Asymmetric predictability in bivariate causal discovery: an information theoretic approach

Soumik Purkayastha and Peter X.K. Song
Department of Biostatistics, School of Public Health

SCAN ME!

A. INTRODUCTION

- **Discovery of causal relationships** from observational data is a cornerstone of scientific research. **Given bivariate** (X, Y)**, a fundamental question is whether X causes Y or, alternatively, if Y causes X.**

- **Even under many simplifying assumptions:** no confounding, no feedback loops, and no selection bias, a structured investigation of causal relationships in bivariate data is a **notoriously hard problem.**

- In absence of a-priori knowledge, we **investigate statistical patterns to find potential causal directions.** In our new framework, **asymmetry is viewed as a low-dimensional representation of causality.**

- **Most measures mask asymmetry** by implicitly assuming that X and Y are equally dependent on each other, which might be false. **Our framework detects association along with asymmetry in (X, Y).**

- Using **Shannon's information theory framework,** we propose a **causal discovery statistic** that quantifies and estimates "**predictive asymmetry**" in (X, Y) in a computationally **fast and robust manner.**

- Our statistic is called the **Directed Mutual Information (DMI)**. DMI scales the popular **mutual information (MI)** by a factor called the **entropy ratio (ER)**, capturing **predictive asymmetry in (X, Y).**

- We establish **key large-sample properties** of our framework by developing a **new data-splitting inference technique** and evaluate its performance through **simulation studies** and a **real data example**.

B. FORMULATION OF THE DMI FRAMEWORK

Mutual information, denoted by $MI(X, Y)$: for bivariate (X, Y) with joint (marginal) densities f_{XY} (f_X and f_Y), is used to investigate the strength of association between (X, Y), given by:

$$MI(X, Y) = E_{XY}\left[\log\left\{\frac{f_{XY}(X, Y)}{f_X(X)f_Y(Y)}\right\}\right].$$

Joint and marginal entropies: the joint entropy is given by $H(X, Y) = -E_{XY}[\log\{f_{XY}(X, Y)\}]$ and the marginal entropies are similarly given by $H(X)$ and $H(Y)$. The conditional entropy of Y on X is given by $H(Y|X) = H(X, Y) - H(X)$.

$$H(X, Y) = \boxed{H(X|Y) + H(Y)} = \boxed{H(X|Y) + MI(X, Y) + H(Y|X)} = \boxed{H(X) + H(Y|X)}$$
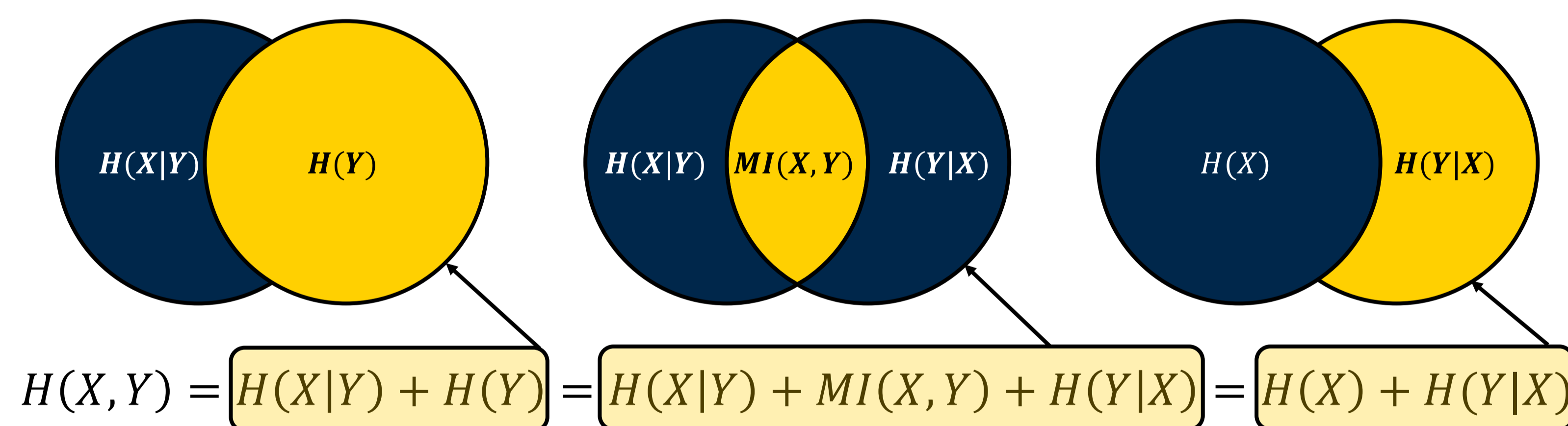
Entropy Ratio, given by $ER(X|Y)$: $H(Y|X)$ measures uncertainty when using X as the predictor and Y as the response. $H(Y|X) < H(X|Y)$ reveals X as the better predictor.

$$ER(X|Y) = \frac{\exp\{H(X|Y)\}}{\exp\{H(X|Y)\} + \exp\{H(Y|X)\}}.$$

$ER(X|Y) > ER(Y|X)$ reveals X as the better predictor than Y.

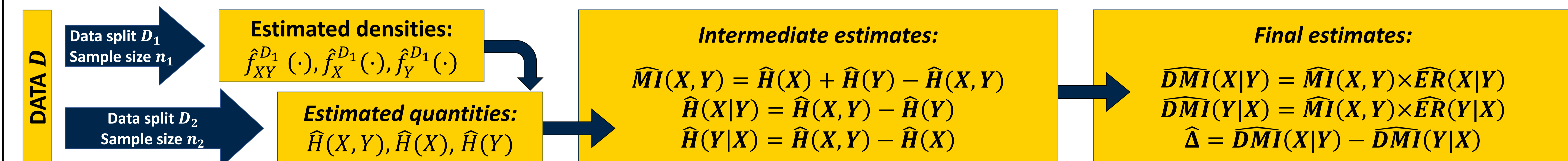**Directed mutual information**: We define $DMI(X|Y) = MI(X, Y) \times ER(X|Y)$ and $\Delta = DMI(X|Y) - DMI(Y|X)$.

- $DMI(X|Y) = 0 = DMI(Y|X) \Longleftrightarrow X$ and Y are independent.
- $\Delta > 0 \Longleftrightarrow ER(X|Y) > ER(Y|X)$. This reveals X as the better predictor than Y.

We use a **Fourier transformation-based method** to estimate the DMI and use **data-splitting for valid statistical inference**.

C. ESTIMATION AND THEORETICAL RESULTS

Estimating densities: Using advances made by *O'Brien et al. (2016)*, we estimate the underlying joint and marginal densities using **Fast-Fourier transformations.** This is **many magnitudes faster** than classical bandwidth-based density estimation, while **maintaining comparable error performance**, making our method **scalable. Estimated densities are used to obtain estimates of DMI and Δ.**

Estimating DMI and Δ: We use a data-splitting technique for estimation and valid inference, shown by the schematic below:

DATA D

Data split D_1 Sample size n_1

Estimated densities: $\hat{f}_{XY}^{D_1}(\cdot), \hat{f}_X^{D_1}(\cdot), \hat{f}_Y^{D_1}(\cdot)$

Data split D_2 Sample size n_2

Estimated quantities: $\hat{H}(X, Y), \hat{H}(X), \hat{H}(Y)$

Intermediate estimates:
$\widehat{MI}(X, Y) = \hat{H}(X) + \hat{H}(Y) - \hat{H}(X, Y)$
$\hat{H}(X|Y) = \hat{H}(X, Y) - \hat{H}(Y)$
$\hat{H}(Y|X) = \hat{H}(X, Y) - \hat{H}(X)$

Final estimates:
$\widehat{DMI}(X|Y) = \widehat{MI}(X, Y) \times \widehat{ER}(X|Y)$
$\widehat{DMI}(Y|X) = \widehat{MI}(X, Y) \times \widehat{ER}(Y|X)$
$\hat{\Delta} = \widehat{DMI}(X|Y) - \widehat{DMI}(Y|X)$

Theoretical results

- **Consistency:** Assuming the density functions are bounded below and above, with $\min(n_1, n_2) \to \infty$, we get **consistent estimates of $DMI(X|Y)$ and $DMI(Y|X)$.**

- **Asymptotic normality:** In addition to the assumptions above, assuming $MI(X, Y) \neq 0$, with $\min(n_1, n_2) \to \infty$, we have
$$\sqrt{n_2}\{\widehat{DMI}(X|Y) - DMI(X|Y)\} \to N(0, \sigma_1^2) \quad and \quad \sqrt{n_2}\{\hat{\Delta} - \Delta\} \to N(0, \sigma_2^2).$$

Both σ_1 and σ_2 may be estimated using standard Monte Carlo tools.

D. DATA APPLICATION: EPIGENETIC CAUSAL DISCOVERY

- We investigate DNA methylation (**DNAm**) and blood pressure (**BP**) in 21 correlated methylation sites of a candidate gene (namely, *ATPB21*) in the *Early Life Exposures in Mexico to Environmental Toxicants* (*ELEMENT*) cohort (Hernandez-Avila et al., 1996).

- As $H_0: DMI = 0$ is a test for independence, we perform a permutation-based test of independence for all 21 methylation sites with Systolic and Diastolic BP (SBP/DBP).

- p-values obtained from *ATPB21* are aggregated using the Cauchy combination test (Liu and Xie, 2019). Overall, **DNAm of *ATP2B1* is associated with DBP (p-value 0.042)** at the 5% level of significance. **One methylation site (*17564205*) is noted to be significantly associated with DBP after applying Bonferroni correction.**

- We examine $DMI(DNAm|DBP) - DMI(DBP|DNAm)$ and note that **DBP exhibits predictive asymmetry over the site *17564205*,** with $\hat{\Delta}(95\% \, CI) = -2.14 \, (-3.85, -0.42)$. Our DMI framework unearths a new causal hypothesis for further investigation.

Association and asymmetric predictability between CpG sites of target gene ATP2B1 and blood pressure in ELEMENT cohort study. Subplots (A1) – (A2) reveal gene-level and CpG site-level association with systolic (SBP) and diastolic blood pressure (DBP) with dotted red line signifying gene-level Bonferroni-adjusted threshold of 5% level of significance. Subplot (B) reveals asymmetric predictability of CpG sites of target gene ATP2B1 that are significantly (Bonferroni-adjusted) associated with SBP or DBP.

(A1) CCT-based combined p-value for association of ATP2B1 gene with DBP: 0.042
(A2) CCT-based combined p-value for association of ATP2B1 gene with SBP: 0.708

CpG site: 17564205 (-log10(p-val): 2.671)

(B) Examining estimated $\hat{\Delta}$ (95% CI) for asymmetric predictability of CpG sites 17564205 in ATP2B1 gene that is significantly associated with diastolic BP.

Asymmetry detected: -2.14 (-3.85, -0.42)

References: (1) O'Brien, T. A., Kashinath, K., Cavanaugh, N. R., Collins, W. D., & O'Brien, J. P. (2016). *A fast and objective multidimensional kernel density estimation method: fastKDE.* **Computational Statistics & Data Analysis.** 101:148–160.
(2) Hernández-Avila M, González-Cossío T, Palazuelos E, Romieu I, Aro A, Fishbein E, et al. (1996). *Dietary and environmental determinants of blood and bone lead levels in lactating postpartum women living in Mexico City.* **Environmental Health Perspectives.** 104:1076–1082.
(3) Liu, Y., & Xie, J. (2019). *Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures.* **Journal of the American Statistical Association.** 115:529:393–402.