Toward Universal Laws of Outlier Propagation

Aram Ebtekar[1], Yuhao Wang[2,3] Dominik Janzing[3]

[1]Independent Researcher, [2]National University of Singapore, [3]Amazon Research Tübingen, Germany

Motivation and Problem Statement

Outliers signal deviations from expected system behavior and often appear in cascades across complex systems.
Traditional anomaly scores (e.g., z-scores, p-values)

1. Rigid Outlier Def. [BMJ+22]

Scores like $\lambda_\tau(x) = -\log P(\tau(X) \geq \tau(x))$ rely on hand-crafted features τ, which may overlook unexpected patterns.

2. High-Likelihood, Still Anomaly

Values like $x = 0$ may have moderate $\lambda_\tau(x)$, but are algorithmically simple.

3. No General Decomposition Rule

Joint scores $\lambda(x_1, x_2, \dots)$ do not decompose across mechanisms $\Rightarrow$ obstructs causal root attribution.

4. Likelihood-Only Reasoning

Compressible but frequent data may have high likelihood yet be missed by density-based methods.

Our Contributions:

Using Algorithmic Information Theory (AIT), we develop a framework to:

- Quantify anomalies via *randomness deficiency*
- Attribute outliers to specific causal mechanisms
- Establish universal statistical bounds on anomaly propagation

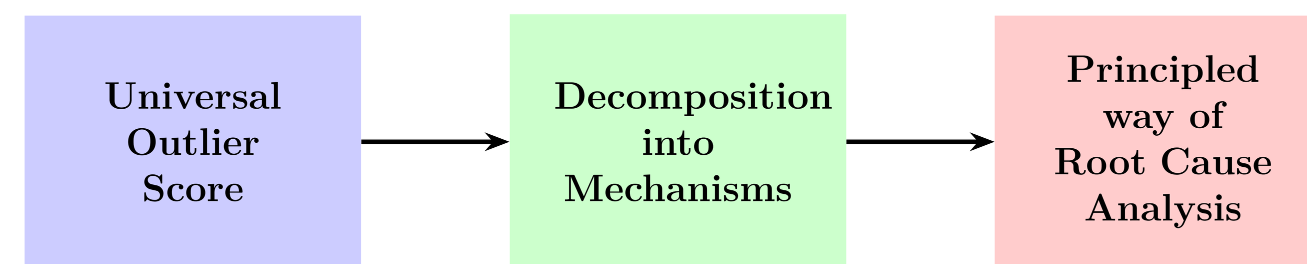
This enables:

- **Decomposability:** Attribute anomalies to specific parts of the system
- **Comparability:** Scores are calibrated across heterogeneous variables or modalities
- **Causal understanding:** Diagnose how anomalies propagate through causal mechanisms

Key Theoretical Tools

- **Kolmogorov Complexity** $K(x)$ [Gács21]: length of the shortest description of x
- **Randomness Deficiency** $\delta(x)$ [Gács21]: deviation from expected complexity
- **Causal Bayesian Networks**:
$$P(X_1, \dots, X_n) = \prod_{j=1}^{n} P(X_j \mid \mathrm{PA}_j)$$
- **Algorithmic Markov Condition** [JS10]:
$$K(x_1, \dots, x_n) \overset{+}{=} \sum_{j=1}^{n} K(x_j \mid \mathrm{pa}_j^*).$$

Conceptual Overview



Main Results

1. Universal Anomaly Score (E-Test)

We define a universal anomaly score using algorithmic information theory:
$$\delta(x) := -\log P(x) - K(x \mid P^*)$$
This semicomputable score captures both statistical rarity and algorithmic simplicity.

2. Decomposition of Randomness Deficiency

Given a causal DAG and joint observation $(x_1, \dots, x_n)$, the randomness deficiency decomposes as:
$$\delta(x_1, \dots, x_n) \approx \sum_j \delta(x_j \mid \mathrm{pa}_j)$$
This enables principled attribution of anomalies to individual causal mechanisms.

3. Randomness deficiency obeys causal monotonicity

Under the independence of mechanisms, randomness deficiency cannot increase along a causal chain. In particular, for a sequence of anomalies $x_1 \to x_2 \to \dots \to x_n$, we have:
$$\delta(x_j) \overset{+}{\leq} \delta(x_{j-1})$$

Experiment: Lempel-Ziv Root Cause

Setup:

- A 4-node causal chain: $X_1 \to X_2 \to X_3 \to X_4$
- Each variable evolves as: $X_j = X_{j-1} + N_j$, where $N_j \sim U[0, 1]$
- One node receives a structured anomaly: all but one digit of N_j set to zero

Method:

- Use Lempel-Ziv compression to approximate local complexity
- Identify the root cause as the node with minimal LZ-compressed residual

Summary of Computable Anomaly Score Examples

Method / Scenario	Interpretation and Theoretical Significance
z-score (Gaussian variable)	• Randomness deficiency $\delta(x)$ grows with the squared z-score: $\delta(x) \gtrsim \frac{\log e}{2} z^2(x) - 2\log\lvert x-\mu\rvert$ • z-scores are calibrated e-tests and computable proxies for algorithmic anomaly scores
Mahalanobis Distance (Multivariate Gaussian)	• Squared Mahalanobis distance lower-bounds randomness deficiency: $\delta(x) \gtrsim \frac{\log e}{2} x^\top \Sigma^{-1} x$ • In linear SEMs, decomposes into a sum of per-mechanism noise scores
Root Cause with Low Marginal Score	• A perturbed variable can have low $\delta(x_j)$ but high $\delta(x_j \mid \mathrm{pa}_j)$ • Accurate attribution requires joint or conditional evaluation (e.g., $\delta(x_1, x_2)$)
Binary Word (m-bit Bernoulli sequence)	• Randomness deficiency lower-bounded by KL divergence: $\delta(w) \gtrsim m \cdot \mathrm{KL}(\hat{p} \parallel p)$ • Captures outliers with either too many or too few 1s, despite high density
Compression Coincidence (Lempel-Ziv)	• Abrupt drops in conditional complexity $K(x_j \mid x_{j-1})$ indicate anomalies • Occurs when injected values yield unusually compressible patterns • Supports detection of structural irregularities even in noiseless or uniform settings

Conclusion and References

- AIT unifies outlier detection and causal attribution.
- Uncomputable information theory can guide practical, computable anomaly scores.

[BMJ+22] Kailash Budhathoki, Lenon Minorics, Patrick Blöbaum, and Dominik Janzing. Causal structure–based root cause analysis of outliers. ICML2022.

[Gács21] Péter Gács. Lecture notes on descriptional complexity and randomness. 2021

[JS10] Dominik Janzing and Bernhard Schölkopf. Causal inference using the algorithmic Markov condition. *IEEE Transactions on Information Theory*, 56(10):5168–5194, 2010.