Graphical Models

David A. Levin

DAIICT

January 2020

Graphical Models

- A **graphical model** specifies a factorization which a joint density must obey.
- Example: If $X_1, \dots, X_n$ are independent, then the joint density completely factors into marginals:

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)\,. \tag{1}$$

- This is too strong a constraint to require of general applicable probability models.
- A graphical model specifies a class of joint densities which obey a factorization which still affords useful simplification but preserves some generality.

Graphical Models

- A **graphical model** specifies a factorization which a joint density must obey.
- Example: If $X_1, \ldots, X_n$ are independent, then the joint density completely factors into marginals:

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)\,. \tag{1}$$

- This is too strong a constraint to require of general applicable probability models.
- A graphical model specifies a class of joint densities which obey a factorization which still affords useful simplification but preserves some generality.

Graphical Models

- A **graphical model** specifies a factorization which a joint density must obey.
- Example: If $X_1, \ldots, X_n$ are independent, then the joint density completely factors into marginals:

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)\,. \tag{1}$$

- This is too strong a constraint to require of general applicable probability models.
- A graphical model specifies a class of joint densities which obey a factorization which still affords useful simplification but preserves some generality.

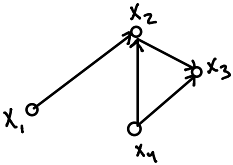
Graphical Models

- A **graphical model** specifies a factorization which a joint density must obey.
- Example: If $X_1, \ldots, X_n$ are independent, then the joint density completely factors into marginals:

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \,. \tag{1}$$

- This is too strong a constraint to require of general applicable probability models.
- A graphical model specifies a class of joint densities which obey a factorization which still affords useful simplification but preserves some generality.

- nodes represent variables
- directed arrows represent dependencies
- Any variable with an arrow pointing to x_i is called a **parent** of x_i; we denote by pa_i all the parents of x_i.



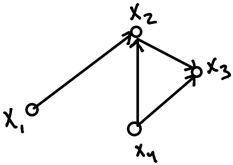
- A joint distribution respects the graphical model encoded by a directed graph if

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i \mid \mathsf{pa}_i)\,.$$

- In the example graph, a joint distribution obeys the graphical model if

$$p(x_1, x_2, x_3, x_4) = p(x_1)\,p(x_2 \mid x_1, x_4)\,p(x_3 \mid x_2, x_4)\,p(x_4)\,.$$

- nodes represent variables
- directed arrows represent dependencies
- Any variable with an arrow pointing to x_i is called a **parent** of x_i; we denote by pa_i all the parents of x_i.



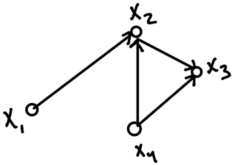
- A joint distribution respects the graphical model encoded by a directed graph if

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid \mathrm{pa}_i)\,.$$

- In the example graph, a joint distribution obeys the graphical model if

$$p(x_1, x_2, x_3, x_4) = p(x_1)\,p(x_2 \mid x_1, x_4)\,p(x_3 \mid x_2, x_4)\,p(x_4)\,.$$

- nodes represent variables
- directed arrows represent dependencies
- Any variable with an arrow pointing to x_i is called a **parent** of x_i; we denote by pa_i all the parents of x_i.



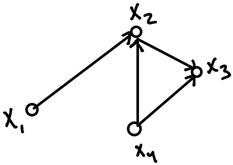
- A joint distribution respects the graphical model encoded by a directed graph if

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid \mathsf{pa}_i) \,.$$

- In the example graph, a joint distribution obeys the graphical model if

$$p(x_1, x_2, x_3, x_4) = p(x_1)\, p(x_2 \mid x_1, x_4)\, p(x_3 \mid x_2, x_4)\, p(x_4) \,.$$

- nodes represent variables
- directed arrows represent dependencies
- Any variable with an arrow pointing to x_i is called a **parent** of x_i; we denote by pa_i all the parents of x_i.



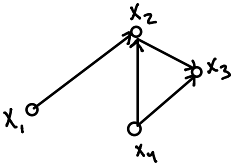
- A joint distribution respects the graphical model encoded by a directed graph if

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid \mathsf{pa}_i)\,.$$

- In the example graph, a joint distribution obeys the graphical model if

$$p(x_1, x_2, x_3, x_4) = p(x_1)\,p(x_2 \mid x_1, x_4)\,p(x_3 \mid x_2, x_4)\,p(x_4)\,.$$

- nodes represent variables
- directed arrows represent dependencies
- Any variable with an arrow pointing to x_i is called a **parent** of x_i; we denote by pa_i all the parents of x_i.



- A joint distribution respects the graphical model encoded by a directed graph if

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid \mathsf{pa}_i)\,.$$

- In the example graph, a joint distribution obeys the graphical model if

$$p(x_1, x_2, x_3, x_4) = p(x_1)\,p(x_2 \mid x_1, x_4)\,p(x_3 \mid x_2, x_4)\,p(x_4)\,.$$

- The graphical model encodes *conditional independence* statements.
- A set of variables X and Y are conditionally independent given Z if

$$\mathbb{P}(X \in A, Y \in B \mid Z) = \mathbb{P}(X \in A \mid Z)\mathbb{P}(X \in B \mid Z)\,.$$

- In terms of pdfs/pmfs,

$$p(x, y \mid z) = p(x \mid z)\, p(y \mid z)\,.$$

- Or equivalently,

$$p(x \mid y, z) = p(x \mid z)\,.$$

- The graphical model encodes *conditional independence* statements.
- A set of variables X and Y are conditionally independent given Z if
$$\mathbb{P}(X \in A, Y \in B \mid Z) = \mathbb{P}(X \in A \mid Z)\mathbb{P}(X \in B \mid Z).$$
- In terms of pdfs/pmfs,
$$p(x, y \mid z) = p(x \mid z)\, p(y \mid z).$$
- Or equivalently,
$$p(x \mid y, z) = p(x \mid z).$$

- The graphical model encodes *conditional independence* statements.
- A set of variables X and Y are conditionally independent given Z if

$$\mathbb{P}(X \in A, Y \in B \mid Z) = \mathbb{P}(X \in A \mid Z)\mathbb{P}(X \in B \mid Z)\,.$$

- In terms of pdfs/pmfs,

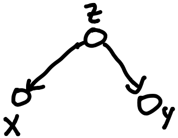
$$p(x, y \mid z) = p(x \mid z)\, p(y \mid z)\,.$$

- Or equivalently,

$$p(x \mid y, z) = p(x \mid z)\,.$$

- The graphical model encodes *conditional independence* statements.
- A set of variables X and Y are conditionally independent given Z if
$$\mathbb{P}(X \in A, Y \in B \mid Z) = \mathbb{P}(X \in A \mid Z)\mathbb{P}(X \in B \mid Z)\,.$$
- In terms of pdfs/pmfs,
$$p(x, y \mid z) = p(x \mid z)\, p(y \mid z)\,.$$
- Or equivalently,
$$p(x \mid y, z) = p(x \mid z)\,.$$

Example 1

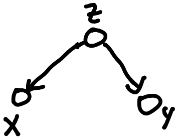


-
- Note that

$$p(x, y \mid z) = \frac{p(x, y, z)}{\sum_{x,y} p(x, y, z)} = \frac{p(x \mid z)\, p(y \mid z)\, p(z)}{\sum_x p(x \mid z) \sum_y p(y \mid z)\, p(z)} = p(x \mid z)\, p(y \mid z)$$

- Any time X and Y are separated by a **tail-to-tail** vertex Z, they are conditionally independent.
- Note that X and Y are not independent unconditionally, in general.
- Exercise: Show by example that X and Y are not necessarily independent.

Example 1

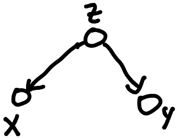


-
- Note that

$$p(x, y \mid z) = \frac{p(x, y, z)}{\sum_{x,y} p(x, y, z)} = \frac{p(x \mid z)\, p(y \mid z)\, p(z)}{\sum_x p(x \mid z) \sum_y p(y \mid z)\, p(z)} = p(x \mid z)\, p(y \mid z)$$

- Any time X and Y are separated by a **tail-to-tail** vertex Z, they are conditionally independent.
- Note that X and Y are not independent unconditionally, in general.
- Exercise: Show by example that X and Y are not necessarily independent.

Example 1

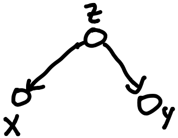


-
- Note that

$$p(x, y \mid z) = \frac{p(x, y, z)}{\sum_{x,y} p(x, y, z)} = \frac{p(x \mid z)\, p(y \mid z)\, p(z)}{\sum_x p(x \mid z) \sum_y p(y \mid z)\, p(z)} = p(x \mid z)\, p(y \mid z)$$

- Any time X and Y are separated by a **tail-to-tail** vertex Z, they are conditionally independent.
- Note that X and Y are not independent unconditionally, in general.
- Exercise: Show by example that X and Y are not necessarily independent.

Example 1

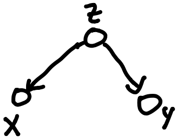


-
- Note that

$$p(x, y \mid z) = \frac{p(x, y, z)}{\sum_{x,y} p(x, y, z)} = \frac{p(x \mid z)\, p(y \mid z)\, p(z)}{\sum_x p(x \mid z) \sum_y p(y \mid z)\, p(z)} = p(x \mid z)\, p(y \mid z)$$

- Any time X and Y are separated by a **tail-to-tail** vertex Z, they are conditionally independent.
- Note that X and Y are not independent unconditionally, in general.
- Exercise: Show by example that X and Y are not necessarily independent.

Example 1

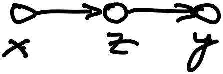


-
- Note that

$$p(x, y \mid z) = \frac{p(x, y, z)}{\sum_{x,y} p(x, y, z)} = \frac{p(x \mid z)\, p(y \mid z)\, p(z)}{\sum_x p(x \mid z) \sum_y p(y \mid z)\, p(z)} = p(x \mid z)\, p(y \mid z)$$

- Any time X and Y are separated by a **tail-to-tail** vertex Z, they are conditionally independent.
- Note that X and Y are not independent unconditionally, in general.
- Exercise: Show by example that X and Y are not necessarily independent.

Example 2



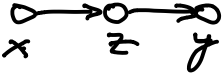
-
-

$$p(x, y \mid z) = \frac{p(x, y, z)}{p(z)} = \frac{p(x)\,p(z \mid x)\,p(y \mid z)}{p(z)}$$

$$= \frac{p(x \mid z)\,p(z)\,p(y \mid z)}{p(z)} = p(x \mid z)\,p(y \mid z)$$

- Thus, if x and y are connected by a **head-to-tail** vertex z, then they are conditionally independent.

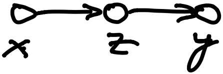
Example 2



$$p(x, y \mid z) = \frac{p(x, y, z)}{p(z)} = \frac{p(x)\,p(z \mid x)\,p(y \mid z)}{p(z)}$$
$$= \frac{p(x \mid z)\,p(z)\,p(y \mid z)}{p(z)} = p(x \mid z)\,p(y \mid z)$$

- Thus, if x and y are connected by a **head-to-tail** vertex z, then they are conditionally independent.

Example 2



$$p(x, y \mid z) = \frac{p(x, y, z)}{p(z)} = \frac{p(x)\,p(z \mid x)\,p(y \mid z)}{p(z)}$$

$$= \frac{p(x \mid z)\,p(z)\,p(y \mid z)}{p(z)} = p(x \mid z)\,p(y \mid z)$$

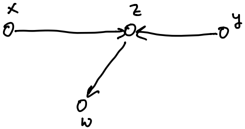
- Thus, if x and y are connected by a **head-to-tail** vertex z, then they are conditionally independent.

-
-

$$p(x,y) = \sum_z p(x,y,z) = \sum_z p(x)\,p(y)\,p(z \mid x,y)$$
$$= p(x)\,p(y) \sum_z p(z \mid x,y) = p(x)\,p(y)\,.$$

so X and Y are independent given $\emptyset$.

- Exercise: Give an example to show that X and Y are not conditionally independent given Z.
- Even more is true:



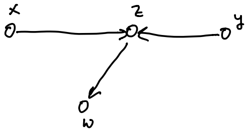
X and Y are not independent given W (in general).

-
-

$$p(x,y) = \sum_z p(x,y,z) = \sum_z p(x)\,p(y)\,p(z \mid x,y)$$
$$= p(x)\,p(y) \sum_z p(z \mid x,y) = p(x)\,p(y)\,.$$

so X and Y are independent given $\emptyset$.

- Exercise: Give an example to show that X and Y are not conditionally independent given Z.
- Even more is true:



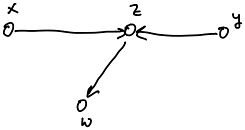
X and Y are not independent given W (in general).

$$p(x,y) = \sum_z p(x,y,z) = \sum_z p(x)\,p(y)\,p(z \mid x,y)$$

$$= p(x)\,p(y) \sum_z p(z \mid x,y) = p(x)\,p(y)\,.$$

so X and Y are independent given $\emptyset$.

- Exercise: Give an example to show that X and Y are not conditionally independent given Z.
- Even more is true:



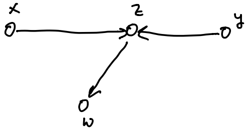
X and Y are not independent given W (in general).

-
-

$$p(x,y) = \sum_z p(x,y,z) = \sum_z p(x)\,p(y)\,p(z \mid x,y)$$
$$= p(x)\,p(y) \sum_z p(z \mid x,y) = p(x)\,p(y)\,.$$

so X and Y are independent given $\emptyset$.

- Exercise: Give an example to show that X and Y are not conditionally independent given Z.
- Even more is true:



X and Y are not independent given W (in general).

- Let A, B, C be sets of vertices. The set C **blocks** A to B if every path from a vertex $a \in A$ to a vertex $b \in B$ contains either
  - a vertex $c \in C$ so that c is head-to-tail or tail-to-tail, or
  - a vertex v which is head-to-head, and which neither belongs to C or has a decendent belonging to C.
-

Theorem 1 (*d*-separation).

If C blocks A and B in the di-graph, then the variables A and the variables B are conditionally independent given C.

- Let A, B, C be sets of vertices. The set C **blocks** A to B if every path from a vertex $a \in A$ to a vertex $b \in B$ contains either
  - a vertex $c \in C$ so that c is head-to-tail or tail-to-tail, or
  - a vertex v which is head-to-head, and which neither belongs to C or has a decendent belonging to C.
-

Theorem 1 (d-separation).

If C blocks A and B in the di-graph, then the variables A and the variables B are conditionally independent given C.

- Let A, B, C be sets of vertices. The set C **blocks** A to B if every path from a vertex $a \in A$ to a vertex $b \in B$ contains either
  - a vertex $c \in C$ so that c is head-to-tail or tail-to-tail, or
  - a vertex v which is head-to-head, and which neither belongs to C or has a decendent belonging to C.
-

Theorem 1 (*d*-separation).

If C blocks A and B in the di-graph, then the variables A and the variables B are conditionally independent given C.

- Let A, B, C be sets of vertices. The set C **blocks** A to B if every path from a vertex $a \in A$ to a vertex $b \in B$ contains either
  - a vertex $c \in C$ so that c is head-to-tail or tail-to-tail, or
  - a vertex v which is head-to-head, and which neither belongs to C or has a decendent belonging to C.
-

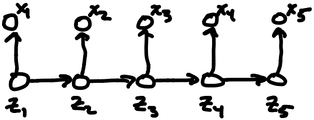
Theorem 1 (d-separation).

If C blocks A and B in the di-graph, then the variables A and the variables B are conditionally independent given C.

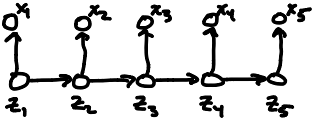
- Let A, B, C be sets of vertices. The set C **blocks** A to B if every path from a vertex $a \in A$ to a vertex $b \in B$ contains either
  - a vertex $c \in C$ so that c is head-to-tail or tail-to-tail, or
  - a vertex v which is head-to-head, and which neither belongs to C or has a decendent belonging to C.
-

Theorem 1 (d-separation).

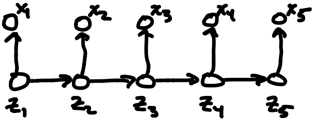
If C blocks A and B in the di-graph, then the variables A and the variables B are conditionally independent given C.



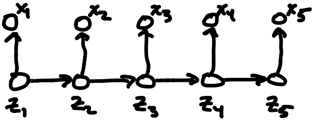
- The above is a **Hidden Markov Model**.
- We can use *d*-separation to establish useful conditional independence relations:
  - $$(X_1, \ldots, X_k) \perp (X_{k+1}, \ldots, X_n) \mid Z_k$$
  - $$X_1, \ldots, X_{k-1} \perp X_k \mid Z_k$$



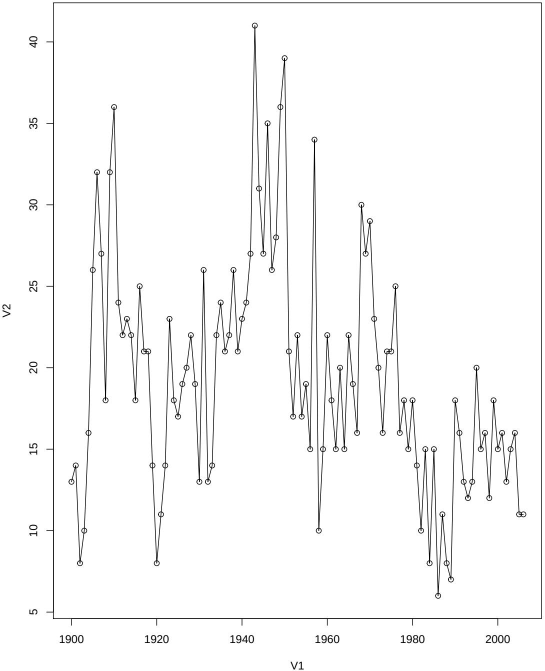
- The above is a **Hidden Markov Model**.
- We can use d-separation to establish useful conditional independence relations:
  - $$(X_1, \ldots, X_k) \perp (X_{k+1}, \ldots, X_n) \mid Z_k$$
  - $$X_1, \ldots, X_{k-1} \perp X_k \mid Z_k$$



- The above is a **Hidden Markov Model**.
- We can use d-separation to establish useful conditional independence relations:
  - $$(X_1, \ldots, X_k) \perp (X_{k+1}, \ldots, X_n) \mid Z_k$$
  - $$X_1, \ldots, X_{k-1} \perp X_k \mid Z_k$$



- The above is a **Hidden Markov Model**.
- We can use *d*-separation to establish useful conditional independence relations:
  - $$(X_1, \ldots, X_k) \perp (X_{k+1}, \ldots, X_n) \mid Z_k$$
  - $$X_1, \ldots, X_{k-1} \perp X_k \mid Z_k$$



40
35
30
25
20
15
10
5
V2
1900
1920
1940
1960
1980
2000
V1