(Conditional) mutual information estimation between mixed type variables for general causal network reconstruction

Vincent Cabeli, Nadir Sella, Guido Uguzzoni and Hervé Isambert, Institut Curie UMR168, Paris

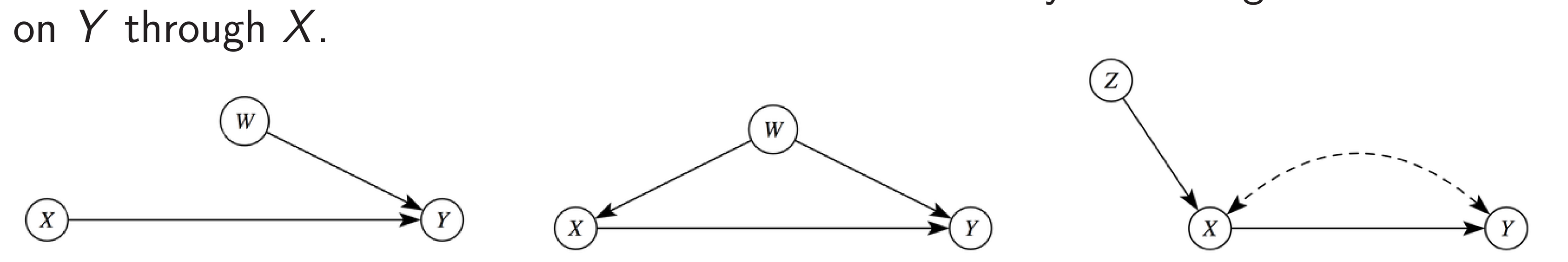
Causal inference from observational data

- Scientific reasoning is often concerned with the underlying causal structure of a data-generating process.
- Many disciplines are ill-equipped to rigorously infer causality (correlation $\neq$ causation). It is possible to gain causal knowledge from purely observational data-set by providing context to observed correlations.
- Causal inference theory offers a way to formally articulate functional assumptions and create causal networks that best fit the observed data.
- We present a novel information-theoretic method (MIIC) which expands constraint-based approaches for causal network reconstruction. MIIC is implemented with a maximum likelihood framework and uses the conditional mutual information as an independence measure to disentangle direct from indirect correlations.

Directed Acyclic Graphs for causal reasoning

A causal network is a Bayesian network with the requirement that the relationships be causal. In other words, if the link $X \rightarrow Y$ exists, then X and Y are conditionally dependent, and intervening on X changes the probability density function of its descendants but not of its ancestors.
Causal networks are powerful tools for reasoning with causality. For example, we can model the assumptions of popular causal inference techniques :

- Randomized Controlled Trials : randomly assign treatment X regardless of other variables W and measure the effects of X on Y directly
- Propensity matching : measure the effects of X on Y taking into accounts the measured confounders W
- Instrumental variables : measure the effect of X on Y by measuring the effects of Z on Y through X.



Our approach tackles the problem of reconstructing the causal network from observations, i.e. find the conditional (in)dependencies and the causality signatures that best fit the data-set.

Constraint-based network reconstruction, MIIC

The mutual information of two random variables X and Y is noted $I(X;Y)$ and it measures the mutual dependence between the two variables.
It is equally sensitive to linear, non-linear or any kind of relationships.
$X \perp\!\!\!\perp Y$ if and only if $I(X;Y)=0$, however for a limited sample size N the mutual information is always >0.
We apply the **minimum description length** principle to derive a complexity term, allowing for robust independence testing in the limited sample case.

General Decomposition of the 2-point information

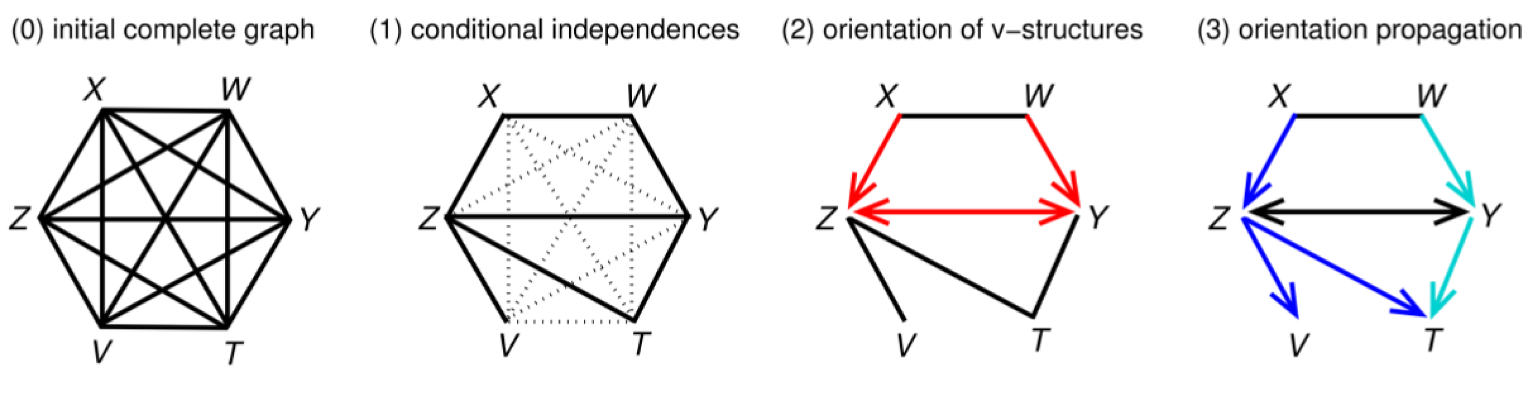
$$\begin{aligned} I(x;y) &= I(x;y;u_1) + I(x;y|u_1) \\ &= I(x;y;u_1) + I(x;y;u_2|u_1) + I(x;y|u_1,u_2) \\ &= I(x;y;u_1) + I(x;y;u_2|u_1) + \dots + I(x;y;z|\{u_i\}) + I(x;y|\{u_i\},z) \end{aligned}$$

MIIC idea

Iteratively **take off positive 3-pt contributions** from **2-pt information**

$$I(x;y|\{u_i\},z) \;=\; I(x;y) - I(x;y;u_1) - I(x;y;u_2|u_1) - \dots - I(x;y;z|\{u_i\})$$

MIIC algorithm



1 **skeleton:**
$\forall\,(x,y)$: if $I(x;y) < k^{MDL}_{x;y}$ **then** $x \not - y$
For top R edge, $\{u_i\} \leftarrow \{u_i\} + z$: **if** $I(x;y|\{u_i\}) < k^{MDL}_{x;y|\{u_i\}}$, **then** $x \not - y$
else search next z for xy and update rank R
until no z can be found $\Rightarrow x - y$

2 **edge filtering:**
$\forall\,(x,y)$: $C_{XY} = \frac{P_{XY}}{\langle P^{\mathrm{rand}}_{XY}\rangle}$ **if** $C_{XY} > \alpha$ **then** $x \not - y$

3 **orientation/propagation:**
$\forall\; x - z - y \;\&\; x \not - y$ sorted by $|I(x;y;z|\{u_i\})| \searrow$
if $I(x;y;z|\{u_i\}) < 0$, $x - z - y \Rightarrow x \rightarrow z \leftarrow y$ (origin of causality)
if $I(x;y;z|\{u_i\}) > 0$, $x \rightarrow z - y \Rightarrow x \rightarrow z \rightarrow y$ (propag. of causality)

Mutual information estimation and independence testing

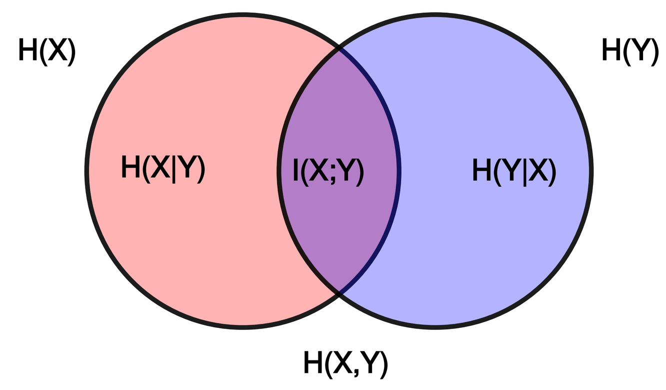
Statistical independence can be formally defined as :

$$X \perp\!\!\!\perp Y \iff P(X,Y) = P(X)P(Y) \iff \mathrm{I}(X;Y) = 0$$

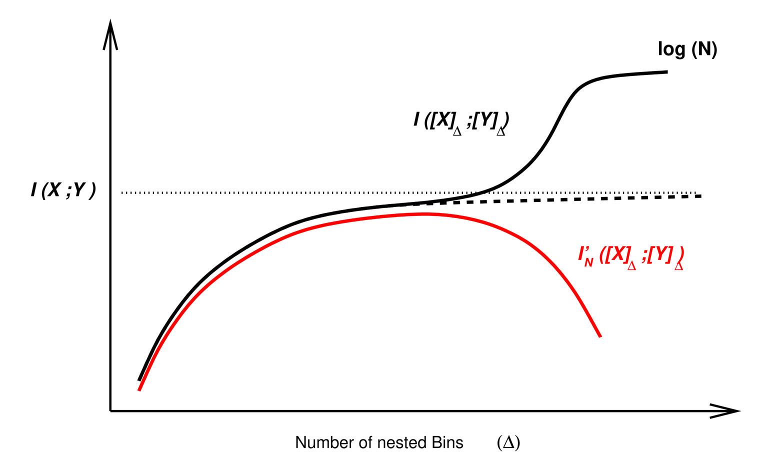
Where $\mathrm{I}(X;Y)$ is the mutual information between X and Y, defined for discrete and continuous variables :

$$\mathrm{I}(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)$$

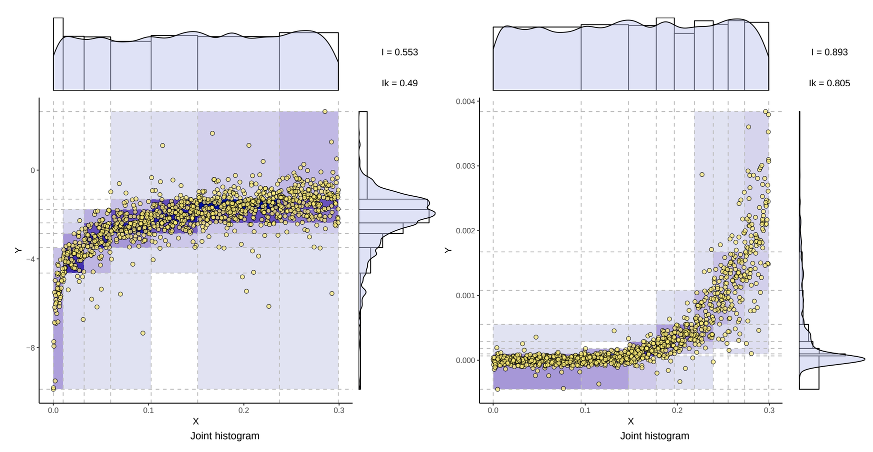
$$\mathrm{I}(X;Y) = \int_{\mathcal{Y}}\int_{\mathcal{X}} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)\,dx\,dy$$



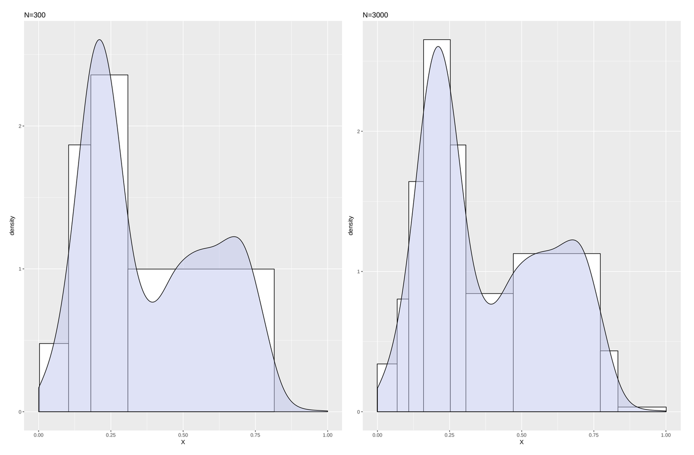
The master definition of mutual information is $\mathrm{I}(X;Y) = \sup_{\mathcal{P},\mathcal{Q}} \mathrm{I}([X]_{\mathcal{P}};[Y]_{\mathcal{Q}})$ where the supremum is over all finite partitions $\mathcal{P}$ and $\mathcal{Q}$. However, refining partitions on finite sample rapidly leads to an overestimation of $\mathrm{I}([X]_\Delta;[Y]_\Delta)$.



In practice we find the optimal cutpoints on X by optimizing $I'(X;Y) = HX + HY - HXY - k^{MDL}_{x;y}$ with a fixed Y, fix X and do the same for Y, repeat until convergence.



MDL-optimal bin count and cutpoint locations in polynomial time (Kontkanen and Myllymäki, 2007)



Note that the optimal cutpoints of a given X depends on the distribution of Y and on their joint distribution XY.
For Conditional mutual information, we estimate several 2-point informations :

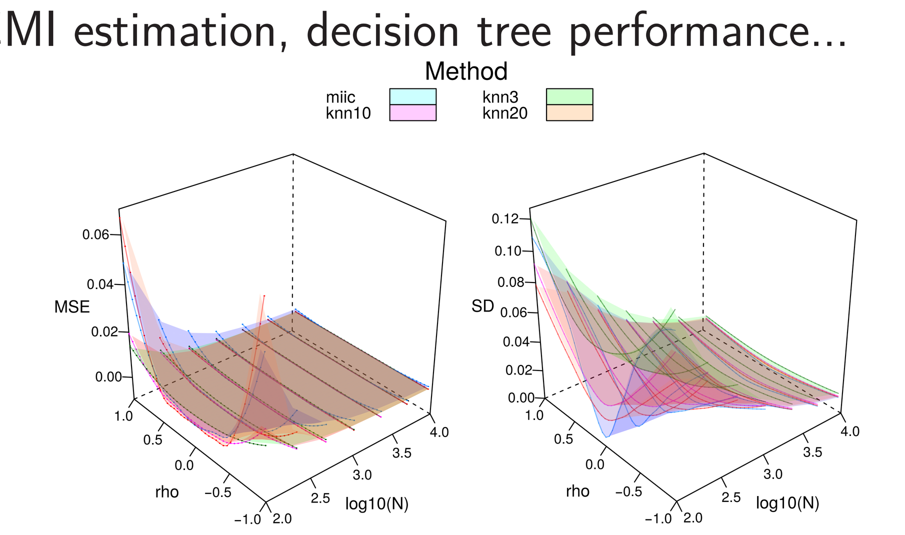
$$I(X;Y|U) = \frac{1}{2}\times\Big(I(X;Y,U) - I(X;U) + I(Y;X,U) - I(Y;U)\Big)$$

$\Rightarrow$ We can generate networks from **discrete**, **continuous** or **mixed** datasets.

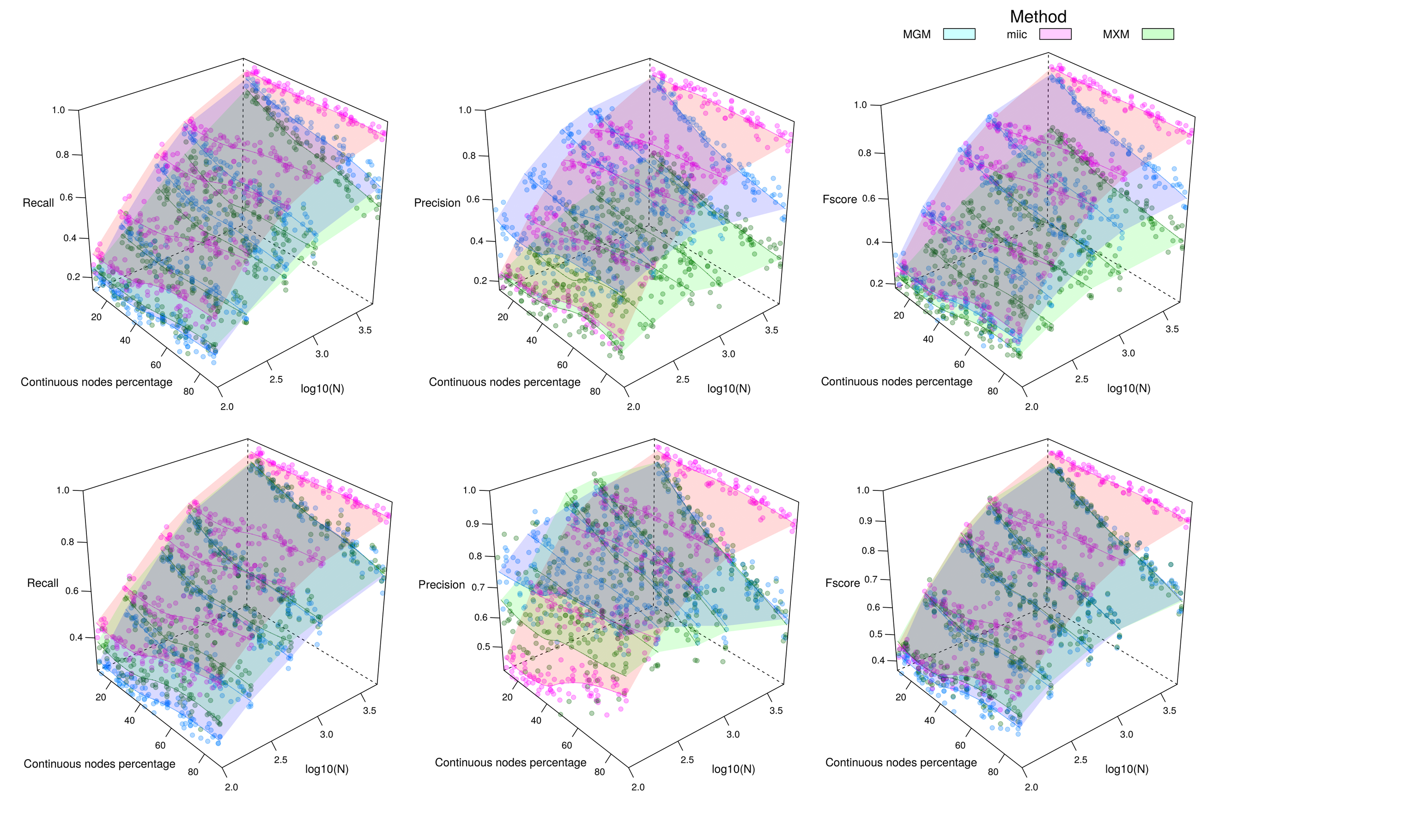
Benchmark results

Discretization benchmarks : correct MI/CMI estimation, decision tree performance...

Our discretization method gives a fair estimation of the mutual information between any two variables and is **parameter-free**. Importantly, it is robust around **independence**.
In the first sense, it is an **independence test in the MDL sense through discretization**.



Simulated network of 100 nodes with varying sample size. F score as a performance measure : $F = 2\cdot\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}$ on skeleton (adjacencies) and oriented CPDAG.



$\Rightarrow$ We report overall better performance on continuous **non-gaussian distributions**, **nonlinear relationships** and **mixed discrete/continuous** networks.

References

- Sella N., Verny L., Uguzzoni G., Affeldt S., Isambert H. *MIIC online: a web server to reconstruct causal or non-causal networks from non-perturbative data*, Bioinformatics, 2017.
- Verny L., Sella N., Affeldt S., Singh PP., Isambert H. *Learning causal networks with latent variables from multivariate information in genomic data*, PLoS Comput. Biol., 2017.

institut Curie SORBONNE UNIVERSITÉ cnrs phy bio