Inference for Directed Acyclic Graph using Deep Generative Learning

Lexin Li

Professor
Division of Biostatistics &
Helen Wills Neuroscience Institute
University of California, Berkeley

Outline

- talk outline:
  - a general overview
  - a case study: hypothesis testing for directed acyclic graph
  - discussion
- thanks:
  - NSF CIF-2102227
  - NIH R01AG061303, R01AG062542
  - Chengchun Shi @ LSE & Yunzhe Zhou @ UC Berkeley

General overview

- **deep generative learning is exploding**
  - extremely rapid and widespread growth and advancements

General overview

- **deep generative learning is exploding**
  - extremely rapid and widespread growth and advancements
  - how to **embrace** it?

General overview

- **deep generative learning is exploding**
  - extremely rapid and widespread growth and advancements
  - how to **embrace** it?
- **how can stat help AI?**
  - data oversampling and data augmentation using LLM – Nakada et al. (2025)
  - detection of data misappropriation in LLM – Cai, Li and Zhang (2025)

General overview

- deep generative learning is exploding
  - extremely rapid and widespread growth and advancements
  - how to **embrace** it?
- how can stat help AI?
  - data oversampling and data augmentation using LLM – Nakada et al. (2025)
  - detection of data misappropriation in LLM – Cai, Li and Zhang (2025)
- how can AI help stat? employ AI as flexible and powerful learning tools to facilitate some classical statistical inference problems

General overview

- **deep generative learning is exploding**
  - extremely rapid and widespread growth and advancements
  - how to **embrace** it?
- **how can stat help AI?**
  - data oversampling and data augmentation using LLM – Nakada et al. (2025)
  - detection of data misappropriation in LLM – Cai, Li and Zhang (2025)
- **how can AI help stat?** employ AI as flexible and powerful learning tools to facilitate some classical statistical inference problems
  - inference for directed acyclic graph – Shi, Zhou and Li (2024, *JASA*) $\Leftarrow$ **this talk**
  - inference for conditional independence – Shi et al. (2021, *JMLR*)
  - inference for the Markov property in time series – Zhou, et al. (2023, *JRSSB*)
  - individual treatment effect inference using diffusion models – Cai, Jin and Li (2025, under review)

Deep generative learning

- **generative adversarial networks** (GANs, Goodfellow et al., 2014):
  - two neural networks, the generator and the discriminator, which are trained simultaneously
  - the generator creates data samples aiming to mimic the real data, while the discriminator evaluates and distinguishes between the generated and real samples, which helps in producing highly realistic outputs
- **variational autoencoders** (Rezende et al., 2014):
  - first encode input data into a latent space representation and then decode this representation back into data
- **normalizing flows** (Dinh et al., 2016):
  - a series of invertible transformations that map data to a simple distribution, like a Gaussian, and back
- **diffusion models** (Sohl-Dickstein et al., 2015):
  - learn to reverse a gradual process of adding noise to data
  - by learning the reverse diffusion process, can generate data starting from noise

Case Study:
Inference for Directed Acyclic Graph

Motivation example

- **brain effective connectivity analysis**:
  - brain is a highly interconnected dynamic system, in which the activity and temporal evolution of neural elements are triggered and influenced by the activities of other elements
  - uncover the **directional influence** that one neural system / region exerts over another
  - a task-evoked functional magnetic resonance imaging (fMRI) dataset from the Human Connectome Project (HCP)
  - analyze the fMRI scans of individuals who undertook a story-math task: $N = 28$ individuals with scores below 65 out of 100, and $N = 28$ individuals with the perfect score of 100
  - **fMRI**: measures blood oxygen level over time, a surrogate measure of brain neural activity; 4D spatial temporal array
  - for each subject, map brain voxels to a list of pre-specified brain regions, then average the time courses of voxels within the same region $\Rightarrow$ **region $\times$ time** matrix, with 316 time points, and 264 brain regions, grouped into 14 functional modules

Problem of interest

- **hypotheses we target**:

$$\mathcal{H}_0(j,k) : k \notin \mathrm{PA}_j, \quad \text{versus} \quad \mathcal{H}_1(j,k) : k \in \mathrm{PA}_j.$$

  - for a given pair of nodes (j,k), $j,k = 1,\ldots,d, j \neq k$
  - d random variables, $X = (X_1,\ldots,X_d)^\top$, that follow **an additive noise model**, $X_j = f_j(X_{\mathrm{PA}_j}) + \varepsilon_j$, with continuous f_j, independent error ε_j
  - the corresponding directed acyclic graph (DAG) is identifiable

- literature review:
  - penalized DAG estimation (Spirtes et al., 2000; van de Geer and Buhlmann, 2013; Zheng et al., 2018; Yuan et al., 2019)
  - Bayesian network (Chickering et al., 2004; Friston, 2011)
  - DAG inference under Gaussian linear DAG (Jankova and van de Geer, 2019; Li et al., 2020)

- challenges:
  - **estimation** is not the same as **inference**
  - Bayesian network is **computationally** intractable for a large network.
  - **linear** DAG; **independent** data

Our proposal

- what we propose (**in a nutshell**):
  - a **general** (not necessarily linear) DAG with **time dependent** observational data $\{X_{i,t}\}_{i=1,t=1}^{N,T} \in \mathbb{R}^d$
  - a testing procedure that integrates three key deep learning ingredients:
    - a DAG structural learning method based on **multilayer perceptron learner (MLP)** to estimate the DAG
    - a supervised learning method based on **MLP** to estimate the conditional mean
    - a distribution generator produced by **GANs** to approximate the conditional distribution
  - the test statistic is **doubly-robust** when either the conditional mean or the distribution generator is well approximated
  - use **data splitting** and **cross-fitting** to ensure a valid type-I error rate control under minimal conditions on the generators
  - employ **multiplier bootstrap** to compute the p-value
  - show the resulting test achieves a valid control of the type-I error and the power approaches one, **asymptotically**, when either N or T diverges to ∞

Our proposal

- **equivalent hypotheses**:

$$\mathcal{H}_0^*(j,k|\mathcal{M}) : X_k \perp\!\!\!\perp X_j \mid X_{\mathcal{M}-\{k\}} \ \text{vs} \ \mathcal{H}_1^*(j,k|\mathcal{M}) : X_k \not\perp\!\!\!\perp X_j \mid X_{\mathcal{M}-\{k\}}$$

  - for a given set of indices $\mathcal{M} \subseteq \{1, \dots, d\}$ such that $j \notin \mathcal{M}$, $\mathrm{PA}_j \subseteq \mathcal{M}$ and $\mathcal{M} \cap \mathrm{DS}_j = \emptyset$
  - when devising a conditional independence test for $\mathcal{H}_0(j,k)$, the conditioning set $\mathcal{M}$ *should* contain the **parents** of node j, but *cannot* contain any **common descendants** of j, k

- key quantity for test statistic construction:

$$S(j,k|\mathcal{M};h) = \mathbb{E}\left\{X_j - \mathbb{E}\left(X_j|X_{\mathcal{M}-\{k\}}\right)\right\} \times \left[h\left(X_k, X_{\mathcal{M}-\{k\}}\right) - \mathbb{E}\left\{h\left(X_k, X_{\mathcal{M}-\{k\}}\right)|X_{\mathcal{M}-\{k\}}\right\}\right].$$

  - use DAG structural learning (Zheng et al., 2020) to learn the set of indices $\mathcal{M}$
  - use MLP to **estimate the conditional mean** $\widehat{\mathbb{E}}\left(X_j|X_{\mathcal{M}-\{k\}}\right)$
  - use GANs to **learn the conditional distribution** of X_k given $X_{\mathcal{M}-\{k\}}$

Our proposal

- testing procedure:
  - the test statistic S is a maximum over B **transformation functions** for improved power; $B = 2000 \Leftarrow$ **where generative AI kicks in**
  - $\mathbb{H} = \left\{ \cos(\omega X_k), \ \sin(\omega X_k) : \omega \in \mathbb{R} \right\}$
  - data splitting and cross-fitting
  - multiplier bootstrap to compute the p-value

- theoretical guarantees:
  - the sample test statistic $\widehat{S}$ is **doubly robust**
  - $\widehat{S}$ converges at the $\sqrt{n}$**-rate**: it suffices to require $\kappa_1 + \kappa_2 > 1/2$, where κ_1, κ_2 is the convergence rate of the conditional mean estimator and the conditional distribution estimator, respectively
  - our proposed test achieves a **parametric** convergence rate and a parametric power guarantee while using **nonparametric** estimators
  - establish the asymptotic size control and power property, when **either** $N \to \infty$**, or** $T \to \infty$

Numerical analysis

- **brain effective connectivity example revisited**:
  - a task-evoked functional magnetic resonance imaging (fMRI) dataset from the Human Connectome Project (HCP)
  - analyze the fMRI scans of individuals who undertook a story-math task: $N = 28$ individuals with scores below 65 out of 100, and $N = 28$ individuals with the perfect score of 100
  - fMRI data summarized as a matrix of time series, with length $T = 316$, and 264 brain regions, grouped into 14 functional modules
  - focus on $d = 127$ brain regions from 4 functional modules: auditory, visual, frontoparietal task control, and default mode, which are generally believed to be involved in language processing and problem solving domains
  - apply the proposed test to the two datasets separately, with the false discovery control at 0.05

Numerical analysis

- **brain effective connectivity example**:

	Auditory (13)		Default mode (58)		Visual (31)		Fronto-parietal (25)	
	low	high	low	high	low	high	low	high
Auditory (13)	20	17	0	0	0	1	2	0
Default mode (58)	0	0	68	46	3	2	11	23
Visual (31)	0	0	3	2	56	46	0	1
Fronto-parietal (25)	2	1	11	23	0	1	22	27

- identify many more within-module connections than the between-module connections; lending data-driven support
- identify more within-module connections for the **frontoparietal** module for the high-performance subjects $\Leftarrow$ known to be involved in sustained attention, complex problem solving and working memory
- identify fewer within-module connections for the **default mode** module and the **visual** module for the high-performance subjects $\Leftarrow$ known to be more active during passive rest and mind-wandering

Discussion

- concluding remarks:
  - AI / DL methods offer a set of highly **flexible and powerful** tools
  - **how to integrate** those methods properly and effectively into a test with desired theoretical guarantees is highly nontrivial
  - our proposed test achieves a **parametric** convergence rate and a parametric power guarantee while using **nonparametric** estimators
  - provide some examples of **harnessing the power of AI** to address classical statistical problems

- reference:

  - Shi, C., Zhou, Y., and Li, L. (2024). Testing directed acyclic graph via structural, supervised and generative adversarial learning. *Journal of the American Statistical Association*, 119, 1833-1846.

Thank You!