統計的機械学習　（応用計量分析２）第8回

因果探索（参考pdf 11章）

振り返り

出欠コード
XXXX

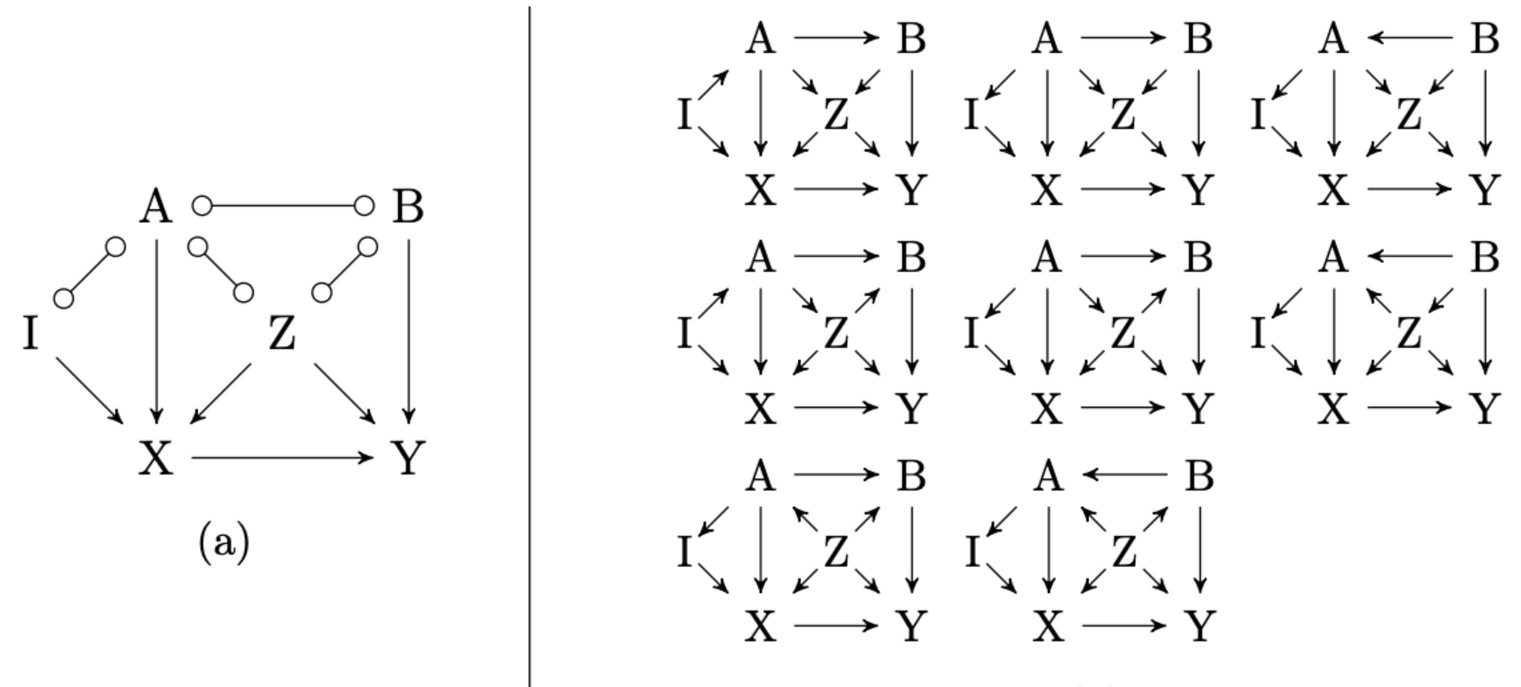
SCMは潜在結果モデルとは異なるもう1つの因果の定式化
因果構造がわかればバックドア基準により調整すべき変数がわかる

- 構造的因果モデルは変数ごとのノイズと関数で変数間の因果関係を表す
- 介入は因果構造および構造方程式の局所的な変更と表せる
- 因果の合流点で条件づけると非因果的相関が生まれる（選択バイアス）
- 調整すべき変数集合を選択するためバックドア基準を用いる
  - 介入対象変数aに入ってくる矢印を含む結果yとの間のパスを遮断するように調整する変数集合Zを選択
  - 合流点をZに含めるとその上流変数間に非因果的相関の経路（バックドアパス）が新たに開くことに注意

DAGの拡張も含めDAGittyは調整変数集合を出力

- Complete Partial DAG (**CPDAG**)
  - マルコフ性のみでデータから検証可能なDAGの候補集合（マルコフ同値類）を表すグラフ
  - 同値類に含まれる矢印ヘッド・矢印テールが複数あるとき ○ で表す
  - 図(a)の X→Y では、とにかく {A, Z} を調整すればよい
  - ということもDAGittyでわかる
- Mixed Ancestral Graph (MAG)
  - MAGでは、因果の向きが特定されない場合、両端矢印ヘッド ←→ を用いる
  - DAGが隠れ変数を含むとき、観測変数のみのMAGが存在する
- Partial Ancestral Graph (PAG)
  - MAGのマルコフ同値類を表すグラフ
  - 同値類に含まれる矢印ヘッド・矢印テールが複数あるとき ○ で表す

CPDAG　　含まれる候補DAG



(a)　　(b)

	DAG	MAG	CPDAG	PAG
Back-door criterion Pearl (1993)	⇒			
Adjustment criterion Shpitser et al. (2010), Shpitser (2012)	⇔			
Adjustment criterion van der Zander et al. (2014)	⇔	⇔		
Generalized back-door criterion Maathuis and Colombo (2015)	⇒	⇒	⇒	⇒
Generalized adjustment criterion Perković et al. (2015)	⇔	⇔	⇔	⇔

Table 1: Graphical criteria for covariate adjustment: ⇒ - sound, ⇔ - sound and complete.

本日の内容

出欠コード
XXXX

因果構造（DAG）の推定方法を学ぶ

- 1. ガイダンス・因果推論と機械学習の概論
- 2. 意思決定理論（期待効用理論）の復習、因果推論との関係
- 3. 潜在結果モデルに基づく因果推論の枠組み
- 4. 平均因果効果の推定法
- 5. 条件付き平均因果効果（CATE）の推定法１：メタ学習器
  - CATEの推定法２：二重機械学習
- 6. CATEの推定法３：決定木と決定森
  - 深層学習に基づく方法
- 7. 構造方程式モデルとバックドア基準
- **8. 因果探索**
- 9. 発展的な因果推論手法：フロントドア調整、操作変数法、回帰不連続デザイン、代理変数法
- 10. 続き
- 11. 発展的な意思決定理論
- 12. 強化学習
- 13. オフライン強化学習
- 14. バンディット
- 15. まとめ

因果構造の推定法：因果探索

因果構造を推定する問題＝因果探索
仮定の置き方と得られる結論の強さにより幾つかのアプローチ

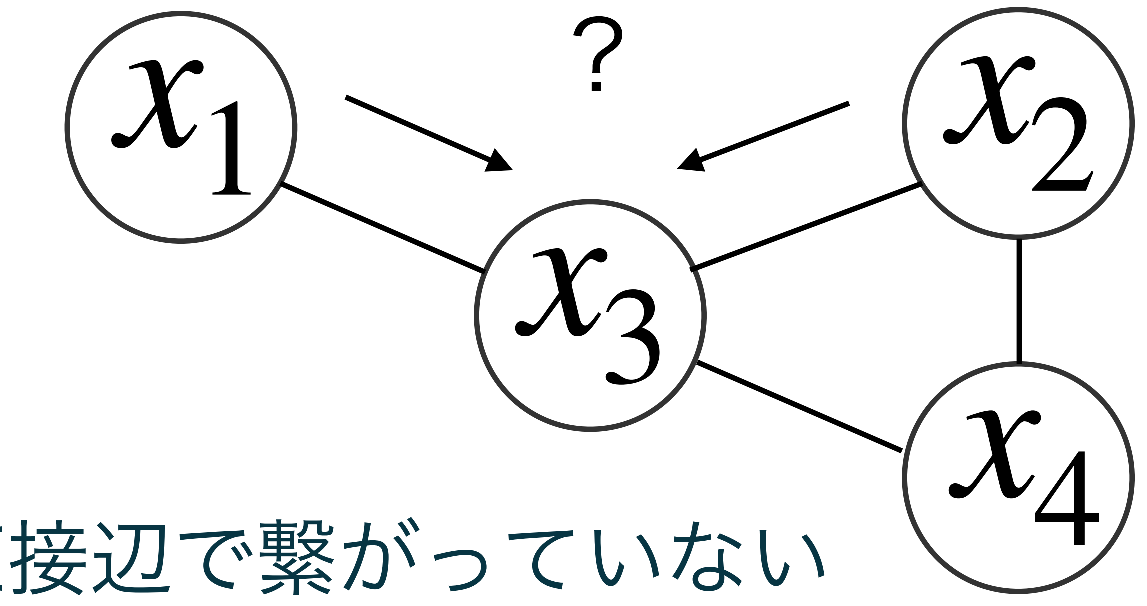
- 複雑な現象で、因果構造がそもそも分かっていない場合にデータから構造を推定する問題を**因果探索**（Causal Discovery）という
- 仮定の置き方によって幾つかのアプローチが存在
  - 制約ベース（ノンパラメトリック）：モデルは仮定せず独立性のみに基づく
  - セミパラメトリック：構造方程式 f のモデルを一部仮定
    - 主に線形性を仮定、ノイズ分布は未知でもよい
  - スコアベース（パラメトリック）：構造方程式 f のモデルをノイズ含めて仮定し、尤度を含む損失関数を最小化

独立性制約に基づくノンパラメトリック法

出欠コード
XXXX

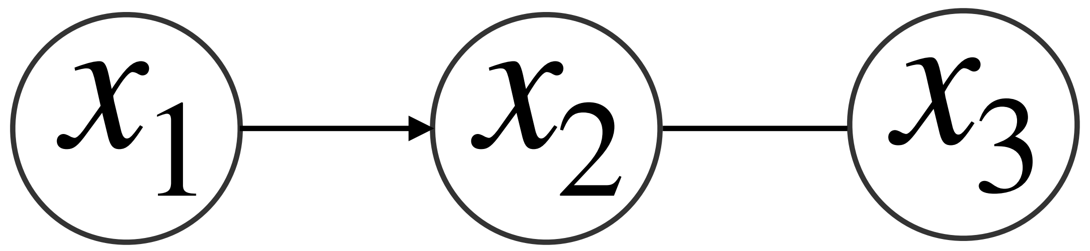
IC法（PC法）は（条件付き）独立性検定のみから可能な限りの構造と向きを特定

- 条件付き独立性検定により向き無しグラフを推定
  - 各変数の組 a, b に対し，条件付き独立 $a \perp\!\!\!\perp b | S$ となる
    変数集合 S （空集合∅を含む） が**存在しない**場合のみa-b間に辺を描く
  - すべての変数間に辺がある完全グラフから始めて、独立性を見つけたら辺を削除
- V構造を全て検出して向き付ける
  - x_3 （を含む任意の変数集合）で条件付けると x_1 と x_2 が従属 ⇒ x_3 は合流点
- 追加で向きを判定するルールを適用（オリエンテーションルール）
- 向きが分からない辺もありうる（CPDAGが出力）

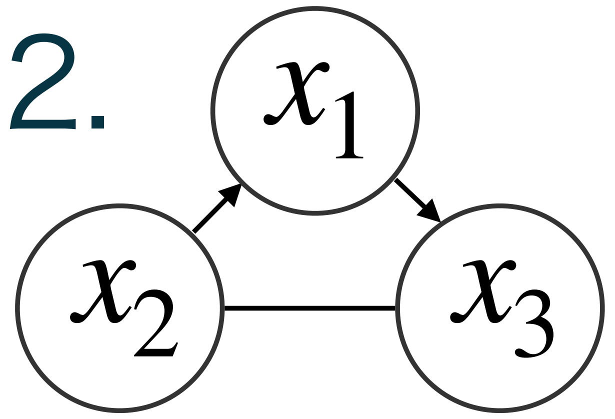


V構造：直接辺で繋がっていない
２変数が合流点を介してつながっている構造

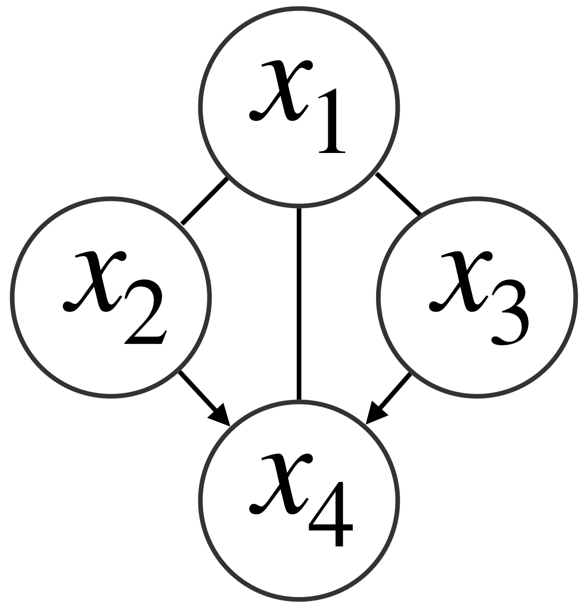
1.



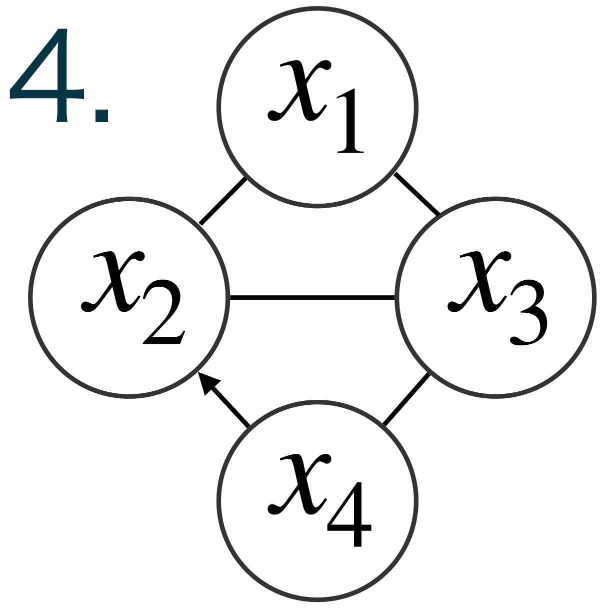
2.



3.



4.



オリエンテーションルール

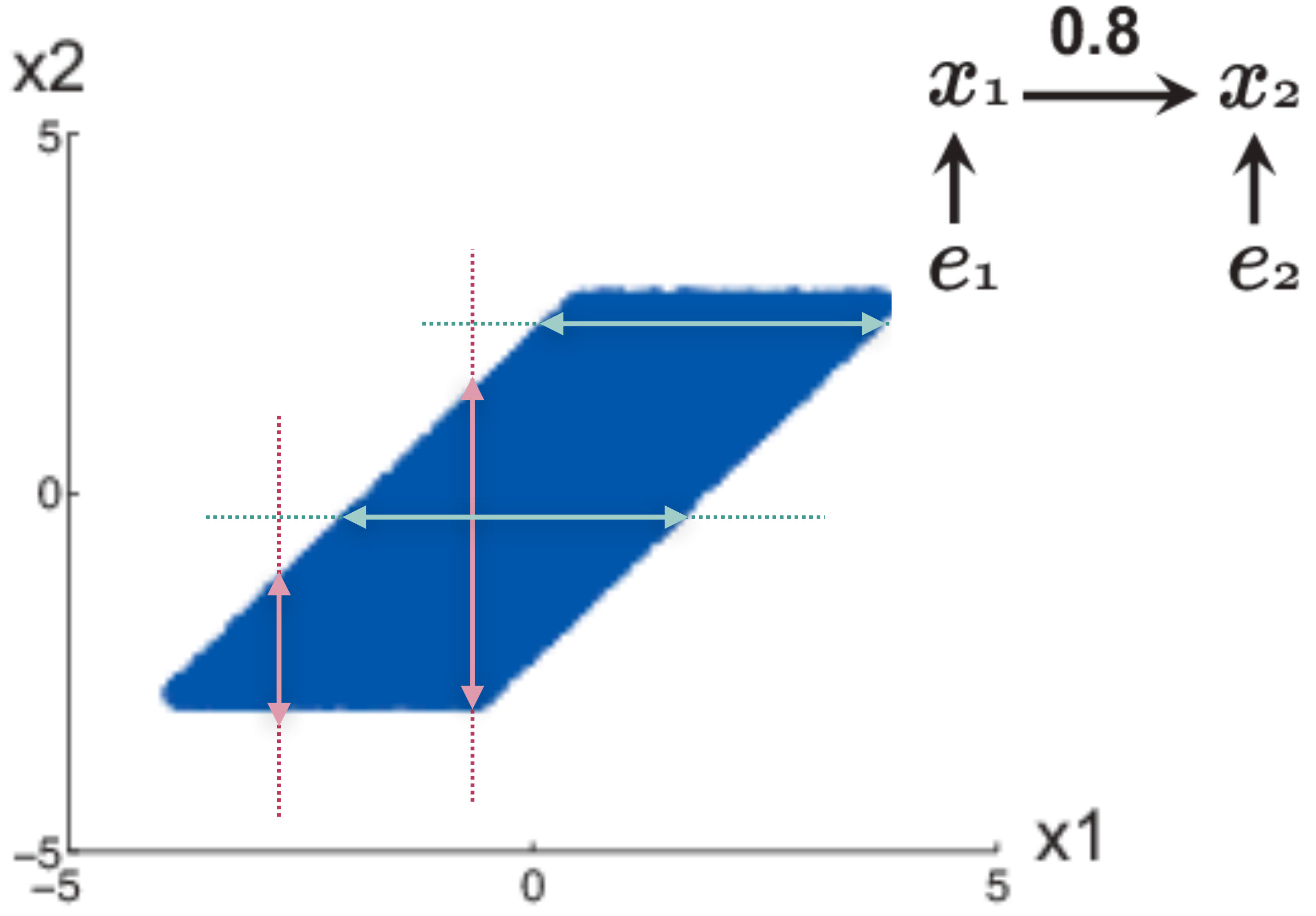
1. x_2は合流点ではないので $x_2 \to x_3$
2. DAG制約から$x_2 \to x_3$
3. $x_4 \to x_1$と仮定するとDAG制約から$x_2 \to x_1$かつ$x_3 \to x_1$となるはず、するとx$_1$は(x_2, x_3)のV構造で矛盾するので$x_1 \to x_4$
4. ルール1から$x_2 \to x_1$、さらに$x_1 \to x_3$を仮定するとルール1から$x_3 \to x_4$となるが非巡回性と矛盾するので$x_3 \to x_1$

モデルに基づく(セミ)パラメトリック法

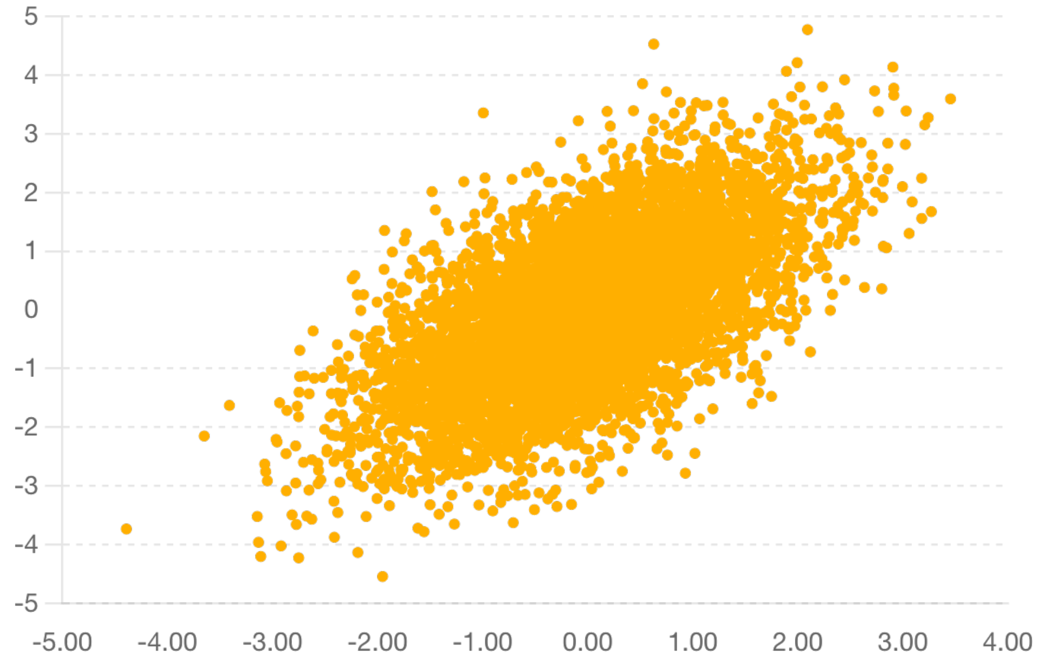
LiNGAMはモデルの線形性とノイズの非ガウス性を仮定
仮定が正しければすべての辺の向きを識別できる

- 線形非ガウス非巡回モデル (LiNGAM)
  - 線形性、ノイズの加法性、非ガウス性を仮定
  - $x_j = f_j(\mathrm{pa}_j, \varepsilon_j) = \beta_j^\top \mathrm{pa}_j + \varepsilon_j$

    まとめると $x = Bx + \varepsilon$
    - 非巡回性⇔Bは並べ替えると厳密三角行列
- ⇒ $(I - B)x = \varepsilon$
  - 上式のεが相互に独立となるような非巡回のBを見つければよい
  - 右図のように向きが違う場合は説明変数側とノイズ（残差）が独立にならない（ガウス分布の場合以外）



Shimizu, Shohei. LiNGAM: Non-Gaussian methods for estimating causal structures. *Behaviormetrika* 41.1 (2014): 65-98.

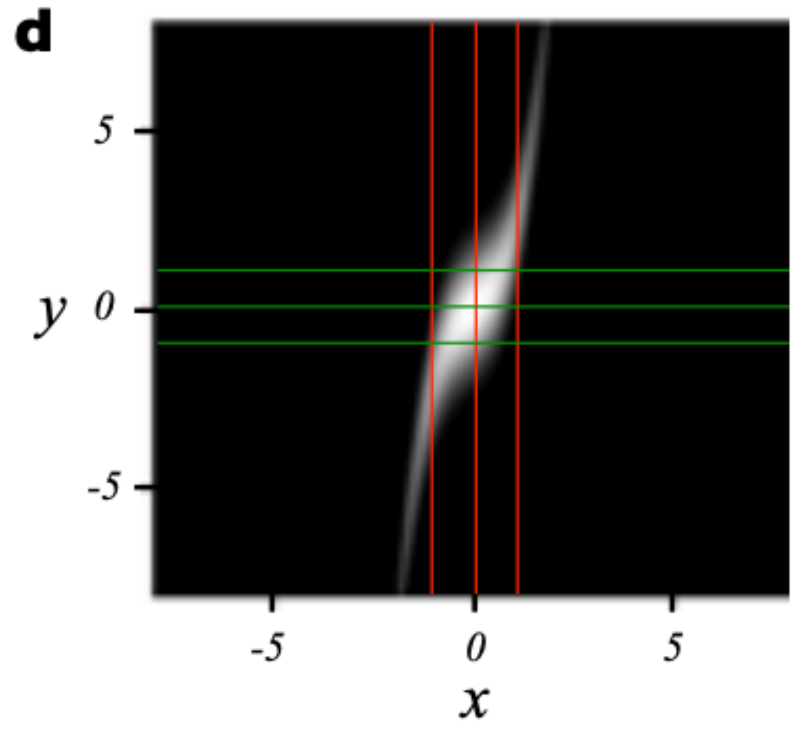


線形+ガウス分布（分散未知）では
どちらが原因か見分けがつかない

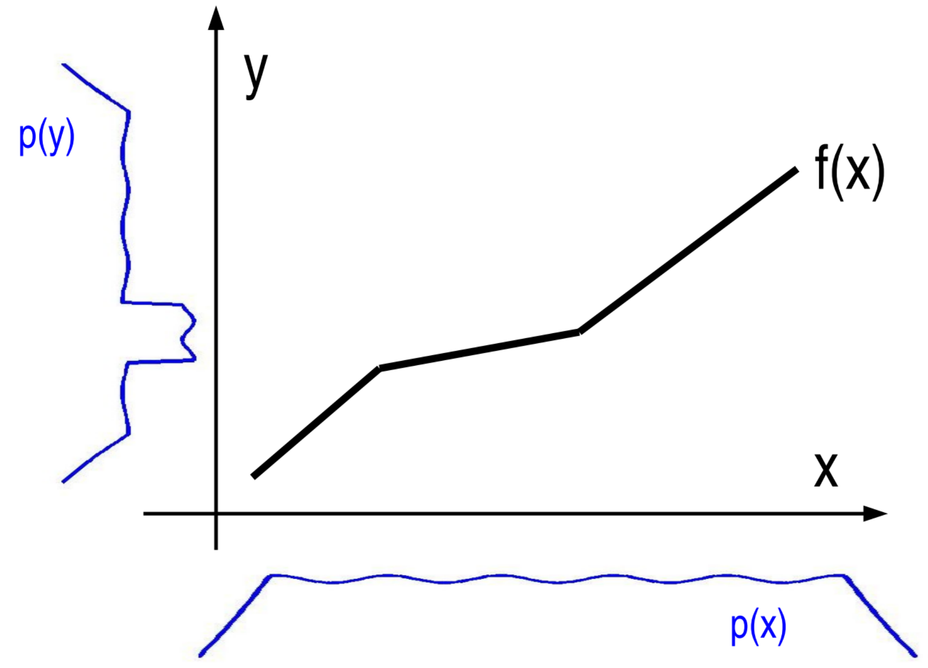
（参考）その他さまざまな因果探索手法

さまざまなアプローチが提案されている

- 加法的ノイズモデル（Additive Noise Model；ANM）
  - $y = f(x) + \varepsilon$
  - ノイズが非ガウスまたはモデルの非線形性が必要
- ポスト非線形モデル
  - $y = g(f(x) + \varepsilon_y) \Leftrightarrow g^{-1}(y) = f(x) + \varepsilon_y$
  - g は可逆な関数
- Greedy Equivalence Search（**GES**）
  - 尤度+パラメタ数による正則化（ベイズ情報量基準BIC）を最小化
  - CPDAGまで識別（尤度が”スコア等価”と呼ばれる性質を満たす場合）
- 情報幾何的因果推論（IGCI）
  - 非線形性＋傾きと密度の独立性を仮定



Hoyer, Patrik, et al. "Nonlinear causal discovery with additive noise models." *Advances in neural information processing systems* 21 (2008).



D. Janzing, J. M. Mooij, K. Zhang, J. Lemeire, J. Zscheischler, P. Daniuˇsis, B. Steudel, and B. Sch¨olkopf. Information-geometric approach to inferring causal directions. Artificial Intelligence, 182-183:1–31, 2012.

非巡回性の制約は組合せ的最適化となり計算量が高い
NOTEARSは非巡回性を連続に緩和した量を正則化として用いる

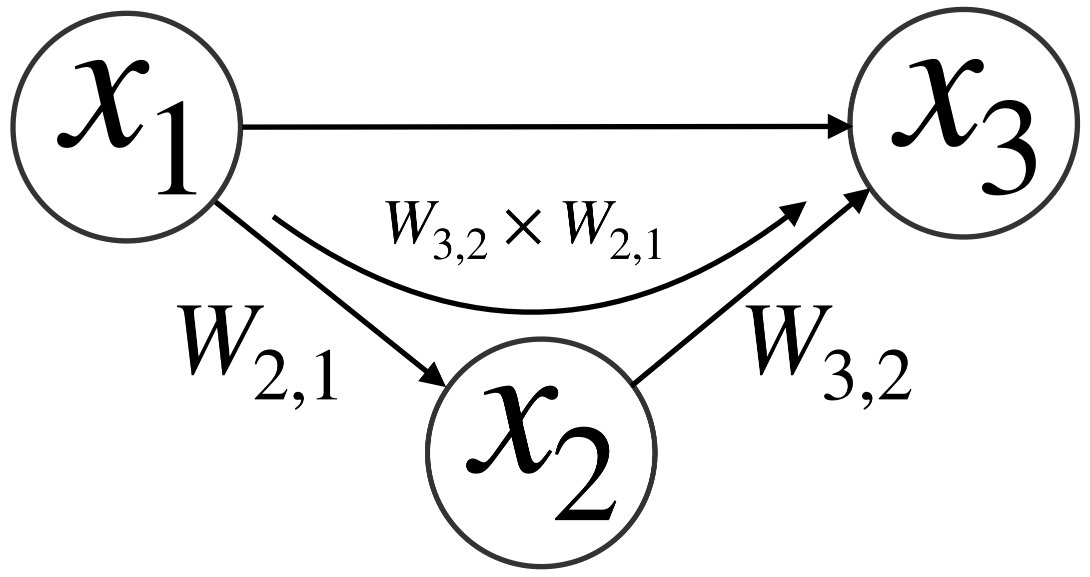
- NOTEARSは深層学習を用いて構造とモデルを同時に推定する
- LiNGAMと同様に線形モデルを仮定
  - $x = Wx + \varepsilon$
- 多くの手法はWの非巡回性を保証した中で最適化を行う
  - $\min_{W \in \mathbb{R}^{d \times d}} F(W)$ subject to $G(W) \in$ DAGs
    - $F(W)$は二乗誤差など
- 巡回性の度合いを表す関数 $h(W)$ があって、 $h(W) = 0 \Leftrightarrow G(W) \in$ DAGs となるならば、上式は以下と同値
  - $\min_{W \in \mathbb{R}^{d \times d}} \max_{\alpha \in \mathbb{R}} F(W) + \alpha h(W) + \rho h(W)^2 \quad (\rho \geq 0)$
  - 非凸なので大域的最適化は困難だが、局所解はWとαの交互最適化によって効率的に解ける（拡張ラグランジュ法）

トレース指数関数によって「巡回した影響の合計」を表せる

- 係数行列Wの非巡回性は以下のトレース指数関数によって表せる
  - $h(W) = \mathrm{tr}\left(e^{W \circ W}\right) - d = 0 \Leftrightarrow G(W) \in$ DAGs　（$W \in \mathbb{R}^{d \times d}$）
    - ただし $\circ$ は行列の要素ごとの積（アダマール積）
    - 行列指数関数は $e^A := \sum_{i=0}^{\infty} \frac{A^i}{i!} = I + A + \frac{A^2}{2!} + \cdots$
- トレース（対角成分の和）は線形：

$$\mathrm{tr}\left(e^S\right) = \mathrm{tr}(I) + \mathrm{tr}(S) + \frac{1}{2!}\mathrm{tr}\left(S^2\right) + \cdots$$

  - $\mathrm{tr}(I) = d$
  - S^kの対角成分は、ある変数の影響が、他の変数を通してkホップで元の変数に戻ってくる度合いを表す → 0であるべき



言語モデルの知識を用いる手法

データだけから因果構造全体を復元するのは難しい問題
→ 知識を言語モデルから抽出する

- データのみから因果構造全体を復元することはかなり難しく、実用的には人間がある程度の構造を事前知識として与えることが多い
- 「（変数名A）を変更すると（変数名B）に変化が生じますか？」と大規模言語モデルに質問
  - 変数名が存在すること、その因果関係が一般的知識として知られていることが前提
    - 大規模言語モデルにその一般的知識が学習され埋め込まれている前提
  - 変数ペアの因果方向判定タスクで既存のデータに基づく手法を凌駕
- ただし新たな知識構築（科学の発展）にはデータも必要

Model	Acc.	Wt. Acc.
Slope (Marx & Vreeken, 2017)	0.75	0.83
bQCD (Tagasovska et al., 2020)	0.68	0.75
PNL-MLP (Zhang & Hyvarinen, 2012)	0.75	0.73
Mosaic (Wu & Fukumizu, 2020)	0.83	0.82
ada	0.50	0.50
text-ada-001	0.49	0.50
babbage	0.51	0.50
text-babbage-001	0.50	0.50
curie	0.51	0.52
text-curie-001	0.50	0.50
davinci	0.48	0.47
text-davinci-001	0.50	0.50
text-davinci-002	0.79	0.79
text-davinci-003	0.82	0.83
LMPrior (Choi et al., 2022)	0.83	-
gpt-3.5-turbo	0.81	0.83
gpt-3.5-turbo (causal agent)	0.86	0.87
gpt-3.5-turbo (single prompt)	0.89	0.92
gpt-4 (single prompt)	**0.96**	**0.97**

Kiciman, Emre, et al. "Causal reasoning and large language models: Opening a new frontier for causality." Transactions on Machine Learning Research (2023).

まとめ

因果構造を推定する問題＝因果探索
仮定の置き方と得られる結論の強さにより幾つかのアプローチ

- 観察データのみから因果構造を完全に特定することは不可能
  - 一部の辺の向きの識別は諦めるか、追加でモデル等に仮定を置くことで識別する
- IC/PC法
  - （条件付き）独立性のみに基づく辺の枝刈りとV構造を用いた向き付け
  - 仮定が少ないが、向きが判定できない辺が残る可能性がある（CPDAGまで識別）
- LiNGAM
  - 線形性を仮定、ノイズ分布はガウス分布でないことを仮定
- NOTEARS
  - 構造方程式 f のモデルをノイズ含めて仮定
  - トレース指数関数により非巡回制約を連続空間で表現
- 言語モデルからの知識抽出
  - 変数ペアの因果関係を問い合わせることで知識を抽出