On the Identifiability of Poisson Branching Structural Causal Model Using Probability Generating Function

Yu Xiang[1], Jie Qiao[1], Zhefeng Liang[1], Zihuai Zeng[1], Ruichu Cai[1,2], Zhifeng Hao[3]

[1]School of Computer Science, Guangdong University of Technology, China
[2]Peng Cheng Laboratory, Shenzhen 518066, China
[3] College of Science, Shantou University, Shantou 515063, China
{thexiang2000, qiaojie.chn, lzfeng011021, zzhuaiiii, cairuichu}@gmail.com,
haozhifeng@stu.edu.cn

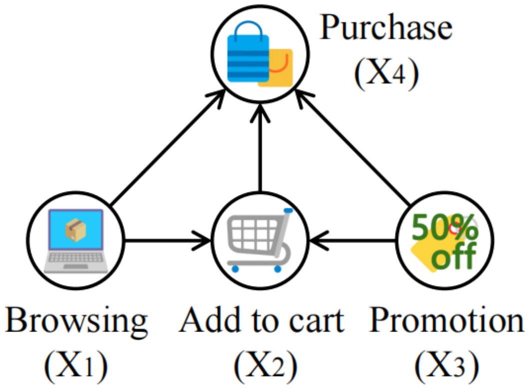
Introduction

Observational count data

	X_1	X_2	X_3	X_4
1	1	2	2	4
2	0	1	1	3
⋮	…	…	…	…
n	1	1	3	5

?

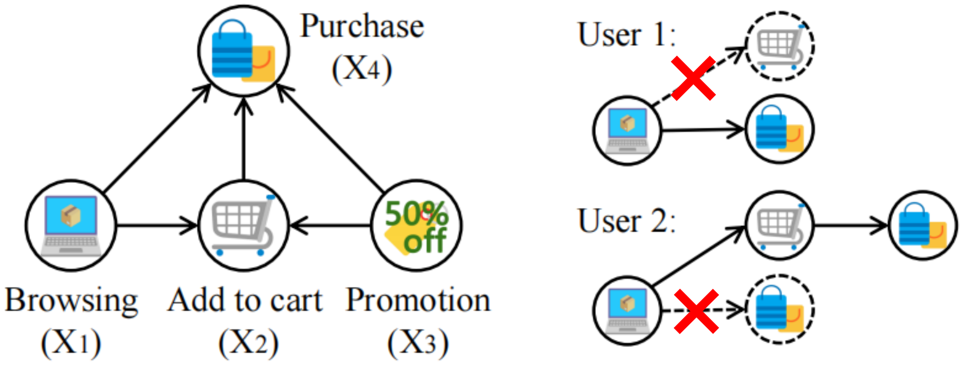
Causal graph in online shopping scenarios



Task: How to learn the causal structure from the observed count data

Recent works: Poisson Branching Structural Causal Model (Qiao 2024). Poisson Bayesian Network (Park and Raskutti 2015), Zero-Inflated Poisson Bayesian Networks(Choi, Chapkin, and Ni 2020), Ordinal Causal Discovery (Ni and Mallick 2022).

Poisson Branching Structural Causal Model (PB-SCM)



(a) Causal Graph (b) Branching Structure

Each event contributes independently to the occurrence of its child event.

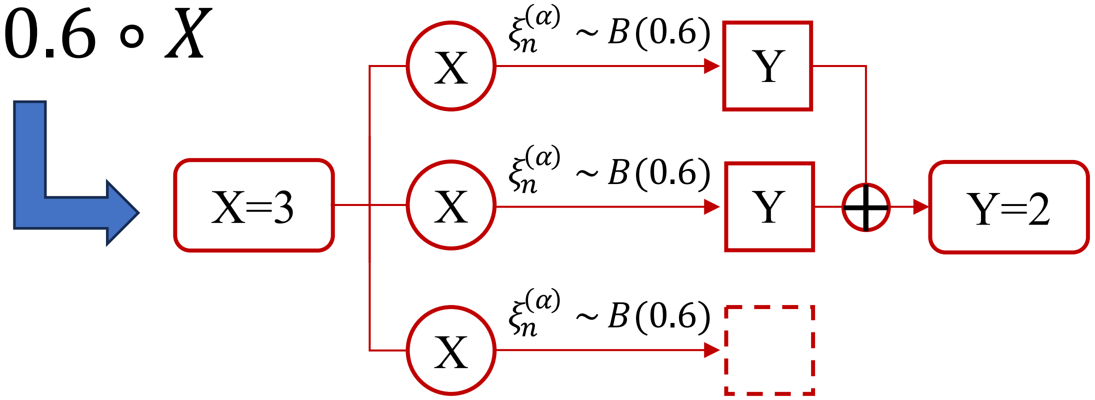
Definition For each random variable $X_i \in X$, let $\epsilon_i \sim Pois(\mu_i)$ be the noise component of X_i, then X_i is generated by:

$$X_i = \sum_{j \in Pa(i)} \alpha_{j,i} \circ X_j + \epsilon_i,$$

where $\alpha_{j,i} \in (0,1]$ is the causal coefficient from X_j to X_i.

Thinning operator "∘" helps model branching structure.

$$Y = 0.6 \circ X$$



$B(0.6)$: Bernoulli distribution with parameter 0.6

$$P(n) = \begin{cases} 0.6, & n = 1 \\ 0.4, & n = 0 \end{cases}$$

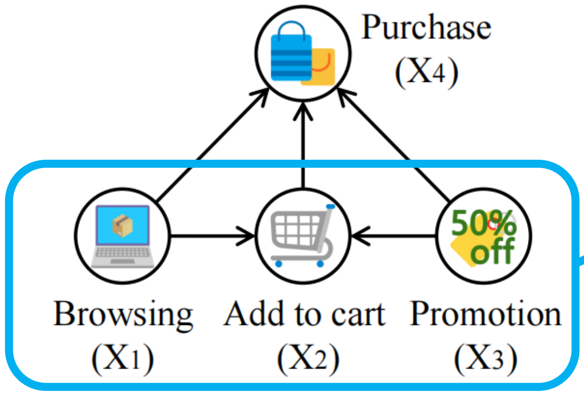
A toy example for the thinning operator, each X has a 60% chance of triggering an occurrence of an event Y.

Limitations of the cumulant-based method for PB-SCM

- Need for estimating higher-order cumulants → 1. High computational cost; 2. Estimation bias

1. High computational cost
2. Estimation bias

- certain identifiability issues



The causal relationships among X_1, X_2 and X_3 are not identifiable

Why Do These Limitations Exist?

The $\widetilde{\Lambda}_k\,(X_i \to X_j)$ constructed by cumulant extract path information from entire graph and only enable causal discovery when they reach the highest non-zero order.

Cumulant-based method can only leverage higher-order information

Probability Generating Function

Definition. Given discrete random vector $\mathbf{X} = [X_1, \ldots, X_d]^T$ taking values in the non-negative integers $\mathbb{Z}^{\geq 0}$, the probability generating function of X is defined as:

$$G_{\mathbf{X}}(\mathbf{z}) = \mathbb{E}[z_1^{X_1} \cdots z_d^{X_d}] = \sum_{x_1,\ldots,x_d=0}^{\infty} p(x_1, ..., x_d) z_1^{x_1} \cdots z_d^{x_d},$$

where p is the probability mass function of $\mathbf{X}$ and $|z_i| \leq 1$.

The expectation form of the PGF doesn't provide any interpretable information.

Theorem 1 (Closed-form solution for PGF of PB-SCM). *Given a random vector $\mathbf{X} = [X_1, ..., X_n]^T$ following PB-SCM, let $\mathbf{z}_{(j)} = \{z_l | l \in Des(j) \cup \{j\}\}$, the PGF of $P(\mathbf{X})$ is given by $G_{\mathbf{X}}(\mathbf{z}) = \prod_{i \in [d]} G_{\epsilon_i}\Big(z_i \times \prod_{j \in Ch(i)} G_{i,j}(\mathbf{z}_{(j)})\Big)$, where*

$$G_{i,j}(\mathbf{z}_{(j)}) = \begin{cases} G_{B(\alpha_{i,j})}\left(z_j \times \prod_{k \in Ch(j)} G_{j,k}(\mathbf{z}_{(k)})\right) & , Ch(j) \neq \emptyset \\ G_{B(\alpha_{i,j})}(z_j) & , Otherwise \end{cases}, \tag{3}$$

in which $G_{\epsilon_i}(\cdot)$ is the PGF of Poisson noise ϵ_i and $G_{B(\alpha_{i,j})}(\cdot)$ is the PGF of Bernoulli distribution with parameter $\alpha_{i,j}$.

The Closed Form of Probability Generating Function

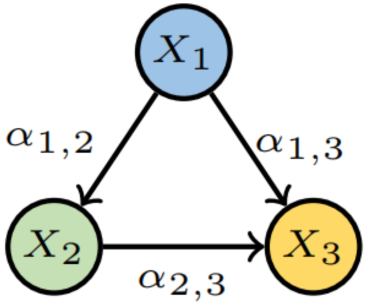
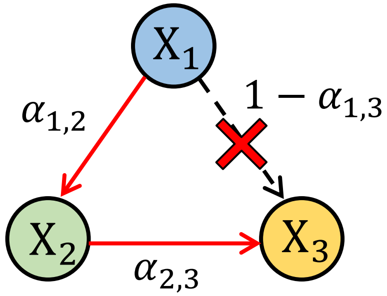


Fig.5 triangular structure.

$$G_{\mathbf{X}}(\mathbf{z}) = E[z_1^{X_1} z_2^{X_2} z_3^{X_3}]$$
$$= \exp\Big[\mu_x[(1-\alpha_{1,2})(1-\alpha_{1,3})z_1 + \alpha_{1,2}(1-\alpha_{2,3})(1-\alpha_{1,3})z_1 z_2 + \alpha_{1,3}(1-\alpha_{1,2})z_1 z_3$$
$$+ \alpha_{1,2}\alpha_{1,3}(1-\alpha_{2,3})z_1 z_2 z_3 + \alpha_{1,2}\alpha_{2,3}(1-\alpha_{1,3})z_1 z_2 z_3 + \alpha_{2,3}\alpha_{1,2}\alpha_{1,3} z_1 z_2 z_3^2] - \mu_x\Big]$$
$$\times \exp\Big[\mu_y \times [(1-\alpha_{2,3})z_2 + \alpha_{2,3} z_2 z_3] - \mu_y\Big] \times \exp\Big[\mu_z z_3 - \mu_z\Big]$$

$$\alpha_{1,2}\alpha_{2,3}(1-\alpha_{1,3})z_1 z_2 z_3$$

One term within the PGF $\mathbb{E}[z_1^{X_1} z_2^{X_2} z_3^{X_3}]$ of the triangular structure.

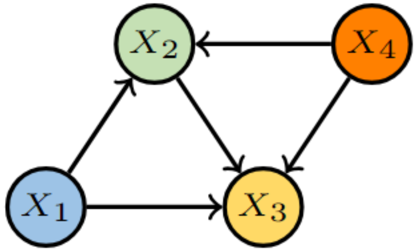


Corresponding branching structure.

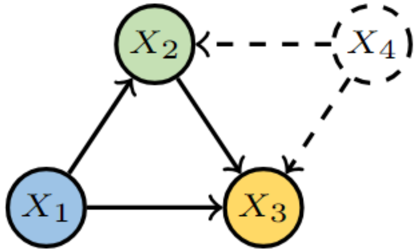
Each term in the closed form of PGF **corresponds to a branching structure.**

Discovery: Each term in the closed form of PGF **corresponds to a branching structure.**

The PGF has a local property, allowing for the efficient identification of **local structures.**



$$\lim_{z_4 \to 0} G_{\boldsymbol{X}}(z_1, z_2, z_3, z_4)$$



Specific **local structures** can be identified
by verifying the existence of corresponding term in $\lim_{z_4 \to 0} G_{\boldsymbol{X}}(z_1, z_2, z_3, z_4)$

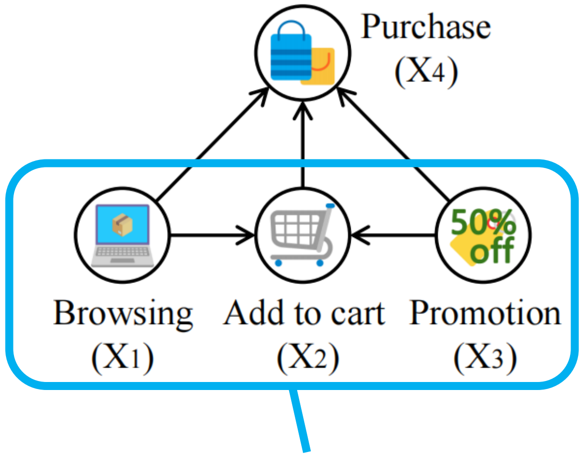
Focus on **local structures**, No need for high-order information

Discovery: Each term in the closed form of PGF **corresponds to a branching structure.**

Identifiability & Method: Specific **local structure** can be identified by verifying the existence of specific items in PGF

PGF leverage low-order information to identify causal direction

- **Certain identifiability issues (√)**



By checking whether the term corresponding to $X_1 \rightarrow X_2 \leftarrow X_3$ exists in PGF $G_{\boldsymbol{X}}(z_1, z_2, z_3, z_4)$.

- **High computational cost (√)**

Total complexity : $\mathcal{O}(\frac{d(d-1)}{2} + \frac{2d(d-1)(d-2)}{3})$

- **Estimation bias (√)**

Identify local structures without relying on high-order cumulants, improving estimation accuracy.

Experiments Result

Sensitivity Experiment

Table 1: Sensitivity to Avg. Indegree Rate.

	F1↑				SHD↓			
Avg. Indegree	2.0	2.5	3.0	3.5	2.0	2.5	3.0	3.5
Ours	**0.74 ± 0.05**	**0.81 ± 0.07**	**0.86 ± 0.03**	**0.89 ± 0.04**	**9.67 ± 1.99**	**8.33 ± 2.28**	**6.40 ± 1.68**	**5.27 ± 1.39**
Cumulant	0.73 ± 0.03	0.77 ± 0.02	0.80 ± 0.04	0.83 ± 0.03	13.40 ± 1.28	14.10 ± 1.51	13.00 ± 2.37	13.20 ± 2.23
PC	0.60 ± 0.17	0.62 ± 0.11	0.54 ± 0.12	0.60 ± 0.12	9.90 ± 3.45	11.80 ± 2.48	15.90 ± 3.91	16.10 ± 3.21
GES	0.48 ± 0.14	0.48 ± 0.11	0.41 ± 0.11	0.37 ± 0.10	14.90 ± 4.48	19.50 ± 4.61	25.90 ± 4.18	30.5 ± 4.06
OCD	0.23 ± 0.22	0.27 ± 0.23	0.28 ± 0.16	0.37 ± 0.14	16.10 ± 3.70	19.40 ± 5.50	23.60 ± 4.62	24.30 ± 4.67

Table 2: Sensitivity to Sample Size.

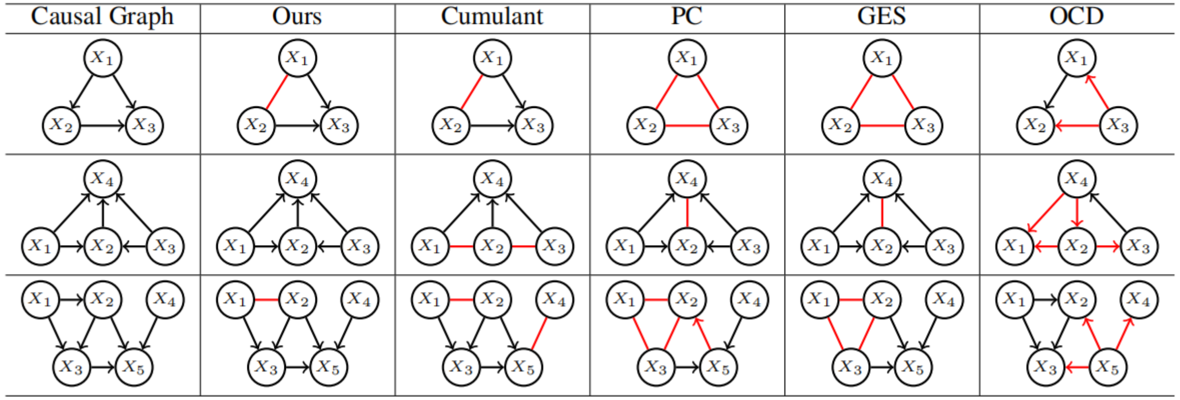
	F1↑				SHD↓			
Sample Size	5000	15000	30000	50000	5000	15000	30000	50000
Ours	**0.75 ± 0.09**	**0.82 ± 0.04**	**0.86 ± 0.03**	**0.87 ± 0.04**	**11.50 ± 3.34**	**9.27 ± 2.37**	**6.40 ± 1.68**	**5.87 ± 1.25**
Cumulant	0.72 ± 0.04	0.78 ± 0.02	0.80 ± 0.04	0.80 ± 0.03	19.90 ± 3.35	15.00 ± 1.41	13.00 ± 2.49	13.60 ± 2.63
PC	0.45 ± 0.11	0.54 ± 0.11	0.54 ± 0.13	0.66 ± 0.09	19.50 ± 4.30	15.70 ± 3.77	15.90 ± 4.12	13.00 ± 2.79
GES	0.39 ± 0.10	0.44 ± 0.20	0.41 ± 0.11	0.43 ± 0.22	23.70 ± 4.14	22.70 ± 7.85	25.90 ± 4.41	24.10 ± 8.84
OCD	0.30 ± 0.12	0.35 ± 0.18	0.28 ± 0.16	0.38 ± 0.20	21.90 ± 3.35	20.80 ± 4.57	23.60 ± 4.62	20.90 ± 5.74

Table 5: Runtime of each method under the default setting.

	Ours	Cumulant	PC	GES	OCD
Runtime (second)	7.94 ± 0.75	77.07 ± 4.94	5.00 ± 1.35	6.90 ± 1.86	9216 ± 1368

Case study

Table 3: Case studies of causal graphs with in total 3, 4, and 5 vertices, respectively. Red undirected edges indicate that adjacency has been learned but the direction cannot be determined, while red directed edges indicate incorrectly learned directions.



Real World Experiment. we orientate in the local triangular structures Y1 – Y2 – F, Y2 – R – F, and the local collider structure Y1 – S – R.

As a result, Our method successfully identifies adjacent vertex Foul – First Yellow Card and other six causal directions, which is consistent with our theoretical result.

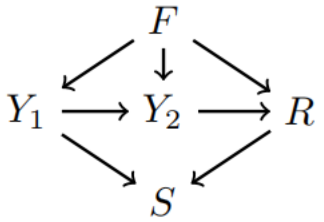


Figure 4: Football Event Graph (F: Foul, Y_1: Yellow card, Y_2: Second yellow card, R: Red card, S: Substitution)

Thanks!