Distribution-free learning of Bayesian network structure

Xiaohai Sun

Max Planck Institute for Biological Cybernetics, Tübingen, Germany

September 16, 2008

MAX-PLANCK-GESELLSCHAFT

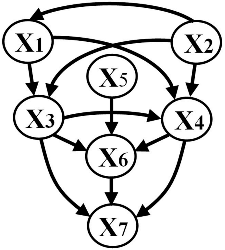
BIOLOGISCHE KYBERNETIK

Learning causal Bayesian network structure

- Purely observational (non-experimental) data on a number of random variables are given
- Estimate direct cause-effect relations between these variables, represented by a directed acyclic graph (causal BN structure)

	X_1	X_2	⋯	X_N
(1)	0	45.2	…	12
(2)	1	1.7	…	62
(3)	1	-3.1	…	90
…	…	…	…	…
(n)	0	19.1	…	22

$\Rightarrow$



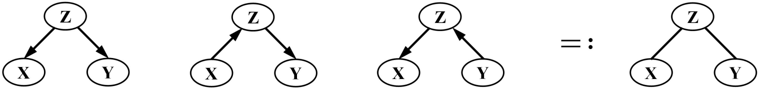
Constraint-based learning

- Based on independence relations
- Weak commitments as to the nature of causal relationships

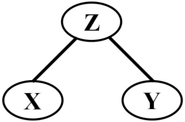
1. Markov assumption states
   "given all its parents, every variable is independent of all its non-descendants".
2. Faithfulness/stability assumption (Spirtes et al. 1993; Pearl 2000) states "*only* the independence relations are true which are implied by the Markov assumption".

From independence constraints to causal BN structure

Identifying $\wedge$-structure by independence constraints



$(X \not\perp\!\!\!\perp Y)$ and $(X \perp\!\!\!\perp Y \mid Z) \Rightarrow \wedge$-structure



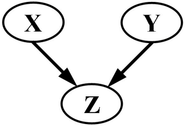
Markov equivalence class of BN structures

$\hookrightarrow$ Learning absence of edges in causal BN structure

$\vee$-structure identification

Markov and faithfulness assumptions lead to a unique graph.

$$(X \perp\!\!\!\perp Y) \text{ and } (X \not\perp\!\!\!\perp Y \mid Z) \Rightarrow \vee\text{-structure}$$



Identification of $\vee$-structure (collider on Z)

$\hookrightarrow$ Learning orientation of edges in causal BN structure

Inductive causation

- Finding conditional independence relations
- Taking Markov and faithfulness assumptions

  1. Learning absence of edges $\hookrightarrow$ skeleton of BN structure
  2. Learning orientation of edges

  $\hookrightarrow$ Inductive causation (IC) algorithm (Pearl 2000)

  Refinement: PC algorithm (Spirtes et al. 1993)

  Using correlation analysis (assumption of normal distribution)

- Our goal:
  non-parametric test of independence on arbitrary domains

Embedding of distributions in RKHS

- $\mathcal{H}_\mathcal{X}$: Hilbert space on measurable space $\mathcal{X}$, spanned by functions $k_\mathcal{X}(x,\cdot)$ $(x \in \mathcal{X})$ with $\langle k_\mathcal{X}(x,\cdot), k_\mathcal{X}(x',\cdot)\rangle = k_\mathcal{X}(x,x')\ \forall x, x' \in \mathcal{X}$.
  X: random variable on $\mathcal{X}$.

- Mean element in RKHS:
  $\mathfrak{M}_X = \mathrm{E}[k_\mathcal{X}(X,\cdot)]$ and $\mathfrak{M}_{XY} = \mathrm{E}[k_\mathcal{X}(X,\cdot)k_\mathcal{Y}(Y,\cdot)]$
- Conditional mean element in RKHS:
  $\mathfrak{M}_{X|Y} = \mathrm{E}[k_\mathcal{X}(X,\cdot)|Y]$ and $\mathfrak{M}_{XY|Z} = \mathrm{E}[k_\mathcal{X}(X,\cdot)k_\mathcal{Y}(Y,\cdot)|Z]$
- Product of mean elements in RKHS:
  $\mathfrak{M}_X\mathfrak{M}_Y = \mathfrak{M}_X \otimes \mathfrak{M}_Y = \mathrm{E}[k_\mathcal{X}(X,\cdot)]\mathrm{E}[k_\mathcal{Y}(Y,\cdot)]$
- Product of conditional mean elements in RKHS:
  $\mathfrak{M}_{X|Z}\mathfrak{M}_{Y|Z} = \mathfrak{M}_{X|Z} \otimes \mathfrak{M}_{Y|Z} = \mathrm{E}[k_\mathcal{X}(X,\cdot)|Z]\mathrm{E}[k_\mathcal{Y}(Y,\cdot)|Z]$

Cross-covariance operator

- Cross-covariance operator in RKHS:

$$\begin{aligned}\langle g, \Sigma_{YX} f\rangle_{\mathcal{H}_\mathcal{Y}} &:= \langle \mathfrak{M}_{XY} - \mathfrak{M}_X \mathfrak{M}_Y, f \otimes g\rangle_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}} \\ &= \mathrm{E}[f(X)g(Y)] - \mathrm{E}[f(X)]\mathrm{E}[g(Y)] \\ &= \mathrm{Cov}[f(X), g(Y)] \quad \forall f \in \mathcal{H}_\mathcal{X}, g \in \mathcal{H}_\mathcal{Y}\end{aligned}$$

- Conditional cross-covariance operator in RKHS:

$$\begin{aligned}\left\langle g, \Sigma_{YX|Z} f\right\rangle_{\mathcal{H}_\mathcal{Y}} &:= \left\langle \mathfrak{M}_{XY} - \mathrm{E}_Z[\mathfrak{M}_{X|Z}\mathfrak{M}_{Y|Z}], f \otimes g\right\rangle_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}} \\ &= \mathrm{E}_{XY}[f(X)g(Y)] - \mathrm{E}_Z[\mathrm{E}[f(X)|Z]\mathrm{E}[g(Y)|Z]] \\ &= \mathrm{E}_Z[\mathrm{Cov}[f(X), g(Y) \,|\, Z]] \quad \forall f \in \mathcal{H}_\mathcal{X}, g \in \mathcal{H}_\mathcal{Y}\end{aligned}$$

HS norm of operator and MMD

- Hilbert-Schmidt (HS) norm of operator Σ: $\mathcal{H}_{\mathcal{X}} \to \mathcal{H}_{\mathcal{Y}}$:

$$\|\Sigma\|_{\mathrm{HS}}^2 = \mathrm{Tr}\left(\Sigma^{\mathsf{T}}\Sigma\right) = \sum_{i,j=1}^{\infty} \langle \varphi_j, \Sigma\phi_i \rangle_{\mathcal{H}_{\mathcal{Y}}}^2 ,$$

$\{\phi_i\}_{i=1}^{\infty}, \{\varphi_j\}_{j=1}^{\infty}$: complete orthonormal systems of $\mathcal{H}_{\mathcal{X}}, \mathcal{H}_{\mathcal{Y}}$.

- kernel Maximum Mean Discrepancy (MMD)

(Borgwardt et al. Bioinformatics 2006)

$$\mathbb{D}_{\mathcal{H}}(\mathcal{P}, \mathcal{Q}) = \sup_{f \in \mathcal{F}} \mathrm{E}_{x \sim \mathcal{P}}[f(x)] - \mathrm{E}_{y \sim \mathcal{Q}}[f(y)] .$$

$\mathcal{P}, \mathcal{Q}$: probability measures. $\mathcal{F}$: unit ball in RKHS $\mathcal{H}$.

$\hookrightarrow \quad \mathcal{P} = \mathcal{Q} \iff \mathbb{D}_{\mathcal{H}}(\mathcal{P}, \mathcal{Q}) = 0$ ($\mathcal{H}$: characteristic RKHS).

(Fukumizu et al. NIPS 2007, 2008; Sriperumbudur et al. COLT 2008)

Unconditional independence with kernel

- HS norm of cross-covariance operator Σ_{YX} corresponds to MMD between P_{xy} and $P_x P_y$

$$\begin{aligned} \|\Sigma_{YX}\|^2_{\mathrm{HS}} &= \langle \mathfrak{M}_{XY} - \mathfrak{M}_X \mathfrak{M}_Y, \mathfrak{M}_{XY} - \mathfrak{M}_X \mathfrak{M}_Y \rangle \\ &= \|\mathfrak{M}_{XY} - \mathfrak{M}_X \mathfrak{M}_Y\|^2_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}} \\ &= \mathbb{D}^2_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}}(P_{xy}, P_x P_y)\,. \end{aligned}$$

- Given characteristic RKHS

$$\begin{aligned} \Sigma_{YX} = O \quad &\Longleftrightarrow \quad \|\Sigma_{YX}\|^2_{\mathrm{HS}} = 0 \\ &\Longleftrightarrow \quad \mathfrak{M}_{XY} = \mathfrak{M}_X \mathfrak{M}_Y \\ &\Longleftrightarrow \quad P_{xy} = P_x P_y \\ &\Longleftrightarrow \quad X \perp\!\!\!\perp Y\,. \end{aligned}$$

Conditional cross-covariance operator and MMD

- HS norm of conditional cross-covariance operator $\Sigma_{YX|Z}$ corresponds to MMD between P_{xy} and $\mathrm{E}_z[P_{x|z}P_{y|z}]$

$$
\begin{aligned}
\|\Sigma_{YX|Z}\|^2_{\mathrm{HS}} &= \left\langle \mathfrak{M}_{XY} - \mathrm{E}_Z[\mathfrak{M}_{X|Z}\mathfrak{M}_{Y|Z}],\, \mathfrak{M}_{XY} - \mathrm{E}_Z[\mathfrak{M}_{X|Z}\mathfrak{M}_{Y|Z}] \right\rangle \\
&= \|\mathfrak{M}_{XY} - \mathrm{E}_Z[\mathfrak{M}_{X|Z}\mathfrak{M}_{Y|Z}]\|^2_{\mathcal{H}_\mathcal{X}\otimes\mathcal{H}_\mathcal{Y}} \\
&= \mathbb{D}^2_{\mathcal{H}_\mathcal{X}\otimes\mathcal{H}_\mathcal{Y}}\left(P_{xy}, \mathrm{E}_z\left[P_{x|z}P_{y|z}\right]\right) .
\end{aligned}
$$

- Given characteristic RKHS

$$
\begin{aligned}
\Sigma_{YX|Z} = O \quad &\Longleftrightarrow \quad \|\Sigma_{YX|Z}\|^2_{\mathrm{HS}} = 0 \\
&\Longleftrightarrow \quad \mathfrak{M}_{XY} = \mathrm{E}_Z[\mathfrak{M}_{X|Z}\mathfrak{M}_{Y|Z}] \\
&\Longleftrightarrow \quad P_{xy} = \mathrm{E}_z[P_{x|z}P_{y|z}] \\
&\underset{\color{red}\not\Rightarrow}{\Leftarrow} \quad X \perp\!\!\!\perp Y \mid Z .
\end{aligned}
$$

Conditional independence with kernel

Define $\dot{X} := (X, Z)$, $\dot{Y} := (Y, Z)$

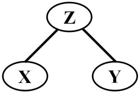
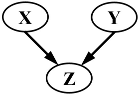
- HS norm / MMD

$$\begin{aligned}\|\Sigma_{\dot{Y}\dot{X}|Z}\|^2_{\mathrm{HS}} &= \mathrm{E}_Z\left[\|\mathfrak{M}_{\dot{X}\dot{Y}|Z} - \mathfrak{M}_{\dot{X}|Z}\mathfrak{M}_{\dot{Y}|Z}\|^2_{\mathcal{H}_{\dot{\mathcal{X}}}\otimes\mathcal{H}_{\dot{\mathcal{Y}}}}\right] \\ &= \mathrm{E}_Z\left[\mathbb{D}^2_{\mathcal{H}_{\dot{\mathcal{X}}}\otimes\mathcal{H}_{\dot{\mathcal{Y}}}}\left(P_{\dot{x}\dot{y}|z}, P_{\dot{x}|z}P_{\dot{y}|z}\right)\right].\end{aligned}$$

- Given characteristic RKHS

$$\begin{aligned}\Sigma_{\dot{Y}\dot{X}|Z} = O \quad &\Longleftrightarrow \quad \|\Sigma_{\dot{Y}\dot{X}|Z}\|^2_{\mathrm{HS}} = 0 \\ &\Longleftrightarrow \quad \mathfrak{M}_{XY|Z} = \mathrm{E}_Z[\mathfrak{M}_{X|Z}\mathfrak{M}_{Y|Z}] \quad \forall Z \\ &\Longleftrightarrow \quad P_{xy|z} = \mathrm{E}_z[P_{x|z}P_{y|z}] \quad \forall z \\ &\Longleftrightarrow \quad X \perp\!\!\!\perp Y \mid Z.\end{aligned}$$

Constraint-based learning of BN structure

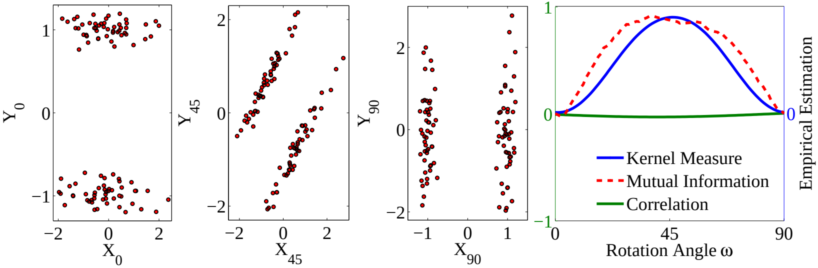
Constraints	$\wedge$-Structure 	$\vee$-Structure 
Independence relations	$X \not\perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y \mid Z$	$X \perp\!\!\!\perp Y$ and $X \not\perp\!\!\!\perp Y \mid Z$
Joint distributions	$P_{xy} \neq P_x P_y$ $P_{xy\mid z} = P_{x\mid z} P_{y\mid z}$	$P_{xy} = P_x P_y$ $P_{xy\mid z} \neq P_{x\mid z} P_{y\mid z}$
Mean element in RKHS	$\mathfrak{M}_{XY} \neq \mathfrak{M}_X \mathfrak{M}_Y$ $\mathfrak{M}_{\dot{X}\dot{Y}} = \mathrm{E}_Z\left[\mathfrak{M}_{\dot{X}\mid Z} \mathfrak{M}_{\dot{Y}\mid Z}\right]$	$\mathfrak{M}_{XY} = \mathfrak{M}_X \mathfrak{M}_Y$ $\mathfrak{M}_{\dot{X}\dot{Y}} \neq \mathrm{E}_Z\left[\mathfrak{M}_{\dot{X}\mid Z} \mathfrak{M}_{\dot{Y}\mid Z}\right]$
HS norm of operators (MMD in RKHS)	$\lVert\Sigma_{XY}\rVert_{\mathrm{HS}}^2 > 0$ $\lVert\Sigma_{\dot{X}\dot{Y}\mid Z}\rVert_{\mathrm{HS}}^2 = 0$	$\lVert\Sigma_{XY}\rVert_{\mathrm{HS}}^2 = 0$ $\lVert\Sigma_{\dot{X}\dot{Y}\mid Z}\rVert_{\mathrm{HS}}^2 > 0$

Alternative measures of dependences

- Correlation coefficient (assumption of normal distribution)
- Mutual information (Kraskov et al. 2004)
  (based on entropy estimates from k-nearest neighbor distances)

$Y_0 \propto \frac{1}{2}\mathcal{N}(1,0.01)+\frac{1}{2}\mathcal{N}(-1,0.01), P(X_0|Y_0<0)\propto\mathcal{N}(0,1), P(X_0|Y_0\geq 0)\propto\mathcal{N}(0,1)$

(X_ω, Y_ω): transformed data with rotation ω in an anticlockwise direction



Summary

- Kernel test of independence for constraint-based learning of causal BN structure

Further issues:

- Connection to mutual information? (Gretton et al. 2005)
- Useful MMD with higher-order tensors?
  e.g., vanishing of

$$\|\mathfrak{M}_{XYZ} - \mathfrak{M}_X \mathfrak{M}_Y \mathfrak{M}_Z\|^2_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Z}} = \mathbb{D}^2_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{Y}}(P_{xyz}, P_x P_y P_z)\,,$$

  indicates mutual independence (more than pairwise independence).

Thanks for your attention!