断り：本レジュメの最後に App. として表を付けている。必要に応じて参照する。

1　ベイジアンネットワーク

前回はグラフィカルモデルとはなんぞやということを学んだ。グラフィカルモデルを学ぶ意義としては「確率モデル、特に確率変数同士の条件付きや独立性を視覚的に捉えることができるようになる」というところにある[*1]。視覚的に、つまりグラフとして図示できるようになり、その図を書くための定義やルールについても前回導入を行った。

今回はそれらの導入した定義やルールを踏まえて、これまで取り扱ってきたような確率分布に従う変数を具体例にとりグラフィカルモデルとして落とし込んでいくことにする[*2]。

1.1　生成モデル

冒頭で書いた内容とは逸れるパートとなる。全くわけわからんので一緒に解読。

1.2　離散変数

親ノードと子ノードが離散変数同士、あるいはガウス変数同士である場合には、ノード同士の関係を階層的に拡張していくことで任意の複雑な有向非循環グラフを構築することができる。特に今回はモデルのパラメタを減らすということがグラフ上においてどのような操作に対応しているのか理解してほしい。まず離散変数の場合を考える。

K 状態の離散変数 $\boldsymbol{x}$ を考える。この $\boldsymbol{x}$ は次のような確率分布に従う。

$$p(\boldsymbol{x} \mid \boldsymbol{\mu}) = \Pi_{k=1}^{K} \mu_k^{x_k} \tag{1}$$

この K 状態離散変数が $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ と M 個ある場合について任意の同時分布を表したい。考えるポイントは次の通りである。

・離散変数同士の独立性、つまりノード変数の関係性は？

・ノード結合の違いによるモデルのパラメタ数の変化は？

・ノード結合が異なる各場合についてグラフィカルモデルの形は？

まず $M = 2$ の簡単な場合について確率分布の式とグラフとの関係性を読み合わせながら確かめよう。そのあとで一般の M に対して上記のポイントを App.1 頁の表【ノード結合とグラフィカルモデル〜離散変数ノード〜】に整理する。各項目について簡潔に結論のみ示しており、細かい背景やロジックについては読み合わせながら確認しよう。

パラメタ数を減らすための手法として

・パラメタの共有

・ディリクレ分布の導入（パラメタの事前分布として）

・シグモイド関数の導入（2 値変数の条件付き分布として）

3 点が紹介されているがいずれも読み合わせとする。

[*1] 知っての通りベイズの定理やら周辺化の操作をこねくり回して計算をするのは大変であった。従来はこの複雑な計算を経て変数同士の関係性を見てきたが、グラフに落とし込むことで「わかりやすく」モデルを理解できるようになる。

[*2] ただしサブセクション「生成モデル」はちょっと別内容になる。

1.3　線形ガウスモデル

今回はノード変数が連続的、特に線形ガウスモデルに従う場合について同時分布とグラフとの関係性を見ていく。

まず D 個の変数上の任意の有向非循環グラフを考える。各ノード i はガウス分布に従う連続確率変数 x_i を表す。このとき線形ガウスモデルにおける各ノードの条件付き分布は（8.11）のように表すことができる。

今我々が知りたいのはこれらの条件付き分布から構成される同時分布について

・分布の形は？

・期待値や分散は？[*3]

というポイントである。前者については条件付き分布の積を取ることで同時分布が多変量ガウス分布になることが簡単に確かめられる。以下では後者について詳しく説明しよう。

今、各ノードがその親ノードよりも大きい番号を持つようにノードが順序づけされていると仮定する。このとき同時分布の期待値も分散も「再帰的な」手続きを踏むことによって計算することができる。期待値を求めるための導入や流れは教科書の説明を参考にされたい。次に、期待値が求められたという前提で共分散行列を求めることを考える。ノード x_i, x_j の共分散行列は以下の形で表せる。

$$
\begin{aligned}
cov[x_i, x_j] &= E[(x_i - E[x_i])(x_j - E[x_j])] \\
&= E[(x_i - E[x_i])\{\sum_{k \in pa_j} w_{jk}x_k + b_j + \sqrt{v_j}\epsilon_j - (\sum_{k \in pa_j} w_{jk}x_k + b_i)\}] \\
&= E[(x_i - E[x_i])\{\sum_{k \in pa_j} w_{jk}(x_k - E(x_k))\} + \sqrt{v_j}\epsilon_j] \\
&= \sum_{k \in pa_j} w_{jk}E[(x_i - E(x_i))]E[(x_k - E(x_k))] + E[(x_i - E(x_i))\sqrt{v_j}\epsilon_j] \\
&= \sum_{k \in pa_j} w_{jk}cov[x_j, x_k] + E[\{\sum_{j \in pa_i} w_{ij}x_j + b_i + \sqrt{v_i}\epsilon_i - (\sum_{j \in pa_i} w_{ij}E(x_j) + b_i)\}\sqrt{v_j}\epsilon_j] \\
&= \sum_{k \in pa_j} w_{jk}cov[x_j, x_k] + E[\{\sum_{j \in pa_i} w_{ij}(x_j - E(x_j)) + \sqrt{v_i}\epsilon_i\}\sqrt{v_j}\epsilon_j] \\
&= \sum_{k \in pa_j} w_{jk}cov[x_j, x_k] + E(\sqrt{v_i v_j}\epsilon_i\epsilon_j) \\
&= \sum_{k \in pa_j} w_{jk}cov[x_j, x_k] + E(\sqrt{v_j^2}\epsilon_i\epsilon_j) \quad \because (i = j \text{ のみ残る}) \\
&= \sum_{k \in pa_j} w_{jk}cov[x_j, x_k] + I_{ij}\sqrt{v_j} \qquad (2)
\end{aligned}
$$

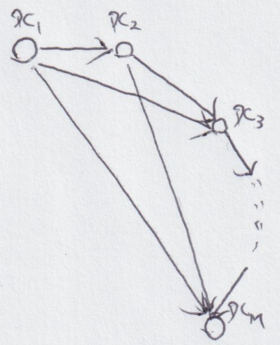
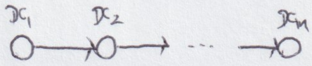
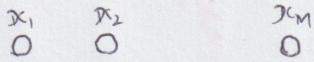
ここで I_{ij} は単位行列 i, j 成分である。このように多変量同時ガウス分布について各成分の期待値や共分散を番号の小さいノードから順に再帰的に計算を行うことで計算することができる。

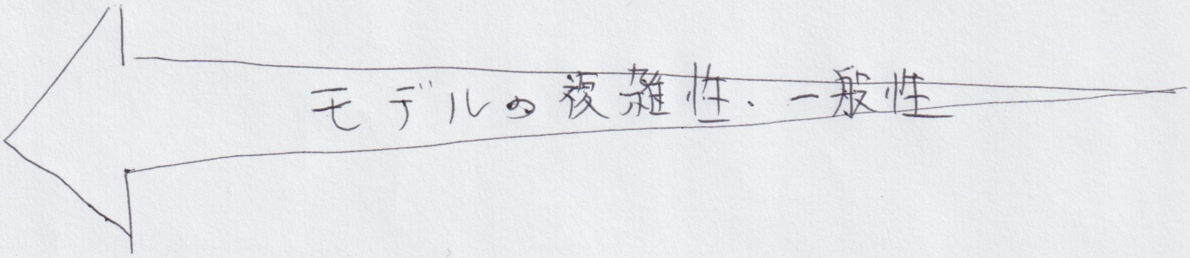
先ほど App. 1 の表で整理したのと同じようにパラメタの取り方とグラフの形との関係性を表に整理しよう (参照：App. 2 表【ノード結合とグラフィカルモデル〜線形ガウスモデル〜】)。

最後に「1 つのノードが多変量変数を持つ場合への拡張」、「階層ベイズモデルの考え方」を読み合わせて終わりとしたい。

[*3] もちろん存在すれば、という前提ではある。

App.1 ノード結合とグラフィカルモデル 〜離散変数ノード〜

	全結合	連鎖	リンクなし
グラフ			
同時分布関数	$p(x_1)p(x_2\mid x_1)p(x_3\mid x_2,x_1)\cdots p(x_M\mid x_{M-1},\cdots,x_1)$	$p(x_1)p(x_2\mid x_1)\cdots p(x_M\mid x_{M-1})$	$p(x_1)p(x_2)\cdots p(x_M)$
同時分布のパラメタ数	K^M-1	$K-1$ + $(M-1)K(K-1)$	$M(K-1)$
パラメタの増大具合	Mに関して指数的	Kに関して2次 Mに関して線形	K, Mに関して線形



App. 2 ノード結合とグラフィカルモデル ～線形ガウスモデル～

	全結合	連鎖※1	リンクなし
グラフ	離散変数の場合と同じ（順番同）		
パラメタ行列W（D×D）	$\begin{pmatrix} 0 & & & O \\ W_{21} & \ddots & & \\ W_{31} & W_{32} & \ddots & \\ \vdots & \vdots & \ddots & \\ W_{D1} & \cdots & W_{D,D-1} & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & & & O \\ W_{21} & \ddots & & \\ 0 & W_{32} & \ddots & \\ \vdots & O & \ddots & \\ 0 & \cdots & W_{D,D-1} & 0 \end{pmatrix}$	Wはない（bとvは存在）
同時分布のパラメタ数	$\frac{1}{2}D(D-1) + D$	$2D-1$ ※2	$2D$

← モデルの一般性・複雑性

※1：教科書で「連鎖」という表現はされていないが、ここでは便宜上離散変数での議論と同じ言葉を使わせてもらう。

※2：間違っていたらゴメン。リンクなしの場合よりパラメタ数が少ない。これは連鎖としたことで行列Wが表内のような形になってしまうことによる。もっと多くのリンクの存在を仮定すればパラメタ数は2D（リンクなしの場合）よりも多くなる。