Session7/8

베이지안 네트워크

Causal Bayesian Network

용어 정의

1. Probability
    - $P(A)$: marginal Prob
    - $P(A|B)$: Conditional Prob
    - $P(A \cap B) = P(A, B) = P(A)P(B|A) = P(B)P(A|B)$
    - Bayes's Th : $P(B|A) = \frac{P(B)P(A|B)}{P(A)}$
2. Marginalize
    - $P(A) = \sum_B P(B)P(A|B) = \int_B P(B)P(A|B)$
3. Independent
    - $A \perp\!\!\!\perp B \Leftrightarrow P(A) = P(A|B)$
    - $P(A, B) = P(A)P(B)$
    - conditional independent : $P(A, B|C) = P(A|C)P(B|C)$

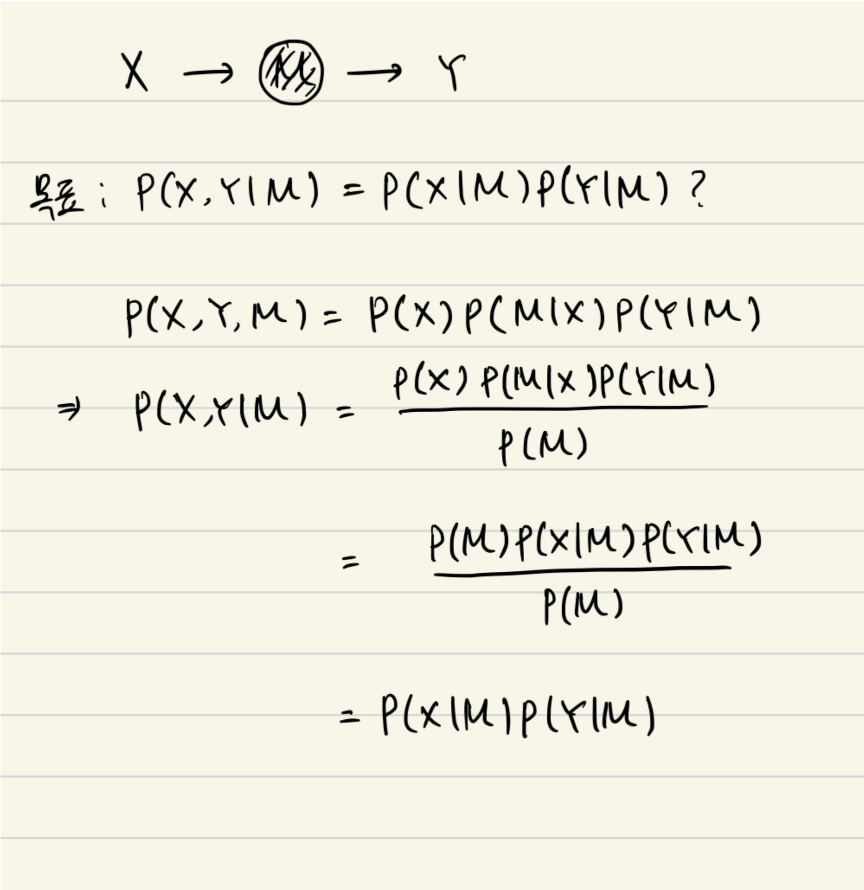
가정

1. Causal Markov Assumption(under DAG)
    - 그래프가 주어졌을때 joint prob을 계산할 때 본인에게 화살표를 주는 변수에만 condition하다.
    - X→Y→Z : $P(X, Y, Z) = P(X)P(Y|X)P(Z|Y)$
    - 위 과정을 Bayesian Network Factorization라고 한다.
    - Bayesian Network Factorization이 Causal Bayesian Network을 이해하는 첫걸음.

Mediator

$$X \to M \to Y$$

- $P(X,Y,M) = P(X)P(M|X)P(Y|M)$
- 목표 : $P(X,Y) = P(X)P(Y)$?
- $P(X,Y) = \sum_M P(X,Y,M) = P(X)\sum_M P(M|X)P(Y|M)$
- $\sum_M P(Y)P(Y|M) = P(Y)$
- 즉, $P(M|X) = P(M)$인지 확인하면 된다.
  - 위 식이 성립하다는 것은 M과 X는 독립이라는 것이지만, M은 X의 영향을 받기 때문에 독립일 수 없다.
- 따라서, X와 Y는 association이 생긴다.



M을 conditioning하면 X와 Y는 독립임을 증명.

Confounder

$$
\begin{tikzcd}
 & C \arrow[dl] \arrow[dr] & \\
X & & Y
\end{tikzcd}
$$

목표 : $P(X,Y) = P(X)P(Y)$?

$$P(X,Y,C) = P(X|C)\,P(Y|C)\,P(C)$$

$$P(X,Y) = \sum_c P(X,Y,c) = \sum_c P(X|c)\underbrace{P(Y|c)P(c)}_{=P(Y,c)\,=\,P(Y)P(c|Y)}$$

$$= P(Y)\sum_c P(X|c)\underline{P(c|Y)}, \quad P(X) = \sum_c \underline{P(c)}P(X|c)$$

$\Rightarrow.\ P(c) = P(c|Y)$?

$\therefore\ X \not\perp Y.$

confounder가 있을 때 X와 Y는 association이 있다.

$$
\begin{tikzcd}
 & C \arrow[dl] \arrow[dr] & \\
X & & Y
\end{tikzcd}
$$

목표 : $P(X,Y|C) = P(X|C)P(Y|C)$

$$P(X,Y,C) = P(X|C)\,P(Y|C)\,P(C)$$

$$\Rightarrow\ P(C)P(X,Y|C) = P(X|C)P(Y|C)P(C).$$

$$\therefore\ P(X,Y|C) = P(X|C)P(Y|C)$$

confounder를 condition했을 때 X와 Y는 조건부 독립이다.

Collider

$$\begin{tikzcd} & C & \\ X \arrow[ur] & & Y \arrow[ul] \end{tikzcd}$$

목표 : $P(X,Y) = P(X)P(Y)$?

$$P(X,Y,C) = P(X)P(Y)P(C|X,Y)$$

$$P(X,Y) = \sum_c P(X,Y,C) = P(X)P(Y)\underbrace{\sum_c P(C|X,Y)}_{=1}$$

$\therefore X \perp\!\!\!\perp Y$

Collider가 있을 때 X와 Y는 독립

$$
\begin{tikzcd}
 & C & \\
X \arrow[ur] & & Y \arrow[ul]
\end{tikzcd}
$$

목표: $P(X,Y|C) = P(X|C)P(Y|C)$. 를 다른시각으로 보면,

$P(X|C) = P(X|C,Y)$. 즉, C라는 조건이 있을때 Y을 측정하더라도 확률이 같다면 C라는 조건 하에서 X,Y는 독립.

$\Rightarrow$ $P(X|C) = \dfrac{P(X)P(C|X)}{P(C)}$, $P(X|C,Y) = \dfrac{P(X|Y)P(C|Y,X)}{P(C|Y)}$.

$\Rightarrow$ X,Y는 조건이 없으면 독립, $P(X) = P(X|Y)$.

$$\frac{P(C|X)}{P(C)} \neq \frac{P(C|Y,X)}{P(C|Y)}$$

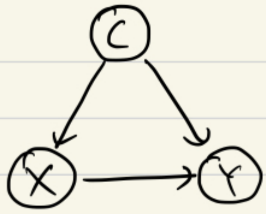
$\because$ $\dfrac{P(C|X)}{P(C)}$: X를 conditioning 하지 않았을 때 대비해서 conditioning 했을때의 C의 확률의 ratio $\Rightarrow$ Effect of X on C.

$\dfrac{P(C|Y,X)}{P(C|Y)}$: Effect of X on C after controlling for Y.

$\Rightarrow$ regression 관점: 통제변수를 추가했을때 메인변수의 계수는 변한다.
즉, X가 C에 미치는 효과와 Y를 control 했을때
X가 C에 미치는 효과는 같지 않다.

Collider를 통제했을 때 X와 Y는 독립이 아니다.

Correlation과 Causation이 다른 이유



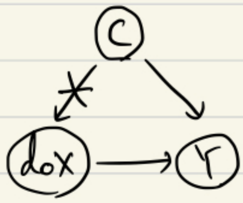
$$P(X,Y,C) = P(X|C)P(Y|X,C)P(C)$$

$$\Rightarrow P(X)P(Y,C|X) = P(X|C)P(Y|X,C)P(C).$$

$$P(Y,C|X) = \frac{\overbrace{P(X|C)P(C)}^{=P(X)P(C|X)}P(Y|X,C)}{P(X)}$$

$$= P(C|X)P(Y|X,C)$$

$$P(Y|X) = \sum_c P(Y,C|X) = \boxed{\sum_c P(C|X)P(Y|X,C)}$$



causal effect $= P(Y|do(X))$

$$P(do(X),Y,C) = P(do(X))P(Y|do(X),C)P(C).$$

$$P(Y,C|do(X)) = P(Y|do(X),C)P(C)$$

$$P(Y|do(X)) = \sum_c P(Y,C|do(X))$$

$$= \boxed{\sum_c P(Y|do(X),C)P(C)}$$

$\Rightarrow P(C|X) = P(C)$?

X와 C는 독립일 수 없기 때문에 같지 않다.

$\therefore$ Correlation $\neq$ causation.