Antal Jakovac
Wigner RC
Corvinus University of Budapest

Causality methods in time series analysis

GPU Day 2025
May 23, 2025, Budapest, Hungary

Introduction: modeling and AI

- Approaches to represent the world
  - we do observation, modeling and solving the models
  - observation tools ⇨ use observation data, we model and solve
  - computational tools ⇨ use data, we model, solve with computer
  - AI tools to set up models! ⇨ no human interaction at all
- What remains as human role?
  - decision: draw the consequences of the results
  - understand the basic motivations that lead to the actual model (symmetries, relations)

Introduction: modeling and AI

- Symmetries and relations
  - translational invariance (e.g. image recognition) ⇨ CNN
  - scale invariance (e.g. image recognition) ⇨ deep networks & pooling
  - Markovian systems (e.g. dynamics) ⇨ only the latest data is relevant
- If no evident symmetry?
  - assume a general model – not efficient (too lot of parameters)
  - attention mechanism, the system learns what is important for the present context (e.g in LLM)
  - generic strategy: reveal causal relations

Causality

- Causality: change in a data channel implies change in another later
- possible causal relations: X, Y are systems
  - $X \perp\!\!\!\perp Y$ (X and Y are independent)
  - $X \rightarrow Y$ (X drives Y), changes in X imply changes in Y
  - $X \Leftrightarrow Y$ (X and Y are interdependent), changes imply change in the other
  - $\exists$ Z that $Z \rightarrow X$ and $Z \rightarrow Y$, but not interdependent (common cause)
- some combinations are logically impossible (e.g. $\exists$ Z that $Z \rightarrow X$, $X \rightarrow Y$)
- the basic causal classes should be distinguishable

Causality

Causality methods

- **Wiener-Granger causality**:
  - assumes a linear stochastic model (AR-like model) between the data series, compares the cases when to predict Y we include the past of X or not (the null hypothesis)
  - great to present predictive precedence, but not really causal relation
  - can not distinguish common cause and bidirectional causal relation
  - assumes linearity and stationarity
  - works in stochastic models
  - very popular, used in lot of applications

Causality

Causality methods

- **Convergent Cross Mapping**: (Sugihara)
  - based on state space reconstruction (Taken's theorem): studying the embedding $[X_t, X_{t-\tau}, \ldots X_{t-(k-1)\tau}]$of the time series allows to reconstruct the complete dynamics
  - we can try to predict Y_t from embedding of X (nearest neighbors) and calculate quality of the reconstruction
  - works in deterministic cases
  - detecting common cause is somewhat heuristic

Causality

Causality methods

- **Dimensional causality**: (Telcs et al.)
  - idea: if X → Y, then the dimension of the attractor of Y is larger than that of X
  - study the dimension of X, Y and (XY), and compare the dimansions
  - appropriate to demonstrate all types of causal relations
  - works in deterministic systems (definition of “dimension”)
  - determining dimension can be tricky

Degrees of freedom method

- idea: the information necessary to reconstruct a data series contains the information of the driving system
- for k-th order differential/differentia equations: $X_n = f(X_{n-1}, \ldots, X_{n-k})$ necessary information: $X_1, \ldots, X_k$
- **degrees of freedom** is the necessary initial conditions
- "causal relations" in recursive equations:
  - X → Y: $X_n = f(X), \quad Y_n = g(X, Y)$
  - X ⇔ Y: $X_n = f(X, Y), \quad Y_n = g(X, Y)$
  - Z common cause: $Z_n = h(Z), \quad , X_n = f(X, Z), \quad Y_n = g(Y, Z)$

Degrees of freedom method

- consequence:
  - X → Y: $X_n = f(X),\ \ Y_n = g(X, Y)$ $\Rightarrow$ $d_X < d_Y$
  - X ⇔ Y: $X_n = f(X, Y),\ \ Y_n = g(X, Y)$ $\Rightarrow$ $d_X = d_Y$
  - Z common cause: $Z_n = h(Z),\ \ , X_n = f(X, Z),\ \ Y_n = g(Y, Z)$ $\Rightarrow$ $d_X + d_Y < d_{XY}$
- how can we determine the degrees of freedom?
  - what comes after a certain $X_1, \ldots, X_a$ series?
  - if $a \geqslant k$, only a single value allowed, for $a < k$ values distribution
  - for stochastic case: distribution stabilizes after k
  - problem fixing all earlier value allows very few data (statistics)

Degrees of freedom method

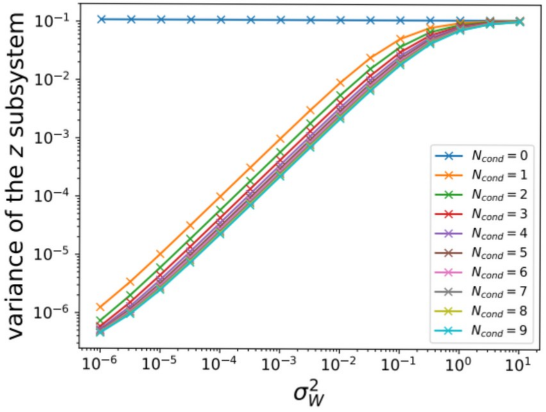
- strategy:
  - X data series, filter it according to some condition posed at generic n
  - C_W → W is the "window size", how many values are allowed
  - use milder conditions, gradually stricter ones (finite size scaling)
  - calculate $P_{i,W} = P(X_{n+i} \mid C_W)$ conditional probabilities
- expectation:
  - for large W $P_{i,W\to\infty} = P(X_n)$ (no information)
  - for W→0: $P_{i,W\to 0} = P(X_n \mid X_{n-1}, \ldots, X_{n-a})$ (no statistics)
  - from the saturation of the curve we read off the information

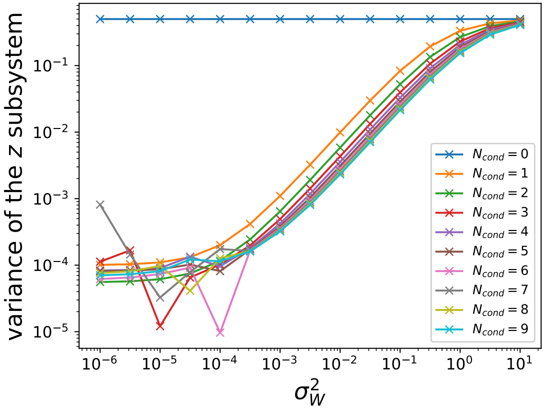
Degrees of freedom method

- stochastic case: all works, but for $W \to 0$ we do not get a sharp distribution
- only those degrees of freedom can be found, where the noise is smaller than the information carried by the given mode
- generic lesson
  - in real world everything affects to everything (Mars on the pendulum)
  - a huge number of modes have tiny effects
  - measurement accuracy is finite
  - only a finite number of modes are taken into account, the rest we treat as “noise”
  - number of degrees of freedom depends on the accuracy!

Degrees of freedom method

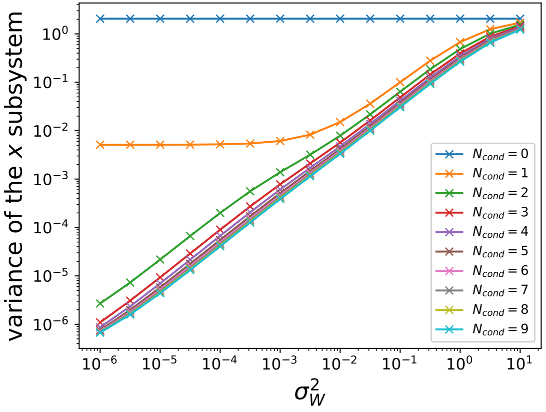
- actual example: linear stochastic system → analytically tractable
- X, Y, Z subsystems, Z is autonomous, it is a common cause for X,Y
- condition: all past elements are in a window [-W,W]
- Z subsystem: one condition fully determines the system → $df_Z = 1$

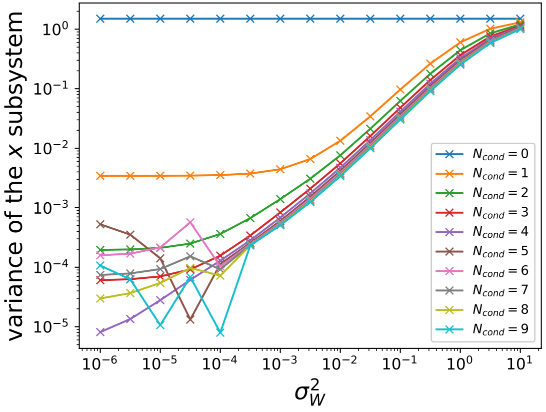




Degrees of freedom method

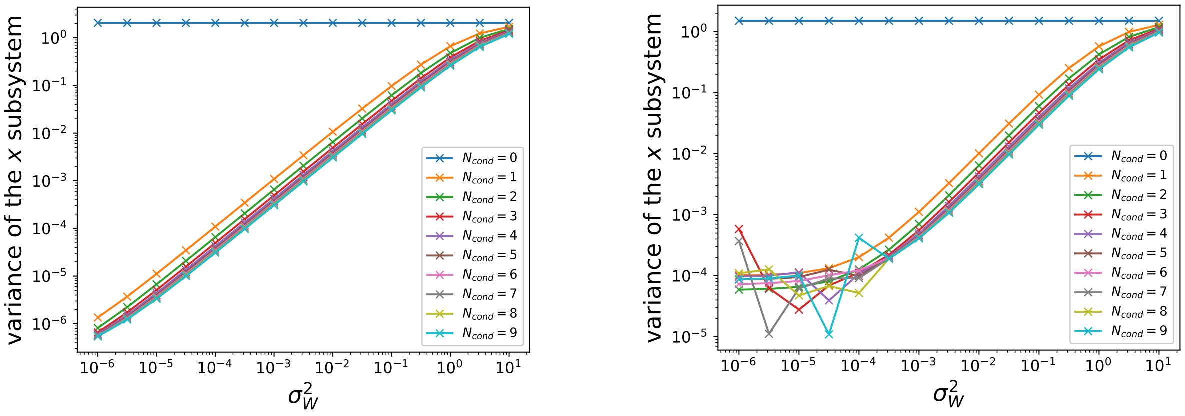
- X (and Y) subsystems: one condition does not determines fully the result, but two already yes: $df_X = 2$
- in the stochastic case: noise level 10^{-4}, signal level 5×10^{-3} → observable





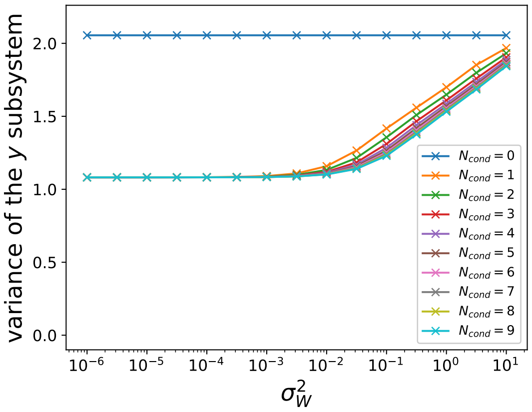
Degrees of freedom method

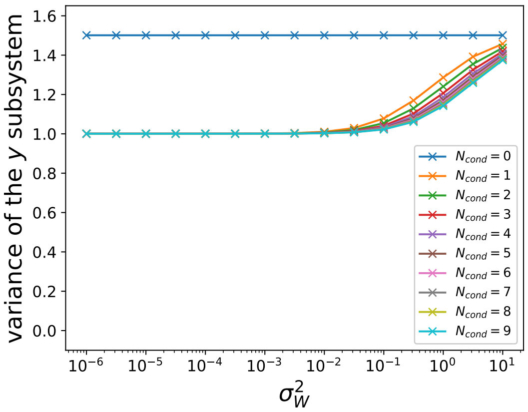
- XZ subsystem: one condition (pair) determines fully both X and Z → $df_{XZ}=2$
- consqeuence: Z→X



Degrees of freedom method

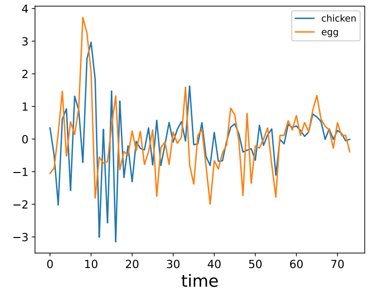
- XZ subsystem: one condition (pair) dos not determine fully Y!
- consqeuence: Y has independent information, so X and Y do not determine each other → Z is a common cause



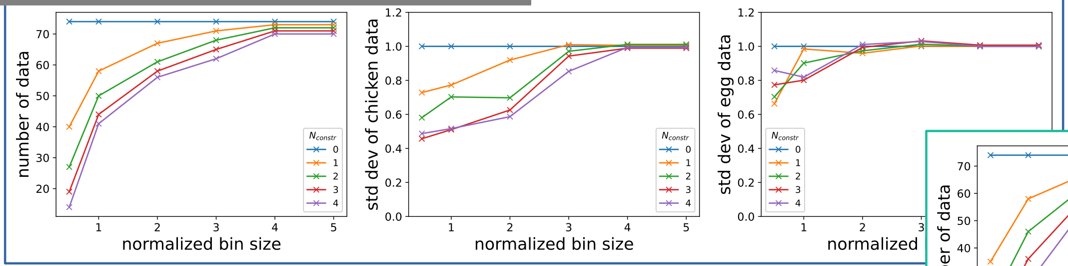


Degrees of freedom method

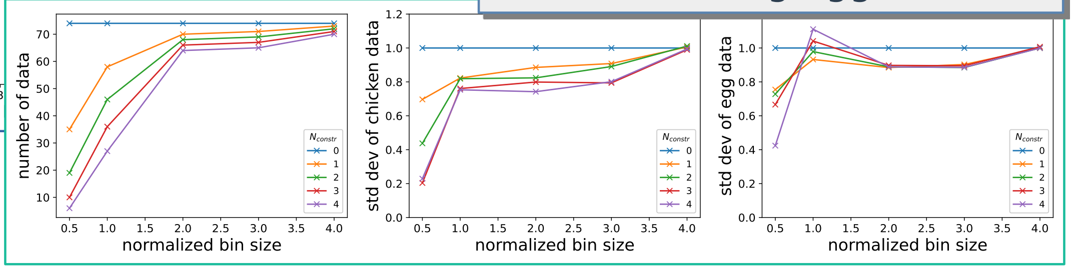
- benchmark example: chicken and egg
  - U.S. Department of Agriculture data, for the period 1930–1980
  - which drives the other?
- eggs affects chicken data, but not reversely: egg→chicken



Constraining chicken data



Constraining eggs data



Conclusion

- causality helps explore the connection between data → first step in setting up a dynamic model
- degrees of freedom method: determine, how many information is needed to fully constrain a time series
- all causal relations can be deducted
- number of df, and so the causality, can depend on the observational accuracy