2023年度インテリジェントシステム
レポート課題＃3（確率モデル・Bayesian Networks：解答例）

以下の問1～ 問 に対する解答をレポートにまとめて（文書ファイルを）eALPSから提出せよ。提出するファイルはpdfであること。文書作成にはlatex, MS-Officeなどを用いることが望ましいが、手書きのレポートをスキャンしてpdfに変換後提出してもよい。

1. 次の(a)～ (f)のうち一般に成立するものを全て選べ。誤っているものについては、どこが誤りか、誤っている箇所について簡単に説明せよ。

   (a) $P(X) \propto \sum_y P(X|Y=y)$

   (b) $P(X|Y) = \frac{P(X,Y)}{\sum_y P(X,Y=y)}$

   (c) $P(X|Y=y) \propto P(Y=y|X)$

   (d) $P(X) = \sum_y \sum_z \sum_w P(X, Y=y, Z=z, W=w)$

   (e) $P(X_1, X_2, \cdots, X_{n-1}, X_n) = P(X_1)\Pi_{i=2}^{n} P(X_i|X_{i-1}, \cdots, X_1)$

   (f) $P(X_1, X_2, \cdots, X_{n-1}, X_n) = P(X_n)\Pi_{i=1}^{n-1} P(X_{n-i}|X_n, \cdots, X_{n-i+1})$

   正しいものは(d), (e), (f)

   誤っているものについての説明例：

   - (a)　$P(X) = \sum_y P(X, Y=y) = \sum_y P(X|Y=y)P(Y=y)$ である。
   - (b)　$P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{P(X,Y)}{\sum_x P(X=x,Y)}$ となる。
     元の式（右辺）$\frac{P(X,Y)}{\sum_y P(X,Y=y)}$ は $\frac{P(X,Y)}{P(X)}$ なので $P(Y|X)$ である。
   - (c)　$P(X|Y=y) = \frac{P(X,Y=y)}{P(Y=y)} \propto P(X, Y=y) = P(Y=y|X)P(X)$ である。

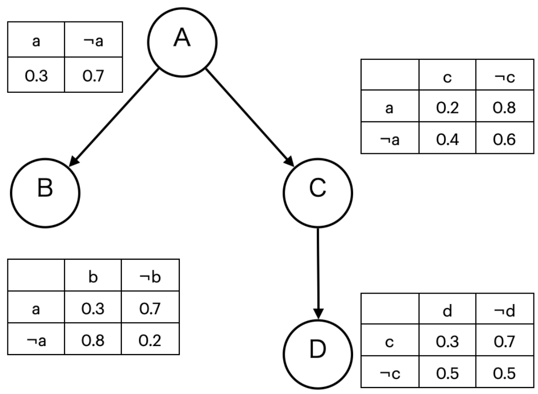
2. 下の Bayesian Network で表現される確率モデルを用いて以下の確率値を求めよ。但し確率変数 A,B,C,D は全て 2 値であり、true または false を値に持つ。例えば A=true を a, A=false を $\neg a$ と表すものとする。解答には有効数字 2 桁の小数を用いること。

(a) $P(\neg b)$

(b) $P(c)$

(c) $P(c|d)$

(d) $P(a|d)$



a	¬a
0.3	0.7

	c	¬c
a	0.2	0.8
¬a	0.4	0.6

	b	¬b
a	0.3	0.7
¬a	0.8	0.2

	d	¬d
c	0.3	0.7
¬c	0.5	0.5

$$\begin{aligned} P(\neg b) &= \sum_A \sum_C \sum_D \boldsymbol{P}(A)\boldsymbol{P}(\neg b|A)\boldsymbol{P}(C|A)\boldsymbol{P}(D|C) \\ &= \sum_A \boldsymbol{P}(A)\boldsymbol{P}(\neg b|A) \sum_C \boldsymbol{P}(C|A) \sum_D \boldsymbol{P}(D|C) \\ &= \sum_A \boldsymbol{P}(A)\boldsymbol{P}(\neg b|A) \\ &= P(\neg b|a)P(a) + P(\neg b|\neg a)P(\neg a) = 0.7 \times 0.3 + 0.2 \times 0.7 = 0.35 \end{aligned}$$

$$\begin{aligned} P(c) &= \sum_A \sum_B \sum_D \boldsymbol{P}(A)\boldsymbol{P}(B|A)\boldsymbol{P}(c|A)\boldsymbol{P}(D|c) \\ &= \sum_A \boldsymbol{P}(A)(\sum_B \boldsymbol{P}(B|A))\boldsymbol{P}(c|A) \sum_D \boldsymbol{P}(D|c) \\ &= \sum_A \boldsymbol{P}(A)\boldsymbol{P}(c|A) = P(c|a)P(a) + P(c|\neg a)P(\neg a) \\ &= 0.2 \times 0.3 + 0.4 \times 0.7 = 0.34 \end{aligned}$$

前の問題の解答より $P(c) = 0.34$ したがって $P(\neg c) = 0.66$

$P(d) = \sum_C \boldsymbol{P}(C,d) = \sum_C \boldsymbol{P}(d|C)\boldsymbol{P}(C) = P(d|c)P(c) + P(d|\neg c)P(\neg c)$

$P(c|d) = \frac{P(d|c)P(c)}{P(d)} = \frac{0.3 \times 0.34}{0.3 \times 0.34 + 0.5 \times 0.66} \approx 0.24$

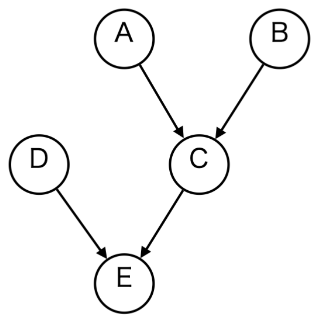
$P(a|d) = \frac{P(a,d)}{P(d)}$ である。$P(d)$ は前の問題で計算済である。

$$
\begin{aligned}
P(a,d) &= \sum_B \sum_C P(a)\boldsymbol{P}(B|a)\boldsymbol{P}(C|a)\boldsymbol{P}(d|C) = P(a)(\sum_B \boldsymbol{P}(B|a))\sum_C \boldsymbol{P}(C|a)\boldsymbol{P}(d|C) \\
&= P(a)(P(c|a)P(d|c) + P(\neg c|a)P(d|\neg c) \\
&= 0.3 \times (0.2 \times 0.3 + 0.8 \times 0.5) = 0.138
\end{aligned}
$$

以上より $P(a|d) = \frac{P(a,d)}{P(d)} = \frac{0.138}{0.432} \approx 0.32$

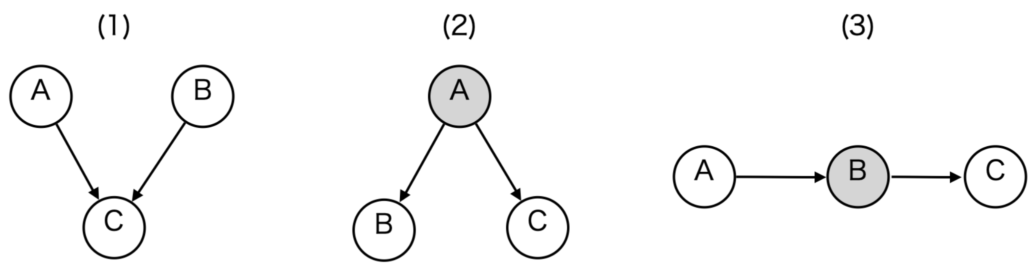
3. 確率変数 X,Y,Z があるとき、X,Y は Z が与えられたとき条件付独立であることを $X \perp Y|Z$ と表すものとする。また、X,Y が（Z を周辺化して消去すると）独立であるとき $X \perp Y$ と表すものとする。下図の Bayesian ネットワークで表すようなモデルが与えられたとき、次の (a)~ (d) のうち正しいものを全て挙げよ。

$$\text{(a) } D \perp E \quad \text{(b) } A \perp B|E \quad \text{(c) } A \perp D \quad \text{(d) } B \perp D|C$$



正しいのは (c), (d)

4. 下図の (1) ～ (3) で示された Bayesian ネットワークについて、図に示された独立性、条件付き独立性が成り立つことを証明せよ。



$A \perp B$　　$B \perp C \mid A$　　$A \perp C \mid B$

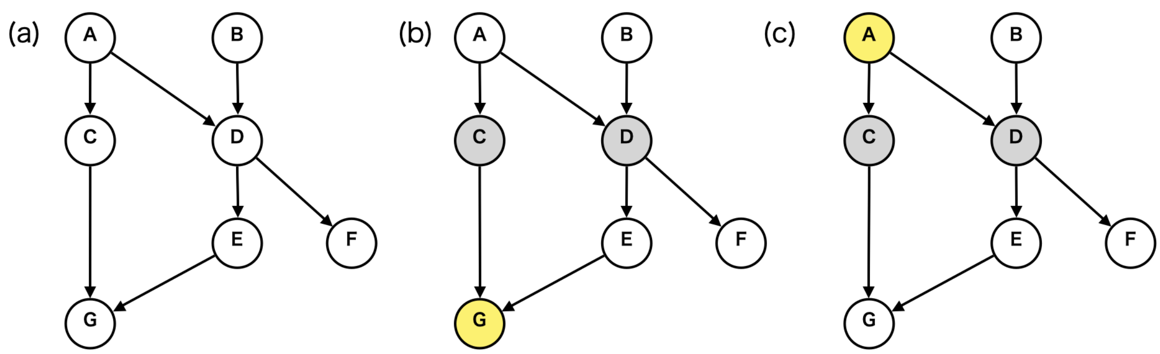
- (1) Bayesian ネットワークより $P(A,B,C) = P(A)P(B)P(C|A,B)$ である。
  $P(A,B) = \sum_c P(A,B,C=c) = P(A)P(B)\sum_c P(C=c|A,B) = P(A)P(B)$ となるから $A \perp B$
- (2) Bayesian ネットワークより $P(A,B,C) = P(A)P(B|A)P(C|A)$ である。
  $P(B,C|A) = \frac{P(A,B,C)}{P(A)}$ だから、
  $P(B,C|A) = \frac{P(A)P(B|A)P(C|A)}{P(A)} = P(B|A)P(C|A)$ であり、したがって
  $B \perp C|A$
- (3)Bayesian ネットワークより $P(A,B,C) = P(A)P(B|A)P(C|B)$ である。
  $P(A,C|B) = \frac{P(A,B,C)}{P(B)} = \frac{P(A)P(B|A)P(C|B)}{P(B)}$ となるが
  $P(A)P(B|A) = P(A,B)$ であり、$P(A|B) = \frac{P(A,B)}{P(B)}$ だから
  $P(A,C|B) = P(A|B)P(C|B)$ すなわち $A \perp C|B$

5. 下図の Bayesian ネットワークに関する問 (a),(b),(c) に解答せよ。

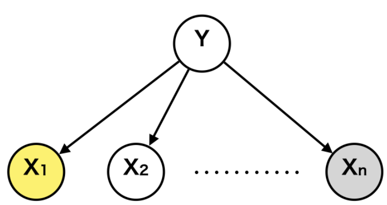
   (a) 下図の Bayesian ネットワークで表されるモデルの場合の同時確率 $P(A,B,C,D,E,F,G)$ を条件付確率の積の形で表せ。
   $P(A,B,C,D,E,F,G) = P(A)P(B)P(C|A)P(D|A,B)P(E|D)P(F|D)P(G|C,E)$

   (b) C,D が観測済であるとき、G と独立であるノードを全て挙げよ。
   A, B, F
   d-分離について調べてみればよい。たとえば、A と G の間の (無向) 経路としては $A \to C \to G$ と $A \to D \to E \to G$ があるが、$A \to C \to G$, $A \to D \to E$ は観測されたノードによって分離されている chain である。
   B については $B \to D \to E$ により d-分離が示せる。
   F については $F \leftarrow D \leftarrow A \to C \to G$ は A の場合に考えたのと同様の chain が含まれている。$F \leftarrow D \to E \to G$ は tent$F \leftarrow D \to E$ を含んでいる。

   (c) C,D が観測済であるとき、A と独立であるノードを全て挙げよ。
   E, F, G
   $B \to D \leftarrow A$ は v-structure であり、D は観測されたノードなので d-分離の条件を満たさない。



6. $X_1, X_2, \cdots, X_n, Y$ は 2 値の確率変数であり、下図の Bayesian ネットワークに示すようなモデルが与えられているものとする。いま確率変数 X_n については観測値 $X_n = x_n$ が与えられ、これに基づいて $P(X_1|X_n = x_n)$ を不要な変数を消去（周辺化）することで求めたい。このとき以下の問 (a),(b),(c) に解答せよ。



   (a) 同時確率 $P(X_1, X_2, \cdots, X_n, Y)$ はどうなるか示せ。
   Bayesian ネットワークより
   $P(X_1, X_2, \cdots, X_n, Y) = P(Y)P(X_1|Y)P(X_2|Y) \cdots P(X_n|Y)$

   (b) 不要な変数を消去するとき以下の 2 つではどちらが適切と考えられるか理由とともに解答せよ。

i. $X_2 \to X_3 \to \cdots \to X_{n-1} \to Y$

ii. $Y \to X_2 \to X_3 \to \cdots \to X_{n-1}$

i の順番（$X_2 \to X_3 \to \cdots \to X_{n-1} \to Y$）が望ましい。この順番で変数を消去していったときまず X_2 を消去したとき得られる factor を f_2 とおくと、

$$f_2(Y) = \sum_{x_2} P(X_2 = x_2|Y)P(Y) = P(Y)\sum_{x_2} P(X_2 = x_2|Y) = P(Y)$$

となる。この後同様に X_j $(j = 3, \cdots, n-1)$ を順番に消去していったとき得られる factor を f_j とおくと

$$f_j(Y) = \sum_{x_j} P(X_j = x_j|Y)P(Y) = P(Y)\sum_{x_j} P(X_j = x_j|Y) = P(Y)$$

となる。X_{n-1} まで消去を終えたとき、残っている factors は $P(Y)$, $P(X_1|Y)$, $P(X_n = x_n|Y)$ である。最後は Y を消去して

$$f(X_1, x_n) = \sum_{y} P(Y = y)P(X_1|Y = y)P(X_n = x_n|Y = y)$$

このように、生成される factor のサイズが増大することはない。

一方、ii の順番（$Y \to X_2 \to X_3 \to \cdots \to X_{n-1}$）で変数を消去（周辺化）していったとき、まず最初 Y を周辺化して以下のような factor g_1 が生成される。

$$g_1(X_1, \cdots, X_{n-1}, x_n) = \sum_{y} P(Y = y)P(X_1|Y = y)\cdots P(X_{n-1}|Y = y)P(X_n = x_n|Y = y)$$

この factor のサイズは 2^{n-1} であるから n が大きいときは非常にサイズが大きくなってしまう。この後、$X_2, \cdots X_{n-1}$ の順番に消去を行なっていったとき生成される factor を g_j $(j = 2, \cdots, n-1)$ とおくと

$$g_j(X_{j+1}, \cdots, x_n) = \sum_{x_j} g_{j-1}(X_j = x_j, X_{j+1}, \cdots, x_n)$$

であるから徐々にサイズは小さくなるものの、比較すると i の順番の方が生成される factor のサイズが小さいため適切であると考えられる。

(c) Bayesian ネットワークとともに与えられる条件付き確率を用いて $P(X_1|X_n = x_n)$ を表せ。（できるだけシンプルな形になるものが望ましい）

上の問題で考えたように $X_2 \to X_3 \to \cdots \to X_{n-1} \to Y$ の順番で変数を消去しいったとき、残る facrtor は

$$f(X_1, x_n) = \sum_{y} P(Y = y)P(X_1|Y = y)P(X_n = x_n|Y = y)$$

であり、これは $P(X_1, X_n = x_n)$ であるから

$$P(X_1|X_n = x_n) = \frac{\sum_y P(Y = y)P(X_1|Y = y)P(X_n = x_n|Y = y)}{\sum_{x_1}\sum_y P(Y = y)P(X_1 = x_1|Y = y)P(X_n = x_n|Y = y)}$$

で求める条件付確率を計算できる。