

グラフィカルモデルと因果探索

阪大 基礎工 数理
鈴木 讓

2025.11.23

R研究集会@統計数理研究所

本日のアジェンダ

- グラフィカルモデルと因果探索
- 新書「グラフィカルモデルと因果探索 with R」(共立出版)
- 依存関係の探索
- 変数順序の探索
- LiNGAMの紹介
- CRAN BNSLパッケージ by Joe Suzuki
- まとめ

渡辺澄夫ベイズ理論

100問

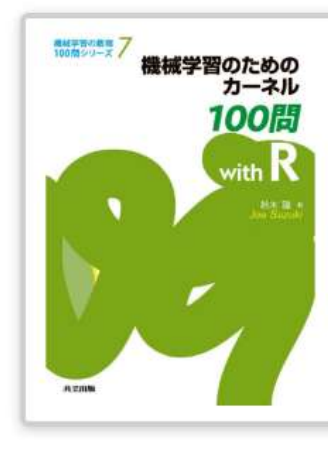
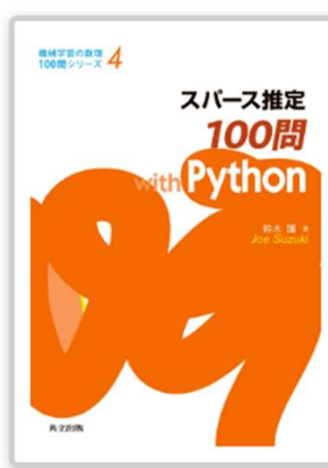
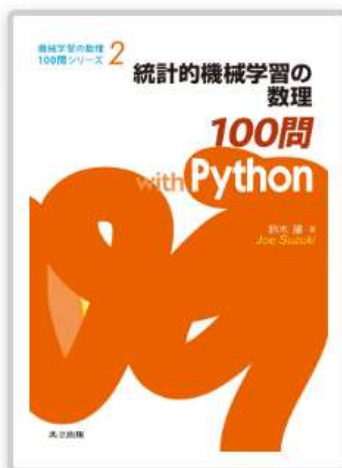
with
R/Stan

鈴木 謙 著
Joe Suzuki

共立出版

機械学習の数理100問シリーズ (共立出版)

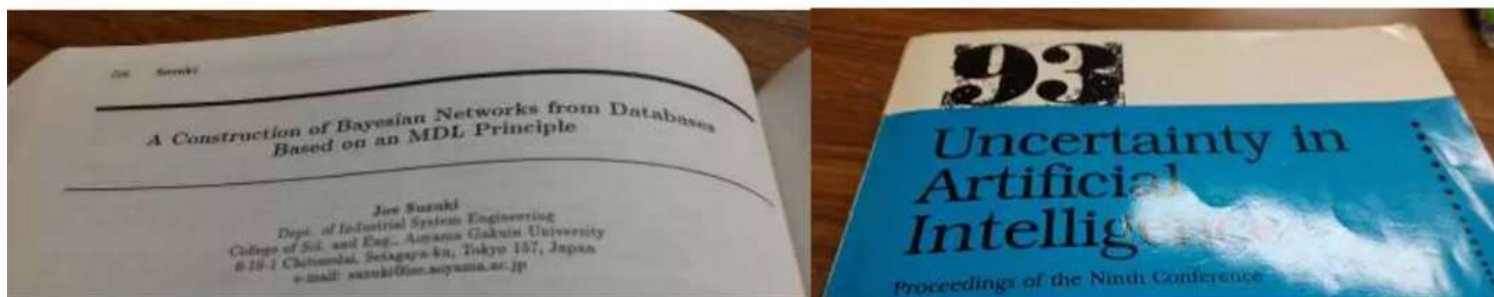
	R	Python	阪大講義
統計的機械学習の数理	2020.3	2020.4	学部3年半期(毎年)
スパース推定	2020.11	2021.2	2021前期
グラフィカルモデルと因果推論	2025.12.11		
機械学習のためのカーネル	2021.11	2021.12	2022前期
渡辺澄夫ベイズ理論	2023.9	2024.6	2023前期
深層学習の数理			2025前期



そもそもの専門: 確率的グラフィカルモデル

Joe Suzuki: “A Construction of Bayesian Networks from Databases Based on the MDL Principle” (UAI-93)

第9回 Uncertainty in Artificial Intelligence (1993)



国内初のベイジアンネットワークの研究者とされる(被引用 270)



第33回 Uncertainty in Artificial Intelligence (2017) 8月11日-15日

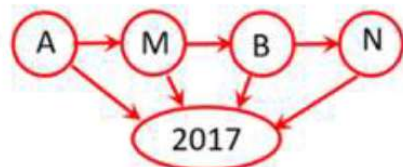




電気通信大学教授
植野真臣氏



産業技術総合研究所
東工大併任教授
本村陽一氏



教育工学などで
博士を多数輩出



The Third Workshop on Advanced
Methodologies for Bayesian Networks
20-22 September 2017
Kyoto, Japan

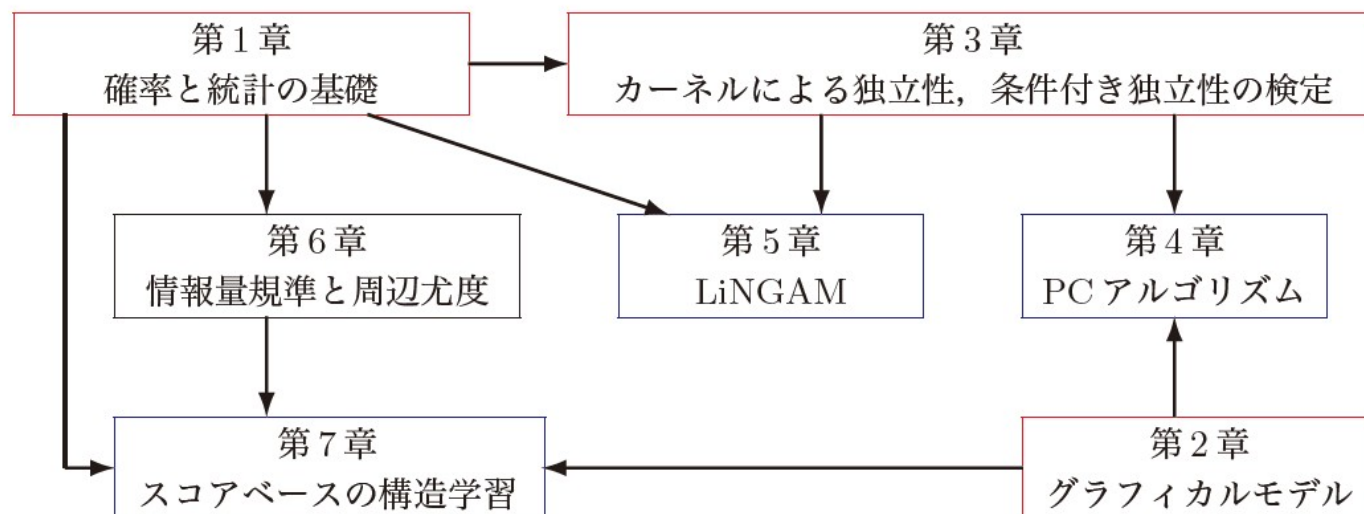
国内唯一のベイジアンネットワーク
運用ツール

グラフィカルモデルと因果探索 with R

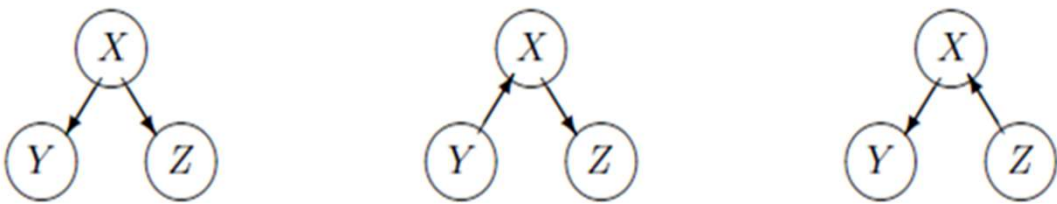
機械学習の数理100問シリーズ(共立出版)

2025年12月11日

因果探索 = 変数順序の探索 + 依存関係の探索



確率的グラフィカルモデル(SEM, BN, MNなども) とネットワーク分析の差異


$$\underbrace{P(X)P(Y|X)P(Z|X) = P(Y)P(X|Y)P(Z|X) = P(Z)P(X|Z)P(Y|X)}_{= \frac{P(XY)P(XZ)}{P(X)}}$$

確率的グラフィカルモデル:

確率的な依存関係を有向、無向グラフで表示



ネットワーク分析:

示したい関係(何でもよい)を有向、無向グラフで表示

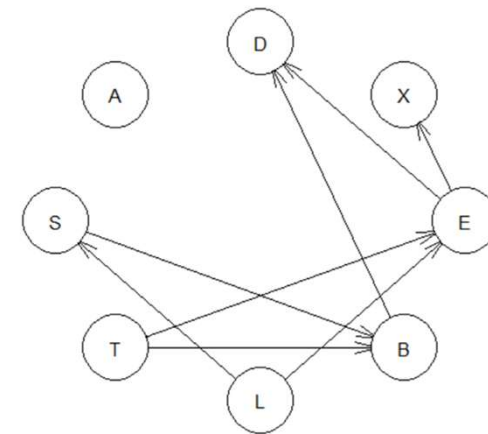
統計学: データから正しい構造を推定

##	A	S	T	L	B	E	X	D
## 1	no	yes	no	no	yes	no	no	yes
## 2	no	yes	no	no	no	no	no	no
## 3	no	no	yes	no	no	yes	yes	yes
## 4	no	no	no	no	yes	no	no	yes
## 5	no	no	no	no	no	no	no	yes
## 6	no	yes	no	no	no	no	no	yes

データ

推定

乱数発生



BN = 分布

- サンプル数 n が大きいときに、正しい構造を推定

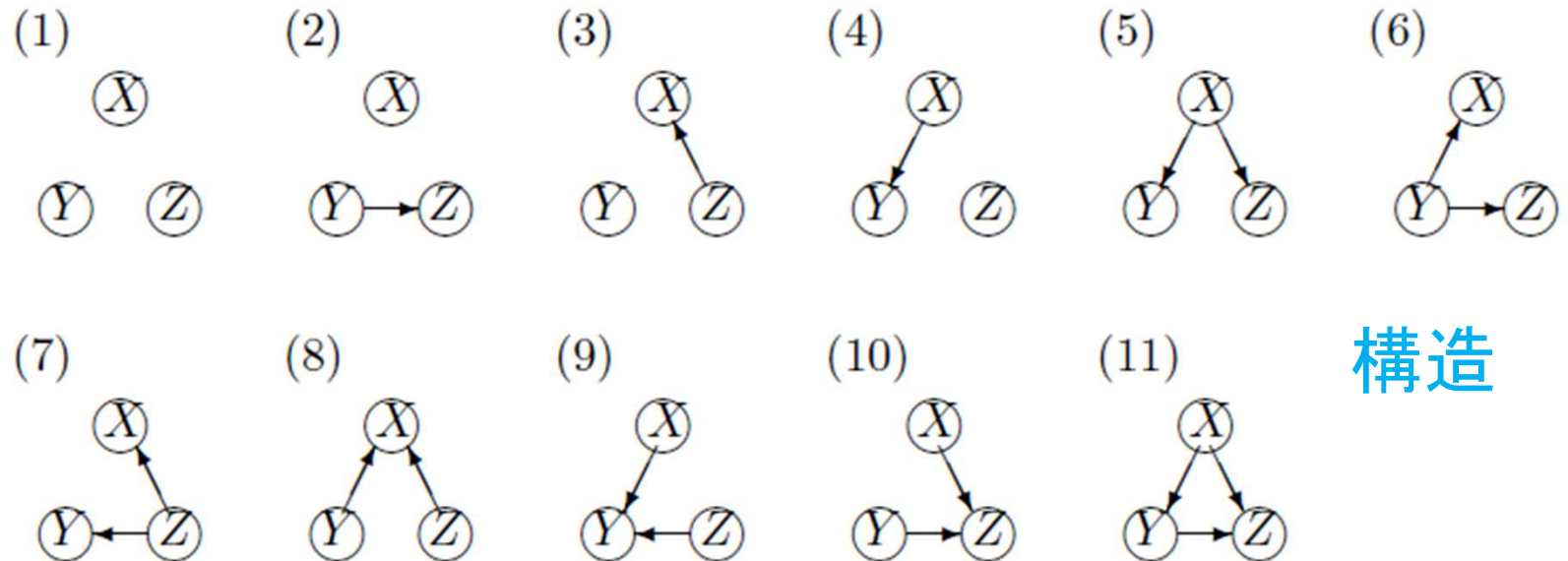
ベイジアンネットワークの構造学習

Bayesian Network Structure Learning (BNSL)

p=3のとき

X	Y	Z
x_1	y_1	z_1
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	z_n

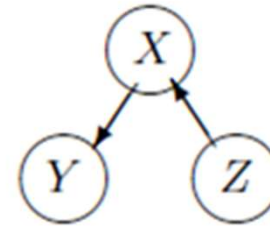
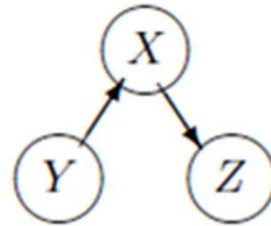
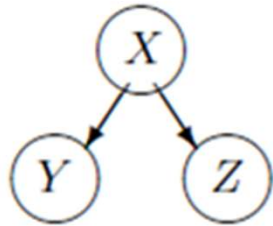
データフレーム



構造

ベイジアンネットワーク (BN)

分布の因数分解を、有向非巡回グラフ(DAG)で表現したもの



$$\underbrace{P(X)P(Y|X)P(Z|X) = P(Y)P(X|Y)P(Z|X) = P(Z)P(X|Z)P(Y|X)}_{= \frac{P(XY)P(XZ)}{P(X)}}$$

(同値)

スコアベースの BNSL

p 変数のデータからなるサンプル $x_1, \dots, x_n$

- 事後確率最大

構造の事前確率 $\times$ その構造での周辺尤度 (パラメータで周辺化)

- 情報量基準 (AIC/BIC)

$$\begin{aligned} & - \sum_{i=1}^n \log P(x_i | \text{パラメータの最尤推定}) + \text{パラメータ数} \\ & - \sum_{i=1}^n \log P(x_i | \text{パラメータの最尤推定}) + \frac{\text{パラメータ数}}{2} \log n \end{aligned}$$

パラメータ: 他変数のもとでの条件付き確率

- 厳密解は、 p に対して指数的
- Hill Climbing など、探索する範囲を限定
- グラフの範囲を限定 (森など)

ベイズ統計学: データのもとでの事後確率最大化

パラメータを周辺化した事後確率

$$P(\text{構造} | \text{データ}) \\ = \underbrace{P(\text{構造})}_{\text{構造の事前確率}} \int \underbrace{P(\text{データ} | \text{構造}, \text{パラメータ}) P(\text{パラメータ} | \text{構造})}_{\text{パラメータの事前確率}} d\text{パラメータ}$$

事前確率としては、標準的なもの

構造	(1)-(11)のどれか
パラメータ	各部分の条件付き確率



Joe Suzuki

友達977人

投稿 基本データ **友達** 写真 動画

友達

すべての友達 最近追加した友達 誕生日



Peter Laurence Spirtes
Carnegie Mellon University

制約ベースの BNSL (PC アルゴリズム)

BN、MN は、 p 変数間の条件付独立性の関係 (複数) を同時に表現したもの

条件付き独立性の検定を繰り返す
完全グラフから枝を切っていく

事象 A, B が独立 ($A \perp\!\!\!\perp B$)

$$P(A \cap B) = P(A)P(B)$$

事象 A, B が事象 C のもとで条件付き独立 ($A \perp\!\!\!\perp B | C$)

$$P(A \cap B | C) = P(A | C)P(B | C)$$

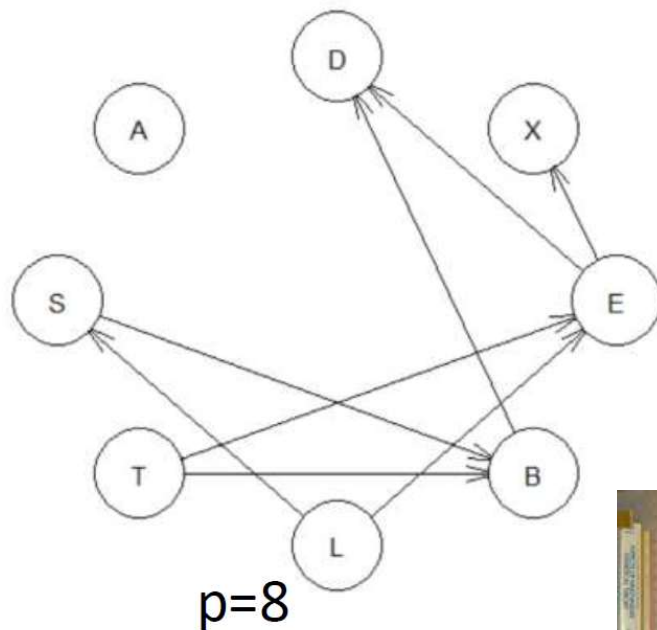


Clark Glymour

BNSLの基本的な使い方(1): 離散 BN

```
library(BNSL)
g=bnsf(asia)
plot(g)
```

ASIA (Lauritzen 1996)



```
##      A      S      T      L      B      E      X      D
## 1 no yes no no yes no no yes
## 2 no yes no no no no no no
## 3 no no yes no no yes yes yes
## 4 no no no no yes no no yes
## 5 no no no no no no no yes
## 6 no yes no no no no no yes
```

変数	意味 (Yes/No)
Asia	最近アジアに行った
Smoking	常習喫煙者
Tuberculosis	結核
Lung Cancer	肺がん
Bronchitis	気管支炎
E	T(結核)またはL(肺がん)の論理和
Chest X-ray	肺の X 線撮影での異常
Dyspnoea	呼吸困難



変数の個数 p とともに指数的な計算?

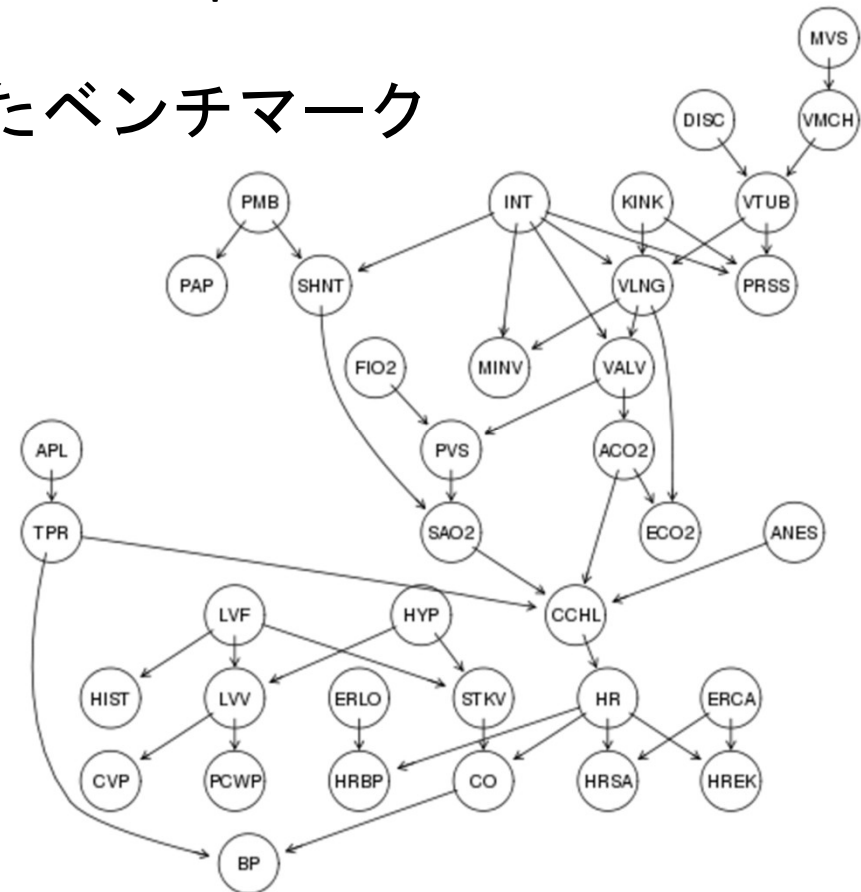
```
library(BNSL)
g=bnsf(alarm)
plot(g)
```

Alarm p=37

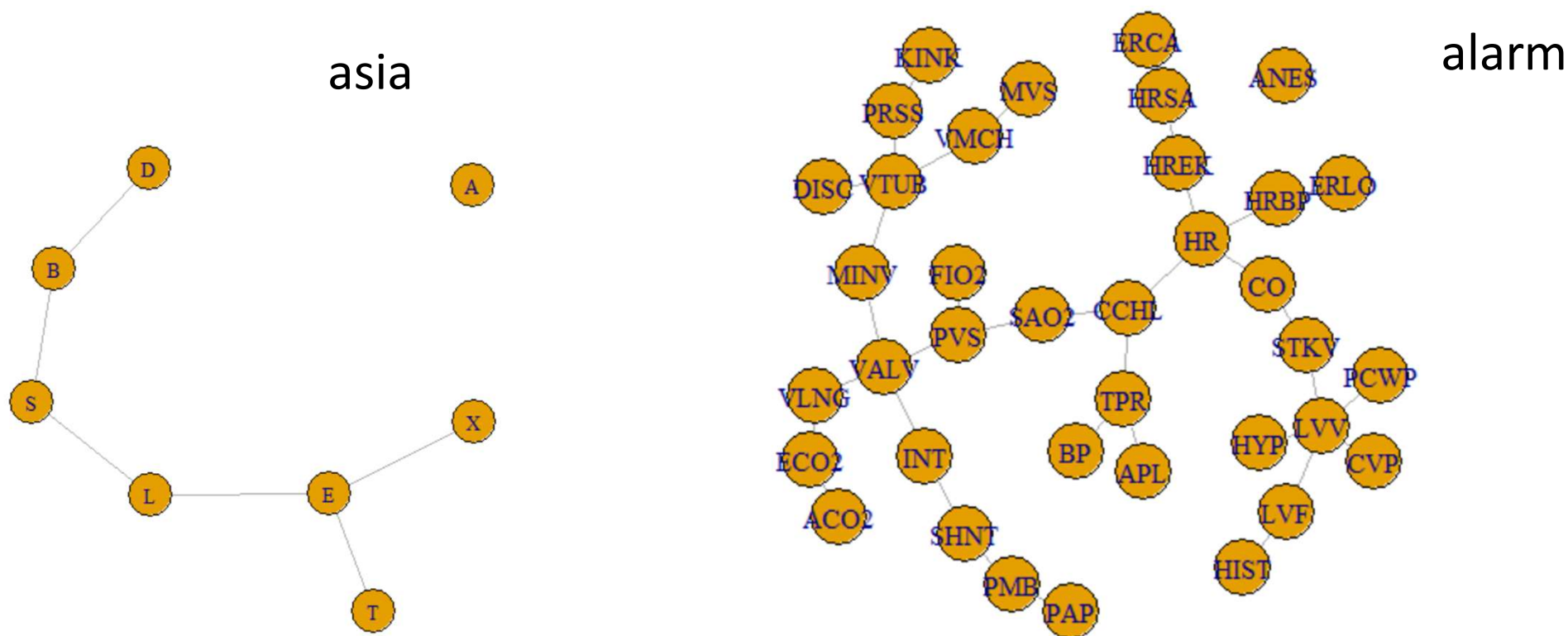
最も知られたベンチマーク

PCで、1日かかる

`g=bnsf(alarm[,1:20])`
 など、変数の個数を限定する



alarmはBNで1日かかるが、森だと5秒で完了



pの2乗の時間

CRAN パッケージ bnlearn www.bnlearn.com

BNの構造学習、確率推論で最もよく使われているパッケージ

Marco Scutari 氏



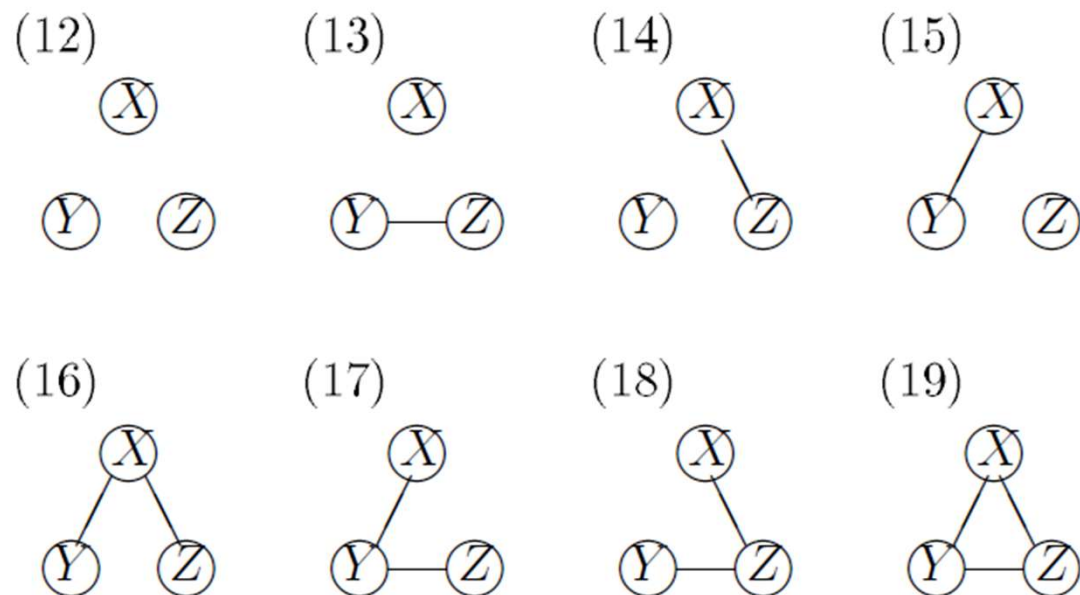
p=3のとき

森の構造学習

Forest Structure Learning (FSL)

X	Y	Z
x_1	y_1	z_1
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	z_n

データフレーム



構造

相互情報量

独立性を仮定した場合
とのK-L情報量

X, Y が離散:

$$I(X, Y) = \sum_x \sum_y P_{XY}(x, y) \log \frac{P_{XY}(x, y)}{P_X(x)P_Y(y)}$$

X, Y が連続:

$$I(X, Y) = \int_x \int_y f_{XY}(x, y) \log \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} dx dy$$

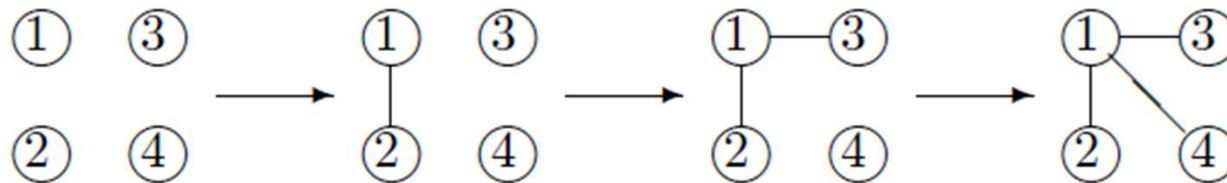
$$I(X, Y) \geq 0$$

$$I(X, Y) = 0 \iff X \perp\!\!\!\perp Y$$

独立であるとき
だけ0になる

データ -> 相互情報量の推定値 -> 森 (Chow-Liu, Kruskal)

分布が森で表現されることを仮定



$$I(1,2) > I(1,3) > I(2,3) > I(1,4) > I(3,4) > I(2,4) > 0$$

相互情報量の推定値大きいものから、
ループができない限り、辺を結んでい
く

ベイズ相互情報量推定量

$$\alpha, \beta \geq 2$$

$$x^n = (x_1, \dots, x_n) \in \{1, \dots, \alpha\}^n, \quad y^n = (y_1, \dots, y_n) \in \{1, \dots, \beta\}^n$$

多項分布にしたがう

$Q(x^n)$: x^n の周辺尤度 , $Q(y^n)$: y^n の周辺尤度 , $Q(x^n, y^n)$: x^n, y^n の周辺尤度

$I(x^n, y^n)$: 相対頻度を用いて計算した相互情報量の推定値 (最尤推定)

$$J(x^n, y^n) := \frac{1}{n} \log \frac{Q(x^n, y^n)}{Q(x^n)Q(y^n)} \approx I(x^n, y^n) - \frac{1}{2}(\alpha - 1)(\beta - 1) \log n$$

確率 1 で

$$J(x^n, y^n) \leq 0 \iff X, Y \text{ が独立}$$

相互情報量の推定: 連続でもできる(新書に記載)

連続、離散混在でもできる(研究レベル)

離散 --- 連続 --- 離散
(困難)

連続 --- 離散 --- 連続
(容易)

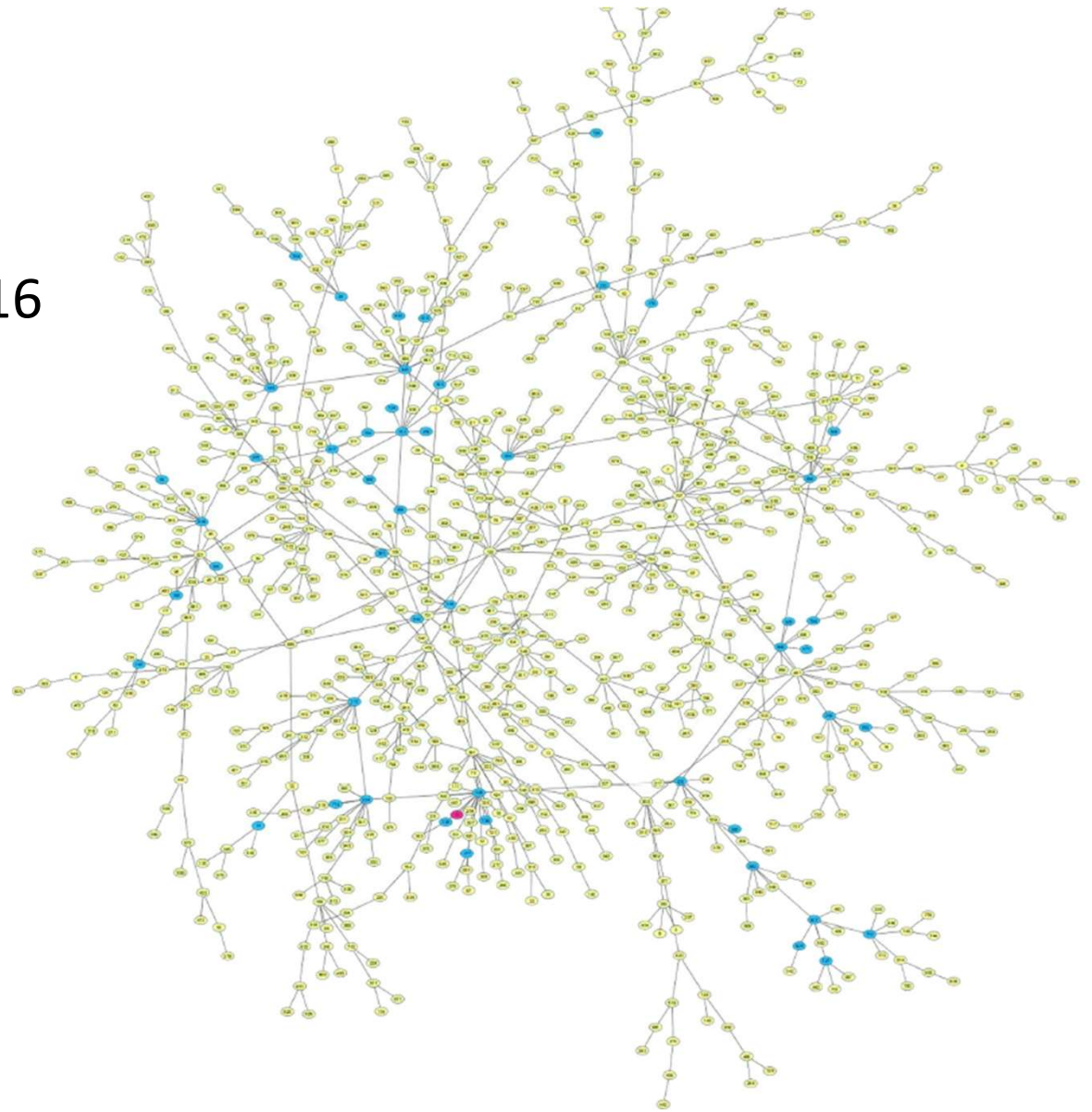
ゲノム解析への応用 (1)

Int. J. Approximate Reasoning, 2016

乳がん患者の遺伝子発現データ
(p51遺伝子をもつサンプル192個、
もたないサンプル58個)

青: p値の大きな遺伝子

赤: 症例/対照
1000個の連続変量
1個の2値変量

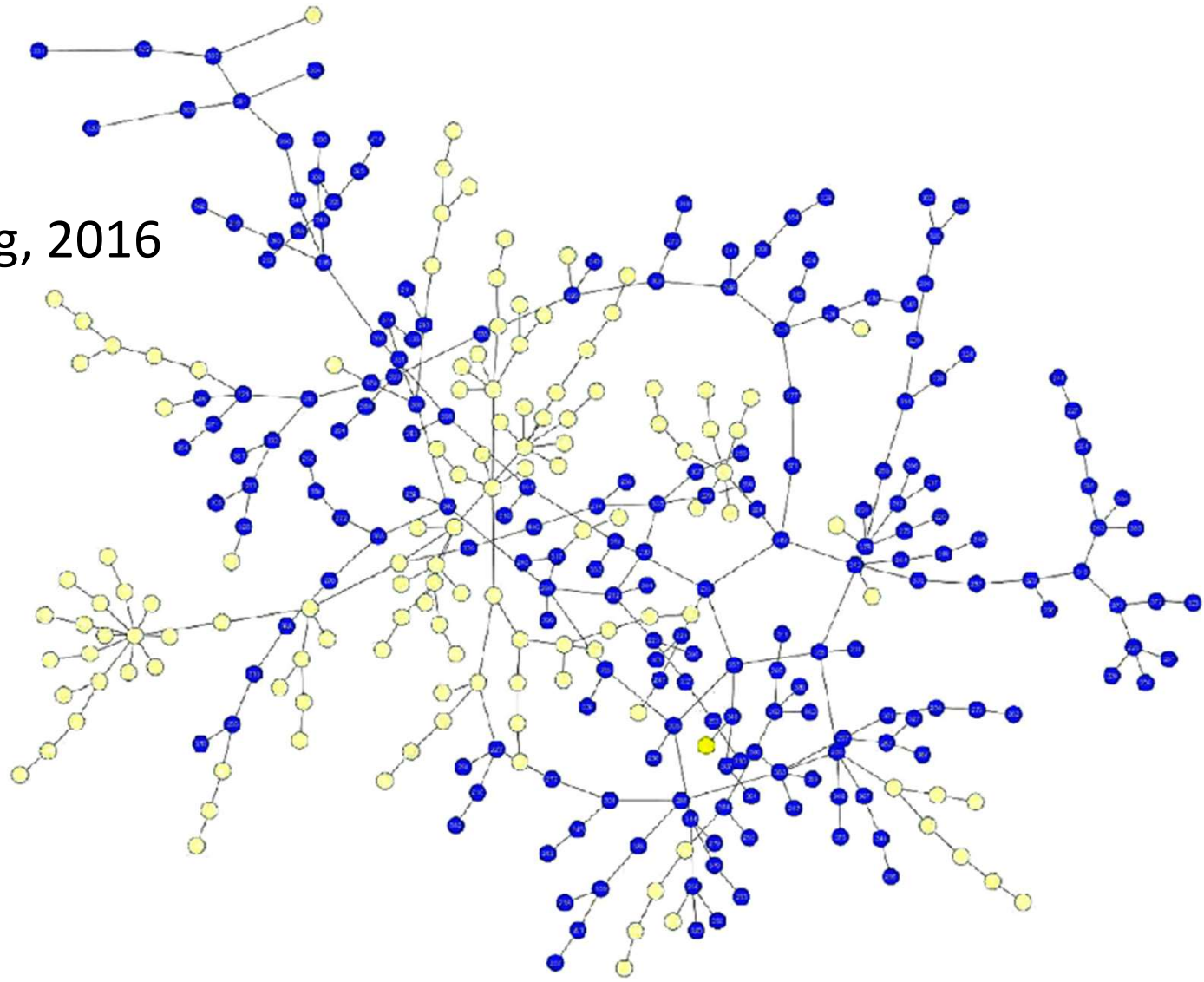


ゲノム解析への応用

(2) J. Approximate Reasoning, 2016

青: 遺伝子発現量

赤: SNP (3値)



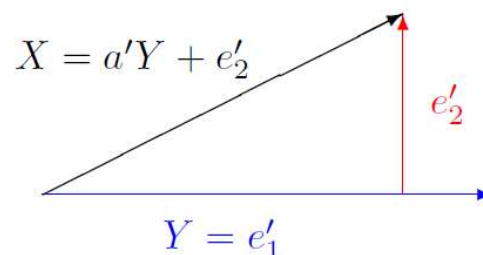
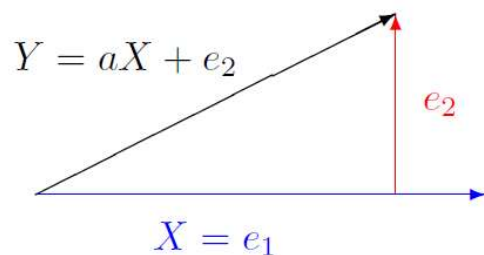
加法雑音モデル (狩野 2003)

X, Y : 線形従属

$$\begin{cases} X = e_1 \\ Y = aX + e_2 \end{cases} \quad a = \frac{\text{Cov}(e_1, e_2)}{V(e_1)} \quad e_1 \perp\!\!\!\perp e_2 \implies X \rightarrow Y$$

$$\begin{cases} Y = e'_1 \\ X = a'Y + e'_2 \end{cases} \quad a' = \frac{\text{Cov}(e'_1, e'_2)}{V(e'_1)} \quad e'_1 \perp\!\!\!\perp e'_2 \implies Y \rightarrow X$$

$e_1 \perp\!\!\!\perp e_2$ と $e'_1 \perp\!\!\!\perp e'_2$ の一方だけを仮定



e_1, e_2 がガウスだと識別できない

$$\begin{cases} U, V : \text{ガウス} & \implies U + V : \text{ガウス} \\ U : \text{ガウス}, b : \text{定数} & \implies bU : \text{ガウス} \end{cases}$$

$$\begin{cases} X = e_1 \\ Y = aX + e_2 \end{cases} \quad \begin{cases} Y = e'_1 \\ X = a'Y + e'_2 \end{cases}$$

$$e_1, e_2 : \text{ガウス} \implies X, Y : \text{ガウス} \implies e'_1, e'_2 : \text{ガウス}$$

$$\begin{cases} e_1, e_2 : \text{ガウスのとき、} & \text{Cov}(e_1, e_2) = 0 & \iff e_1 \perp\!\!\!\perp e_2 \\ e'_1, e'_2 : \text{ガウスのとき、} & \text{Cov}(e'_1, e'_2) = 0 & \iff e'_1 \perp\!\!\!\perp e'_2 \end{cases}$$

$$e_1, e_2 : \text{ガウス} \implies e_1 \perp\!\!\!\perp e_2, e'_1 \perp\!\!\!\perp e'_2$$

(識別不可)

具体的な手順 (2変数)

$$\left. \begin{array}{l} x^n = (x_1, \dots, x_n) \\ y^n = (y_1, \dots, y_n) \end{array} \right\} \text{観測データ } (n: \text{標本サイズ}), \quad \left. \begin{array}{l} y_x^n := y^n - \frac{c(x^n, y^n)}{v(x^n)} x^n \\ x_y^n := x^n - \frac{c(x^n, y^n)}{v(y^n)} y^n \end{array} \right\} \text{として、}$$

$x^n \perp\!\!\!\perp y_x^n$ と $y^n \perp\!\!\!\perp x_y^n$ のどちらが真か

$v(\cdot)$: 標本分散, $c(\cdot, \cdot)$: 標本共分散

$$\left\{ \begin{array}{l} X = e_1 \\ Y = aX + e_2 \end{array} \right\} \quad \left\{ \begin{array}{l} Y = e'_1 \\ X = a'Y + e'_2 \end{array} \right\} \quad \text{の } e_1 \perp\!\!\!\perp e_2 \text{ と } e'_1 \perp\!\!\!\perp e'_2 \text{ の比較}$$

具体的な手順 (3 変数): 上流からの影響をのぞいて、残差の独立性を評価

$$\left. \begin{array}{l} x^n = (x_1, \dots, x_n) \\ y^n = (y_1, \dots, y_n) \\ z^n = (z_1, \dots, z_n) \end{array} \right\} \text{観測データ (} n: \text{標本サイズ)}$$

第 1 ステージ

$x^n \perp\!\!\!\perp \{y_x^n, z_x^n\}$ と $y^n \perp\!\!\!\perp \{z_y^n, x_y^n\}$ と $z^n \perp\!\!\!\perp \{x_z^n, y_z^n\}$ のどれが真か

$$\left. \begin{array}{l} \underline{x^n \perp\!\!\!\perp \{y_x^n, z_x^n\} \text{ が真のとき、}} \\ y_{xz}^n := y_x^n - \frac{c(y_x^n, z_x^n)}{v(z_x^n)} z_x^n \\ z_{xy}^n := z_x^n - \frac{c(y_x^n, z_x^n)}{v(y_x^n)} y_x^n \end{array} \right\} \text{として、}$$

第 2 ステージ

$y_x^n \perp\!\!\!\perp z_{xy}^n$ と $z_x^n \perp\!\!\!\perp y_{xz}^n$ のどちらが真か

具体的な手順 (3変数): まとめて

$$x^n, y^n, z^n \Rightarrow \left\{ \begin{array}{l} x^n \perp\!\!\!\perp \{y_x^n, z_x^n\} \Rightarrow \\ y^n \perp\!\!\!\perp \{z_y^n, x_y^n\} \Rightarrow \\ z^n \perp\!\!\!\perp \{x_z^n, y_z^n\} \Rightarrow \end{array} \right. \left\{ \begin{array}{l} y_x^n \perp\!\!\!\perp z_{xy}^n \Rightarrow X \rightarrow Y \rightarrow Z \\ z_x^n \perp\!\!\!\perp y_{xz}^n \Rightarrow X \rightarrow Z \rightarrow Y \\ z_y^n \perp\!\!\!\perp x_{yz}^n \Rightarrow Y \rightarrow Z \rightarrow X \\ x_y^n \perp\!\!\!\perp z_{yx}^n \Rightarrow X \rightarrow Z \rightarrow Y \\ x_z^n \perp\!\!\!\perp y_{zx}^n \Rightarrow Z \rightarrow X \rightarrow Y \\ y_z^n \perp\!\!\!\perp x_{zy}^n \Rightarrow Z \rightarrow Y \rightarrow X \end{array} \right.$$

HSIC による独立性検定

$\mathcal{X}$: 集合

$k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$: 正定値カーネル

H : k に対応する RKHS (再生核ヒルベルト空間)

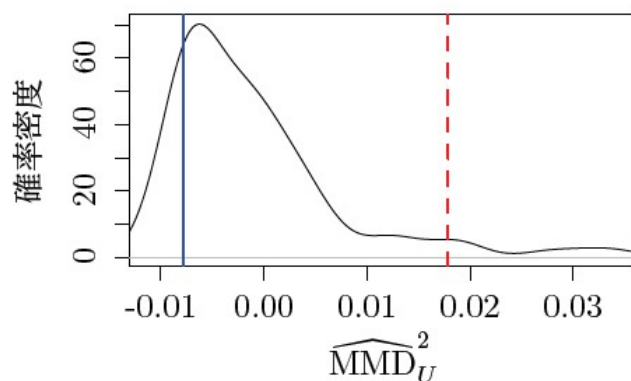
X, Y : $\mathcal{X}$ を標本空間とする確率変数

HSIC (Hilbert-Schmidt Information Criterion)

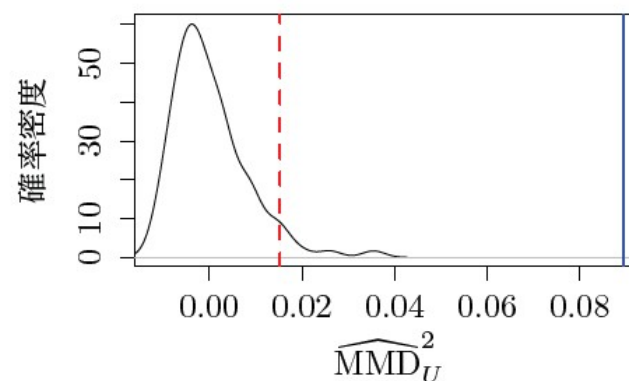
$X, Y \in \mathcal{X}$ の独立性を $k(X, \cdot), k(Y, \cdot) \in H$ の独立性で評価

$$\text{HSIC}(X, Y) := \|\mathbb{E}[k(X, \cdot)k(Y, \cdot)] - \mathbb{E}[k(X, \cdot)]\mathbb{E}[k(Y, \cdot)]\|_H^2 = 0 \iff X \perp\!\!\!\perp Y$$

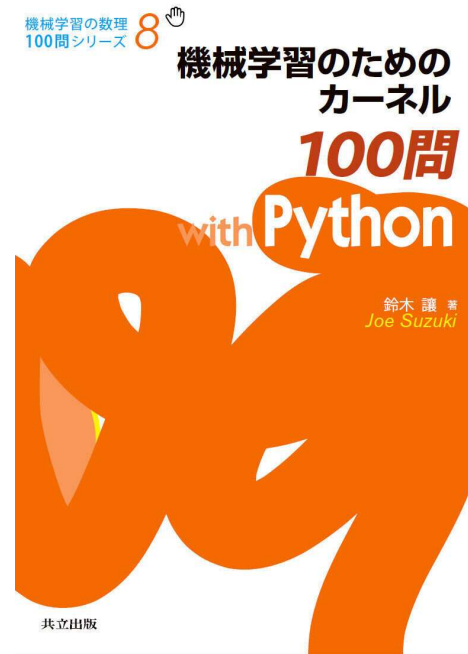
X, Y の分布が同じ場合 (並べ替え)



X, Y の分布が異なる場合 (並べ替え)



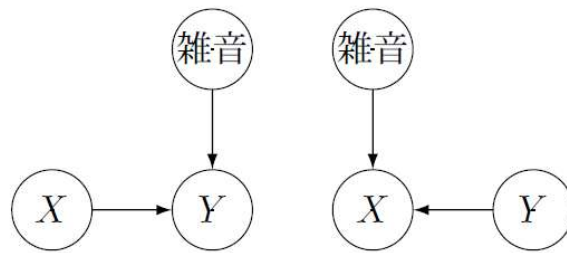
LiNGAMとHSICのプログラム (第5章)



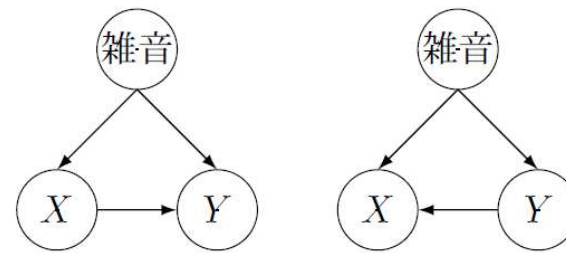
4.4	スプライン	112
4.5	Random Fourier Features	117
4.6	Nyström 近似	122
4.7	不完全 Cholesky 分解	124
付録: 命題の証明		129
問題	46~64	129
第 5 章	MMD と HSIC	135
5.1	RKHS における確率変数	135
5.2	MMD と 2 標本問題	139
5.3	HSIC と独立性検定	145
5.4	特性カーネルと普遍カーネル	154
5.5	経験過程入門	157
付録: 命題の証明		164
問題	65~83	164
第 6 章	Gauss 過程と関数データ解析	171
6.1	関数	171

最近の成果: 交絡の存在を許容

$(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R} \times \mathbb{R}$ から、 $X \rightarrow Y$ or $Y \rightarrow X$ を同定したい



いずれかを真と仮定 (LiNGAM)



交絡があるのがふつう (本研究)

貢献

1. KL情報量で交絡を定量化し、それを最小にする定式化
2. 最短経路問題を用いた因果順序同定の計算方法
3. 解の正確さ、計算の少なさの保証

基準の提案

まとめ

1. 確率的グラフィカルモデルの概要
2. 機械学習の数理100問シリーズ「確率的グラフィカルモデルとLiNGAM with R」
3. CRAN BNSL
4. LiNGAMの概要

ブラックボックスにしないで、中味を理解してもらいたい