edges of rank 2 to a node identified by a fraction or a cycle (Lemma 4 in the paper by Gupta and Bläser). In this case, the list additionally contains an entry `nodes` that describes the path to this node. The first node in the list is the current node and the type of the last node in the list is fraction or cycle

fraction This means that this node can be identified by a missing bidirected edge to the root (Section 2.1 in the paper by Gupta and Bläser). In this case, the list additionally contains two entries `numerator` and `denominator`, each of the format `{"what": "sigma", "i": i, "j": j}`

- **cycle** This means that this node can be identified by a cycle of missing bidirected edges of rank 2 (Definition 10 in the paper by Gupta and Bläser). In this case, the list additionally contains an entry `nodes` that describes the cycle. The first and last node in the list are the current node. If the identifiability of the current node is 1, the list further contains an entry `reason` explaining why the identifiability is 1. It can take the following values:
    - **a_is_zero** This means that a = 0 in the equation for some lambda_i,j (case 2 in Lemma 8 in the paper by Gupta and Bläser). In this case, the list additionally contains an entry `reason_edge` of the format `{"what": "lambda", "i": i, "j": j}`
    - **discriminant_is_zero** This means that the discriminant is 0 (case 1 in Lemma 8 in the paper by Gupta and Bläser)
    - **only_one_option** This means that the equation for a missing bidirected edge {i, j} is only satisfied by one of the options (last step in Algorithm 2 in the paper by Gupta and Bläser). In this case, the list additionally contains an entry `reason_edge` of the format `{"what": "missing_bidirected", "i": i, "j": j}`

seed The seed used for identification for reproducibility

prime The prime used for identification for reproducibility

Note

The algorithm is randomized. The error probability is below 4.1e-6 for n = 200 and below 0.0026 for n = 1000 (the estimated running time for n = 1000 is more than two hours). If the prime is not given, using the same seed twice will generate the same prime twice. Whether or not a prime is given doesn't influence the other random computations.

References

Yasmine Briefs and Markus Bläser. Faster generic identification in tree-shaped structural causal models (2025)

Aaryan Gupta and Markus Bläser. Identification for tree-shaped structural causal models in polynomial time (2024)

Benito van der Zander et al. Identification in tree-shaped linear structural causal models (2022)

Examples

```
### This defines the example in Figure 1 in the paper by Gupta and Bläser
bidirected <- matrix(c(2, 3), byrow = TRUE, ncol = 2)
directed <- c(1, 2)
fasttreeid_identify(bidirected, directed,
seed="1909956540882291621", prime = "499562279898941057")
```

Index