

Efficient Ensemble Conditional Independence Test Framework for Causal Discovery

Zhengkang Guan Kun Kuang

College of Computer Science and Technology, Zhejiang University

ICLR 2026

Motivation: The Computational Bottleneck

- **Causal Discovery relies on CITs:** Constraint-based algorithms require numerous Conditional Independence Tests (CITs).
- **The Bottleneck:** Traditional CITs suffer from prohibitive computational costs with high time complexity regarding sample size.
- **Goal:** How to generally reduce the computational cost of CITs while preserving testing power across diverse conditional dependence structures?

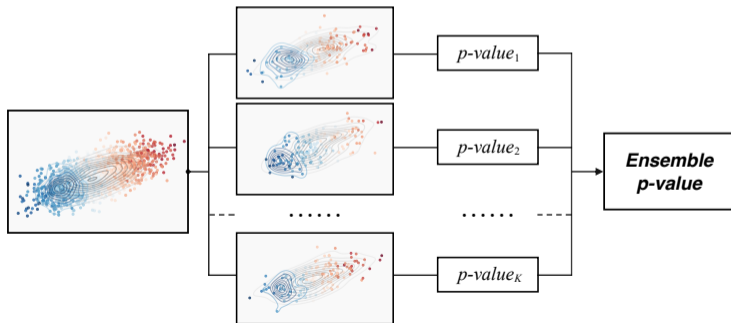
Proposed Method: E-CIT Framework

A plug-and-play Ensemble Conditional Independence Test (E-CIT) framework.

- **Divide-and-Aggregate Strategy:**

Partition data \rightarrow Parallel base CITs $(p_1, \dots, p_K) \rightarrow$ Aggregation.

- **Efficiency:** Complexity is reduced to **linear** in sample size, when subset size is fixed.



How to combine diverse p-values generated by complex CIT mechanisms?

- We propose a novel combination method based on the closure property of **Stable Distributions** $\mathbf{S}(\alpha, \beta, \gamma, \delta)$.
- **Ensemble Test Statistic** (T_e):

$$T_e = \frac{1}{K} \sum_{k=1}^K F_S^{-1}(p_k)$$

where F_S^{-1} is the inverse CDF of the stable distribution.

- Flexibility via parameter α (tail heaviness) handles non-parametric CIT contexts, outperforming traditional normal-based methods.

E-CIT provides robust theoretical properties **without** imposing strict parametric assumptions on the base test:

Theorem 1: Basic Properties

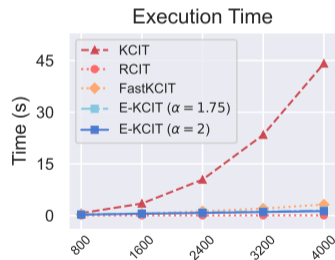
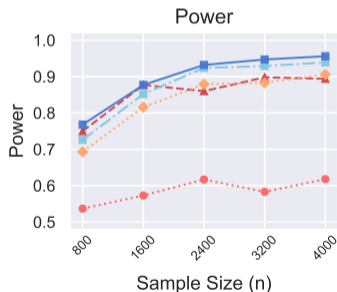
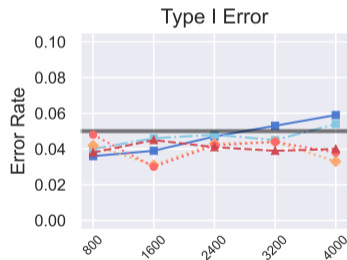
- **Validity:** Under H_0 , $p_{ensemble} \sim U[0, 1]$, ensuring exact Type I error control.
- **Admissibility:** No other test uniformly outperforms E-CIT in error rates and optimality.
- **Unbiasedness:** If subtests are unbiased ($P(reject|H_1) \geq \alpha$), the ensemble preserves it.

Theorem 2: Power Consistency

- The power of E-CIT approaches 1 as $K \rightarrow \infty$, under mild and practically verifiable conditions (e.g., basic effectiveness of subtests).

Experimental Results

- **Significant Speedup:** Runtime is drastically reduced, enabling massive sample tests.
- **Competitive Performance:** Improves power in complex scenarios, especially under challenging heavy-tailed noise.



- **Flow-Cytometry Dataset:** E-CIT enhances the performance of most CIT methods on real-world single-cell signaling data.
- **Causal Discovery:** Outperforms original methods in SHD when applied to the PC algorithm.

CIT	Type	F1-score \uparrow
KCIT	Orig.	0.624
	E-CIT	0.695
RCIT	Orig.	0.665
	E-CIT	0.687
LPCIT	Orig.	0.691
	E-CIT	0.741
CCIT	Orig.	0.605
	E-CIT	0.646
FisherZ	Orig.	0.737
	E-CIT	0.767

Summary: E-CIT is a modular, linear-time, theoretically sound framework that resolves the core computational bottleneck of constraint-based causal discovery.

Thanks for your attention!

Contact

Zhengkang Guan (zhengkang.guan@zju.edu.cn)

Kun Kuang (kunkuang@zju.edu.cn)