Causal inference and LOB

Wentao Wang, Yuhang Guo

Présentation du cours APM_52112_EP

17 Feb 2026

Temporal data are divided into 2 types:

- Multivariate Time Series **(MTS)**: $x_t = (x_t^1, \ldots, x_t^d)$, objective is to study $x_i^{t-k} \to x_j^t$;
- Event Sequence **(ES)**: (t_n, e_n), study if one event can trigger another.

The goal is to unify temporal series methods, covering MTS and ES, comparing the connections and differences among various approaches.

Hypothesis

Causal sufficiency: No unobserved common cause (no hidden confounding variables);
Faithfulness: Statistical independence entirely determined by graph structure;
Temporal Priority: The cause must precede the effect.

Here we define the SCM by MTS:

$$x_i = f_i(Pa(x_i), u_i) \Rightarrow x_i^t = f_i(Pa(x_i^t), u_i^t).$$

Three types of casual graphs are proposed in reality:

- Full-time causal graph: Each point in time unfolds as a node;
- Window causal graph: Limit maximum lag length K;
- Summary causal graph: Indicates only if a causal relationship exists, without distinguishing temporal lags.

Causal Discovery from Multivariate Time Series

MTS-CD considers the realization of $x_i^{t-k} \to x_j^t$, and judges casual relationships between variables. Methods are resumed as following 4 types:

- **Constraint-based:** conditional independence based, if $X \perp Y \mid Z$, delete side between variable. Unstable under high dimension, sensible to sample. Ex: PCMCI(+), CTPC;
- **Score-based:** best model fit based, find G that $\hat{G} = \arg\min_G S(D, G)$, by combinatorial search (hill climbing, NP-hard) or continuous optimization (NOTEARS, transform into $h(A = 0) +$ grad, depend on model hypothesis);
- **SCM-based:** data mechanism based, suppose $x_i^t = f_i(Pa(x_i^t), u_i^t)$, identify causal direction through noise independence/non-Gaussianity, depend on model hypothesis. Ex: VAR-LiNGAM;
- **Granger causality:** predition capacity based, consider $x_i\ Granger - causes\ x_j$ if passed x_i improves prediction to x_j, not equivalent to casual relationship. Ex: NN-Granger casuality.

Causal Discovery from Event Sequences

Consider event sequence as $\{(t_n, e_n)\}_{n=1}^{N}$, realizes $A_{ij} = 1 \Rightarrow i$ *caused* j. Difference from MTS-CD: uneven time intervals, variable values non continuous, causality manifests as "triggered probability changes".

We define conditional intensity: $\lambda_j(t)$, probability of event j occurs at time t. If past event i increases the probability of future event j occurring, then a causal relationship is considered to exist. Methods classified into 4 types:

- Granger causality: most typically Hawkes process.

  $$\lambda_j(t) = \mu_j + \sum_i \sum_{t_k^i < t} \alpha_{ij}\, g\left(t - t_k^i\right).$$

  Each past event temporarily increases the probability of future events occurring, at intensity α_{ij}.
- Constraint-based: difficult to realize independency test.
- Score-based: mostly based on likelihood maximization.
- Extension of SCM ideas.

Evaluation Resources and Metrics

Causal discovery requires comparing whether the structure is correct. Datasets of evaluation is classified as 3 types:

- **Synthetic Data:** known ground-truth, establish the true causal graphgenerate data with SCM or Hawkes and compare structure;
- **Semi-synthetic Data:** use real data statistical properties with injection of known casual relationship;
- **Real-world Data:** causal validity cannot be rigorously verified.

Evalutation metrics are also classified as 3 types:

- **Structural Recovery Metrics:** Precision / Recall, F1-score, SHD (compare missing/extra sides, wrong directions).
- **Predictive Metrics**: used in Granger-based. Ex: likelihood, prediction error.
- **Time-relevant Metrics:** assess lag identification and causal delay.

Metrics are uncomparable, without uniform standard.

Perspectives

Open issues:

- Hidden confounders difficult to detecte;
- Non-stationarity of casual relationships, regime change possible;
- Conditional independency test and NN method decays at high dimension;
- Unification of continuous/discrete cases (MTS/ES);
- Trade-off of expressive power with interpretability for DL methods.

New Perspectives:

- Unifying causal modeling frameworks across different types of temporal data;
- Combining representation learning with causal discovery;
- Using interventional data to improve identifiability;
- Developing scalable methods for large-scale causal discovery;
- Establishing standardized evaluation benchmarks.

ADF test

H_0 : the series contains a unit root (non-stationary);
H_1 : stationary.

ADF statistic: 1.497, higher than all critical values (1%, 5%, 10%).
p-value $\approx$ 0.535, the unit root hypothesis cannot be rejected.

Dynamic kernel density: time-sliding & recursion

- Observations: $X_1, \ldots, X_t$ (in the paper: daily returns).
- From t_0 on, estimate a time-varying density by exponentially discounting the past.

$$\hat{f}_t^{h,\omega}(x) = \frac{1}{h} \sum_{i=1}^{t} w_{t,i}\, K\left(\frac{x - X_i}{h}\right), \qquad \sum_{i=1}^{t} w_{t,i} = 1.$$

$$w_{t,i} = \begin{cases} \dfrac{1-\omega}{1-\omega^{t_0}}\,\omega^{\,t-i}, & i \leq t_0, \\ (1-\omega)\,\omega^{\,t-i}, & i > t_0, \end{cases} \qquad 0 < \omega < 1.$$

- Key advantage: the estimator admits an online recursion (no full re-computation).

$$\hat{f}_{t+1}^{h,\omega}(x) = \omega\, \hat{f}_t^{h,\omega}(x) + \frac{1-\omega}{h} K\left(\frac{x - X_{t+1}}{h}\right), \qquad t \geq t_0.$$

CIG (PIT-based): Uniformity & Independence

- Probability Integral Transform (PIT) using the one-step-ahead predictive CDF:

$$Z_t^{h,\omega} = \hat{F}_{t-1}^{h,\omega}(X_t), \qquad t = t_0 + 1, \dots, T.$$

- If the predictive densities are well calibrated, then $\{Z_t^{h,\omega}\}$ should be iid $U(0,1)$.
- Hence the CIG checks: **(i) uniformity** and **(ii) independence (via lags)**.

(i) Univariate uniformity (KS gap)

$$k = \sup_{z \in [0,1]} \left| \hat{F}_Z(z) - z \right|, \qquad \hat{F}_Z(z) = \frac{1}{T - t_0} \sum_{t=t_0+1}^{T} \mathbf{1}\{Z_t \le z\}.$$

(ii) Independence via lag-τ bivariate discrepancy

$$\hat{F}_\tau(u, v) = \frac{1}{T - \tau - t_0} \sum_{t=t_0+1}^{T-\tau} \mathbf{1}\{Z_t \le u,\ Z_{t+\tau} \le v\}, \qquad k_\tau = \sup_{u,v \in [0,1]} \left| \hat{F}_\tau(u, v) - uv \right|.$$

Parameter selection: bandwidth h & discount ω

- Define $k_0^\circ := k$ (uniformity) and $k_\tau^\circ := k_\tau$ for $\tau \geq 1$ (dependence).
- KS-type statistics depend on sample size, so use a size-adapted aggregation:

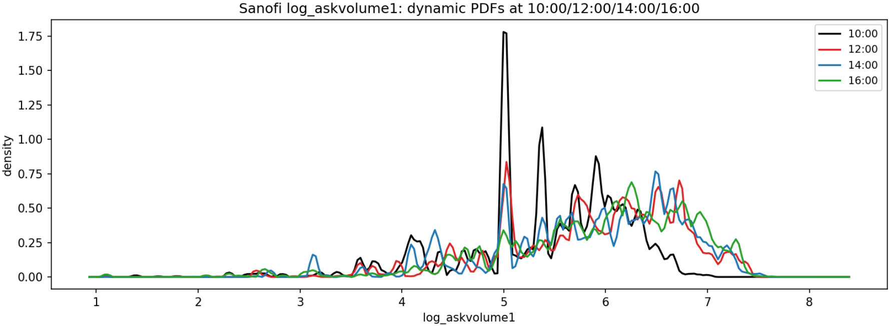
$$d_\nu\left(Z_{t_0+1}^{h,\omega}, \ldots, Z_T^{h,\omega}\right) = \max_{0 \leq \tau \leq \nu}\left(\sqrt{T - \tau - t_0}\, k_\tau^\circ\right).$$

- **Unconstrained optimum** (jointly enforces uniformity + independence):

$$(h^*, \omega^*) = \arg \min_{h>0,\ 0<\omega<1} d_\nu\left(Z_{t_0+1}^{h,\omega}, \ldots, Z_T^{h,\omega}\right).$$

- **Constrained version**: limit day-to-day distribution jumps, yielding $\omega > 1 - \nu^{-1}$:

$$(h_c^*, \omega_c^*) = \arg \min_{h>0,\ 1-\nu^{-1}<\omega<1} d_\nu\left(Z_{t_0+1}^{h,\omega}, \ldots, Z_T^{h,\omega}\right).$$


Sanofi log_askvolume1: dynamic PDFs at 10:00/12:00/14:00/16:00
10:00
12:00
14:00
16:00
density
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
log_askvolume1