Probabilistic Modeling 1

Given: Variables $x_1, x_2, ..., x_N$, N is large
Task: Build a joint distribution function $Pr(x_1, ...x_N)$
Goal: Efficiently represent, estimate, and answer inference queries on the distribution

Bayesian Networks 2

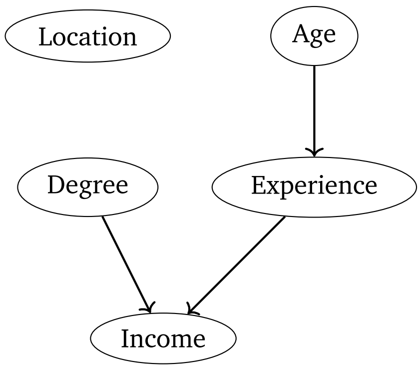
Graph G : Directed Acyclic Graph(DAG)
Potentials : defined at each node in terms of its parents

$$\psi_i(x_i, Pa(x_i)) = Pr(x_i|Pa(x_i))$$

where $Pa(x_i) =$ set of nodes in G pointing to x_i Probability Distribution

$$Pr(X_1..X_N) = \prod_{i=1}^{n} Pr(x_i|Pa(x_i))$$

Example:



Conditional Independencies 3

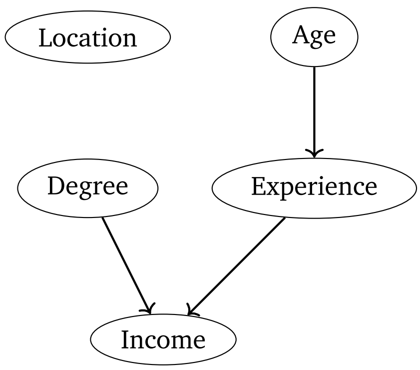
Given three sets of variables X,Y,Z , X is conditionally independent of Y given Z iff ($X \perp\!\!\!\perp Y|Z$) iff

$$Pr(X|Y, Z) = Pr(X|Z)$$

Local Conditional Independencies in BN: for each x_i

$$x_i \perp\!\!\!\perp Nd(x_i)|Pa(x_i)$$

where $Nd(x_i) = non-descendants$ of x_i
Example:



For the above graph , the Local CIs are

$$L \perp\!\!\!\perp A, D, E, I$$

$$A \perp\!\!\!\perp L, D$$

$$D \perp\!\!\!\perp A, L, E$$

$$E \perp\!\!\!\perp L, D|A$$

$$I \perp\!\!\!\perp A, L|D, E$$

Drawing a Bayesian Network Starting from Distribution 4

Drawing a BN starting from a distribution Given a distribution $P(x_1, ..., x_n)$ to which we can ask any CI of the form "Is $X \perp\!\!\!\perp Y|Z$?" and get a yes/no answer.
Goal: Draw a minimal, correct BN G to represent P.
1) A DAG G is correct if all Local-CIs that are implied in G hold in P.
2) A DAG G is minimal if we cannot remove any edge(s) from G and still get a correct BN for P.

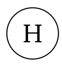
4.a Algorithm for drawing a BN from CIs

Let $x_1, \ldots, x_n$ = Choose an ordering of the variables.
For $i = 1, \ldots, n$, define S as the smallest subset of $Q_i = \{x_1, \ldots, x_{i-1}\}$ such that

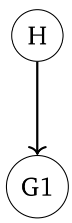
$$x_i \perp\!\!\!\perp Q_i - S|S$$

Make each variable in S a parent of x_i.
Example :
H = hardworking or not
I = intelligent or not
G1 = Grade in Course 1
G2 = Grade in Course 2

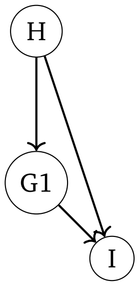
Consider order = H, G1, I, G2
Step 1: Add node H.



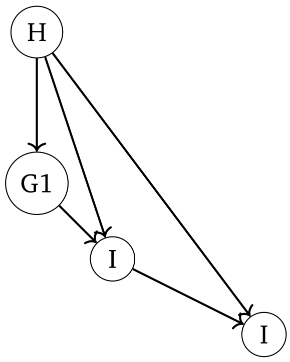
Step 2:
Is $G1 \perp\!\!\!\perp H|\phi$? $\rightarrow$ No
Is $G1 \perp\!\!\!\perp \phi|H$? $\rightarrow$ Yes
Make H the parent of G1.



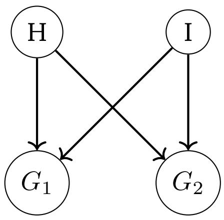
Step 3:
Is $I \perp\!\!\!\perp H, G_1|\phi$? $\rightarrow$ No
Is $I \perp\!\!\!\perp G_1|H$? $\rightarrow$ No
Is $I \perp\!\!\!\perp H|G_1$? $\rightarrow$ No
Is $I \perp\!\!\!\perp \phi|H, G_1$? $\rightarrow$ Yes
Make H,G_1 the parent of I.



Step 4:
Is $G_2 \perp\!\!\!\perp H, G_1, I | \phi$? $\rightarrow$ No
Is $G_2 \perp\!\!\!\perp \phi | H, G_1, I$? $\rightarrow$ Yes
Is $G_2 \perp\!\!\!\perp H, I | G_1$? $\rightarrow$ No
Is $G_2 \perp\!\!\!\perp I | H, G_1$? $\rightarrow$ No
Is $G_2 \perp\!\!\!\perp G_1 | H, I$? $\rightarrow$ Yes
Is $G_2 \perp\!\!\!\perp H | I, G_1$? $\rightarrow$ No
Is $G_2 \perp\!\!\!\perp H, G_1 | I$? $\rightarrow$ No
Is $G_2 \perp\!\!\!\perp I, G_1 | H$? $\rightarrow$ No
Make H,I the parents of G_2.



Alternate Ordering : H , I , G_1 , G_2



CIs and Factorization 5

Theorem: Given a distribution $P(x_1, \ldots, x_n)$ and a directed acyclic graph (DAG) G, if P satisfies the Local-CI induced by G, then P can be factorized according to G.

$$\text{Local-CI}(P, G) \implies \text{Factorize}(P, G)$$

Proof:

- Let $X_1, X_2, \ldots, X_n$ be topologically ordered such that parents appear before their children in the directed acyclic graph (DAG) G.
- For every X_i, the set $X_1, \ldots, X_{i-1}$ consists of all non-descendants of X_i, i.e.,

  $$X_1, \ldots, X_{i-1} = \text{Pa}_G(X_i) \cup \text{ND}'(X_i),$$

  where $\text{Pa}_G(X_i)$ are the parents of X_i in G and $\text{ND}'(X_i)$ are the non-descendants of X_i in G.
- From the assumption of local conditional independence ($\text{Local-CI}(P, G)$), we have:

  $$P(X_i \mid X_1, \ldots, X_{i-1}) = P(X_i \mid \text{Pa}_G(X_i) \cup \text{ND}'(X_i)) = P(X_i \mid \text{Pa}_G(X_i)).$$

- Using the chain rule of probability, we can write the joint distribution as:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}).$$

- Substituting the local conditional independence result into the chain rule, we get:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}_G(X_i)).$$

- This shows that the joint distribution P can be factorized according to the structure of the DAG G, as required.

$$\therefore \text{Local-CI}(P, G) \implies \text{Factorize}(P, G).$$

Theorem: Given a distribution $P(x_1, \ldots, x_n)$ and a directed acyclic graph (DAG) G, if P can be factorized according to G, then P satisfies the (Local-CI) induced by G.

$$\text{Factorize}(P, G) \implies \text{Local-CI}(P, G)$$

CI properties 6

- **Symmetry:**
$$(X \perp Y \mid Z) \implies (Y \perp X \mid Z).$$

- **Decomposition:**
$$(X \perp Y, W \mid Z) \implies (X \perp Y \mid Z).$$

- **Weak Union:**
$$(X \perp Y, W \mid Z) \implies (X \perp Y \mid Z, W).$$

- **Contraction:**
$$(X \perp W \mid Z, Y) \wedge (X \perp Y \mid Z) \implies (X \perp Y, W \mid Z).$$
$$(X \perp W | Y) \wedge (X \perp Y) \implies (X \perp Y, W)$$

Global CIs in a Bayesian Network 7

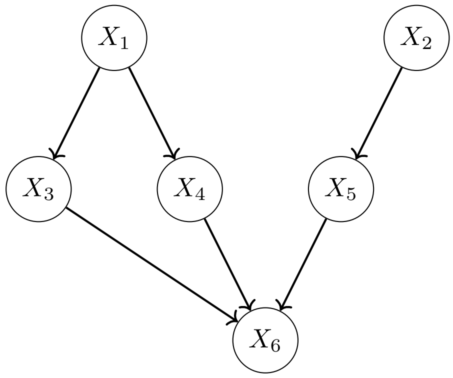
- Let X, Y, and Z be three sets of variables. If Z d-separates X from Y in the Bayesian Network (BN), then:
$$X \perp Y \mid Z.$$

- In a directed graph H, Z d-separates X from Y if all paths P from X to Y is blocked by Z.

- A path P is blocked by Z if any of the following four conditions hold:

  1. $X_1 \leftrightharpoons X_2 \leftrightharpoons \cdots \leftrightharpoons X_{i-1} \rightarrow X_i \rightarrow X_{i+1} \leftrightharpoons \cdots \leftrightharpoons X_n$ and $X_i \in Z$.
  2. $X_1 \leftrightharpoons X_2 \leftrightharpoons \cdots \leftrightharpoons X_{i-1} \leftarrow X_i \leftarrow X_{i+1} \leftrightharpoons \cdots \leftrightharpoons X_n$ and $X_i \in Z$.
  3. $X_1 \leftrightharpoons X_2 \leftrightharpoons \cdots \leftrightharpoons X_{i-1} \rightarrow X_i \leftarrow X_{i+1} \leftrightharpoons \cdots \leftrightharpoons X_n$ and $X_i \in Z$.
  4. $X_1 \leftrightharpoons X_2 \leftrightharpoons \cdots \leftrightharpoons X_{i-1} \rightarrow X_i \leftarrow X_{i+1} \leftrightharpoons \cdots \leftrightharpoons X_n$ where $X_i \notin Z$ and $Desc(X_i) \notin Z$

Theorem:
The d-separation test identifies the complete set of conditional independencies that hold in all distributions that conform to a given Bayesian network.
Example:



Let $X = \{X_3, X_5\}$, $Z = \emptyset$, and $Y = \{X_4\}$. We aim to show that X and Y are not marginally independent.
Reasoning:

- The graph structure involves the following paths:
  - $X_3 \to X_6 \leftarrow X_4$: This path is blocked because X_6 is a collider, and $Z = \emptyset$ (no descendants of X_6 are in Z).
  - $X_5 \to X_6 \leftarrow X_4$: Similarly, this path is blocked for the same reason X_6 is a collider and $Z = \emptyset$.
- However, X_3 and X_4 share a common parent X_1, which means they are not marginally independent.
- Therefore, $X = \{X_3, X_5\}$ and $Y = \{X_4\}$ are not marginally independent due to the common cause X_1.
  Additionally: For conditional independence, if we condition on X_1, then $X_3 \perp X_4 | X_1$, because conditioning on X_1 blocks the path through the common cause. However, this does not apply to the marginal independence scenario described above.

Note: In a Bayesian Network, the set of CIs combined with axioms of probability can be used to derive the Global CIs.

Equivalence of BNs 8

Definition: Two Bayesian Network DAGs are said to be equivalent if they express the same set of CIs.
Example:
Consider DAGs $\mathbf{G_1}$ and $\mathbf{G_2}$ representing BNs of the distributions $P_1(X, Y, Z)$ and $P_2(X, Y, Z)$ respectively as follows

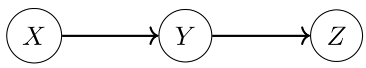


Figure 1: $\mathbf{G_1}$

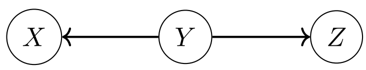


Figure 2: $\mathbf{G_2}$

Now set of CIs represented by $\mathbf{G_1}$ is $\{X \perp\!\!\!\perp Z \mid Y\}$ using d-separation algorithm.
Set of CIs expressed by $\mathbf{G_2}$ is also $\{X \perp\!\!\!\perp Z \mid Y\}$ using Local CI rule.
Hence, as set of CIs expressed by both the DAGs are same, $\mathbf{G_1}$ and $\mathbf{G_2}$ are equivalent BNs.

$$\mathbf{G_1} \equiv \mathbf{G_2} \tag{1}$$

Now, consider DAG $\mathbf{G_3}$ as follows

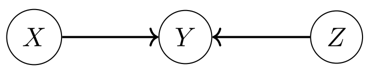


Figure 3: $\mathbf{G_3}$

Now set of CIs represented by $\mathbf{G_3}$ is $\{X \perp\!\!\!\perp Z \mid \phi\}$, that is, $\{X \perp\!\!\!\perp Z\}$ using d-separation algorithm. We see that CIs set for $\mathbf{G_3}$ is not same as that of $\mathbf{G_1}$ and $\mathbf{G_2}$.
Hence, BN $\mathbf{G_3}$ is not equivalent to that of $\mathbf{G_1}$ and $\mathbf{G_2}$ by definition.

$$\mathbf{G_3} \not\equiv \mathbf{G_1} \tag{2}$$

$$\mathbf{G_3} \not\equiv \mathbf{G_2} \tag{3}$$

Now, consider DAG $\mathbf{G_4}$ as follows

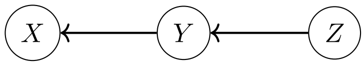


Figure 4: $\mathbf{G_4}$

By a similar argument we get

$$\mathbf{G_4} \equiv \mathbf{G_1} \equiv \mathbf{G_2} \tag{4}$$

$$\mathbf{G_4} \not\equiv \mathbf{G_3} \tag{5}$$

Consider BN DAGs $\mathbf{G_5}$ and $\mathbf{G_6}$ as follows

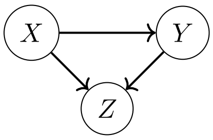


Figure 5: $\mathbf{G_5}$

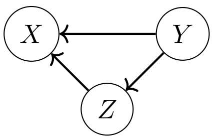


Figure 6: $\mathbf{G_6}$

We can see that CIs set for both $\mathbf{G_5}$ and $\mathbf{G_6}$ are empty sets. Hence, they will be equivalent BNs by definition.

$$\mathbf{G_5} \equiv \mathbf{G_6} \tag{6}$$

8.a Relation between BN equivalence and BN structure

Theorem: Two BNs G_1 and G_2 are equivalent if and only if they have the same skeleton and the same set of immoralities.
Note: An ***immorality*** is a structure of the form $\mathbf{x} \to \mathbf{y} \leftarrow \mathbf{z}$, with no edge between **x** and **z**.

Using this theorem we can easily identify equivalence between BNs in the previous example.
$\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_3}$ and $\mathbf{G_4}$ have the same skeleton structure, while $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{G_4}$ have empty immorality set while $\mathbf{G_3}$ has a singleton immorality set. So, using the theorem, we get $\mathbf{G_1} \equiv \mathbf{G_2} \equiv \mathbf{G_4} \not\equiv \mathbf{G_3}$ which verifies with our earlier result.

Optimal BN Algorithm 9

Note: The following algorithm, also known as PC Algorithm, is valid when a BN exists for the distribution

Given a distribution $P(x_1, ..., x_n)$ to which we can ask any CI of the form "Is $X \perp\!\!\!\perp Y|Z$?" and get a yes/no answer.

9.a The Algorithm

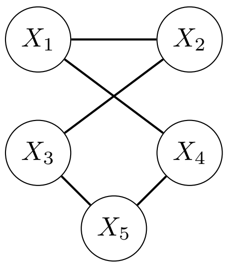
- **Step 1:** Initialise **G** to be a complete graph (All nodes have undirected edged beween them)
- **Step 2:** For each pair of variables x_i and x_j, if there exists a subset Z such that $x_i \perp\!\!\!\perp x_j \mid Z$
  - Remove edge between x_i and x_j
  - Set $W(x_i, x_j) = Z$
- **Step 3:** For each triplet x_i, x_j, x_k such that there is an edge between x_i-x_k and x_j-x_k, no edge between x_i,x_j and $x_k \notin W(x_i, x_j)$
  - Orient the edges between them as x_i –¿ x_k ¡– x_j (Add immorality)
- **Step 4:** When no such triplet remain, orient the rest of the edges in any way such that **G** is a DAG and no new immorality is added.

9.b An example with PC Algorithm

Consider a distribution with 5 random variables X_1, X_2, X_3, X_4, X_5 with the following CIs
$X_1 \perp\!\!\!\perp X_3 \mid X_2$;
$X_1 \perp\!\!\!\perp X_5 \mid X_3, X_4$;
$X_2 \perp\!\!\!\perp X_5 \mid X_3, X_4$;
$X_2 \perp\!\!\!\perp X_4 \mid X_1$;
$X_3 \perp\!\!\!\perp X_4 \mid X_1, X_2$;

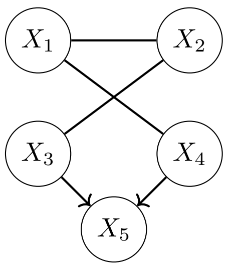
Step 1: Graph G with 5 nodes and all interconnected

Step 2: For each given CI of the form $X \perp\!\!\!\perp Y \mid Z$, we remove an edge between X and Y and set $W(X, Y) = Z$. So, starting with $X_1 \perp\!\!\!\perp X_3 \mid X_2$, we remove the edge between X_1 and X_3 and set $W(X_1, X_3) = X_2$. We do the same with all other four CIs as well to get the following graph

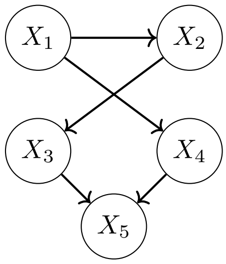


With
$W(X_1, X_3) = X_2$;
$W(X_1, X_5) = \{X_3, X_4\}$;
$W(X_2, X_5) = \{X_3, X_4\}$;
$W(X_2, X_4) = X_1$;
$W(X_3, X_4) = \{X_1, X_2\}$;

Step 3: To find triplets which has immorality
Consider the triplet X_3, X_5, X_4, so $X_3 - X_5$ exists, $X_4 - X_5$ exists, $X_3 - X_4$ does not and $X_5 \notin W(X_3, X_4)$, so there will be an immorality as $X_3 - > X_5 < -X_4$
Consider the triplet X_1, X_2, X_3, so $X_1 - X_2$ exists, $X_3 - X_2$ exists, $X_1 - X_3$ does not but $X_2 \in W(X_1, X_3)$, so there will not be an immorality.
Similarly, we can do this for other triplets which results in only one immorality which we discussed above. So we orient that triplet as immorality as follows



Now we can orient rest of the edges in any direction such that there is no added immorality and **G** is still a DAG. One such possible orientation of edge directions is as follows



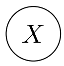
Hence, this is the final BN according to PC algorithm.

Shortcomings of Bayesian Networks 10

All probability distributions cannot be represented by Bayesian Networks

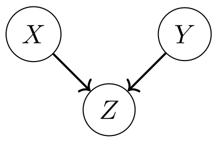
Example 1
Consider the distribution of 3 random variables X, Y and Z with the following CIs given
$X \perp\!\!\!\perp Y$,
$X \perp\!\!\!\perp Z$,
$Y \perp\!\!\!\perp Z$,
$X \not\perp\!\!\!\perp \{Y, Z\}$

Now when we construct Bayesian Network for this distribution using the order X, Y and then Z
Step 1: Add node X.



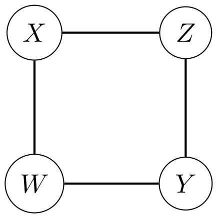
Step 2:
"Is $Y \perp\!\!\!\perp X|\phi$?" -¿ Yes
"Is $Y \perp\!\!\!\perp \phi|X$?" -¿ Yes
No parent added for Y.

Step 3:
"Is $Z \perp\!\!\!\perp X, Y|\phi$?" -¿ No
"Is $Z \perp\!\!\!\perp Y|X$?" -¿ No
"Is $Z \perp\!\!\!\perp X|Y$?" -¿ No
"Is $Z \perp\!\!\!\perp \phi|X, Y$?" -¿ Yes
Make X,Y the parent of Z.

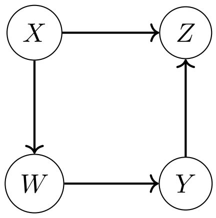


This is the final Bayesian Network. We can observe that as there is a direct edge between X and Z, as well as, Y and Z, the CIs in the originial distribution $X \perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z$ get lost in the BN. Hence, we can conclude that this original distribution cannot be represented by a BN.

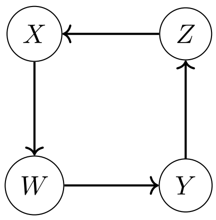
Example 2 (Symmetric Dependencies)
Consider the distribution of 4 random variables X, Y, Z and W with the following CIs given
$X \perp\!\!\!\perp Y \mid Z, W$
$Z \perp\!\!\!\perp W \mid X, Y$
Constructing BN using PC Algorithm, we get the following skeleton structure with $W(X, Y) = \{Z, W\}$ and $W(Z, W) = \{X, Y\}$ with an empty immorality set. So according to the algorithm, we are free to direct the edges any way we want given that the graph is a DAG and has no immoralities.



We can see that, if condition of DAG is imposed, for all direction permutations of edge will lead to non-empty immorality set. (as follows)



And if condition of no immorality is imposed, for all permutations, then the graph will be cyclic, therefore will not be DAG. (as follows)



Hence, we see that we cannot construct a BN for this distribution.

Other than this, BNs are also not great in for representing symmetric interactions among variables due the directed nature of the edges.