

GaussDetect-LiNGAM:Causal Direction Identification without Gaussianity test

Ziyi Ding^{*1} Xiao-Ping Zhang¹

Abstract

We propose GaussDetect-LiNGAM, a novel approach for bivariate causal discovery that eliminates the need for explicit Gaussianity tests by leveraging a fundamental equivalence between noise Gaussianity and residual independence in the reverse regression. Under the standard LiNGAM assumptions of linearity, acyclicity, and exogeneity, we prove that the Gaussianity of the forward-model noise is equivalent to the independence between the regressor and residual in the reverse model. This theoretical insight allows us to replace fragile and sample-sensitive Gaussianity tests with robust kernel-based independence tests. Experimental results validate the equivalence and demonstrate that GaussDetect-LiNGAM maintains high consistency across diverse noise types and sample sizes, while reducing the number of tests per decision (TPD). Our method enhances both the efficiency and practical applicability of causal inference, making LiNGAM more accessible and reliable in real-world scenarios.

1. Introduction

Causal inference plays a critical role in social sciences and many other applied fields, especially when dealing with complex causal relationships. Structural Causal Models (SCM) provide a theoretical framework for such inference, capturing the causal relationships between variables through explicit structural equation models(Pearl, 2009; Spirtes et al.).

The Linear Non-Gaussian Acyclic Model (LiNGAM) was introduced to address this issue(Shimizu et al., a; Hyvärinen & Oja; Shen et al.). By breaking the assumption of Gaussian noise and utilizing the non-Gaussian nature of noise, LiNGAM effectively distinguishes causal relationships be-

tween variables, allowing for a more accurate identification of causal directions.

However, the original formulation of the LiNGAM model relies on a constructive algorithm and does not provide strict mathematical theorems to prove its identifiability. DirectLiNGAM further addresses the identifiability issue in multivariate cases, but the proofs of these theorems strictly follow the assumptions of non-Gaussianity, linearity, acyclicity, and exogeneity(Shimizu et al., b). Although pairwise-LiNGAM effectively uses the likelihood-ratio approach to determine the causal direction between two variables, it still requires prior knowledge of the four key assumptions: non-Gaussianity, linearity, acyclicity, and exogeneity(Hyvärinen & Smith). The challenge is that, in many real-world situations, noise may be non-Gaussian but cannot be effectively detected, which significantly limits the applicability of the LiNGAM model.

Specifically, this paper focuses on the analysis of the bivariate LiNGAM model, as it is often difficult to ensure that all variables have non-Gaussian noise in practical applications. Compared to the multivariate case, the bivariate model is more tractable in handling local causal relationships and offers better operational feasibility.

This study addresses this issue by proving that,in the bivariate LiNGAM model, the Gaussianity test for the noise in the forward model and the independence test between the independent variable and noise in the reverse model are consistently equivalent.

A key contribution of this study is the demonstration that, as long as the bivariate variables satisfy the assumptions of linearity, acyclicity, and exogeneity, the non-Gaussian nature of the noise can be directly assessed through the independence test in the reverse model. This approach eliminates the need for prior non-Gaussianity tests, allowing researchers to skip the step of checking for non-Gaussian noise distribution. Instead, by focusing solely on the independence test, the process is simplified, making it more efficient. This enables the use of various robust and reliable independence testing methods,such as kernel independence testing(Akaho; Bach & Jordan), significantly improving the practicality and applicability of the LiNGAM model in

^{*}Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. Correspondence to: Ziyi Ding <dingzy25@mails.tsinghua.edu.cn>, Xiao-Ping Zhang <xpzhang@ieee.org>.

real-world causal inference tasks.

2. Notation

Let X and Y be random variables. The joint probability density function of X and Y is denoted by $f_{X,Y}(x,y)$, and the marginal probability density functions of X and Y are denoted by $f_X(x)$ and $f_Y(y)$, respectively. That is, $f_{X,Y}(x,y)$ is the joint probability density function of X and Y , $f_X(x)$ is the marginal probability density function of X , $f_Y(y)$ is the marginal probability density function of Y .

In such situations, if the noise terms are non-Gaussian, we can utilize the LiNGAM model to identify the causal direction between the variables. This is possible because non-Gaussian noise helps distinguish between the direction of causality, even when the relationship between the two variables is complex or bidirectional.

Definition 2.1 (Bivariate LiNGAM). In the case of two observed variables X and Y , the Linear Non-Gaussian Acyclic Model (LiNGAM) assumes the following linear relationships with non-Gaussian noise:

$$Y = aX + \epsilon_Y \quad (1)$$

$$X = bY + \epsilon_X \quad (2)$$

where: X and Y are the observed variables, ϵ_X and ϵ_Y are the non-Gaussian noise terms for X and Y , respectively. The model assumes that the noise terms ϵ_X and ϵ_Y are non-Gaussian.

In the context of causal inference, two models are often discussed: the forward model and the reverse model.

Forward model: In this model, we assume that the independent variable X influences the dependent variable Y , typically expressed as $Y = aX + \epsilon_Y$, where a is the regression coefficient, and ϵ_Y is the noise term.

Reverse model: In the reverse model, we assume that the dependent variable Y influences the independent variable X , typically expressed as $X = bY + \epsilon_X$, where b is the regression coefficient, and ϵ_X is the noise term.

The key challenge is to determine the correct causal direction between the variables, especially when both models seem plausible. The LiNGAM model helps resolve this by leveraging the non-Gaussianity of the noise terms to identify the true direction of causality.

In massive inference and reproducibility studies, the same null hypothesis is often examined by more than one procedure—frequentist p-value versus Bayes factor, Wald test versus likelihood-ratio test, or an original codebase versus its re-implementation. If two such procedures always reach identical reject/retain decisions, the simpler, faster, or more

numerically stable one can be deployed without any risk of contradiction. To justify this substitution rigorously, we first need a sharp, sample-level guarantee that no future data set will ever make the two tests disagree. Therefore, Definition 2.2 is formulated as follows:

Definition 2.2 (Consistent Equivalence of Hypothesis Tests). Let there be two tests $\varphi_1(S)$ and $\varphi_2(S)$, both used to test the same hypothesis. If the null hypothesis is H_0 and the alternative hypothesis is H_1 , the test function $\varphi_i(X)$ (where $i \in \{1, 2\}$) indicates whether the null hypothesis H_0 is rejected based on the sample data S . Specifically:

$$\varphi_i(S) = \begin{cases} 1 & \text{if the null hypothesis is rejected,} \\ 0 & \text{if the null hypothesis is not rejected.} \end{cases} \quad (3)$$

The two tests $\varphi_1(S)$ and $\varphi_2(S)$ are said to be equivalent (in the sense of consistent conclusions) if for all S ,

$$\varphi_1(S) = \varphi_2(S) \quad (4)$$

The definition of the equivalence of hypothesis tests is well-defined because it provides a clear and unambiguous description of two hypothesis tests $\varphi_1(X)$ and $\varphi_2(X)$, including their decision rules (reject or fail to reject H_0). The equivalence condition requires that the two tests yield the same result for all sample data X , ensuring consistency. The condition is both sufficient and necessary: if the tests are equivalent, they must give the same outcome for every possible X . Furthermore, the definition is mathematically rigorous, using standard notation with no ambiguity, and it is universally applicable to any hypothesis test. Thus, this definition satisfies the criteria for being a well-defined concept.

3. Problem Formulation

Consider two competing bivariate linear acyclic models. The goal is to determine the causal direction **without prior knowledge of whether the noise is non-Gaussian**. Forward model: $Y = aX + \epsilon_Y$, $X \perp \epsilon_Y$. and Reverse model: $X = bY + \epsilon_X$, $Y \perp \epsilon_X$.

Two sets of hypothesis tests are defined: one for independence tests (IT) and one for Gaussianity tests (GT): Independence Tests (IT):

$$H_{IT_{10}} : Y \perp \epsilon_X, H_{IT_{11}} : Y \not\perp \epsilon_X \quad (5)$$

$$H_{IT_{20}} : X \perp \epsilon_Y, H_{IT_{21}} : X \not\perp \epsilon_Y \quad (6)$$

Gaussianity Tests (GT):

$$H_{GT_{10}} : \epsilon_Y \sim N(0, \sigma_Y^2), H_{GT_{11}} : \epsilon_Y \not\sim N(0, \sigma_Y^2) \quad (7)$$

$$H_{GT_{20}} : \epsilon_X \sim N(0, \sigma_X^2), H_{GT_{21}} : \epsilon_X \not\sim N(0, \sigma_X^2) \quad (8)$$

Here, $\varphi_{IT_1}, \varphi_{IT_2}$ are the decision variables for the independence tests, and $\varphi_{GT_1}, \varphi_{GT_2}$ are the decision variables for the Gaussianity tests, both taking values from $\{0, 1\}$ (where 1 means do not reject, and 0 means reject).

If the data generation is assumed to follow the forward model, then $H_{IT_{10}}$ and $H_{GT_{10}}$ should be tested. On the other hand, if the data generation follows the reversed model, then $H_{IT_{20}}$ and $H_{GT_{20}}$ should be tested.

Next, the expected equivalence will be discussed. Under the concept of consistent equivalence, it is expected that:

$$\begin{aligned}\varphi_{IT_1} &= \varphi_{GT_1} \\ \varphi_{IT_2} &= \varphi_{GT_2}\end{aligned}\tag{10}$$

This issue will be explored further, with the expectation that within this model class, independence tests (IT) and Gaussianity tests (GT) will be considered equivalent at the sample level. If this equivalence holds, independence tests can be used to replace Gaussianity tests.

Based on the equivalence, the direction identification via independence tests is considered. The question to be explored is whether a criterion or algorithm can be developed to determine the direction (forward or reversed) under the assumptions of linearity, no cycles, and exogeneity. If the direction cannot be determined, the issue is how the algorithm should handle cases of "undetermined" or "inconclusive."

It is worth mentioning that this does not remove the role of non-Gaussianity for identifiability: if the noise is Gaussian, the direction is not identifiable. Rather, it removes the need to *know in advance* whether the noise is non-Gaussian; the independence test outcomes themselves indicate whether one is in the identifiable (non-Gaussian) or non-identifiable (Gaussian) regime.

4. Consistent Equivalence of Gaussianity Test and Independence Test in the LiNGAM

This section focuses on the relationship between the Gaussianity test and the independence test in the LiNGAM model. First, we present Lemma 1 (Darmois; Lukacs & King):

Lemma 4.1 (Skitovich–Darmois Theorem). *Let X_1, X_2, \dots, X_n be independent random variables. If there exist non-zero constants $c_1, c_2, \dots, c_n, d_1, d_2, \dots, d_n$ such that the linear combinations $Y_1 = \sum_{i=1}^n c_i X_i$ and $Y_2 = \sum_{j=1}^n d_j X_j$ are independent, then X_1, X_2, \dots, X_n are normally distributed.*

The conclusion of Lemma 4.1 is non-trivial. Generally, we can infer independence from the properties of the distribution, but Lemma 4.1 (the Skitovich-Darmois theorem) allows us to infer the distributional properties of random variables based on their independence.

Theorem 4.2. *If X and Y satisfy the forward model $Y = aX + \epsilon_Y$ and the reverse model $X = bY + \epsilon_X$, a sufficient and necessary condition for $(\begin{smallmatrix} Y \\ \epsilon_X \end{smallmatrix})$ to be independent is that $(\begin{smallmatrix} X \\ \epsilon_Y \end{smallmatrix})$ is independent and both follow a normal distribution.*

Proof. **Sufficiency:** Assume that the components of $(\begin{smallmatrix} X \\ \epsilon_Y \end{smallmatrix})$, namely X and ϵ_Y , are independent, and both X and ϵ_Y follow normal distributions. Furthermore, we have the following relationship between $(\begin{smallmatrix} Y \\ \epsilon_X \end{smallmatrix})$ and $(\begin{smallmatrix} X \\ \epsilon_Y \end{smallmatrix})$:

$$\left(\begin{array}{c} Y \\ \epsilon_X \end{array}\right) = \begin{pmatrix} a & 1 \\ 1 - ba & -b \end{pmatrix} \left(\begin{array}{c} X \\ \epsilon_Y \end{array}\right)$$

Since $(\begin{smallmatrix} X \\ \epsilon_Y \end{smallmatrix})$ consists of independent components, and both components are normally distributed, it follows that $(\begin{smallmatrix} Y \\ \epsilon_X \end{smallmatrix})$ also follows a normal distribution.

Next, consider the covariance matrix $\text{Cov}(Y, \epsilon_X)$:

$$\text{Cov}(Y, \epsilon_X) = \begin{pmatrix} a & 1 \\ 1 - ba & -b \end{pmatrix} \text{Cov}(X, \epsilon_Y) \begin{pmatrix} a & 1 \\ 1 - ba & -b \end{pmatrix}^T$$

For simplicity, assume that the components of $(\begin{smallmatrix} X \\ \epsilon_Y \end{smallmatrix})$ have unit variances (if not, we can standardize the variables). Thus, we have:

$$\text{Cov}(X, \epsilon_Y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Since $\begin{pmatrix} a & 1 \\ 1 - ba & -b \end{pmatrix}$ is an orthogonal matrix, we know that:

$$\begin{pmatrix} a & 1 \\ 1 - ba & -b \end{pmatrix} \begin{pmatrix} a & 1 \\ 1 - ba & -b \end{pmatrix}^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Thus, we conclude that:

$$\text{Cov}(Y, \epsilon_X) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Since $(\begin{smallmatrix} Y \\ \epsilon_X \end{smallmatrix})$ is a multivariate normal distribution and its components are uncorrelated, they must also be independent. Therefore, we have shown that the components of $(\begin{smallmatrix} Y \\ \epsilon_X \end{smallmatrix})$ are independent.

Necessity: Since the components of $(\begin{smallmatrix} Y \\ \epsilon_X \end{smallmatrix})$, namely Y and ϵ_X , are independent, we now look at the relationship between $(\begin{smallmatrix} X \\ \epsilon_Y \end{smallmatrix})$ and $(\begin{smallmatrix} Y \\ \epsilon_X \end{smallmatrix})$:

$$\left(\begin{array}{c} X \\ \epsilon_Y \end{array}\right) = \begin{pmatrix} a & 1 - ba \\ 1 & -b \end{pmatrix} \left(\begin{array}{c} Y \\ \epsilon_X \end{array}\right)$$

Since (ϵ_X^Y) is independent and the linear combination (ϵ_Y^X) is based on it, by the Skitovich-Darmois theorem, we conclude that (ϵ_X^Y) must follow a normal distribution.

Furthermore, by using the same reasoning as in the sufficiency proof, we can show that the components of (ϵ_X^Y) are also independent.

Thus, the proof is complete. \square

Theorem 4.2 specifies the conditions under which causal direction can and cannot be identified, and rigorously proves the identifiability of the bivariate LiNGAM model. Specifically, if $Y \perp \epsilon_X$, then ϵ_Y must follow a normal distribution. In the forward model, if the noise does not follow a normal distribution, the causal direction can be identified. However, if the noise satisfies the assumption of Gaussianity, the causal direction cannot be determined.

From a theoretical perspective, as long as the noise does not follow a normal distribution under certain conditions, the LiNGAM model can identify the causal direction. However, in practice, the only data available is usually observational data, and it is impossible to know in advance whether the noise follows a normal distribution. Therefore, hypothesis testing must be used. This leads to the following Theorem 4.3.

Theorem 4.3. *In the causal direction identification problem of the Bivariate LiNGAM , let $\varphi_{IT_i}(S), i \in \{1, 2\}$ be the independence test mentioned in (5) and (6), and $\varphi_{GT_i}(S), i \in \{1, 2\}$ be the Gaussianity test mentioned in (7) and (8). Then:*

$$\varphi_{IT_i}(S) = \varphi_{GT_i}(S), \quad \forall S, i \in \{1, 2\} \quad (11)$$

This means that both tests yield consistent conclusions for all sample data.

Proof. We first show $\varphi_{IT_1}(S) = \varphi_{GT_1}(S)$. When $\varphi_{IT_1}(S) = 1$, it means that Y and ϵ_X are independent. According to Theorem 1 (with X and Y interchanged), if Y and ϵ_X are independent, then ϵ_Y must follow a normal distribution. Therefore, the Gaussianity test $\varphi_{GT_1}(S)$ will conclude that $\epsilon_Y \sim \mathcal{N}(0, \sigma_Y^2)$, i.e., $\varphi_{GT_1}(S) = 1$. Thus, when $\varphi_{IT_1}(S) = 1$, we also have $\varphi_{GT_1}(S) = 1$, and both tests give consistent conclusions.

When $\varphi_{IT_1}(S) = 0$, it means that Y and ϵ_X are not independent, indicating a dependency between them. In this case, ϵ_Y cannot follow a normal distribution, because if ϵ_Y were Gaussian, it would be independent of Y . Therefore, the Gaussianity test $\varphi_{GT_1}(S)$ will conclude that ϵ_Y does not follow a normal distribution, i.e., $\varphi_{GT_1}(S) = 0$. Hence, when $\varphi_{IT_1}(S) = 0$, we also have $\varphi_{GT_1}(S) = 0$, and both tests yield consistent conclusions.

For $i = 2$, the same argument applies after interchanging X and Y , which shows $\varphi_{IT_2}(S) = \varphi_{GT_2}(S)$. Combining both cases yields the claim. \square

Gaussianity tests often lack accuracy and power: they are sample-size sensitive and falter under misspecification. Independence tests are more mature and robust; Theorem 4.3 proves their equivalence in bivariate LiNGAM, so the latter can replace the former when Gaussianity checks underperform.

Specifically, Shapiro–Wilk (SW) excels with small samples but over-rejects in large ones and weakens under leptokurtic distributions; Anderson–Darling (AD) shares the small-sample merit yet loses accuracy and speed as data grow; Jarque–Bera (JB), though universally applicable, becomes unstable with few observations and loses power when tails alone deviate. These traditional methods hinge on specific sample sizes and distributional forms, rendering them unreliable in complex scenarios, whereas independence tests remain flexible across varying data conditions.

By Theorem 4.3, for each $i \in \{1, 2\}$ the independence and Gaussianity tests are equivalent, $\varphi_{IT_i}(S) = \varphi_{GT_i}(S)$. In particular, let

$$\begin{aligned} \varphi_1 &= \begin{cases} 1, & \text{if } Y \perp \epsilon_X \text{ (equivalently } \epsilon_Y \sim \mathcal{N}(0, \sigma_Y^2)), \\ 0, & \text{otherwise,} \end{cases} \\ \varphi_2 &= \begin{cases} 1, & \text{if } X \perp \epsilon_Y \text{ (equivalently } \epsilon_X \sim \mathcal{N}(0, \sigma_X^2)), \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Hence, the joint decision (φ_1, φ_2) can be interpreted as follows:

- $(\varphi_1, \varphi_2) = (1, 0) \Rightarrow Y \rightarrow X$. Accepting $Y \perp \epsilon_X$ (reverse regression residual independent of the predictor) and rejecting $X \perp \epsilon_Y$ (forward regression residual dependent on the predictor) is exactly the LiNGAM exogeneity pattern for the causal direction $Y \rightarrow X$.
- $(\varphi_1, \varphi_2) = (0, 1) \Rightarrow X \rightarrow Y$. Symmetric to the previous case: independence holds only in the forward regression, indicating $X \rightarrow Y$.
- $(\varphi_1, \varphi_2) = (1, 1) \Rightarrow \text{Undetermined (noise effectively Gaussian)}$. Both independence tests are accepted, and by Theorem 4.3 both Gaussianity tests are also accepted, so ϵ_X, ϵ_Y behave as Gaussian. In the linear-Gaussian setting the direction is not identifiable, thus we abstain.
- $(\varphi_1, \varphi_2) = (0, 0) \Rightarrow \text{Inconclusive}$. Both independence tests are rejected (and, equivalently, both Gaussianity tests are rejected). This offers insufficient evidence for either direction and may indicate finite-

Table 1. LiNGAM Variants Comparison

Variant	L, A, E	NG	RoV	Return
ICA	Must	Must	Global	DAG
Direct	Must	Must	Global	DAG
Pairwise	Must	Must	Local	DAG
GaussDetect	Must	Optional	Local	DAG,ND

Note: All methods are LiNGAM variants. L = Linearity, A = Acyclicity, E = Exogeneity, NG = Non-Gaussianity, RoV = Range of Variables (Global/Local), DAG = Directed Acyclic Graph, ND = Noise Distribution. Must = All three assumptions (L, A, E) must be satisfied.

sample power issues, model misspecification (e.g., non-linearity, hidden confounding), or near-Gaussian disturbances.

5. Algorithms

Based on the conclusion in Theorem 4.3, the GaussDetect-LiNGAM algorithm is proposed.

By Theorem 4.3, the independence tests and Gaussianity tests are equivalent decision rules for the bivariate case. This allows us to release the *a priori* knowledge requirement on the noise non-Gaussianity: GaussDetect-LiNGAM does not assume that the disturbance terms are known to be non-Gaussian in advance; instead, it diagnoses (non-)Gaussianity from the data via the tests and proceeds accordingly. Importantly, this does not mean that non-Gaussian noise is unnecessary for identifiability; it only means we do not need to know it *a priori*. When both residuals are Gaussian (both tests accepted), the algorithm abstains and declares LiNGAM inapplicable.

In contrast, classical LiNGAM methods require the full set of assumptions to hold *ex ante* and simultaneously: (i) linearity, (ii) acyclicity, (iii) exogeneity (no hidden confounding), and (iv) non-Gaussian disturbances. All four are indispensable for identifiability in the traditional framework. GaussDetect-LiNGAM keeps (i)–(iii) as standing assumptions and replaces the *a priori* imposition of (iv) with a data-driven diagnostic, as summarized in Algorithm 1.

The advantage of the GaussDetect-LiNGAM algorithm lies in its ability to automatically detect the Gaussian nature of the noise during causal inference, without requiring *a priori* assumptions about the noise distribution. By performing independence tests, GaussDetect-LiNGAM can handle data with varying noise types, avoiding the dependency on non-Gaussian noise assumptions typical of traditional LiNGAM algorithms. This feature makes GaussDetect-LiNGAM more adaptable in scenarios with unknown or changing noise, providing more accurate and reliable causal inference results.

Algorithm 1 GaussDetect-LiNGAM

Input: Dataset $\mathbf{X} = \{X, Y\}$ (two continuous variables)

Output: Causal direction: $X \rightarrow Y$, $Y \rightarrow X$, “Gaussian noise”, or “Inconclusive”

Step 1: Standardize variables X and Y

Step 2: Fit regression models

$$\text{Forward: } Y = aX + \epsilon_Y$$

$$\text{Reversed: } X = bY + \epsilon_X$$

Step 3: Perform independence tests

$$H_{10}: \text{Test } X \perp \epsilon_Y$$

$$H_{20}: \text{Test } Y \perp \epsilon_X$$

Step 4: Determine causal direction

if H_{10} accepted and H_{20} rejected **then**

$$X \rightarrow Y$$

else if H_{20} accepted and H_{10} rejected **then**

$$Y \rightarrow X$$

else if H_{10} and H_{20} both accepted **then**

“Gaussian noise” (LiNGAM not applicable)

else

“Inconclusive”

end if

return causal direction

6. Experiments

6.1. Experimental Setup

Experiment 1 aims to validate the correctness of Theorem 4.3 by evaluating its accuracy through the consistency rate (Consistency Rate). Experiment 2 aims to compare GaussDetect-LiNGAM and Pairwise-LiNGAM across different noise types (Gaussian and Non-Gaussian) and varying sample sizes (400, 800, 1600). The core goal of this experiment is to evaluate the differences between these two algorithms in terms of tests per decision (TPD).

6.1.1. SAMPLE GENERATION

We generated two sets of samples: 1.Gaussian Noise: Samples drawn from the standard Gaussian distribution. 2.Non-Gaussian Noise: Including Exponential, Laplace, and Poisson distributions. Each of these non-Gaussian noise types generated the same number of samples as the Gaussian noise. The generated datasets (X and Y) were divided into two parts: 1. X is a random sample generated from the standard normal distribution. 2. Y is generated by the formula $Y = 2X + \epsilon_Y$, where the noise is composed of either Gaussian or non-Gaussian noise.

6.2.2. RESULTS

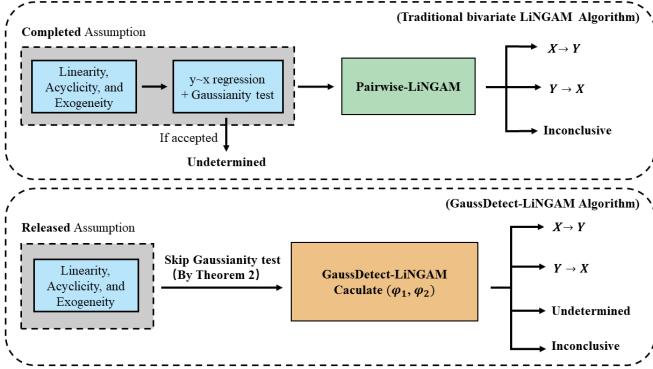


Figure 1. This figure compares the Traditional Pairwise LiNGAM and GaussDetect-LiNGAM algorithms. Pairwise LiNGAM requires completed assumptions (linearity, acyclicity, exogeneity, and non-Gaussianity) and involves two steps: a Gaussianity test followed by an independence test. In contrast, GaussDetect-LiNGAM only requires the released assumptions (linearity, acyclicity, and exogeneity) and can infer causality with just one independence test, without needing the Gaussianity test.

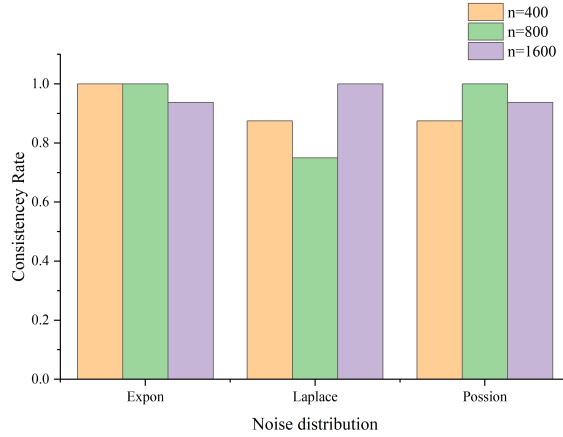


Figure 2. The figure illustrates the Consistency Rate across different noise distributions (Exponential, Laplace, and Poisson) and sample sizes ($n=400$, $n=800$, $n=1600$). The bars in different colors represent the performance of the consistency rate for each sample size.

Each set of samples (Gaussian and non-Gaussian noise) was divided into multiple small batches. Each batch underwent Gaussianity tests and independence tests, and we calculated the tests per decision (TPD) and consistency rate (Consistency Rate).

6.2. Consistency Rate Experiment

The consistency rate (Consistency Rate) measures whether the results of the Gaussianity test and independence test in each batch are consistent. This experiment aims to evaluate whether GaussDetect-LiNGAM can consistently determine the validity of the data under different noise types and sample sizes.

6.2.1. EXPERIMENT PROCEDURE

1. For each dataset, the data is divided into multiple small batches.
2. First, a Gaussianity test is performed.
3. For samples that pass the Gaussianity test, we perform regression on the residual noise and perform the independence test between the residual noise and the predictor variable.
4. Record whether each batch yields consistent conclusions (whether the Gaussianity test and independence test results are consistent).
5. Calculate the consistency rate for each noise type and sample size.

As shown in the figure 2, GaussDetect-LiNGAM consistently demonstrates a high consistency rate, especially in larger sample sizes, with many cases achieving a consistency rate of 1. This indicates that GaussDetect-LiNGAM is effective at providing consistent judgments across different noise types.

In the case of non-Gaussian noise (such as Laplace and Poisson distributions), we also observe fairly consistent results. Although the Gaussianity test does not always pass, when the noise type is close to Gaussian, GaussDetect-LiNGAM still achieves a high consistency rate.

6.3. Tests per Decision (TPD) Experiment

TPD(Tests per Decision) refers to the average number of tests required by GaussDetect-LiNGAM and Pairwise-LiNGAM to make a decision. This experiment compares the decision-making efficiency of both algorithms across different sample sizes and noise types.

6.3.1. EXPERIMENT PROCEDURE

1. For each dataset, the data is divided into multiple small batches.
2. For GaussDetect-LiNGAM, only independence tests are performed, and the number of tests per decision is calculated.
3. For Pairwise-LiNGAM, we first perform a Gaussianity

test, and if the Gaussianity test passes, we proceed with two independence tests.

4. Calculate TPD (tests per decision) and compute the average number of tests for each sample size and noise type.

6.3.2. RESULTS

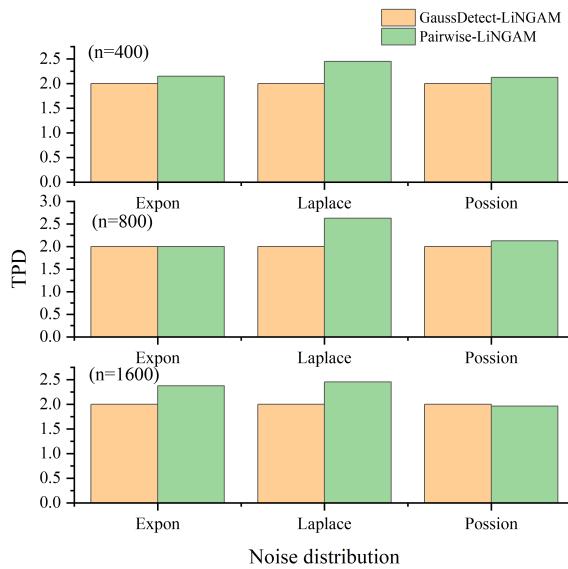


Figure 3. This figure illustrates the Tests per Decision (TPD) required by GaussDetect-LiNGAM and Pairwise-LiNGAM across different noise distributions (Exponential, Laplace, and Poisson) and sample sizes ($n=400$, $n=800$, $n=1600$). The bars in different colors represent the performance of GaussDetect-LiNGAM (in orange) and Pairwise-LiNGAM (in green).

As shown in the figure 3, GaussDetect-LiNGAM requires only one independence test per decision, so its TPD is generally low, around 2 tests. In contrast, Pairwise-LiNGAM requires first performing the Gaussianity test, followed by two independence tests, which results in a higher TPD, typically greater than 2 tests. This suggests that GaussDetect-LiNGAM is more efficient in terms of the number of tests required per decision, while Pairwise-LiNGAM has a higher computational cost due to the additional Gaussianity test.

6.4. Experiment Summary

Consistency Rate: The results demonstrate the correctness of Theorem 4.3 and the accuracy of GaussDetect-LiNGAM. It consistently achieves a high consistency rate across all noise types, particularly in cases involving non-Gaussian noise such as Laplace and Poisson distributions. This validates the effectiveness of GaussDetect-LiNGAM in handling non-Gaussian noise while maintaining reliable causal-

ity inference.

Tests per Decision (TPD): GaussDetect-LiNGAM requires only independence tests per decision, which results in a lower TPD of around 2. On the other hand, Pairwise-LiNGAM requires an additional Gaussianity test and two independence tests, resulting in a higher TPD, generally greater than 2.

These results indicate that GaussDetect-LiNGAM is more efficient, especially when handling non-Gaussian noise, with a lower computational cost, while Pairwise-LiNGAM provides more detailed analysis but requires more tests per decision and incurs higher computational overhead.

7. Conclusion

We have shown that, in the bivariate LiNGAM setting, the Gaussianity of noise can be equivalently assessed through an independence test in the reverse model. This theoretical result eliminates the need for explicit non-Gaussianity assumptions and enables the use of more flexible and powerful independence testing methods. The proposed GaussDetect-LiNGAM algorithm simplifies causal direction identification while maintaining robust performance under diverse data conditions. These contributions make LiNGAM more accessible and reliable for practical causal inference applications.

References

- Akaho, S. A kernel method for canonical correlation analysis. 3:1–48. ISSN 1532-4435. doi: 10.1162/153244303768966085. URL <https://dl.acm.org/doi/10.1162/153244303768966085>.
- Darmois, G. Analyse générale des liaisons stochastiques: Etude particulière de l'analyse factorielle linéaire. 21(1/2):2–8. ISSN 0373-1138. doi: 10.2307/1401511. URL <https://www.jstor.org/stable/1401511>.
- Hyvärinen, A. and Oja, E. Independent component analysis: Algorithms and applications. 13(4–5):411–430. ISSN 0893-6080. doi: 10.1016/s0893-6080(00)00026-5.
- Hyvärinen, A. and Smith, S. M. Pairwise Likelihood Ratios for Estimation of Non-Gaussian Structural Equation Models. 14:111–152. ISSN 1532-4435.
- Lukacs, E. and King, E. P. A Property of the Normal Distribution. 25(2):389–394. ISSN 0003-4851, 2168-8990. doi: 10.1214/aoms/1177728796. URL <https://projecteuclid.org/journals/annals-of-mathematical-statistics/>

volume-25/issue-2/
A-Property-of-the-Normal-Distribution/
10.1214/aoms/1177728796.full.

Pearl, J. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.

Shen, H., Jegelka, S., and Gretton, A. Fast Kernel ICA using an Approximate Newton Method. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, pp. 476–483. PMLR. URL <https://proceedings.mlr.press/v2/shen07a.html>.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., and Jordan, M. A linear non-Gaussian acyclic model for causal discovery. 7 (10), a. URL <http://www.jmlr.org/papers/volume7/shimizu06a/shimizu06a.pdf>.

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvarinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., and Bollen, K. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model, b. URL <http://arxiv.org/abs/1101.2489>.

Spirites, P., Glymour, C., and Scheines, R. Causation, Prediction, and Search.