本レジュメの最後に付録として Appendix のページを添付している。適宜参照されたい。

1　条件付き独立性

1.1　有向分離（D 分離）

ノード変数の観測によってグラフィカルモデルの経路が遮断される、あるいは遮断が解かれることを確かめた。この遮断と条件付き独立との関係性を見た。ノード集合 A に属する任意のノードから B に属する全てのノードへの経路が C により遮断されているとき、ノード集合 A, B は C によって有向分離されていると呼ぶのであった。

この有向分離されているグラフィカルモデルの 1 例としてナイーブベイズモデルを思い出そう。

ナイーブベイズモデルにおける重要な仮定はクラス $\boldsymbol{z}$ で条件付けたとき、入力変数 $x_1, \ldots, x_D$ の分布が互いに独立となることである。逆に $\boldsymbol{z}$ を与えずこれについて周辺化すると x_i から x_j への遮断は解かれる。これは周辺分布 $p(\boldsymbol{x})$ が一般に $\boldsymbol{x}$ の成分について因数分解できなくなるということである。

ナイーブベイズモデルの簡単なアプリケーション例については上巻の p.44-45 を振り返ろう。ここでの事例は画像データと血液データをクラスで条件してしまえばそれぞれ独立として取り扱ってしまえるという話であった。

ナイーブベイズはその名の通り「ナイーブ[*1]」なモデルだがそれでもなお有用だという話。図 1.27[*2]キャプション内で言っている「条件付き密度の 2 つのピークは事後確立には影響しない」とはどういうことだっけ？ベイズの定理で条件付き密度から事後確率を求めるはずなのに「影響しない」ってなんでだっけ？

有向グラフは同時確率から条件付き確率の積への因数分解を表現する（8.1.3-4)。一方で有向分離規準を通じて条件付き独立性も表現する（p.91 中段の主張)。有向分離定理って何？紹介あったっけ？フィルタの話はやっぱり訳わからん。

極端な場合の説明は当たり前のことしか書いていないがここで何を伝えたい？

最後にマルコフブランケットの概念を紹介。あるノード $\boldsymbol{x}_i$ を孤立させるためのノード集合の最小単位である。共同親ノードの観測を与えることで遮断が解かれることを防ぐことができる。でもだから何？

2　マルコフ確率場

本節では無向グラフィカルモデル（マルコフ確率場、マルコフネットワークとも呼ぶ）を導入する。このモデルではノード間のリンクに向きが存在しないため、条件付き独立と変数観測による遮断の振る舞いが有向グラフの場合と異なる[*3]。また同時分布が局所的な関数の積で表現される際に、その積が有向グラフの場合とどのように異なるかについても議論する。

[*1] 条件付き独立の仮定が強いことによる。これは図 8.24 を見れば明らかである。全結合のグラフとは程遠いほどリンクが少ない。
[*2] 上巻 p.43
[*3] 独立性とグラフ構造の意味づけのうち後者を工夫してやるという話が続いていく。

2.1　条件付き独立性

有向グラフでは head-to-head の存在により「遮断」の定義がちょっとめんどくさかった[*4]。そこで無向グラフを用いる、すなわちリンクの方向性を除去することで head-to-head ノードのような直感に反する現象が起こらなくなるのである[*5]。

無向グラフにおいてノード集合 A, B 及び C を考える。条件付き独立性

$$A \perp B \mid C \tag{1}$$

が成立するかどうか判断したい[*6]。これが成立するためには「集合 A に含まれるノードと集合 B に含まれるノードを結ぶ全ての可能な経路を考えたとき、それらの経路が全て集合 C に含まれるノードの少なくとも 1 つを必ず通ること」が必要になる。逆に遮断されない経路が 1 つでもあれば条件付き独立性は[*7]満たされなくなってしまう。このことを図 8.27 をベースに確認しよう（See App.1)。

さらに条件付き独立性をもっと簡単に調べる方法がありこれも併せて確認したい（See App.2)。

2.2　分解特性

ここでは条件付き独立性と因数分解との対応性について考える。同時分布 $p(\boldsymbol{x})$ がグラフの局所的な変数集合上の関数の積としてどのように表せるか考える必要がある。まず以下では「局所的」の意味を理解してもらう。

1 つのリンクで直接接続されていないノード変数 x_i と x_j を考える。これらの変数は他の全変数が与えられたもとで条件付き独立でなければならない。この 2 ノード間には接続がないため他の変数を観測することで全経路について遮断が起こり条件付き独立が導かれる。この条件付き独立性は次の表式で書ける。

$$p(x_i, x_j \mid \boldsymbol{x}_{\setminus\{i,j\}}) = p(x_i \mid \boldsymbol{x}_{\setminus\{i,j\}})p(x_j \mid \boldsymbol{x}_{\setminus\{i,j\}}) \tag{2}$$

$\boldsymbol{x}_{\setminus\{i,j\}}$ は全変数集合から x_i, x_j を除いたものである。x_i, x_j が同じ因子に含まれないように因数分解をすれば条件付き独立性が成り立つ。

ここでクリークと呼ばれる概念を導入すると理解の助けになる。この概念について図 8.29 を使って考えてみよう。

同時分布を因数分解した時の各因子をクリークが含む変数集合の関数にすれば良い。もっと言えば極大クリークに限定してしまっても一般性を失わない。

クリークを C と書きクリーク内の変数集合を $\boldsymbol{x}_C$ と書く。このとき同時分布はグラフの極大クリーク上の

[*4] 変数を観測することで独立性が失われてしまうという話。head-to-tail と tail-to-tail では変数の観測によって条件付き独立性が生まれていたのに。

[*5] 所感。初め無向グラフに対して難しい概念として捉えようとしてしまっていた。しかしながら実際に手を動かしてグラフを書いてみるともはや当たり前のことしか主張していないことに気付ける。もはや直観的を飛び越えて直感的でさえある。理解するためのハードルを勝手に上げてはならないという反省をここに記しておく。初め「リンクがつながっていないので独立性があるって当たり前じゃないか、そんなレベルの理解で良いのか」と疑ってしまったが「良いんですよ」と言ってあげたい。

[*6] 独立マークが教科書式のスタイルで出力できないため「垂直」マークで代用させてもらう。読み替えて理解してほしい。

[*7] 集合 A と集合 B の間について、という意味。

ポテンシャル関数[*8]$\psi_C(\boldsymbol{x}_C)$ の積の形で表現できる。

$$p(\boldsymbol{x}) = \frac{1}{Z}\Pi_C\psi_C(\boldsymbol{x}_C) \tag{3}$$

ここで Z は規格化定数であり以下のように与えられる。

$$Z = \sum_{\boldsymbol{x}}\Pi_C\psi_C(\boldsymbol{x}_C) \tag{4}$$

これは分配関数とも呼ばれる[*9]。規格化定数の存在が計算負荷のうえでモデルの弱点となりうる話は読み合わせ。ただし局所的な条件付き分布を計算する際には分配関数の計算は不要になる。その理由を具体例で確かめよう。ここは具体例を作って示したかったができず...

ここでようやく無向グラフの条件付き独立性と因数分解の関係について定式化を行う。と書いたものの教科書の定式化って別に定式化になってなくね？という話。教科書が言いたいことは「グラフの見た目から読み取れる条件付き独立性を満たす変数で作った分布の集合」イコール「(8.39) で書ける分布の集合」であると理解している。これってただ「Hammersley-Clifford の定理」を使えばそれが言えるってだけのことしか主張していないように見えるんだけど私間違ってる？要するに教科書内で閉じた説明、定式化にはなってないでしょというツッコミです。

ポテンシャル関数を狭義に正と約束しておく。このとき指数関数で表現してやると物理的に自然な表式が得られる。

$$\Pi_C\psi_C(\boldsymbol{x}_C) = \exp\{-E(\boldsymbol{x}_C)\} \tag{5}$$

このように表現するとポテンシャル関数の積を取ったときに指数の肩内はエネルギーの和で書くことができてこれは物理的に自然な記述となる。ちなみにこの指数関数の表現はボルツマン因子[*10]と呼ばれる。

ポテンシャル関数の選択には任意性がある。それゆえに対象とするアプリケーションごとに適切なポテンシャル関数を持ってくる必要がある[*11]。次回は画像ノイズ除去の事例において無向グラフの考え方を適用してみる。Ising model のハミルトニアンを導入することでこの問題を解決する手法について学んでいく[*12]。

[*8] ポテンシャル関数は物理学でよく現れる量となっている。ポテンシャルは基本的には距離の関数であって代表的なものでいうと万有引力ポテンシャル（力学）やクーロンポテンシャル（電磁気学）などが存在する。これら 2 つは距離に関して逆べきの 2 乗に比例する形式となっている。

[*9] 分配関数は統計力学で頻繁に顔を出す量である。しかしながら今回はあまり物理的な意味を考察することはしない。ここでもおそらく規格化の手法として導入したに過ぎないと思っている。ただし先の注釈で述べたようにポテンシャル関数は物理的な現象を記述するための関数であって確率密度分布の表式ではない。ゆえに今回は明示的に規格化定数として導入されたものと解釈している。有向グラフの場合は条件付き分布（＝確率密度分布）が規格化されることにより自動的に同時分布も規格化されていた。無向と有向とでは事情が異なる、つまり取り扱う因子が違うよという話である。

[*10] これは教科書の表現が誤りと思う。ボルツマン分布は $\frac{\text{（ボルツマン因子）}}{\text{（ボルツマン因子の総和）}}$ のように比として定義するはず。なお因子の総和はエネルギー状態全てについて和を取ったものと理解されたい。

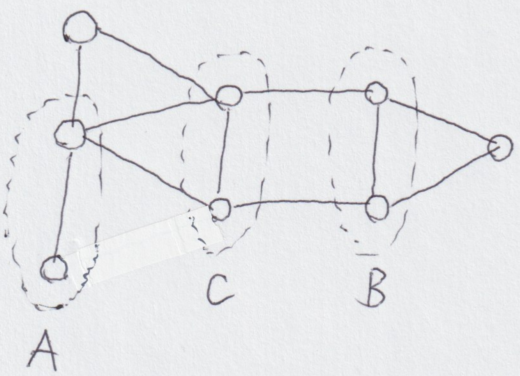
[*11] 繰り返しになるが同時に適切な規格化定数を持ってくる必要もある。

[*12] もうこれ物理学科出身じゃないと何言ってるかさっぱりだと思う。

App. 1 無向グラフにおける条件付き独立性

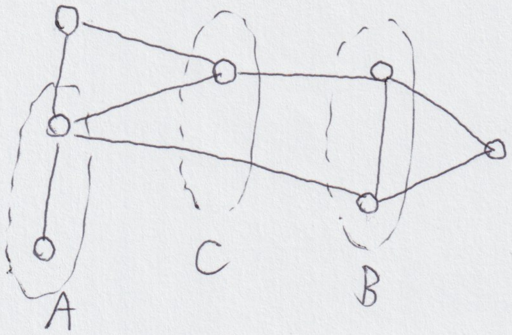
ノード集合AとBについて条件付き独立性をグラフの形から判断する。

Case. 1



AB間の経路上で必ず
Cに含まれるノードを少なくとも1つ通る。
このときA, Bは「遮断される」と考える。
A, Bの独立性が満たされる。

Case. 2

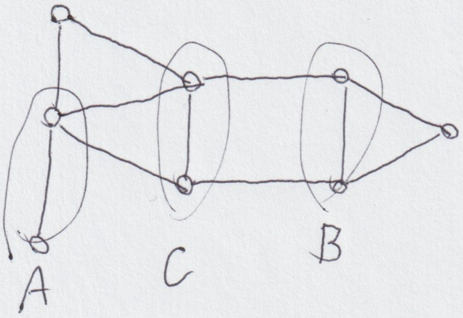


AB間が直接結ばれるような
経路が存在する。
このときA, Bは「遮断されない」と考える。
A, Bの独立性が満たされない。

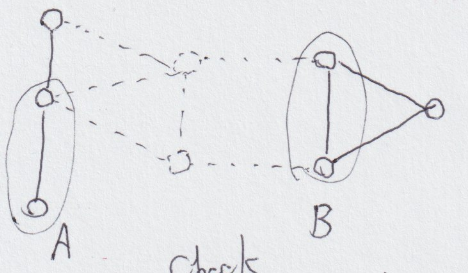
App.2 無向グラフにおける条件付き独立性を簡便に確かめる方法

ノード集合A, B, Cとあったときに Cを消去したとき AとBのつながりが
どうなっているか見てやる.

Case.1

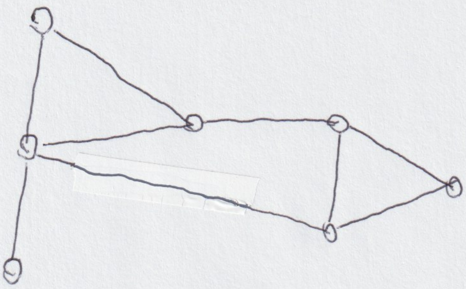


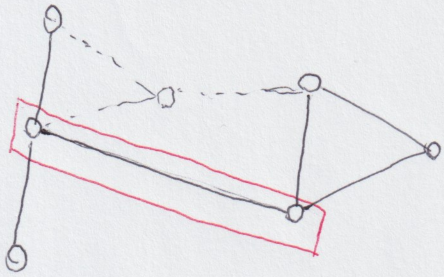
消去
✓ C内のノード全て
✓ これに接続するリンク全て



Check
◎A, B内の任意のノードどうしを
つなぐリンクはあるか?
▷ No! 条件付き独立性OK!

Case.2





Check
◎リンクある?
▷ Yes! 条件付き独立性NG!