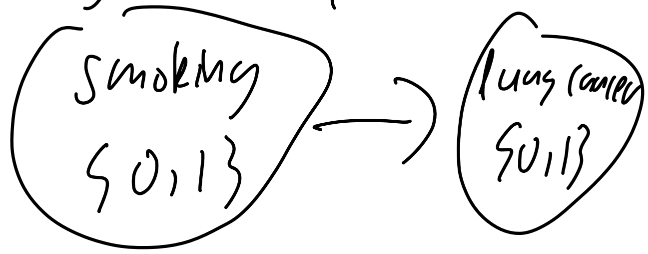
Causal Structure Discovery

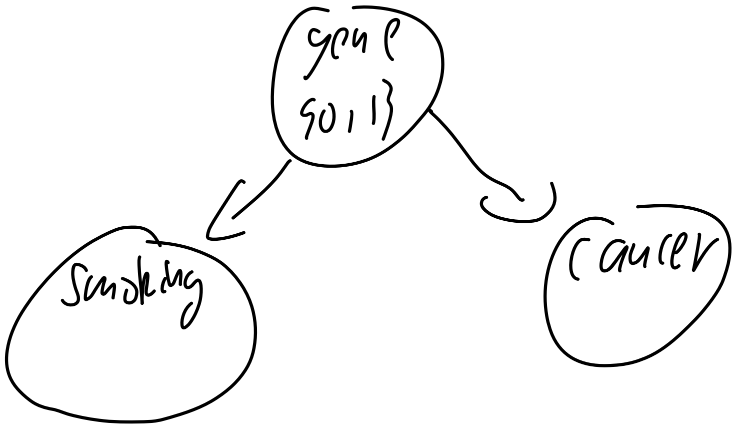
- Does smoking cause lung cancer?



observation data

		cancer Yes	No
smoking	Yes	15%	85%
	No	6%	94%

correlation does not imply causality

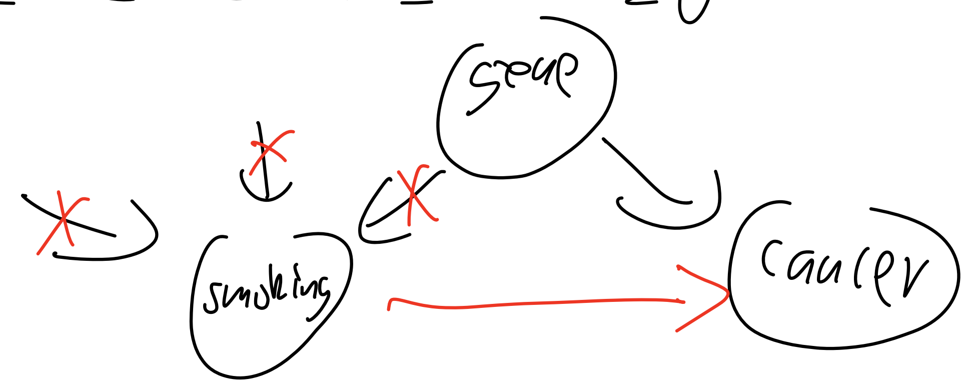


Interventional Data

Random trials

randomly pick 50% with smoking=1

— — — — — — — — — =0



Assumption: All nodes are observed with observation data

* we have infinite data

Recall: BN $G=(V,E)$ is a DAG

$$P(X) = \prod_{i=1}^{n} P_i(X_i \mid X_{\pi_i})$$

Def: Markov Equivalence class (MEC)

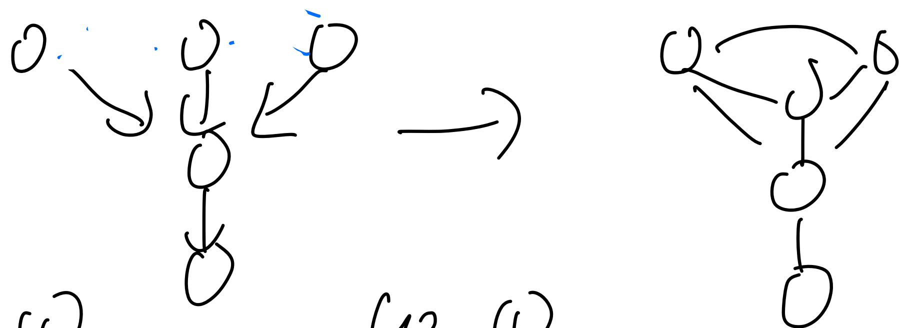
$G_1 \sim G_2 \iff \begin{cases} \text{skeleton is the same} \\ \text{Moral graph is the same} \end{cases}$

Def Moral graph of a graph G:

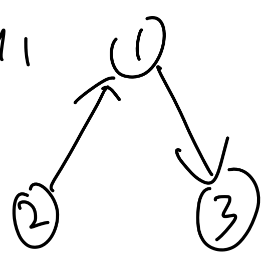
add edges between all pairs of non-adjacent nodes with a common child

$\Rightarrow$ undirected graph

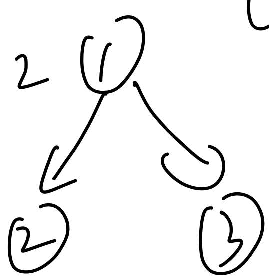
G:



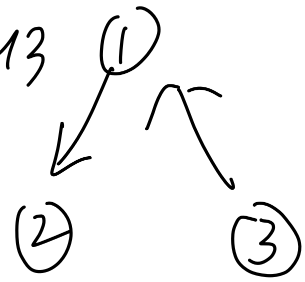
G_1



G_2



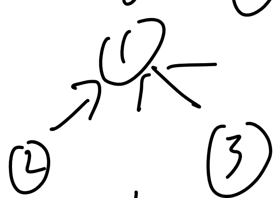
G_3



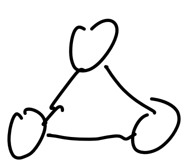
Moral graph:

claim: From observation data, we can only recover G up to its MEC

G_4



Moral graph:



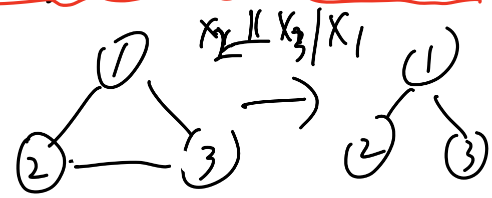
Constraint-based Algorithm [SGS-Algorithm 2001]

Step 1: start with a complete $G=(V,E)$ undirected

Step 2: using observation data, for $(i,j) \in V \times V$ remove (i,j) from E, if $\exists\, S \subset V$ s.t.

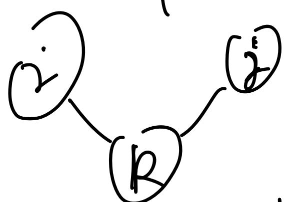
skeleton identification

$$X_i \perp\!\!\!\perp X_j \mid X_S$$



Step 3: for all triplets $(i,j,k) \in V \times V \times V$

edge orientation

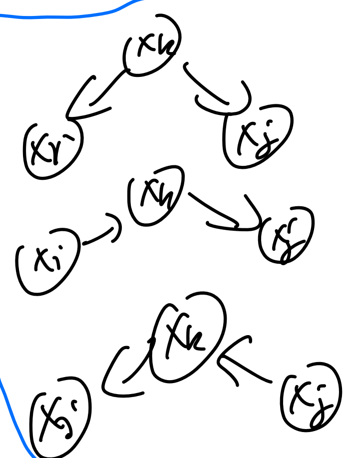


check if $X_i \perp\!\!\!\perp X_j \mid X_{rest} \setminus \{X_k\}$

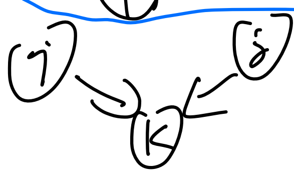
if $X_i \perp\!\!\!\perp X_j \mid X_k \longrightarrow$

$X_i \not\!\perp\!\!\!\perp X_j \mid X_k$

$\Downarrow$



direct edges as



Step 4 orient remaining undirected edges by consistency (BN is DAG)

Claim: If $P(X)$ is faithful w.r.t. G then SGS algorithm: $\lim_{N \to \infty} P(\hat{G}_{SGS} \neq G^*) \to 0$ wrt MEC

Problem: need to decide conditional independence

limited data $\rightarrow$ inaccurate

$\hat{I}(X_i, X_j \mid X_k)$

Score-based Algorithm

Recall: log-likelihood score of a DAG G given $x^{(1)}, \dots, x^{(N)}$

$$\text{score}(G) = N \cdot \underbrace{\sum_{i=1}^{n} I_{\hat{p}}(X_i; X_{\pi_i})}_{\text{depend on } G} - \underbrace{N \sum_{i=1}^{n} H_{\hat{p}}(X_i)}_{\text{does not depend on } G}$$

if we want to maximize score w.r.t. G

score $\uparrow$ if G is larger (more edges)

Def: Bayesian Information Criterion (BIC) Score

$$\text{score}_{BIC}(G) \triangleq \text{score}(G) - \frac{\log N}{2} \cdot \dim(G)$$

where $\dim(G) = \sum_{i=1}^{n} (|\mathcal{X}|-1) \cdot |\mathcal{X}|^{|\pi_i|}$, $|\mathcal{X}|$: # of possible values

follows Minimum Description Length (MDL) of each node

score $(G) \sim N$

penalty $\sim \log N$ $\Rightarrow$ $\dim(G) \lesssim \frac{N}{\log N}$

Properties: ① Score equivalent:

$$G_1 \sim G_2 \iff \text{Score}_{BIC}(G_1) = \text{Score}_{BIC}(G_2)$$

② Consistency: If G^* is PG for $P(X)$ then $N \to \infty$, G^* is the unique maximizer up to MEC

③ Decomposable

$$\text{Score}_{BIC} = \sum_{i=1}^{n} \widetilde{\text{Score}} (X_i, X_{\pi_i})$$

Algorithm [Greedy Equivalence Search] (GES)

Initialize empty graph $G^{(1)} = (V, E = \phi)$

Phase I: $t = 1, \dots, T$ (till until no improvement)

add an edge that maximizes $\text{Score}_{BIC}(G^{(t+1)})$

Phase II: $t = T+1, \dots$ (till no improvement)

remove an edge that maximizes $\text{Score}_{BIC}(G^{(t+1)})$

Claims (2003): As $N \to \infty$ GES correctly finds MEC under faithfulness.

* Permutation-based Greedy Search Algorithm

Initialize: $\pi^{(1)}$ as arbitrary ordering

$$(1,2,3,4) \to 1243, \quad 1324, \quad 2134$$

Repeat $t=1, \dots$

for each permutation π in the neighborhood of $\pi^{(t)}$

neighbor of a permutation: diff only in two adjacent positions

2, 5, 3, 1, 4

2, 3, 5, 1, 4

construct G_π by

$(\pi_i, \pi_j) \in E_\pi$,

if $X_{\pi_i} \not\perp X_{\pi_j} \mid X_{\pi_1}, \dots \pi_{i-1}, \pi_{i+1} \dots \pi_{j-1}$,

Evaluate $\text{Score}_{BIC}(G_\pi)$

$\pi^{(t+1)} \leftarrow \pi$ with the highest Score_{BIC}

Claim: if $n \to \infty$, permutation-based also finds G^* up to MEC.