数理統計 補助資料 ～ベイジアンネットワーク～

2023年度2学期: 月曜1限, 水曜3限

担当教員: 石垣 司

ベイジアンネットワーク

複数変数間の依存関係をネットワーク構造により表現することで，その確率的関係を理解・発見するための手法

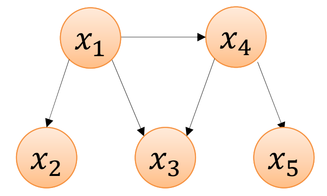
1. 変数間の依存関係をわかりやすく可視化できる
2. 注目する事象の条件付き確率を推論できる
3. 離散変数間の非線形関係・交互作用を扱うことが得意

本授業では，離散変数のベイジアンネットワークのみを扱う

数理的内容よりも，使い方&意義の紹介が主

多変量データ

ID	離散変数1	離散変数2	…	離散変数 P
1	x_{11}	x_{12}	…	x_{1P}
2	x_{21}	x_{21}	…	x_{2P}
⋮	⋮	⋮	⋱	⋮
i	x_{i1}	x_{i2}	…	x_{iP}
⋮	⋮	⋮	⋱	⋮
N	x_{N1}	x_{N2}	…	x_{NP}



多変量データをよく説明できるネットワーク構造

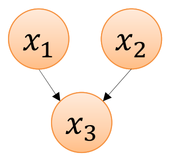
交互作用の発見と事後確率の推論

例: 車の購入の要因分析

x_1: ダイレクトメール A or B の送付

x_2: 店内キャンペーンの有無

x_3: 車の購入・非購入



ネットワーク構造の可視化

条件付き確率表

$\Pr(x_1)$

$x_1 = A$	0.5
$x_1 = B$	0.5

$\Pr(x_2)$

$x_2 = 0$	0.8
$x_2 = 1$	0.2

$\Pr(x_3|x_1, x_2)$

	x_1 = A, x_2 = 0	x_1 = A, x_2 = 1	x_1 = B, x_2 = 0	x_1 = B, x_2 = 1
$x_3 = 0$	0.98	0.98	0.98	0.75
$x_3 = 1$	0.02	0.02	0.02	0.25

交互作用

事後確率

$\Pr(x_1 = A|x_3 = 1) = 10/43 \fallingdotseq 0.23$

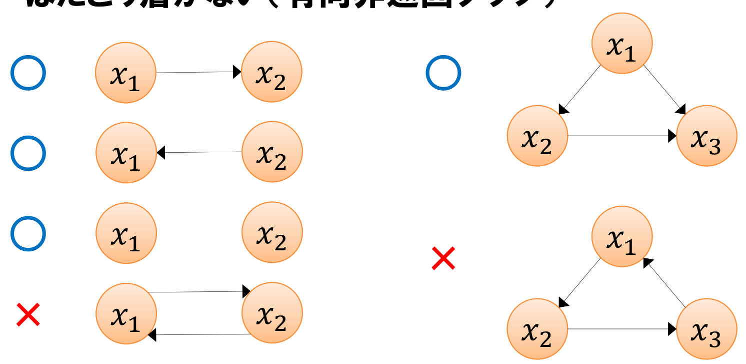
$\Pr(x_1 = B|x_3 = 1) = 33/43 \fallingdotseq 0.77$

車を購入した顧客の77%はダイレクトメールBを受け取っていた

ベイジアンネットワークの線の引き方

ベイジアンネットワークで変数をつなげるルール

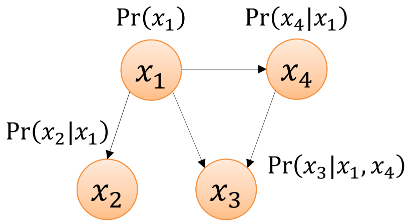
- 依存関係のある変数間には矢線を引く．そのとき，矢線の向きは必ず一方向のみに向いている（有向グラフ）
- ある一つの変数から矢線をたどり進んでいっても元の変数にはたどり着かない（有向非巡回グラフ）



ベイジアンネットワークの数理モデル

表記方法

- $\mathrm{Pa}(x_i)$：**変数 x_i の親の集合**
- $\mathrm{Pr}(x_i|\mathrm{Pa}(x_i))$：**依存関係（条件付き確率）**
- **同時確率**：$\mathrm{Pr}(x_1,\ldots,x_P)=\prod_{i=1}^{P}\mathrm{Pr}(x_i|\mathrm{Pa}(x_i))$



各変数の親の集合

$\mathrm{Pa}(x_1)=\emptyset,\ \mathrm{Pa}(x_2)=\{x_1\},$
$\mathrm{Pa}(x_3)=\{x_1,x_4\},\ \mathrm{Pa}(x_4)=\{x_1\}$

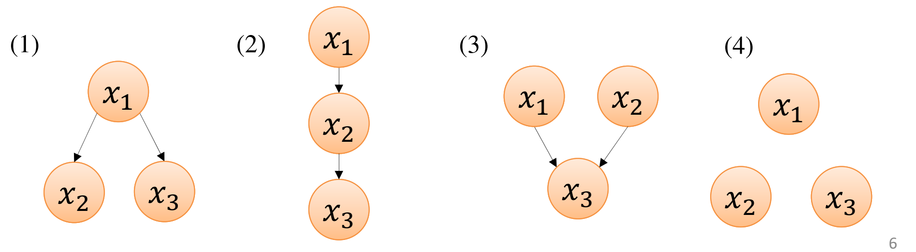
同時確率

$$\mathrm{Pr}(x_1,x_2,x_3,x_4)=\mathrm{Pr}(x_1)\,\mathrm{Pr}(x_2|x_1)\,\mathrm{Pr}(x_3|x_1,x_4)\,\mathrm{Pr}(x_4|x_1)$$

異なる依存関係の表現 #1

例：3変数での代表的な依存関係の表現

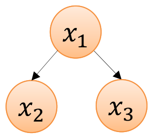
1. **変数 x_1 が変数 x_2 と x_3 へ影響を与えている**
2. **変数 x_1 が変数 x_2 へ影響を与え，変数 x_2 が変数 x_3へ影響を与えている**
3. **変数 x_1 と x_2 が変数 x_3 へ影響を与えている**
4. **変数 x_1, x_2, x_3 の間に関係性はない**



異なる依存関係の表現 #2

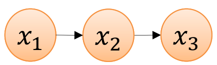
ネットワーク構造が異なると条件付き確率も異なる

(1)



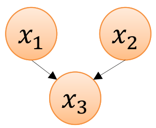
$$\mathrm{Pr}(x_1|x_3)=\frac{\mathrm{Pr}(x_1,x_3)}{\mathrm{Pr}(x_3)}=\frac{\sum_{x_2}\mathrm{Pr}(x_1)\mathrm{Pr}(x_2|x_1)\mathrm{Pr}(x_3|x_1)}{\sum_{x_2}\sum_{x_1}\mathrm{Pr}(x_1)\mathrm{Pr}(x_2|x_1)\mathrm{Pr}(x_3|x_1)}$$

(2)



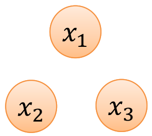
$$\mathrm{Pr}(x_1|x_3)=\frac{\mathrm{Pr}(x_1,x_3)}{\mathrm{Pr}(x_3)}=\frac{\sum_{x_2}\mathrm{Pr}(x_1)\mathrm{Pr}(x_2|x_1)\mathrm{Pr}(x_3|x_2)}{\sum_{x_2}\sum_{x_1}\mathrm{Pr}(x_1)\mathrm{Pr}(x_2|x_1)\mathrm{Pr}(x_3|x_2)}$$

(3)



$$\mathrm{Pr}(x_1|x_3)=\frac{\mathrm{Pr}(x_1,x_3)}{\mathrm{Pr}(x_3)}=\frac{\sum_{x_2}\mathrm{Pr}(x_1)\mathrm{Pr}(x_2)\mathrm{Pr}(x_3|x_1,x_2)}{\sum_{x_2}\sum_{x_1}\mathrm{Pr}(x_1)\mathrm{Pr}(x_2)\mathrm{Pr}(x_3|x_1,x_2)}$$

(4)

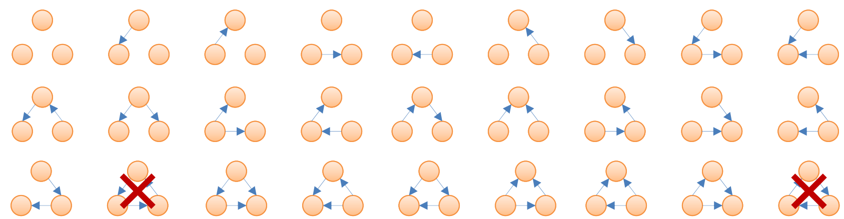


$$\mathrm{Pr}(x_1|x_3)=\frac{\mathrm{Pr}(x_1,x_3)}{\mathrm{Pr}(x_3)}=\mathrm{Pr}(x_1)$$

ネットワーク構造の探索

様々な基準に準拠してネットワーク構造を自動的に決定

- **しかし，変数の種類 P が大の時，厳密解の計算が困難**
  - **P に依存して計算量が指数的に増大**
- **ソフトウェアに搭載されているアルゴリズムの多くは近似解**
  - **例：$P=3$ のときの探索すべきパターン数は約 3^P**



#メモ 数理的事項は計算機科学に関する内容が主となるため本授業ではこれ以上扱わない。これ以降の学習の道筋。ネットワーク構造の推定に関しては、スコア基準のためのAIC, BIC, MDL、条件付き独立に準拠したPCアルゴリズムなど. 確率推論に関しては、グラフ構造、木構造、確率伝搬法、Loopy BPなど

重回帰分析とベイジアンネットワークの比較

	重回帰分析	ベイジアンネットワーク
目的変数と説明変数	１つの目的変数を複数の変数を用いて説明する一義的な依存関係の分析	複数の変数間の依存関係を分析。必ずしも目的変数を設定する必要はない
非線形関係	線形関係の分析が主。非線形関係は明示的に閾値や交互作用項を設定	変数間の条件付き確率表と交互作用の意味で非線形関係の取り扱いが可能
連続変数と離散変数	どちらも説明変数へ用いることが可能	連続変数は閾値などを用いて離散変数に変換する必要がある
推定精度の統計的理論	データの量と質が十分であるとき、説明変数の種類が多くてもパラメータ（回帰係数）は安定的に推定できる。また、統計的に推定精度を議論できる	データの量と質が十分であっても、説明変数の種類が多いと最適なネットワーク構造の探索と確率推論の計算コストが高い. 重回帰分析のような係数の検定はできない

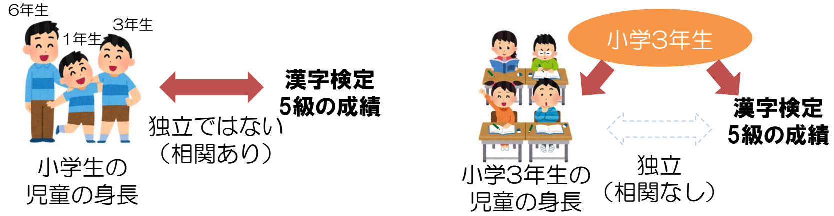
条件付き独立

変数 z の値が与えられた時, 変数 x, y が独立になる関係

$$\Pr(x, y|z) = \Pr(x|z)\Pr(y|z)$$

z の情報が所与のとき, x に関する情報を得ても, y に関する情報を得ることができない
z の情報が所与のとき, x の情報が変化しても, y に対して影響を与えない

- **例：小学生の児童の身長(x)と漢字力(y)は独立か？** $\Rightarrow$ No
  児童の学年・年齢が身長と学力の両方に影響
- **例：小学3年生(z)の児童の身長と漢字力は独立か？** $\Rightarrow$ Yes
  学年の情報が所与のとき条件付き独立

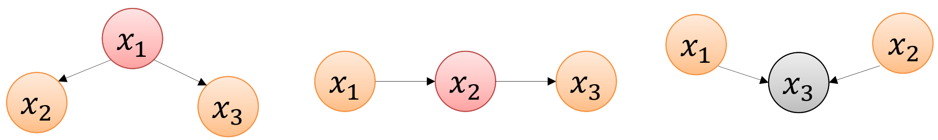


有向分離

複雑なベイジアンネットワーク（より一般にはグラフィカルモデル）**でも, 効率的かつ直感的に変数集合間の独立性を見つけ出すことに利用できるネットワーク構造の性質**

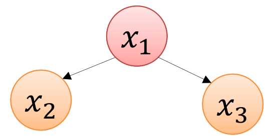
- **条件付き独立の関係を知ることで, お互いに影響を与えない変数群を視覚的に理解することができる**
  （多変数の中から注目すべき変数を絞ることができる）

Tail-to-tail **型**　　Head-to-tail **型**　　Head-to-head **型**



有向分離 #1

Tail-to-tail **型**

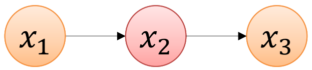


- **x_1 の情報が所与のとき, x_2 と x_3 は条件付き独立**
  （x_1 の情報が無いとき, x_2 と x_3 は独立とは限らない）
  例：小学生の身長(x_2)と語彙力(x_3)は学年(x_1)が所与で条件付き独立。
  また, 学年(x_1)の情報が無いと身長(x_2)と語彙力(x_3)は独立ではない
- **同時確率：**$\Pr(x_1, x_2, x_3) = \Pr(x_1)\Pr(x_2|x_1)\Pr(x_3|x_1)$
- **条件付き確率**

$$\Pr(x_2, x_3|x_1) = \frac{\Pr(x_1, x_2, x_3)}{\Pr(x_1)} = \Pr(x_2|x_1)\Pr(x_3|x_1)$$

有向分離 #2

Head-to-tail 型

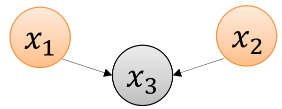


- x_2 の情報が所与のとき, x_1 と x_3 は条件付き独立
  （x_2 の情報が無いとき, x_1 と x_3 は独立とは限らない）
  例：小学生の学年(x_1)と文章読解力(x_3)は語彙力(x_2)が所与で条件付き独立。また, 語彙力(x_2)の情報が無いと学年(x_1)と文章読解力(x_3)は独立ではない
- 同時確率：$\Pr(x_1, x_2, x_3) = \Pr(x_1)\Pr(x_2|x_1)\Pr(x_3|x_2)$
- 条件付き確率

$$\Pr(x_1, x_3|x_2) = \frac{\Pr(x_1, x_2, x_3)}{\Pr(x_2)} = \Pr(x_1|x_2)\Pr(x_3|x_2)$$

有向分離 #3

Head-to-head 型

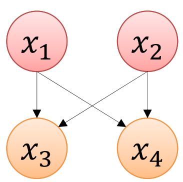


- x_3 の情報が無いとき, x_1 と x_2 は独立
  （x_3 の情報が所与のとき, x_1 と x_2 は独立とは限らない）
  P3の例：ダイレクトメールA or Bの送付(x_1)と店内キャンペーンの有無(x_2)は購買・非購買の情報(x_3)が無いときは独立。また, x_3 の情報が所与のとき x_1 と x_2 は独立とは限らない。例えば, ダイレクトメールB&店内キャンペーンの両方の施策を受けることで購買傾向が高まる交互作用があるとき, $x_3 = 1$ の顧客は $x_1 = B$ かつ $x_2 = 1$ となる傾向がある
- 同時確率：$\Pr(x_1, x_2, x_3) = \Pr(x_1)\Pr(x_2)\Pr(x_3|x_1, x_2)$
- 周辺確率：$\Pr(x_1, x_2) = \sum_{x_3}\Pr(x_1, x_2, x_3) = \Pr(x_1)\Pr(x_2)$

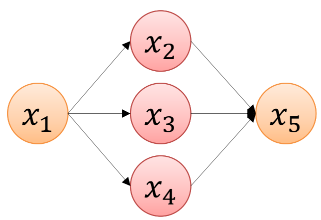
有向分離 #4

有向分離は変数の集合でも同様の性質をもつ

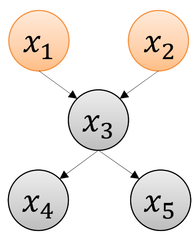
- 例 Tail-to-tail 型
  x_1, x_2 の情報が全て所与のとき, x_3 と x_4 は条件付き独立
- 例 Head-to-tail 型
  x_2, x_3, x_4 の情報が全て所与のとき, x_1 と x_5 は条件付き独立
- 例 Head-to-head 型
  x_3, x_4, x_5 の情報が全て無いとき, x_1 と x_2 は独立



Tail-to-tail 型



Head-to-tail 型



Head-to-head 型

ベイジアンネットワーク分析をやってみよう

ポルトガルの金融機関のダイレクトマーケティングデータ

- 2008年から2013年の実データ ($N = 45{,}211$)
- 消費者にマーケティングプロモーションを実施
- y：プロモーション成果(加入・非加入の結果), contact：連絡方法(自宅電話or携帯電話), job：職業(12種), marital：結婚(未婚, 既婚, 離婚), education：学歴(4種), default：債務不履行の経験の有無, housing：家のローンの有無, loan：ローンの有無
- Rのbnlearnパッケージを用いて分析
  プロモーション成果(y)への因果関係を分析したので, 変数 y からその他の変数へは矢線を引かないように設定
  BICスコアによるネットワーク探索

#メモ　UCI Machine Learning Repository (http://archive.ics.uci.edu/ml)でBank Marketing Data Setとして無料でダウンロード可能

Rによるベイジアンネットワーク分析

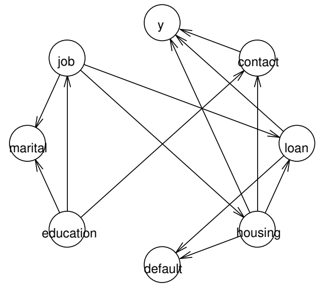
```
install.packages("bnlearn")
library(bnlearn)
d = read.csv("bank-full.csv",header=T, stringsAsFactors =
TRUE)

#利用する変数の抽出
data =
d[,c("job","marital","education","default","housing","loan",
"contact","y")]

#矢線をひかない変数の指定
#"y","job"はyからjobへ矢線を引かないという意味。それの
組合せの記述
bl =
matrix(c("y","job","y","marital","y","education","y","default
","y","housing","y","loan","y","contact","job","education"),
ncol = 2, byrow = TRUE)

#ネットワーク構造の可視化
res=hc(data,score="bic",blacklist=bl)
plot(res)
fitted = bn.fit(res, data)

#条件付き確率の計算
#Pr(loan="no"|y="no")の計算
cpquery(fitted, event = (loan == "no"), evidence = (y ==
"no"))
```

※Rではbnlearnやdealパッケージが有名である. ただし, ネットワーク構造を推定するアルゴリズムが異なるため, 利用するパッケージで推定結果が異なる場合もある

※Rのbnlearnパッケージ内のcpquery関数はベイジアンネットワークの条件付き確率を求める関数であるが, 乱数を用いた近似アルゴリズムのため実行する度に出力が異なる

Rによるベイジアンネットワーク分析の結果

Rの出力



ネットワーク構造の可視化

交互作用の発見

$\Pr(y = \text{yes}|loan = \text{yes}, housing = \text{yes}) \fallingdotseq 0.06$
$\Pr(y = \text{yes}|loan = \text{yes}, housing = \text{no}) \fallingdotseq 0.08$
$\Pr(y = \text{yes}|loan = \text{no}, housing = \text{yes}) \fallingdotseq 0.07$
$\Pr(y = \text{yes}|loan = \text{no}, housing = \text{no}) \fallingdotseq 0.18$

条件付き確率の推論

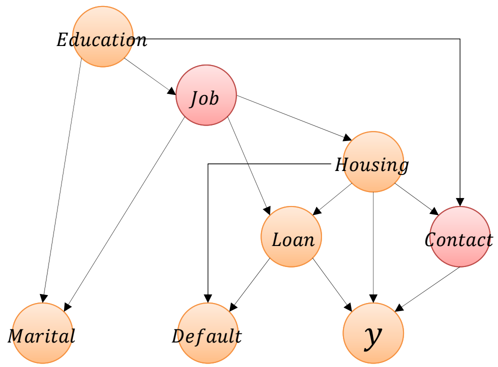
$\Pr(marital = \text{married}|y = \text{yes}) \fallingdotseq 0.59$
$\Pr(marital = \text{single}|y = \text{yes}) \fallingdotseq 0.31$
$\Pr(marital = \text{divorced}|y = \text{yes}) \fallingdotseq 0.12$

– ネットワークをたどることで, 無数にある変数の組合せの中から少ない労力で成果(y)に対する交互作用の発見に成功

#メモ　ポルトガルの金融機関について詳しい人であれば当然の知識の可能性もある。しかし、それ以外の人にとって、データだけからしらみつぶしにこの関係を見つけ出すのは相当の労力が必要。ベイジアンネットワークは知識の候補となる変数間の関係の発見をサポートしてくれる

分析結果と有向分離の使い方

有向分離の例



– Job と Contact の情報が所与の下では, プロモーション成果 y と Education の関係は条件付き独立となることがネットワーク構造から効率的かつ直感的に判断できる

補足：ベイジアンネットワークと因果関係

「因果関係をモデル化したベイジアンネットワークは変数間の確率的な因果関係を分析できる」は正しい。

「ベイジアンネットワークは変数間の因果関係のモデルであり, 変数間の確率的な因果関係を分析できる」は正しいとは限らない

重要な注意事項(再掲)

– <u>数式や確率表自体は原因と結果を区別&規定できない</u>
– 特に, データから自動的に推定されたベイジアンネットワークのネットワーク構造が因果関係のモデルである保証は無い. 分析者による解釈が必要&重要

演習問題

問題：プロモーション成果(y)とその他の変数を有向分離する変数群を答えなさい。また, その結果は, プロモーション成果(y)と有向分離された変数群とはどのような関係になるか答えなさい