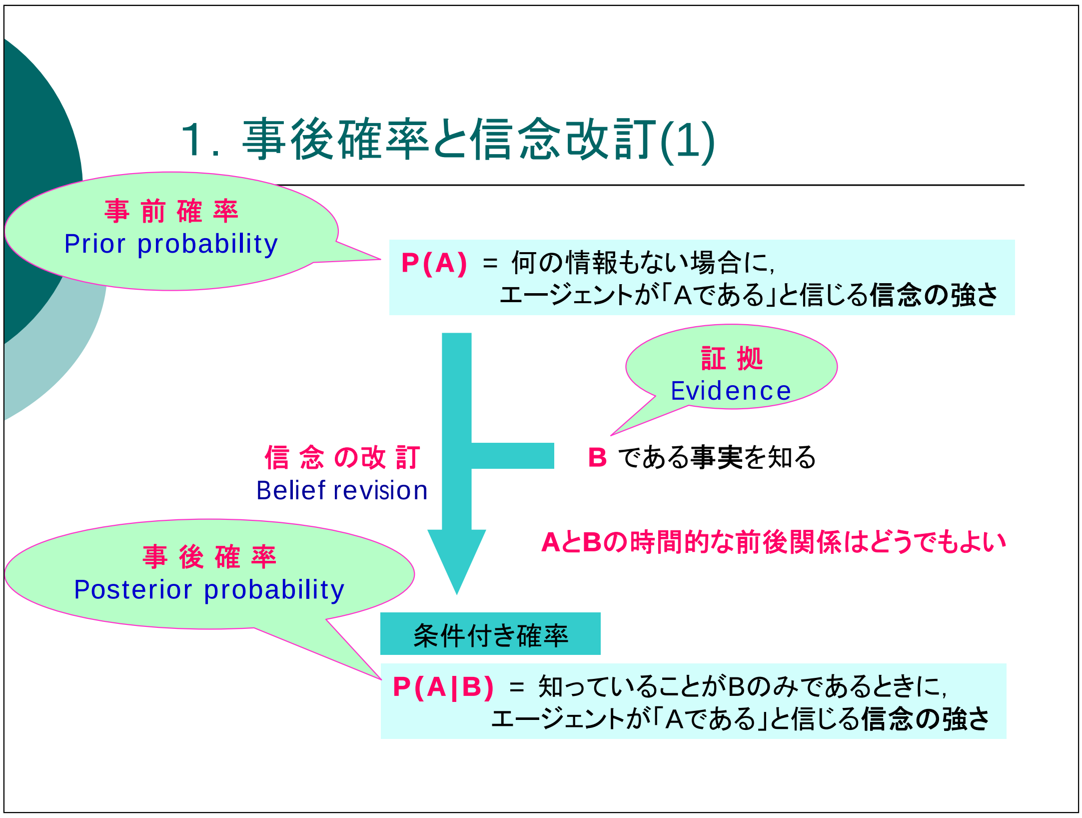
知能情報処理　不確実性推論（2）
不確実な情報に基づいて行動するエージェント

ベイジアンネット

- 事後確率と信念改訂
- 結合分布からの事後確率計算
- ベイズルール
- ベイジアンネット
- λ - π メッセージ転送アルゴリズム



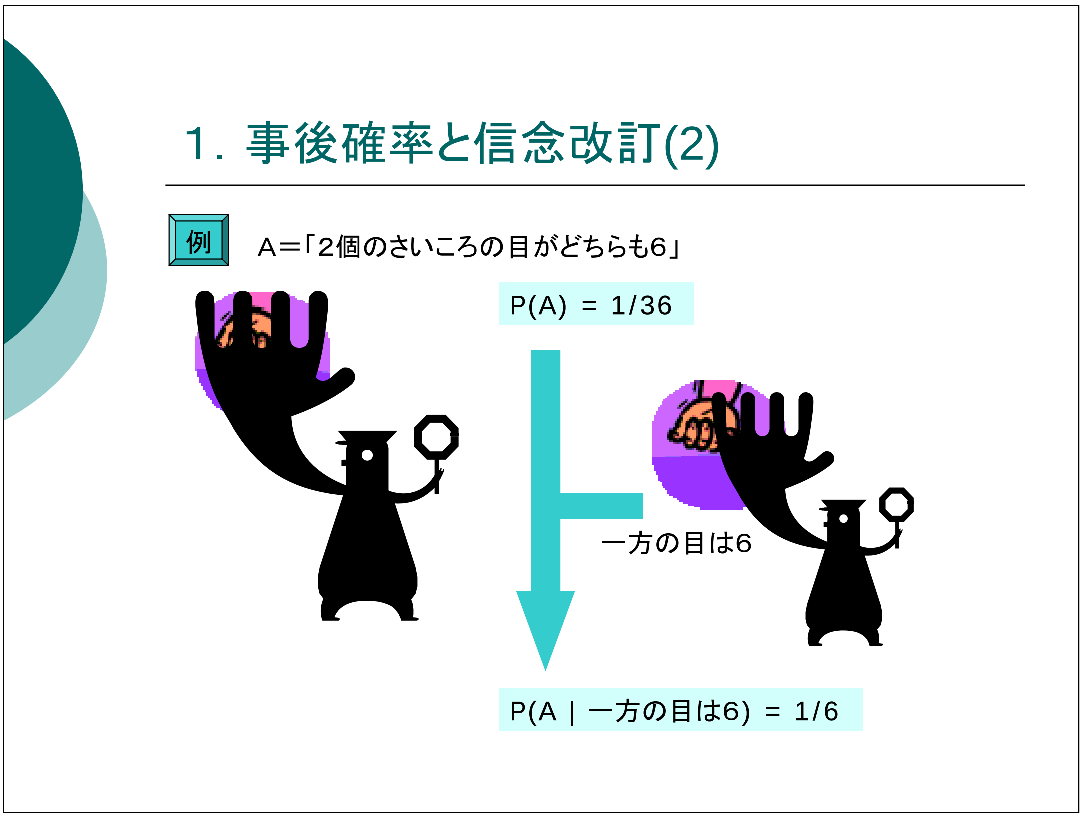
不確実な情報を確率を用いてモデル化し，エージェントの意思決定に結びつけるための基礎技術を学ぼう．

純粋数学においては，「確率」はある種の公理を満たす数値として抽象的に定義されているだけで，応用する上での解釈は示されていない．よくある解釈は，サイコロの1の目が出る確率が1/6であるというのは，「6回のうち1回はその目が出ること」というように，試行を何度も繰り返して得られる統計的な解釈と結びつけている．しかし，「明日，雨が降る確率は20％」という場合，明日という日は歴史上でただ1日しかなく，試行を繰り返せるものではないし，全く同じ気象条件の日がこれまで何度もあったという解釈もやや苦しい．

人工知能で確率を応用する場面では，確率P(A)は，何の情報もない場合に，事前にエージェントが「Aである」ことを信じている**信念の強さ**を表していると解釈する．いま，エージェントが「Bである」ことを知ったとしよう．すると，その知識を知った事後においては，エージェントがもつ「Aである」と信じる信念の強さが変わってくるかもしれない．これを**信念の改訂**という．その新しい信念の強さを，**条件付き確率**P(A|B)によってモデル化する．すなわち，

P(A|B) = 知っていることがBのみであるときに，エージェントが「Aである」と信じる信念の強さ

P(A)を**事前確率**，Bを**証拠**，条件付き確率P(A|B)のことを**事後確率**と呼ぶ．



1つ前のスライドの事後確率P(A|B)の解釈においては，AとBが生起する時刻の前後関係については何も言っていないことに注意しよう．良くある状況設定は，Bは過去に生起した事象であり，Aは将来に生起するかもしれない事象である．Aは将来のことなので，Bという知識を利用して，確率的に予測するしかないわけである．

反対に，Aが過去に生起したかもしれない事象で，Bが現時点で知り得た事象であるような応用も多い．つまり，現在の知識から過去を推定するわけである．たとえば，犯行現場に残されたBという証拠を知ったうえで，犯行を犯した者がAであることの信念の強さをこれでモデル化できる．Bという証拠が発見されたとき，この条件付き確率の計算に基づいて，エージェント(探偵かもしれない)はAが犯人であるかどうかについての**信念を改訂**することができる．

これから学ぶ理論はどちらのタイプの応用にも使うことができるが，例題では後者のような応用を設定している．

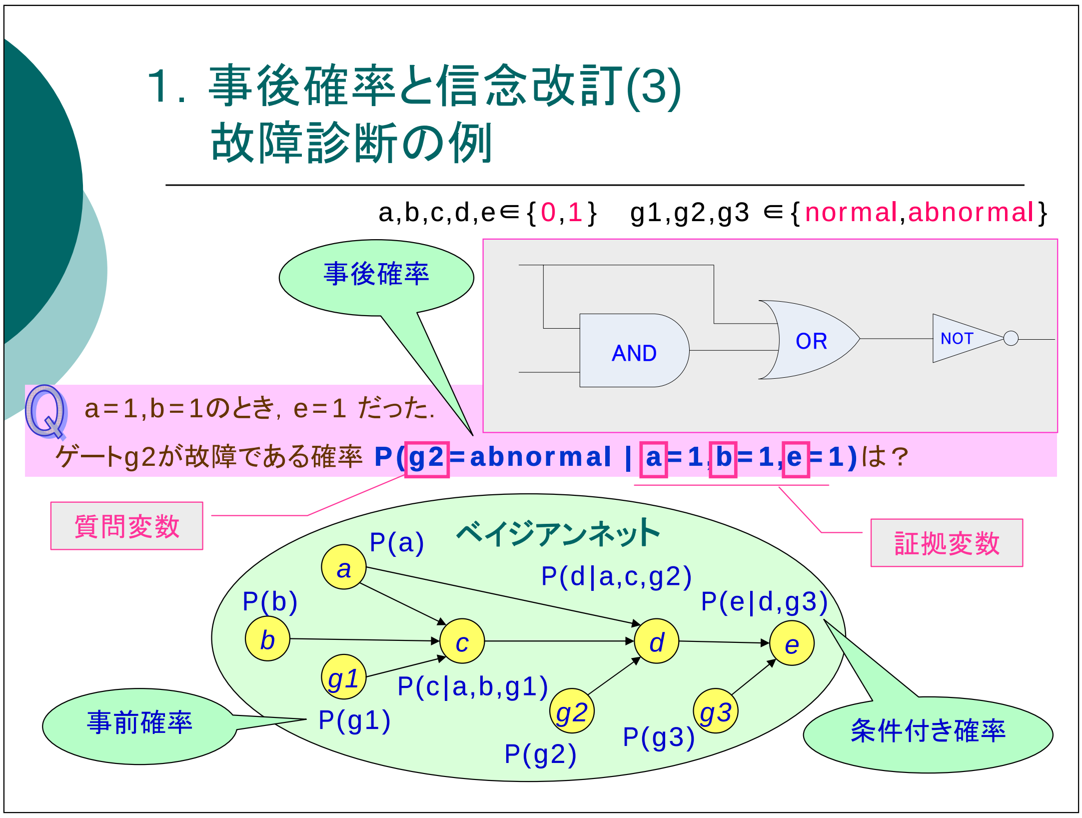
最初の簡単な例として，「2個のさいころの目がどちらも6である」という事象をAとしよう．

何も情報がないとき，Aの事前確率はP(A)=1/36である．

ここで，一方の目が6であることをエージェントが知ったとする．このとき，Aの事後確率は，

$$P(A \mid \text{一方の目は6}) = 1/6$$

となり，信念が改訂される．



システムの確率モデルの例として，簡単な論理回路の故障診断を考えてみる．

スライドの回路で，a,b,c,d,eは信号の値を表す確率変数で，0または1の値をとる．g1,g2,g3は，各ゲートが正常か異常かを表す確率変数である．

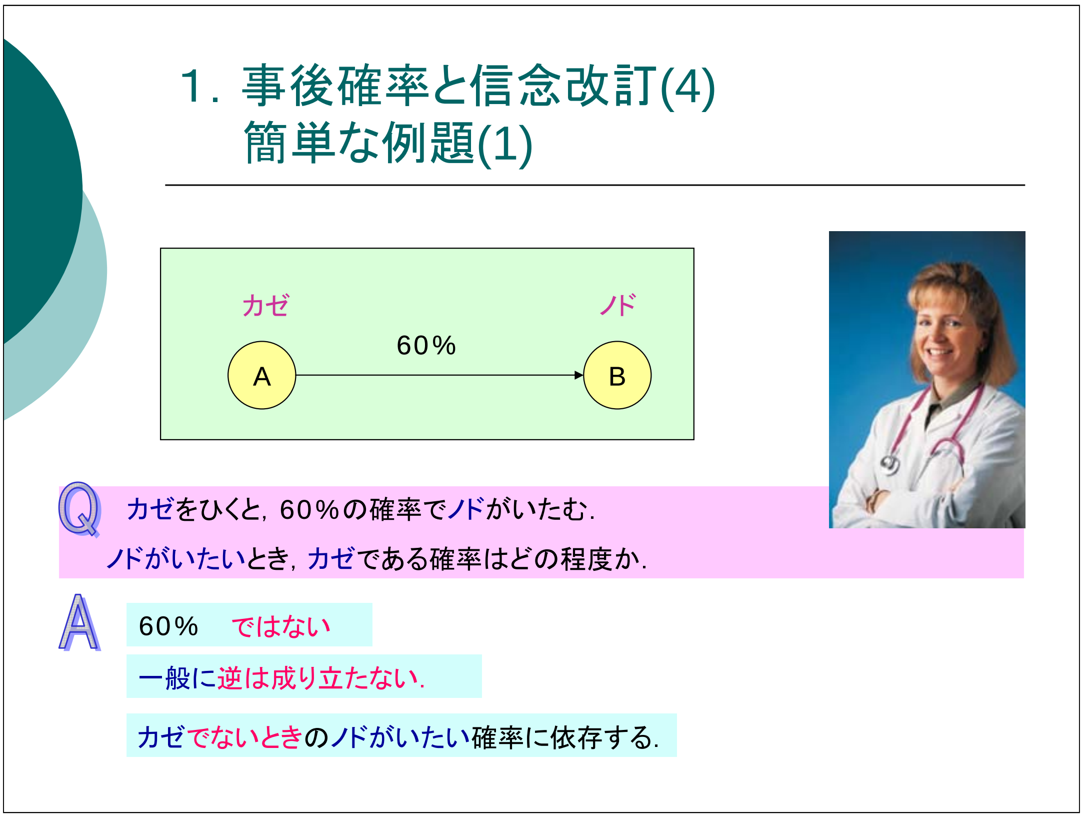
これらの確率変数をノード（頂点）とし，それらの間の**因果関係**をアーク（有向辺）で表したグラフを描くと図のようになる．たとえば，eはdとg3の値に依存して決まるので，dとg3からeにアークが付けられている．

さらに，この構造に確率的な情報を付与する．具体的には，いま見たような直接的な依存関係ごとに条件付き確率分布を与える．たとえば，いま見たeとd,g3の依存関係は，条件確率分布P(e|d,g3)で表現する．すなわち，d=0,1とg3=normal.abnormalの4通りの組合せごとにe=0,1である確率をそれぞれ与えるわけである．

さらに，入ってくるアークが1本もないノードについては，事前確率を与えることとする．

このように，グラフと条件付き確率分布によって構築された確率モデルが**ベイジアンネット**と呼ばれるものである．その正式な定義は後に学ぶ．

ベイジアンネットが与えられているとき，新たにわかった証拠から事後確率を計算するアルゴリズムが知られている．それを用いると，たとえば，スライドに書かれたような質問に対する解を効率良く求めることができる．証拠を表す確率変数を**証拠変数**，事後確率を知りたい確率変数を**質問変数**という．



事後確率の計算の基礎を理解するために，簡単な例題を解いてみよう．(ただし，データは仮想的なものなので，現実生活の知識とはしないでください．)

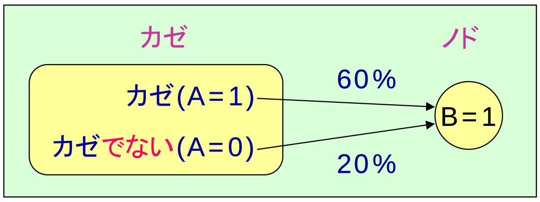
Q．カゼをひくと，60%の確率でノドがいたむ．ノドがいたいとき，カゼである確率はどの程度か．

この答えは，60%ではない．

「カゼならば，ノドが痛い」が成り立っても，一般に，その逆「ノドが痛いならば，カゼである」が成り立たないのと同じ理由である．ノドが痛いからといって，カゼの確率は60%とはならない．

このQは，「解答不能」である．少なくとも，「カゼでないときにノドが痛い確率」と知らないと，答えを出せないのである．極端な場合，この確率が0なのか1なのかで結論がかなり異なることを直観的に想像してほしい．

1. 事後確率と信念改訂(5)
簡単な例題(2)



Q カゼのときには，60%の確率でノドがいたい．
カゼでないときには，20%の確率でノドがいたい．

ノドがいたいとき，カゼである確率はどの程度か．

A 60÷(60+20)=3/4=75%　ではない

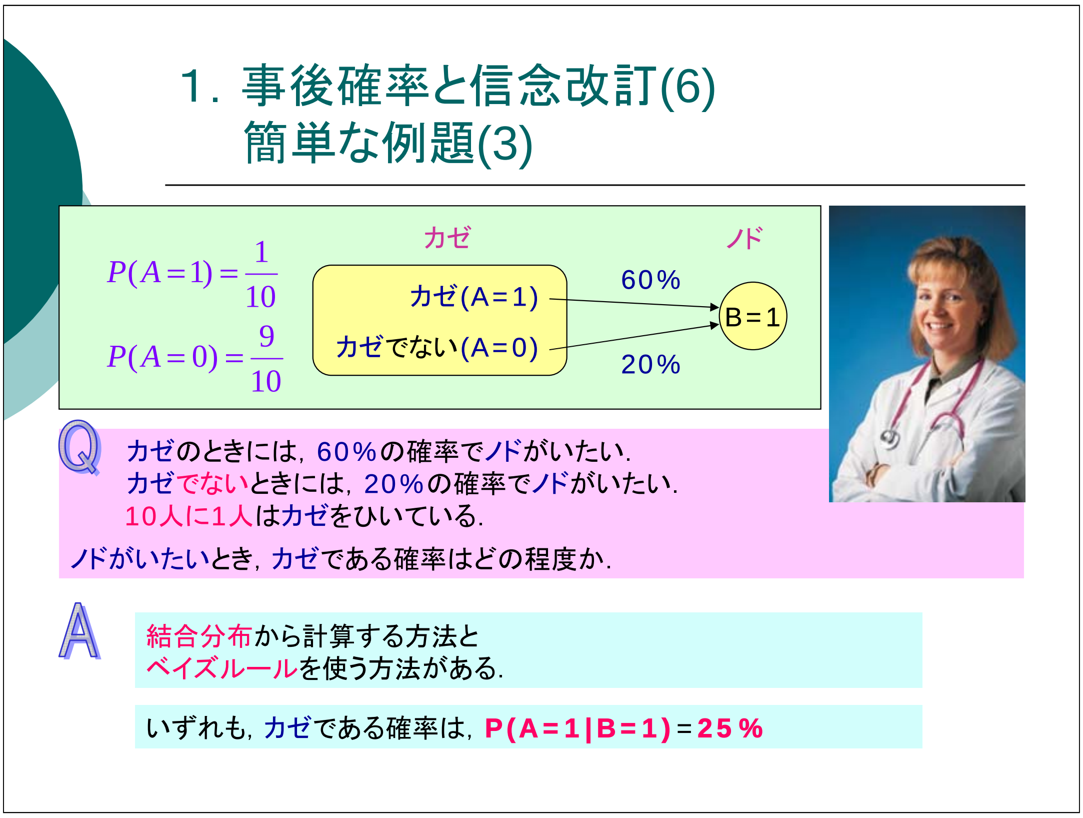
ここは南国の楽園で「カゼは希有な病」かもしれない．
カゼの事前確率に依存する．

この例題に，「カゼでないときには，20%の確率でノドがいたい」という情報を追加した．今度は答えを出せるだろうか．

何となく 60÷(60+20)=3/4=75%とする答えは，確率論的な根拠がなく，誤りである．

しかし，ここは南国の楽園で，カゼはめったに見られない希有な病で，ここ数百年，カゼをひいた住人はいないとしたらどうだろう．カゼにかかると60%の確率でノドが痛いといっても，そんな希有な病になったとは誰も思わないだろう．カゼでなくても20%の確率でノドが痛むものだからある．

つまり，このQにはまだ答えられない．カゼである事前確率を知らないと答えを計算できないのである．



そこで，「10人に1人はカゼをひいている」というカゼの事前確率を追加した．

これで答えを計算することができる．その方法として，**結合分布**から計算する方法と**ベイズルール**を使う方法がある．次のスライド以降でその両方を見ていくことにする．

2. 結合分布からの事後確率の計算(1)
結合分布

X

確率変数 (random variable)
授業では「**命題**」を表わすのに限定.
値は **0** または **1**

$$P(X=0)+P(X=1)=1$$

$$P(X)=(p_0,p_1)$$

確率分布 (probability distribution)
Xについての関数. 表で与えられる.

$$P(X=x,Y=y)$$

結合確率 (joint probability)
同時確率 (simultaneous probability)

$$\underbrace{P(X,Y)}$$

結合分布
(joint probability distribution)

すべての変数の取り得る値の
すべての組合せに対する確率の表

このスライドは確率論の基本事項をまとめたものである. 特に重要なのは, **結合分布**である. これは, いま考察の対象としている確率モデルに現れているすべての変数の取りうる値のすべての組合せごとに確率を関数または表で与えるものである.

2. 結合分布からの事後確率の計算(2)　周辺分布

$$P(X) = \sum_{Y} P(X, Y)$$

周辺分布
(marginal distribution)

結合分布

P(X,Y)	Y=0	Y=1	P(X)
X=0	0.1	0.2	0.3
X=1	0.3	0.4	0.7
P(Y)	0.4	0.6	

周辺分布 P(X) は，パラメータで指定された変数 X だけを残し，それ以外の変数 Y の取りうる値の組合せのすべてについての結合確率の和である．スライドのような2次元の表の場合には，行方向の和，あるいは列方向の和がそれに当たる．通常，それらの和はこのスライドのように表の右端や下端などの「周辺」に表示されるので，周辺分布と呼ばれている．

XやYが複数個あるような多次元の場合も，これを自然に拡張して考える．

2. 結合分布からの事後確率の計算(3) 条件付き確率

条件付き確率 (conditional probability)

$$P(X=x\mid Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}$$

条件付き確率分布

$$P(X\mid Y)=\frac{P(X,Y)}{P(Y)}$$

結合分布

P(X,Y)	Y=0	Y=1
X=0	0.1	0.2
X=1	0.3	0.4
P(Y)	0.4	0.6

周辺分布

$$P(X=0\mid \boxed{Y=1})=\frac{\boxed{P(X=0,Y=1)}}{\boxed{P(Y=1)}}=\frac{0.2}{0.6}=\frac{1}{3}$$

このスライドは，**条件付き確率**を説明している．

単に公式で覚えるのではなく，結合分布の表と関連付けて理解してほしい．たとえば，P(X=0|Y=1)の場合，条件部がY=1となっているので，結合分布表のY=1の全エントリだけを考察の対象とし，調べたい事象(X=0)の確率 P(X=0,Y=1)と条件を満たす全事象の確率P(Y=1)との比を求めるのである．

2. 結合分布からの事後確率の計算(4) チェーンルール

条件付き確率

$$P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$$

⇩

チェーンルール (chain rule)

$$P(X,Y) = P(X \mid Y)P(Y)$$

$$P(X,Y,Z) = P(X \mid Y,Z)P(Y \mid Z)P(Z)$$

条件付き確率の式から，**チェーンルール**と呼ばれる公式が得られる．むしろ，こちらの方が見慣れているかもしれない．

この公式は，変数が3個以上の場合にも自然に拡張できる．

2. 結合分布からの事後確率の計算(5)
簡単な例題の解(1)

$$P(B=1 \mid A=1)=\frac{3}{5} \quad P(B=0 \mid A=1)=\frac{2}{5}$$

$$P(A=1)=\frac{1}{10} \qquad P(A=0)=\frac{9}{10}$$

カゼ

カゼ(A=1)

カゼでない(A=0)

60%

20%

ノド

B=1

$$P(B=1 \mid A=0)=\frac{1}{5} \quad P(B=0 \mid A=0)=\frac{4}{5}$$

A チェーンルールを使って，結合分布を計算する．

$$P(A=1,B=1)=\frac{1}{10}\times\frac{3}{5}=\frac{3}{50} \qquad P(A=1,B=0)=\frac{1}{10}\times\frac{2}{5}=\frac{2}{50}$$

$$P(A=0,B=1)=\frac{9}{10}\times\frac{1}{5}=\frac{9}{50} \qquad P(A=0,B=0)=\frac{9}{10}\times\frac{4}{5}=\frac{36}{50}$$

では，さきほどの例題に戻って，答えを計算してみよう．

ここでは結合分布を用いて，すなおに事後確率（＝条件付き確率）を求める方法を使う．

まず，チェーンルールを使って，結合分布の4つのエントリを計算する．

2. 結合分布からの事後確率の計算(6) 簡単な例題の解(2)

結合分布

単位:1/50

P(A,B)	B=1	B=0	P(A)
A=1	3	2	5
A=0	9	36	45
P(B)	12	38	50

条件付き確率を計算する.

$$P(A=1 \mid B=1) = \frac{P(A=1, B=1)}{P(B=1)}$$

$$= \frac{3}{12} = \frac{1}{4}$$

Ans. 25%

いま求めた結合分布を表にしてみた．ただし，表は実際の確率を50倍した値を示している．

ここで先ほど学んだ方法を使って条件付き確率を求めれば，答え（ノドが痛いときにカゼである確率）は25%となる．

2. 結合分布からの事後確率の計算(7) この方法の問題点

結合分布

X_1	X_2	…	X_n	確率
0	0	…	0	
0	0	…	1	
…	…	…	…	…
1	1	…	0	
1	1	…	1	

記憶領域の問題

- 結合分布表のサイズが指数的

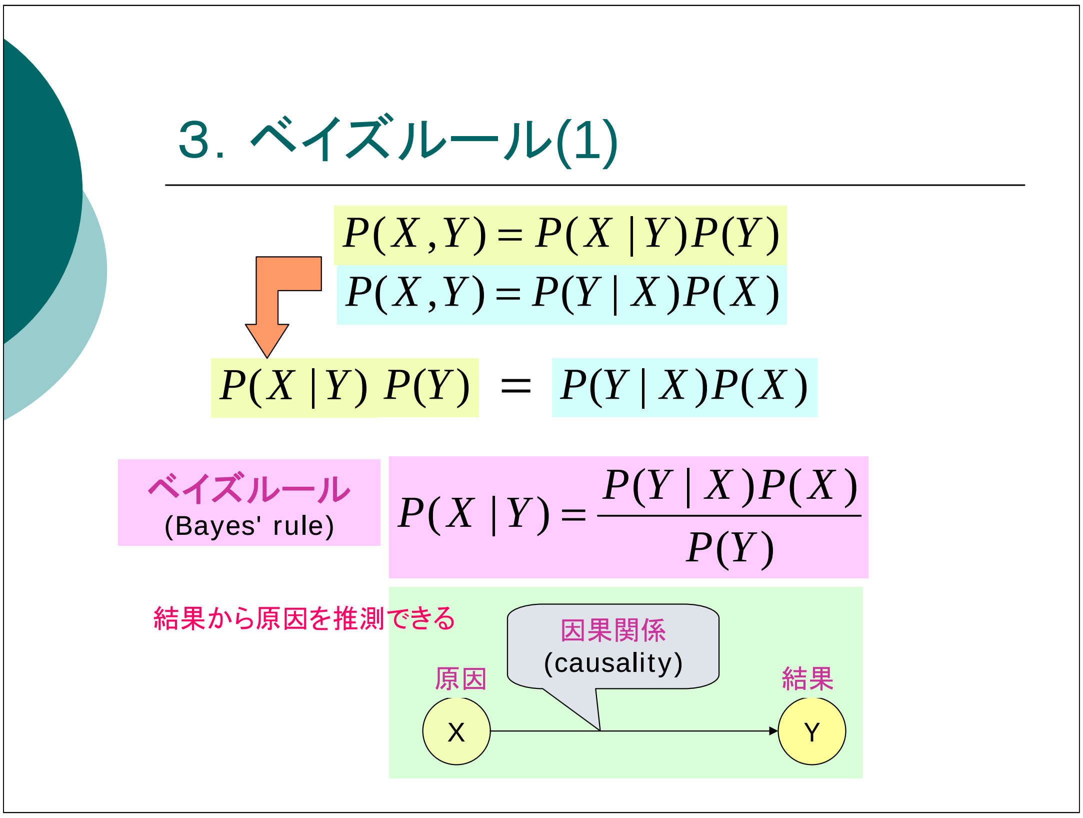
$$O(2^n)$$

計算時間の問題

- 周辺分布の計算時間が指数的

$$O(2^n)$$

いまの例題で学んだような結合分布を用いて事後確率を計算する方法は，考え方はすなおでわかりやすいのだが，効率の面で問題がある．必要なメモリの量と計算時間が，いずれも**指数オーダー**となるのである．したがって，変数の数が多いときには事実上使用できない．たとえば，変数が10個なら2の10乗＝1キロという単位なのでまだだいじょうぶだろうが，変数が20個で2の20乗＝1メガ，変数が30個で2の30乗＝1ギガとなるあたりで，現在のコンピュータの能力ではほぼ限界に近いだろう．



いま見た効率の問題を解決するのが，以降で学ぶベイジアンネットである．その数学的基礎は**ベイズルール**と呼ばれる簡単な公式である．

このスライドでは，チェーンルールからベイズルール

P(X|Y) = P(Y|X) P(X) / P(Y)

を導いている．

確率変数XとYの間に，原因と結果の因果関係があるとしよう．ふつうは，原因が与えられたときに，結果の条件付き確率P(Y|X)が与えられていることが多い．

ベイズルールは，それを用いて逆に，結果を知って原因を推定するための**事後確率**P(X|Y)を計算する公式となっている．

3. ベイズルール(2) 正規化

$$P(X|Y)=\frac{P(Y|X)P(X)}{P(Y)}$$

分子だけ計算

X の値には無関係な一定値　*Y*を固定

正規化定数 $\alpha = 1/P(Y)$

$$\boxed{P(X|Y)=\alpha P(Y|X)P(X)}$$

$$P(X=0|Y)=\alpha P(Y|X=0)P(X=0)$$

$$P(X=1|Y)=\alpha P(Y|X=1)P(X=1)$$

$$P(X=0|Y)+P(X=1|Y)=1$$

となるように，定数倍
（正規化）

ベイズルールで証拠変数Yを固定し，質問変数Xの値をいろいろ(といってもこの授業では0と1)変えることを想定しよう．すると，分母のP(Y)はXに無関係な定数となるので，

P(X|Y) = α P(Y|X)P(X)

と書くことができる．

したがって，Xの各値0,1に対して，もともとのベイズルールの分子だけ計算し，その和が確率の総和である1になるように，それぞれを定数倍(α倍)すればよい．この操作を**正規化**という．

3. ベイズルール(3) 簡単な例題の解

$P(A=1)=\frac{1}{10}$

$P(A=0)=\frac{9}{10}$

カゼ

カゼ(A=1)

カゼでない(A=0)

$P(B=1|A=1)=\frac{3}{5}$ 60%

$P(B=1|A=0)=\frac{1}{5}$ 20%

ノド

B=1

A

ベイズルールを使って，事後確率を計算する．

$$P(A=1 \mid B=1) = \alpha P(B=1 \mid A=1) P(A=1)$$
$$= \frac{3}{5} \times \frac{1}{10}\alpha = \frac{3}{50}\alpha$$

正規化 → 25%

3/(3+9)

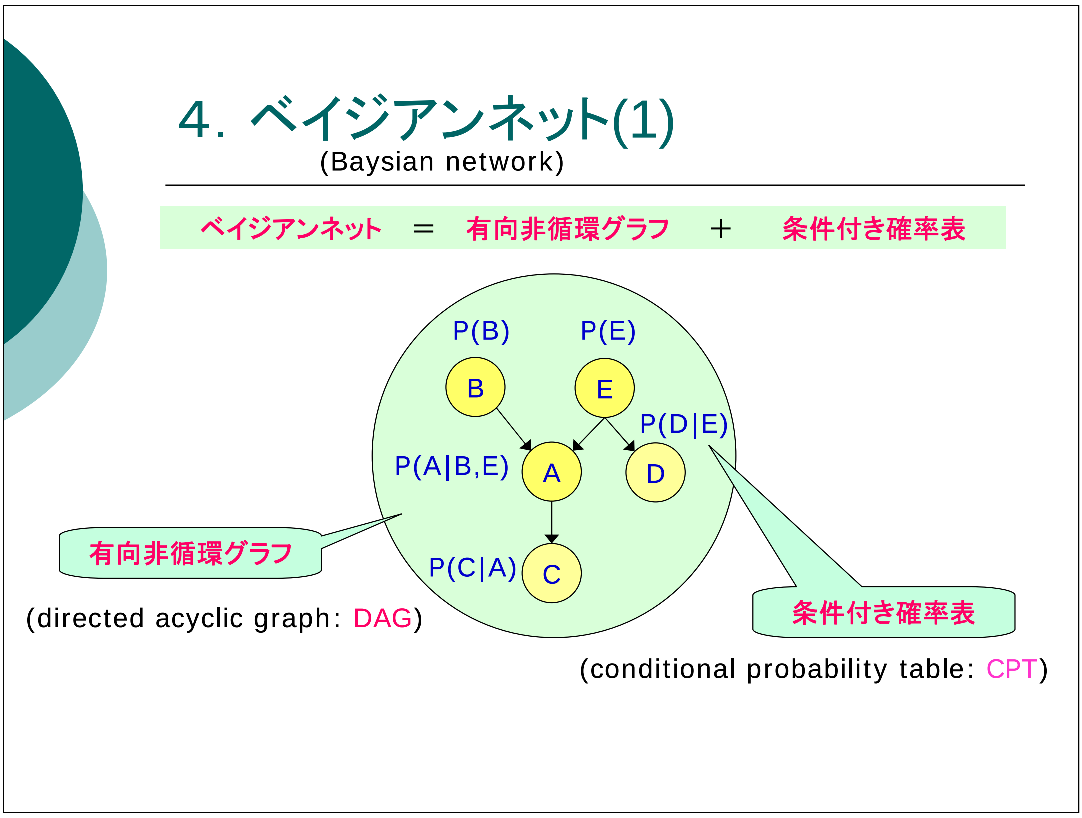
Ans.

$$P(A=0 \mid B=1) = \alpha P(B=1 \mid A=0) P(A=0)$$
$$= \frac{1}{5} \times \frac{9}{10}\alpha = \frac{9}{50}\alpha$$

→ 75%

カゼとノドの例題に戻って，ベイズルールを使って事後確率を計算すると，このスライドのように，ノドが痛いという証拠のもとで，カゼか否かの確率は 3:9 の比となるので，正規化すると，カゼの確率は25%となる．このように，実際には正規化定数 α の値を直接求めなくても，ベイズルールの分子の値に比例して，確率の総和である1を按分することで事後確率が求められる．

休憩



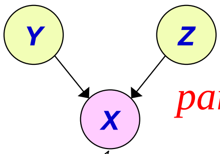
ベイジアンネットは，確率モデルをグラフィカル（グラフ理論的）に表現するためのモデルともいえるもので，**条件付き独立**と呼ばれる仮定を積極的に導入することによって，比較的アーク（有向辺）の数が少ないグラフによって確率モデルをビジュアルにわかりやすく表現できる．また，効率良いアルゴリズムによって事後確率の計算をすることができる．

数学的には，ベイジアンネットはつぎに述べる(V,E,P)の3つの要素からなるものとして定義される．

Vはグラフのノードの集合で，それぞれのノードは1つの確率変数を表している．Eは因果関係のある2つの確率変数を関連づけるアーク（有向辺）の集合である．ベイジアンネットでは，グラフG=(V,E)に巡回路（サイクル）を含むことは許されず，Gは**有向非循環グラフ**(directed acyclic graph: **DAG**)である．

4. ベイジアンネット(2) 条件付き確率表

$parents(Y) = \{\}$



Xの親ノードの集合

$parents(X) = \{Y, Z\}$

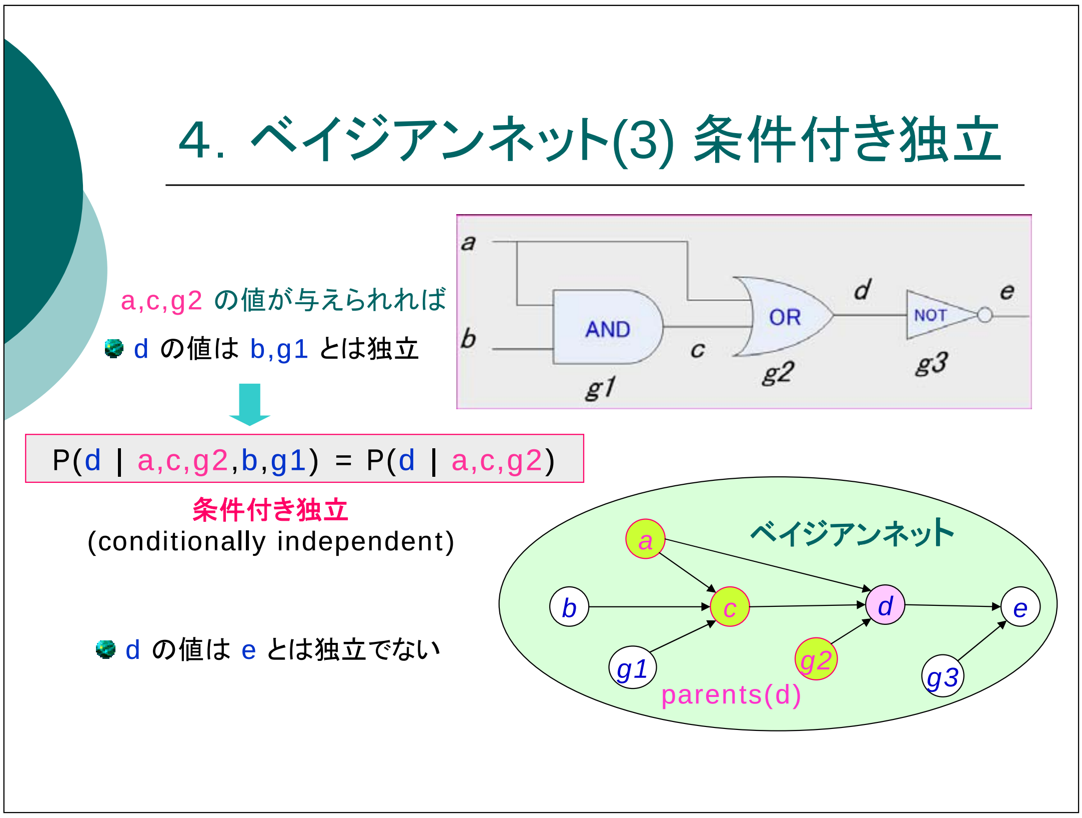
各ノード X ごとに
その親ノードの集合を条件とする
条件付き確率表をつくる

$P(X \mid parents(X))$

$P(X \mid Y, Z)$

Y	Z	X=1	X=0
1	1	0.95	0.05
1	0	0.9	0.1
0	1	0.3	0.7
0	0	0.01	0.99

Pは**条件付き確率表**の集合である．これは，すべてのノードXごとに，P(X | parents(X))という形式の条件付き確率を与えるものである．ここで，parents(X)はXの**親ノード**（Xに入ってくるアークの根もとにあるノード）の集合である．親ノードの表す確率変数の取りうる値のすべての組合せに対して，X=1である確率を指定する必要がある．ちなみに，各組合せに対してX=0である確率は，1からX=1である確率を引くことによって求められる．



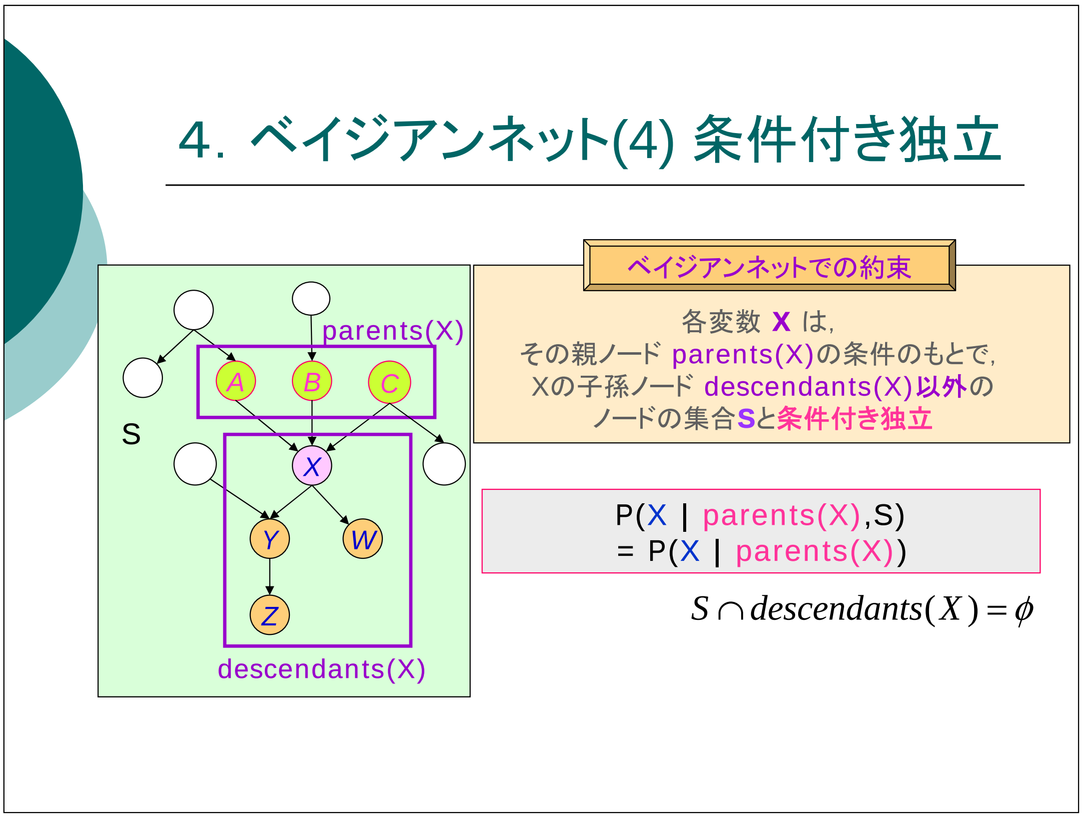
ベイジアンネットで確率論的にもっとも重要な概念は**条件付き独立**というものである．

この回路では，明らかに，cの親ノード{ a,c,g2 }の値が与えられれば d の値は親以外の {b,g1} とは独立である．これを確率の式で書くと

P(d | a,c,g2,b,g1) = P(d | a,c,g2)

となる．このとき，d と {b,g1} は a,c,g2 の条件のもとで条件付き独立であるという．

なお，{a,c,g2} の値が知られていても，その条件のもとで d は e とは独立でないことに注意しておこう．g3が正常ならば，dとeは値が反転（一方が1で他方が0)しているはずなので，値に相関があり，独立ではない．



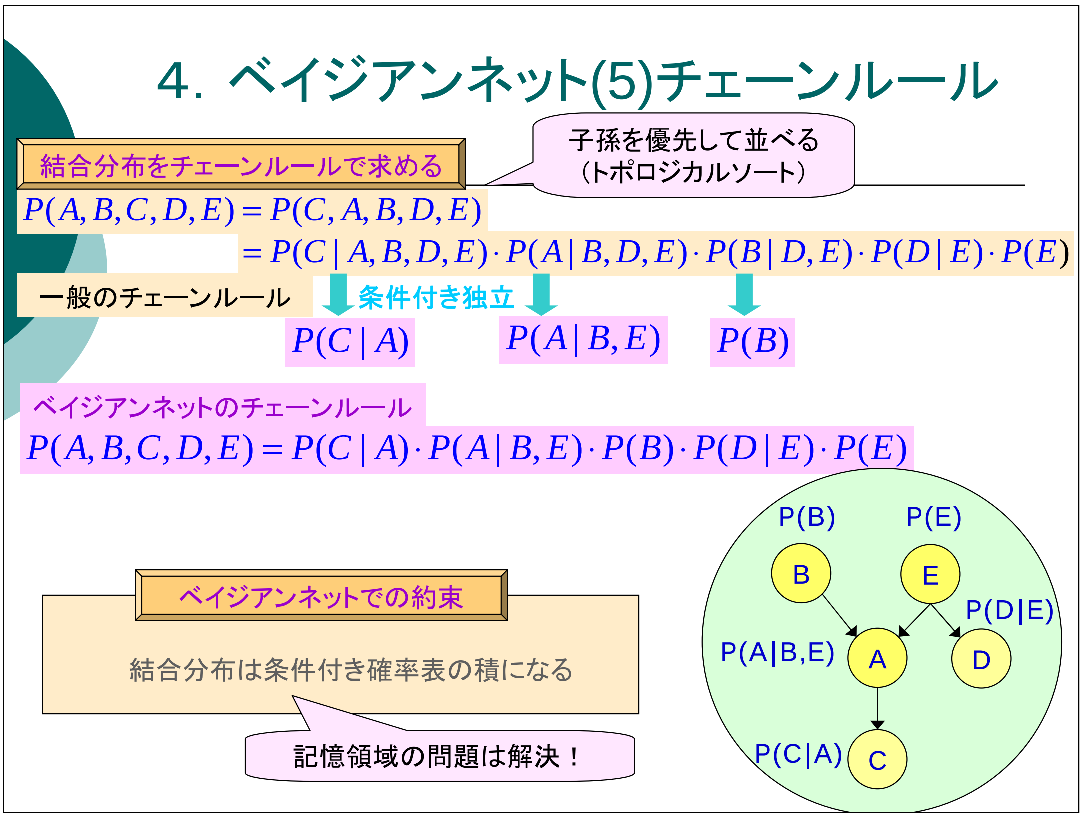
1つ前のスライドの例をもっと一般化すると，ベイジアンネットでは，親ノード parents(X)の条件のもとで，変数X は，Xの子孫ノード descendants(X)以外のノードの集合Sと条件付き独立である．Xの**子孫**とは，Xからアークに沿って到達できるノードのこと(X自身も含む)である．逆に子孫から見ると，Xは**祖先**である．したがって，この図では，A,B,Cの条件のもとで，Xは白丸のノードと条件付き独立である．

これを一般的な式で書くとつぎのようになる．

P(X | parents(X), S) = P(X | parents(X))

つまり，条件付き確率表の条件部にすべての親 parents(X) が含まれているときには，条件部から子孫以外のノードSを削除してよいのである．

これはベイジアンネットの**意味論**(セマンティクス)として天下り的に決めておく約束ごとなので，「なぜ？」と聞かれても困る．応用分野でそういうことが成り立つときに，こういうベイジアンネットを使えということなる．この約束の範囲で現実的な多くの応用がわかりやすくモデル化でき，また事後確率の計算が簡単になるのである．

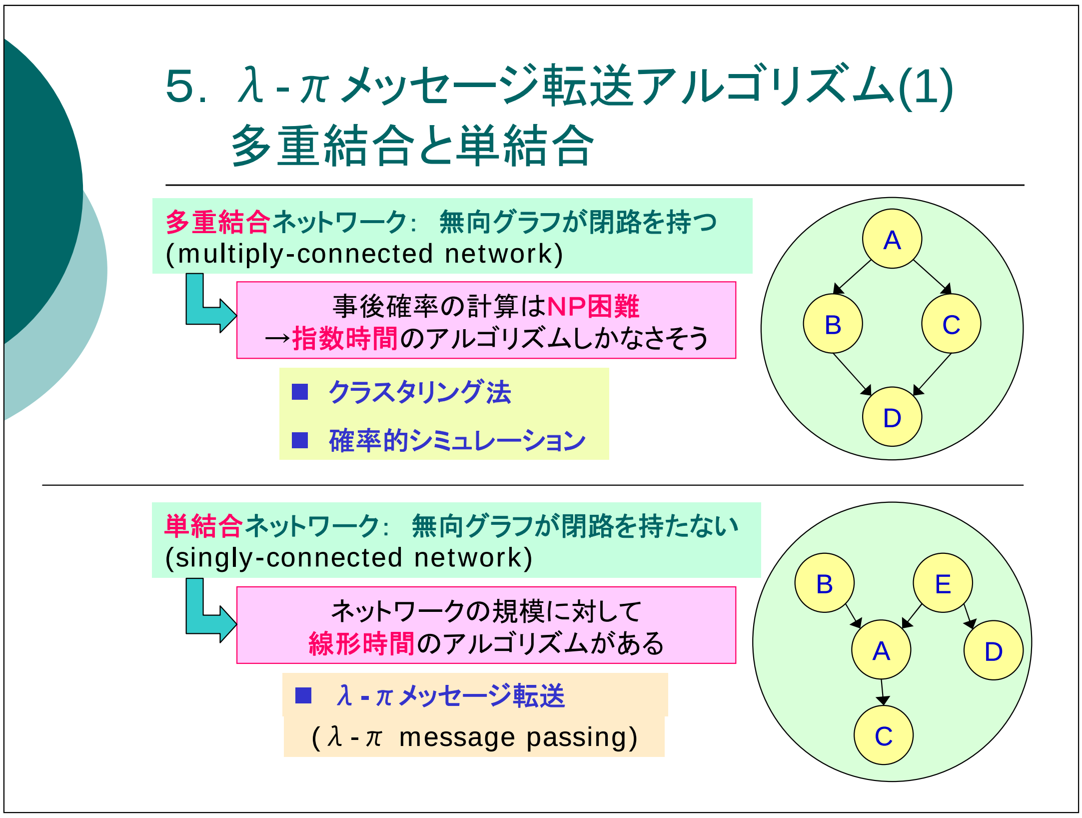


ベイジアンネットの条件付き独立の仮定を利用すると，**チェーンルール**が非常に簡単になる．

この例のように，5個の変数からなる結合確率を考えよう．まず，P(A,B,C,D,E)の引数をトポロジカルソートという考えで，子孫を優先して(祖先より左の方に来るように)並べかえる．そのような並べ方は複数あるが，任意のものでよいので，ここではP(C,A,B,D,E)としよう．その後，一般のチェーンルールを適用すると，図に示す5つの確率の積になる．

しかし，条件付き独立を使うと，そのうち3つが簡単な形となる．まず，ノードがトポロジカルソートしてあることとその後のチェーンルールの積の作り方を観察することによって，たとえばノードAの条件付き確率表P(A|B,D,E)のように，表の条件部にはこの表の主ともいえるノードAのすべての親parents(A)={B,E}が含まれる．さらに，やはりトポロジカルソートによって，条件部にはAの子孫は決して含まれない．したがって，ベイジアンネットの条件付き独立の仮定により，条件部には親{B,E}だけ残し，それ以外のノードである D は削除してよいことになる．

結果として得られたチェーンルールを観察すると，ベイジアンネットにおいては，**<u>結合分布は条件付き確率表の積になる</u>**という重要な性質がわかる．結合分布さえ得られれば，周辺分布だろうが条件付き確率だろうが，定義どおりに計算することができるので，これはうれしい．



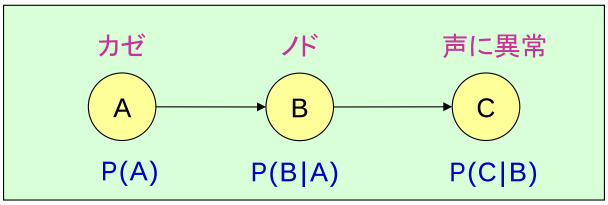
このように結合分布が得られれば，どんな事後確率でも定義にしたがって計算できる．特に，全変数を網羅した結合分布を記憶するための大規模な表は必要がなく，各ノード毎の（小さな）条件付き確率表だけを記憶しておけばよい点で，記憶領域の問題が解決されている．しかし，事後確率の計算過程で必要となる周辺分布の計算のために，すでに見たような大量の計算時間を必要とするので，計算時間の問題はまだ解決されておらず，変数の数が多いときには実用的ではない．

そこで，ベイズルールを利用した効率よい事後確率の計算法が研究されてきている．ただし，その効率はベイジアンネットのグラフの構造に依存する．

グラフが**多重結合ネットワーク**（無向グラフにしたときに閉路を持つ有向グラフ）のときには，その計算は**NP困難**であることが知られている．これは，コンピュータサイエンスの基礎的な部分に革命でも起こらない限り，指数時間のアルゴリズムしか存在しないと考えられている問題であることを意味する．それでも，**クラスタリング法**，**確率的シミュレーション**などと呼ばれる方法が開発され，少しでも効率よく計算できるように工夫されている．

グラフが**単結合ネットワーク**（無向グラフにしたときに閉路を持たない有向グラフ）のときには，ネットワークの規模に対して**線形時間**の効率良いアルゴリズムがある．すなわち，ノードやアークの数を増やしてベイジアンネットの規模を拡大しても，それに比例した程度の計算時間しかかからない．そのようなアルゴリズムとして，**λ-π メッセージ転送**アルゴリズムと呼ばれるものがある．

5. λ-π メッセージ転送アルゴリズム(2) 枝分かれのない例



A	P(A)
1	1/10
0	9/10

A	P(B=1|A)
1	3/5
0	1/5

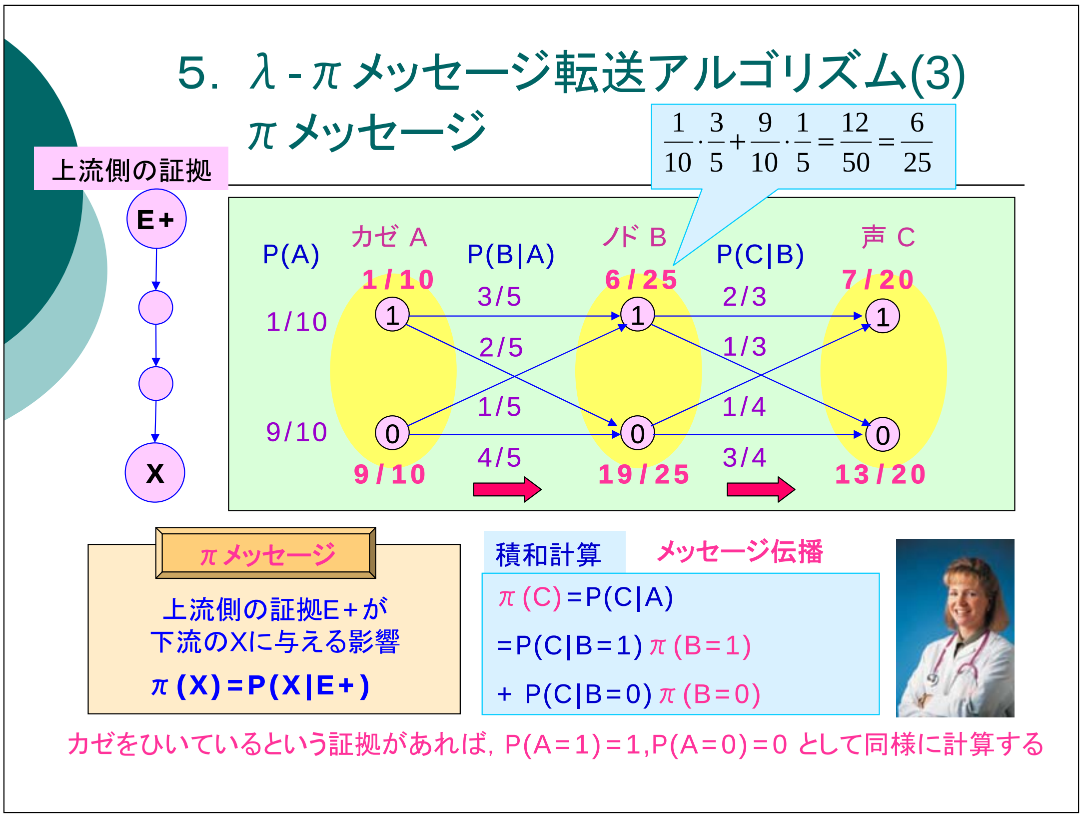
B	P(C=1|B)
1	2/3
0	1/4

Q 声に異常があるとき，カゼの確率は？

λ-π メッセージ転送アルゴリズムは，かなり複雑なものなので，この授業では**グラフに枝分かれがない**場合だけを考える．すなわち，一直線上にノードが並んでいる場合である．

これまでの例に，「声に異常がある」ことを表す変数Cを追加して，図のようなベイジアンネットを構成しよう．ノドが痛いときには3人に2人の割合で声に異常があり，ノドが痛くないときには4人に1人の割合で声に異常がある．ベイジアンネットの条件付き独立性により，ノドが痛いかどうかが条件として与えられれば，声の異常はカゼかどうかと独立であるとみなす．

この設定のもとで，声に異常があると知ったとき，カゼの確率を求めたい．



λ-π メッセージ転送アルゴリズムを理解するために，ベイジアンネットで与えられた設定をこのスライドの図のように整理しなおしておく．このアルゴリズムは，まず，上流側の証拠E+が下流のXに与える影響を考察するために，

π(X)=P(X | E+)

で定義される**π メッセージ**と呼ばれる確率分布を計算する．これは確率分布なので，Xの取りうるそれぞれの値(0,1)毎に計算する必要がある．

上流側の証拠E+というのが，たとえばAだとしよう．A=1であることが知られていれば，定義より

π (A=1)=P(A=1 | E+) = 1

π (A=0)=P(A=0 | E+) = 0

のように，知られている状況を表すπ メッセージを1，はっきりと否定できる状況のものを0とする．しかし，この例題のように，上流側に証拠がないときは，E+は空集合なので，

π (A=1)=P(A=1 | E+) = P(A=1)

π (A=0)=P(A=0 | E+) = P(A=0)

のように，事前確率がそのままπ メッセージとなる．

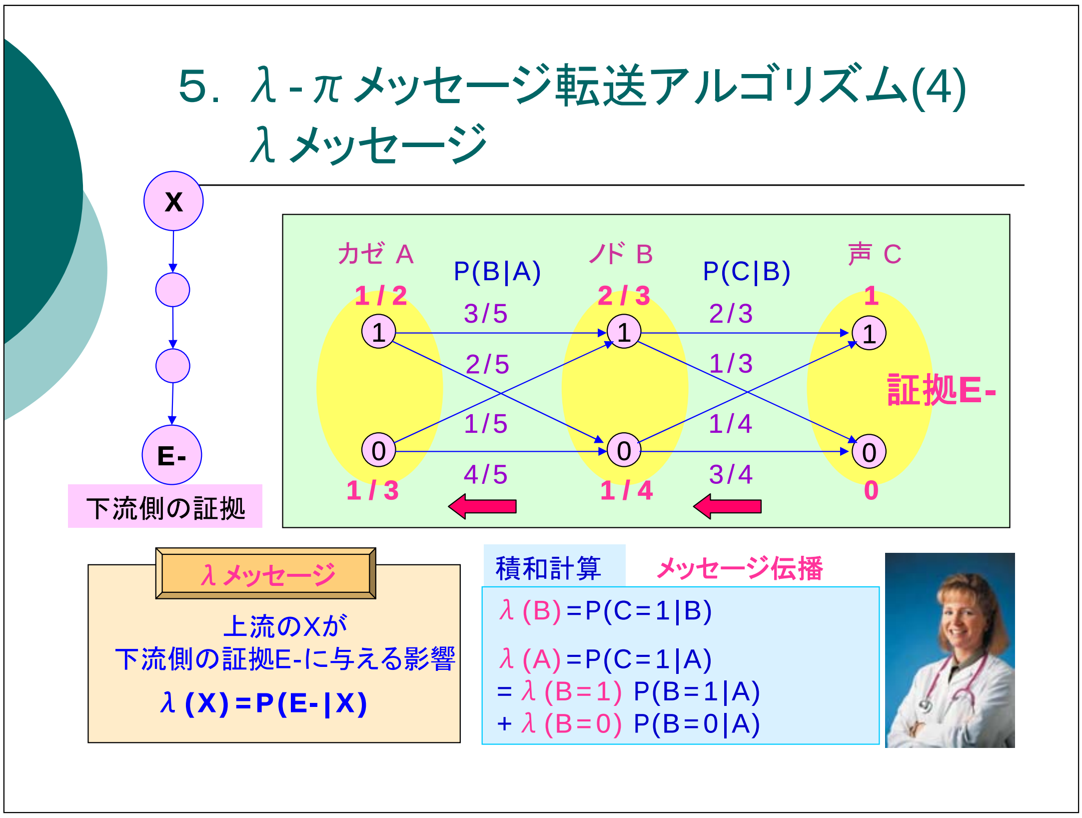
事前確率が直接ベイジアンネットによって与えられていないノードについては，上流側からどのような経路でXに影響を与えるかを経路毎にチェーンルールを用いて計算し，それら排反事象の確率の和をπ メッセージとする．たとえば，B=1のノードには，A=1からとA=0からの影響があるので

π (B=1)=P(B=1 | E+) = P(B=1)

=P(B=1|A=1)P(A=1) + P(B=1|A=0)P(A=0)

=P(B=1|A=1) π (A=1) + P(B=1|A=0) π (A=0)

となり，条件付き確率表P(B|A)と上流側から**伝播**してくるπ メッセージπ (A)を用いた**積和計算**により，その値は6/25となる．



つぎに，アルゴリズムは，上流のXが下流側の証拠E-に与える影響を表す**λメッセージ**と呼ばれる確率分布を計算する．その定義は

λ(X)=P(E-|X)

である．すなわち，もしXが条件として与えられたとき，下流側の証拠はどの程度の確率で生じるものなのかを計算する．これもやはり確率分布なので，Xの取りうるそれぞれの値(0,1)毎に計算する必要がある．

この例題では，下流側の証拠E-というのは，C=1である．したがって，定義より

λ(C=1)=P(E- | C=1) = P(C=1 | C=1) = 1

λ(C=0)=P(E- | C=0) = P(C=1 | C=0) = 0

のように，知られている状況を表すλメッセージを1，はっきりと否定できる状況のものを0とする．

他のノードのλメッセージは，ちょうど，πメッセージの伝播と同様の積和計算を下流から上流に逆流させながら行う．まず，ノードBについては，直接 C=1 への影響を計算できる．

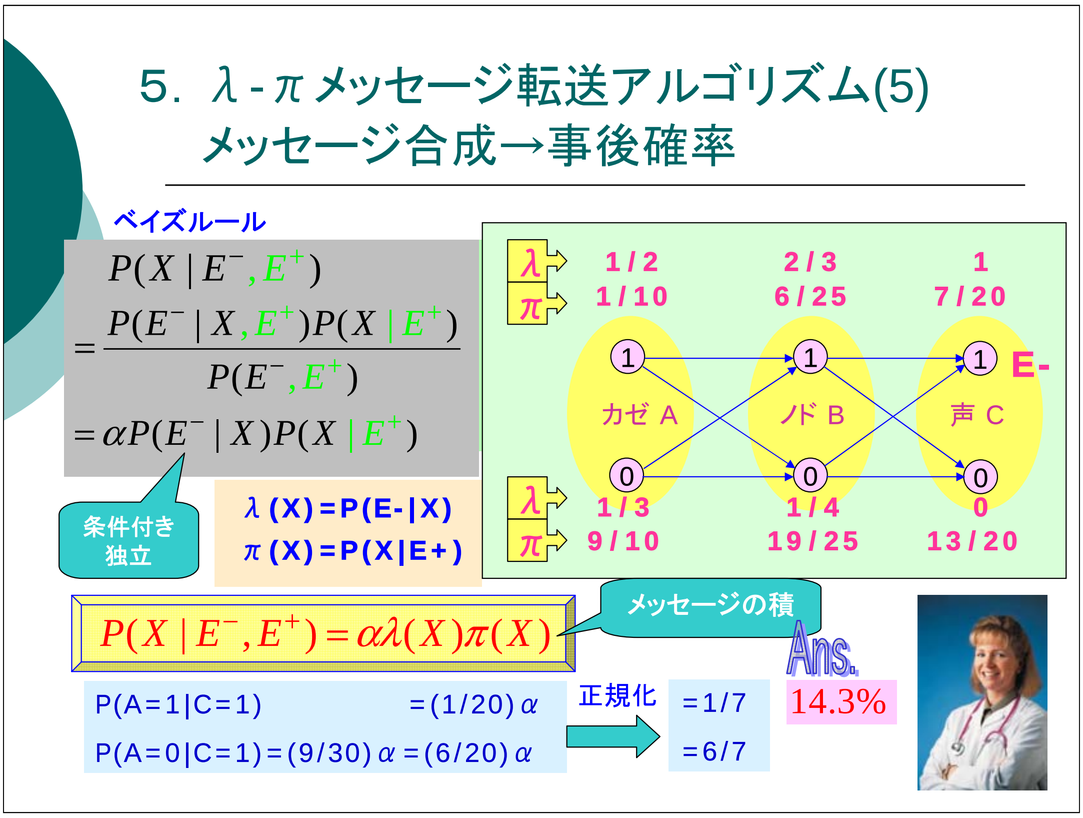
λ(B=1) = P(C=1|B=1) = 2/3

λ(B=0) = P(C=1|B=0) = 1/4

つぎに，ノードAについては，B=1経由とB=0経由の影響に分けてチェーンルールを利用した積和計算となる．

λ(A=1) = P(C=1|A=1)
= P(C=1|B=1)P(B=1|A=1) + P(C=1|B=0)P(B=0|A=1)
= λ(B=1) P(B=1|A=1) + λ(B=0) P(B=0|A=1)

このように，下流側から伝播してくるλメッセージλ(B)と条件付き確率表P(B|A)を用いた積和計算により， λ(A=1) = 1/2となる．



最後に，これまで求まったπメッセージとλメッセージを**合成**すると，各ノードの事後確率となる．

まず，P(X | E-) に対するベイズルールの条件部に E+ も付け加えた一般的な状況を表す P(X | E-, E+) を考え，ややベイズルールを拡張した式を考えると，

P(X | E-, E+) = α λ(X) π(X)

が得られる．

したがって，<u>λメッセージとπメッセージの積を正規化したものが事後確率となる</u>．すなわち，Xの各値(1,0)ごとに，両**メッセージの積**を計算し，その比に応じて1を按分したものが事後確率となる．

この例題の場合，声に異常があるときにカゼである確率は P(A=1|C=1)=1/7 となる．なお，この例題で要求されたP(A | C=1) だけを求めればよいのなら，すべてのπメッセージを計算する必要はなかったことに気が付いてほしい．したがって，実際にはもっと早く計算できることになる．

（演習問題）

1. 上記の問題を結合分布を用いた方法で解いて，結果が一致することを確かめよ．
2. 声に異常があるときに，ノドが痛い確率は？
3. 声に異常があることに加えて，カゼである新たな証拠が得られたとする．このとき，上で求めたノドが痛い確率を改訂せよ．