■ 8.2 - 8.2.2（途中まで）の概要

8章の冒頭をおさらいすると、グラフィカルモデルを扱う利点は次のようなものだった。

1. 確率モデルの構造を視覚化でき、新しいモデルの設計方針を決めるのに役立つ。
2. グラフの構造を調べることで、条件付き独立性などの性質がわかる。
3. 複雑なモデルにおける学習や推論をグラフ上の操作として簡潔に表現できる。

このうち、1. の内容は8.1の節で学んだ。すなわち、複雑性・一般性の高い確率モデルであるほどリンクの本数が多く全結合に近いグラフィカルモデルになり、独立性などの制約が多い確率モデルであるほどリンクが少ないグラフィカルモデルになることを学んだ。

本節では、2. の条件付き独立性について学ぶ。グラフ構造を調べることで、解析的な調査なしで変数間の条件付き独立の性質を読み取ることができる。これを実現する枠組みは有向分離と呼ばれ、本節の最後で解説する。条件付き独立性を知ることにより、推論や学習に必要な計算を効率化することができる。これについては本節以降で学ぶこととなる。

■ 8.2 条件付き独立性

（概要と被る部分が多いのでパス）

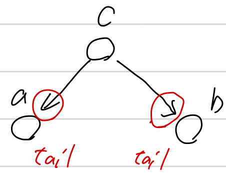
c が条件づけられた時の a と b の条件付き独立を次のような記法で表すこととする。

$$a \perp\!\!\!\perp b \mid c$$

■ 8.2.1 3つのグラフの例

本項ではノードが3つの場合に絞り、ノード a 、 b 、c がどのようなリンク構造の時に条件付き独立が成立するかについて解説する。そのために a から b へのリンクの経路を3つのケース（ tail-to-tail, head-to-tail, head-to-head ）に分類し、それぞれにおける独立性を解析的に検証する。

tail-to-tail の例



右図のグラフィカルモデルを考える。ノード a からノード b への経路を考えると、 c から伸びているリンクの矢先 (tail) がノードを繋いでいるため tail-to-tail と呼ばれる。このとき、同時分布は次のように分解できる。

$$p(a,b,c) = p(a|c)p(b|c)p(c)$$

周辺化をすると、

$$p(a,b) = \sum_c p(a|c)p(b|c)p(c)$$

これは一般に $p(a)p(b)$ と一致しないので、

$$a \not\perp\!\!\!\perp b \mid \emptyset$$

が言える。ここで、 $\emptyset$ は空集合, $\not\perp\!\!\!\perp$ は独立性が「一般には」成立しないことを意味する。

次に c の条件付けを考えると,

$$\begin{aligned} p(a,b|c) &= p(a,b,c)/p(c) \\ &= p(a|c)p(b|c) \end{aligned}$$

$$\Leftrightarrow \quad a \perp\!\!\!\perp b \mid c$$

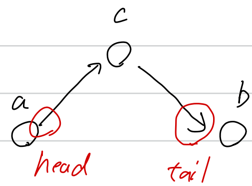
これより、条件付き独立が成り立つ.

このとき、ノード c は a から b への経路を遮断 (block) していると言う。

head-to-tail の例

同様の解析的操作によって、次の結果が得られる。

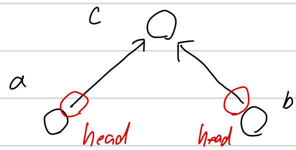
（証明は読み合わせ）



$$a \not\perp\!\!\!\perp b \mid \emptyset$$
$$a \perp\!\!\!\perp b \mid c$$

head-to-head の例

この例では前の2つと異なり、条件づけなかった場合のみ独立性が成り立つ。（証明は読み合わせ）



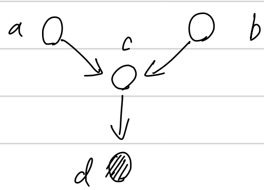
$$a \perp\!\!\!\perp b \mid \emptyset$$
$$a \not\perp\!\!\!\perp b \mid c$$

このとき、 c の条件付けによって a から b への経路の「遮断が解かれる(unblock)」と言う。

head-to-head については更に特徴的な性質があり、 **head-to-head ノードかあるいはその子孫のうちのいずれか**が観測されれば経路の遮断は解かれることを示すことができる。

（証明；演習8.10）

簡単のため、c の直結の子孫である d がのみが条件づけられた時の a と b の独立性を確かめる。



$$p(a,b) = \sum_c \sum_d p(a,b,c,d)$$
$$= \sum_c \sum_d p(a)p(b)p(c|a,b)p(d|c)$$
$$= p(a)p(b) \sum_c \sum_d p(c,d|a,b)$$
$$= p(a)p(b)$$

これより、何も条件付けないとき、経路 a, b は遮断されている。
次に d を条件付けたときを考える。

$$
\begin{aligned}
p(a,b|d) &= \sum_c p(a,b,c|d) \\
&= \sum_c p(a,b,c,d) / p(d) \\
&= \sum_c \frac{p(a)p(b)\,p(c|a,b)\,p(d|c)}{p(d)} \\
&= p(a)p(b) \sum_c \frac{p(c,d|a,b)}{p(d)}
\end{aligned}
$$

これは $p(a|d)p(b|d)$ に分解できないため、遮断が解かれている。 □

教科書では車の燃料装置の例をもって、「遮断が解かれる」例の直観的な解説をしている。

これまでのノードとの対応関係を整理すると

バッテリの状態B -> これまでのノード a

燃料タンクの状態F -> これまでのノード b

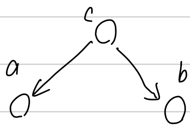
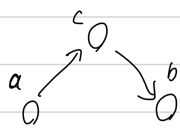
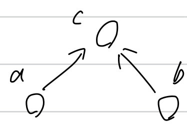
電動燃料計G -> これまでのノード c

として、

$$p(F|G) \neq p(F|G,B) \quad (\text{つまり } F \not\perp\!\!\!\perp B \mid G)$$

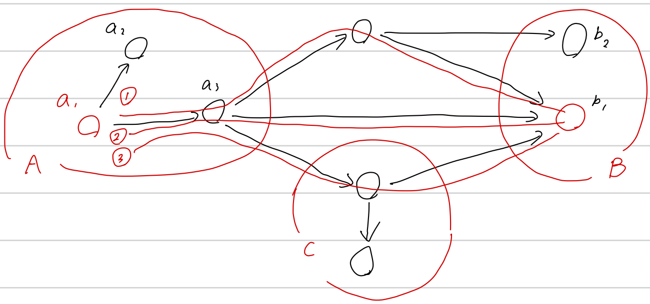
を示している。

これまでの例を表にまとめると次のようになる。

	tail-to-tail	head-to-tail	head-to-head
グラフ			
条件なし	$a \not\!\perp\!\!\!\perp b \mid \emptyset$	$a \not\!\perp\!\!\!\perp b \mid \emptyset$	$a \perp\!\!\!\perp b \mid \emptyset$
cの条件付き	$a \perp\!\!\!\perp b \mid c$	$a \perp\!\!\!\perp b \mid c$	$a \not\!\perp\!\!\!\perp b \mid c$

8.2.2 有向分離（D分離）

これまで3ノード a, b, c に関する条件付き独立を扱っていたが、この項ではノード集合 A, B, C についての条件付き独立 $A \perp\!\!\!\perp B \mid C$ を扱う。Aに属する任意のノードからBに属する任意のノードへの全ての経路が、Cにより遮断されているとき、CによってAとBは有向分離されているという。



a_1 から b_1 への全ての経路で遮断されているかを調べる。

a_1 から b_2 〃

⋮

a_3 から b_2 〃

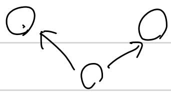
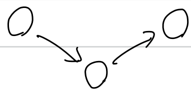
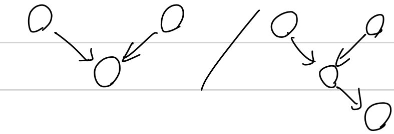
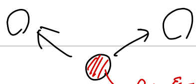
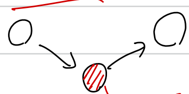
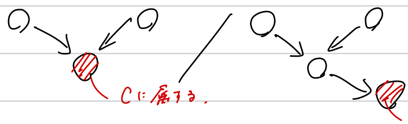
全て遮断されているとき、Cによって AとB は有向分離され、$A \perp\!\!\!\perp B \mid C$ である。

経路が遮断されているとは、経路上に次の条件を満たすノードが1つでも存在することである。

(a) 自身が集合Cに含まれ、経路において head-to-tail か tail-to-tail である。

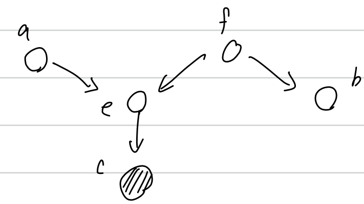
(b) 自身と子孫がいずれも集合Cに含まれず、経路において head-to-head である

次の表の中では、3つのパターンで経路が遮断されることとなる。

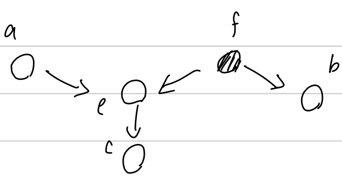
tail-to-tail	head-to-tail	head-to-head
		遮断 
遮断 	遮断 	

実際のグラフを見ながら、AとBがCによって有向分離されているかを検証してみる。

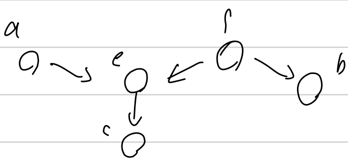
ここで、影のついたノードはCに含まれるノードで、$A \equiv \{a\}$, $B \equiv \{b\}$ とする。



: 有向分離されない

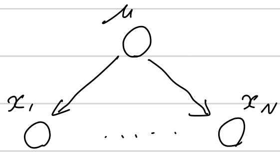


: 有向分離される.
eとfで遮断



: 有向分離される.
eで遮断.

1変量ガウス分布のモデルについて考える。平均 μ に従うガウス分布からデータ $D=\{x_1, x_2, \cdots, x_N\}$ が得られたとする。この時のグラフィカルモデルは右図のようになる。x_i , x_j の関係は μ において tail-to-tail となるので、μ が与えられたもとでの x_i と x_j は独立であるが、μ が観測されていないもとでの x_i と x_j は一般に独立ではない。

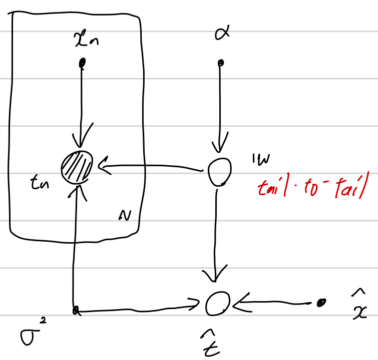


最後にベイズ多項式回帰についても考える。学習データのラベル t_n と未知データのラベル $\hat{t}$ の経路は、重みベクトル $\mathbf{w}$ において tail-to-tail のため、$\mathbf{w}$ が得られたもとでは

$\hat{t} \perp\!\!\!\perp t_n \mid \mathbf{w}$ が成立する。このことから、

$$p(\hat{t} \mid \mathbf{w}, \{t_n\}) = p(\hat{t} \mid \mathbf{w})$$

が成り立ち、一度 $\mathbf{w}$ の事後分布が求まれば、新しいデータのラベルの予測に訓練データが必要ないことがわかる。



次回の 8.2.2 の続きでは、条件付き独立の関係からモデル構造が簡単化されることをナイーブベイズモデルによって確かめる。また、8.3 節以降では学習や推論をグラフ上の操作として簡潔に表現する方法を学ぶ。