BITS Pilani

Pilani | Dubai | Goa | Hyderabad

Probabilistic Graphical Model

Session # 2 : Mathematical Preliminaries

Seetha Parameswaran

seetha.p@pilani.bits-pilani.ac.in

innovate achieve lead

The instructor is gratefully acknowledging the authors who made their course materials freely available online.

Table of Contents

Uncertainty

- We select a course of actions among many possibilities.
- Decisions may be based on the information obtained from the environment, previous knowledge and the objectives.
- Eg: It looks cloudy. Should I carry an umbrella?
- The information and knowledge is incomplete or unreliable. So the decisions made are not certain. We make decisions under uncertainty.
- One of the goals of AI is to develop systems that can reason and make decisions under uncertainty.

Uncertainty

$T = f(V_1, V_2, \ldots V_N)$ → cannot be observed

→ compute T exactly

- Due to
  - Partial observability
  - Non-determinism
- Complexity increases
  - Each piece of knowledge may not be independently used to arrive at decisions.
  - Deduced facts are maintained along with new facts. This increases the knowledge base.
- Examples
  - A medical doctor in an emergency.
  - An autonomous vehicle that detects what might be an obstacle in its way.
  - A financial agent needs to select the best investment.

Uncertain Reasoning

$P(P_1/C), P(P_2/C) \ldots P(P_n/C)$

innovate achieve lead

Example

- Diagnosing a dental patients' toothache.
- Toothache may have different causes. Equation using propositional logic:

$P \Rightarrow Q,\ Q \Rightarrow R$

$\rightarrow P \Rightarrow R$

$$\boxed{Toothache \implies Cavity \lor GumProblem \lor Abscess \lor \ldots}$$

- Change to a causal rule.

GumProblem ⇒ Toothache

$$Cavity \implies Toothache$$

Causal rules

But not all cavity cause toothache.

- So make logically exhaustive.

Cavity1 ⇒ Toothache

Cavity2 ⇒ Toothache

Uncertain Reasoning

3 reasons for failure when using logic in Judgmental domains [medical diagnosis, law, business, design, automobile repair, gardening, ...]

- Laziness – complete set of antecedents and consequences
- Theoretical ignorance – no complete theory
- Practical ignorance – not all tests can be run

Belief and Degree of Belief

- Belief State is a representation of a set of all possible world states.
- Agent’s knowledge can provide only a degree of belief.
- Tool to deal with Degree of Belief is Probability Theory.

Belief is derived from

1. statistical data.
2. some general knowledge.
3. combination of evidence sources.

Probabilistic Statements

- Probability statements instead of propositional logic.
- Probability statements are made with respect to knowledge state.

Example

1. Probability that a patient has a cavity, given that she has toothache is 0.8.
2. Probability that a patient has a cavity, given that she has toothache and a history of gum disease is 0.4.

$P(C/T) = 0.8 \qquad P(C/T, G) = 0.4$

why?

Table of Contents

Sample Space

- A sample space Ω specifies set of all possible outcomes that we want to consider.

Coin toss $\quad \Omega = \{H, T\}$

Die Roll $\quad \Omega = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$

Discrete Sample space

- Probability of an outcome $P(\omega)$ specifies the chance or probability with each possible outcome.

$$P(H) = 0.5$$

$$P(⚄) = \frac{1}{6}$$

Measurable Event

- An event space S or Φ is a subset of outcomes, where probabilities can be assigned.
- We are interested in the set of outcomes.

Even die roll $E = \{⚁, ⚃, ⚅\}$

Prime die roll $M = \{⚁, ⚂, ⚄\}$

$\Omega = \{⚀, ⚁, \ldots, ⚅\}$

S = event space i.e.

$S = \{\phi, \Omega, \{⚀\}, \{⚁\}, \ldots, \{⚀, ⚁\}, \ldots\}$

- Properties of Event Space
  - Event space contains empty event ϕ and the trivial event Ω.
  - It is closed under union.

    If $\alpha, \beta \in S$, then $\alpha \cup \beta \in S$

    $\alpha \rightarrow \bar{\alpha}$

  - It is closed under complementation.

    If $\alpha \in S$, then $\Omega - \alpha \in S$

Probability of Event

- Probability of an event is given by the sum of the probabilities of the outcomes it contains.

$$P(a) = \sum_{\omega \in a} P(\omega) \tag{1}$$

Even die roll $\quad P(E) = \frac{3}{6} = 0.5$

Prime die roll $\quad P(M) = \frac{3}{6} = 0.5$

$$P(E) = \frac{|E|}{|\Omega|} = \frac{3}{6}$$

$$P(⚀) = \frac{1}{|\Omega|}$$

$$P(⚁) = \frac{1}{|\Omega|}$$

$$P(\{⚀, ⚁, ⚄\}) = P(⚀) + P(⚁) + P(⚄)$$

Prior Probability

- Prior or Unconditional probabilities refer to degree of belief in the absence of any other information.

$$P(DieTotal = 11) = P((5,6)) + P((6,5)) = 1/18$$

$$\frac{1}{36} + \frac{1}{36} = \frac{2}{36} = \frac{1}{18}$$

$$P(DieTotal = 11) = \frac{1}{18}$$

$$P(DieTotal = 11 / Die1 = 1) = 0$$

Evidence

- The information that has already been revealed is called evidence.
  - She is having toothache. $Toothache = True$ or $toothache =$ false
  - We roll a dice and we get 5. $Die_1 = 5$

Posterior Probability

- Conditional or Posterior probability refer to the probability of some event occurring given a particular condition.

$$P(cavity\,|toothache) = 0.6$$

- Condition on all evidences that has been observed.

$$P(\alpha|\beta) = \frac{P(\alpha \wedge \beta)}{P(\beta)} \qquad \text{where } P(\beta) > 0 \tag{2}$$

Probability Model

- Associate a numerical probability $P(\omega)$ with each event S.
- Axioms of probability theory

axioms →

(1) $$\boxed{P(\omega) \geq 0} \tag{3}$$

(2) $$P(\Omega) = \sum_{\omega \in \Omega} P(\omega) = 1 \tag{4}$$

(3) if $\alpha, \beta \in S$ and $\alpha \wedge \beta = \phi$, $P(\alpha \cup \beta) = P(\alpha) + P(\beta)$

can be derived → $$P(\varphi) = 0 \tag{5}$$

$$P(a \mid \beta) = \frac{P(a \wedge \beta)}{P(\beta)} \quad \text{where } P(\beta) > 0 \tag{6}$$

can be derived → $$P(a \vee \beta) = P(a) + P(\beta) - P(a \wedge \beta) \tag{7}$$

$P(\alpha) = P(\alpha\bar{\beta}) + P(\alpha\beta)$ $\alpha\bar{\beta} = \alpha_{only}$

Derivation from the axioms

$P(\alpha) = P(\alpha_{only}) + P(\alpha \cap \beta)$ [By 3rd axiom]

$P(\beta) = P(\beta_{only}) + P(\alpha \cap \beta)$ [By 3rd axiom]

$P(\alpha \cup \beta) = P(\alpha_{only}) + P(\beta_{only}) + P(\alpha \cap \beta)$ [By 3rd axiom]

$= [P(\alpha) - P(\alpha \cap \beta)] + [P(\beta) - P(\alpha \cap \beta)] + P(\alpha \cap \beta)$

$$\boxed{P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)}$$

Bayes Rule

- Conditional probabilities can be derived from the prior given the evidence.

$$P(\alpha|\beta) = \frac{P(\beta|\alpha)P(\alpha)}{P(\beta)} \quad \text{where } P(\beta) > 0 \tag{8}$$

$$P(\alpha \wedge \beta) = P(\alpha)\,P(\beta/\alpha) = P(\beta)\,P(\alpha/\beta)$$

Example 1

Smart = "All Students are Smart"
Grade A = "All students have got grade A"

- Consider the student population, and let Smart denote smart students and GradeA denote students who got grade A. Based on estimates from past statistics assume that $P(GradeA|Smart) = 0.6$, the probability for students being smart is 0.3 and the prior probability of students receiving high grades is 0.2. Estimate the probability that the student is smart given grade A.

P(Smart) P(GradeA/Smart) → P(Smart/GradeA)

P(GradeA)

Solution

$$P(Smart / Grade = A) = \frac{P(Grade = A / Smart)\, P(Smart)}{P(Grade = A)}$$

$$\text{Given,} P(Smart) = 0.3$$

$$P(GradeA) = 0.2$$

$$P(GradeA|Smart) = 0.6$$

According to Bayes' rule

$$P(Smart|GradeA) = \frac{0.6 * 0.3}{0.2} = 0.9$$

Example 2

- Suppose that a tuberculosis (TB) skin test is 95 percent accurate. Suppose that 1 in 1000 of the subjects who get tested is infected. What is the probability of getting a positive test result?

95% accurate means if the patient is TB-infected then the test will be +ve with probability 95%, if the patient is not TB-infected the test will be -ve with probability 95%.

Solution

$P(\text{positive}) = P(\text{Positive} \,\&\, TB) + P(\text{Positive} \,\&\, \neg TB)$

$\neg \rightarrow \text{not}$

$P(A) = P(AB) + P(A\bar{B})$

$P(TB \,\&\, Pos)$

$$\text{Given}, P(TB) = 0.001$$

$$P(\text{infected subjects get a positive result}) = 0.001 * 0.95$$

$P(TB) \times P(Pos/TB)$

$$P(\text{uninfected subjects get a positive result}) = 0.999 * 0.05$$

$P(\neg TB) \times P(Pos/\neg TB)$

$P(TB \,\&\, \text{positive})$

$$P(Positive) = 0.001 * 0.95 + 0.999 * 0.05 = 0.0509$$

$P(\neg TB \,\&\, Pos)$

According to Bayes' rule,

$$P(TB|Positive) = \frac{0.001 * 0.95}{0.0509} = 0.0187$$

$< 2\%$

$= \frac{P(Pos \,\&\, TB)}{P(Pos)}$

Table of Contents

Random Variable

- Variables used in probability theory.
- Uppercase letter
- A random variable X is defined by a function that associates with each outcome in Ω a value or a state.

$f(\omega)$

$$\mathbf{P}(attribute = value) = \mathbf{P}(\omega \in \Omega : f_{attribute} = value)$$

$$\mathbf{P}(X = x) = \mathbf{P}(\omega \in \Omega : X(\omega) = x) \tag{9}$$

- Use $P(x)$ as a shorthand for $P(X = x)$.

$$\boxed{\sum_{x \in Val(X)} P(X = x) = \sum_{x} P(x) = 1} \tag{10}$$

Domain of Random Variable

- Every random variable has a domain, the set of possible values it can take.
- Finite random variable or Infinite random variable.
- Domain can be discrete or continuous (integer or real).
  Weather random variable – Domain $= \{sunny, overcast, rainy, cloudy, snow\}$
- Boolean Random variable
  - $Domain = Val(X) = \{true, false\}$
  - x^1 to denote *true* and x^0 to denote *false*.
  - The distribution of binary random variable is called a Bernoulli distribution.

Joint Distribution

- The distribution over several random variables are described using joint distribution.
- Set of random variables $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$
- Joint Distribution $\mathbf{P}(\mathcal{X}) = P(X_1, X_2, \ldots, X_n)$.
- Eg: Suppose random variable *Grade* reports the final grade of a student and the student's intelligence is given by *Intelligence*. Then the joint distribution $\mathbf{P}(Intelligence, Grade)$

$P(\text{Intelligence} = \text{low})$
$= P(I = \text{low}, G = A) + P(I = \text{low}, G = B) + P(I = \text{low}, G = C)$

		Intelligence	
		low	high
Grade	A	0.07	0.18
	B	0.28	0.09
	C	0.35	0.03

Marginal Distribution

- The distribution over events that can be described using X is often referred to as the marginal distribution over the random variable X.
- Summing out
- Marginal distribution is denoted by $P(X)$.
- Row-wise or Column-wise summations in the JD gives MD.

$$\mathbf{P}(Y) = \sum_{z \in Z} \mathbf{P}(Y, z) \qquad (11)$$

$$f_x(x) = \int f(x,y)\,dy$$

- Sum over all the possible combinations of values of the set of varaibles of Z.

Marginal Distribution

- Eg: Suppose random variable *Grade* reports the final grade of a student . Then the marginal distribution of *Grade*

$$P(Grade = A) = 0.25$$
$$P(Grade = B) = 0.37$$
$$P(Grade = C) = 0.38$$

$$P(Grade) = 1$$

Conditional Probability Distribution

- The conditional distribution over a random variable given an observation of the value of another random variable is referred to as Conditional Probability Distribution.
- Compute conditional probability of some variable given evidence about others.

$P(Y \& z_1) + P(Y \& z_2)$

$$\boxed{\mathbf{P}(Y) = \sum_z \mathbf{P}(Y \mid z) P(z)} \qquad (12)$$

- Eg: What is the probability for the student to have high intelligence given that the grade scored is *A*.

$= P(Y)$

$$P(Intelligence = high \mid Grade = A) = \frac{0.18}{0.25} = 0.72$$

$\leftarrow \dfrac{P(Grade = A,\ Intelligence = high)}{P(Grade = A)}$

$P(Intelligence)$

$\rightarrow \neq P(Intelligence)$

$\rightarrow \neq$ Marginal Distribution

Exercise

Given Joint Distribution $\mathbf{P}(Cavity,\ Toothache, Catch)$ of 3 binary random variables.

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Joint Distribution Entries

P(¬toothache, cavity, ¬catch)

1. Compute $P(toothache)$.
2. Compute $P(cavity)$.
3. Compute $P(cavity\ / toothache)$ given the evidence of toothache .

innovate achieve lead

Solution

$$P(toothache) = P(toothache, \underline{cavity, catch}) + P(toothache, \underline{\neg cavity, catch}) + P(toothache, \underline{cavity, \neg catch}) + P(toothache, \underline{\neg cavity, \neg catch})$$

$$= 0.108 + 0.016 + 0.012 + 0.064$$

$$= 0.20$$

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

$$P(cavity \mid toothache) = \frac{P(cavity \land toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.20} = \frac{6}{10}$$

Independence of Events

- An event α is independent of event β in P,

$$P(\alpha \perp \beta) \quad \text{if} \quad \begin{array}{ll} P(\alpha \mid \beta) = P(\alpha) & \text{or} \\ P(\beta \mid \alpha) = P(\beta) & \text{or} \\ \boxed{P(\alpha \wedge \beta) = P(\alpha)P(\beta)} & \end{array} \tag{13}$$

$$P(\alpha, \beta) = P(\alpha / \beta) P(\beta)$$

$$P(\alpha, \beta) = P(\alpha) P(\beta)$$

- $\alpha \perp \beta$ implies $\beta \perp \alpha$
- Eg: Tossing two coins, Rolling a die.

Independence of Random Variables

- Two random variables X and Y can be independent of each other.
- A random variable X is independent of another random variable Y,

$$P(X \perp Y) \quad \text{if} \quad \begin{cases} \boxed{P(X \mid Y) = P(X)} & \text{or} \\ \boxed{P(Y \mid X) = P(Y)} & \text{or} \\ \boxed{P(X \wedge Y) = P(X)P(Y)} & \end{cases} \qquad (14)$$

$$P(X \wedge Y) = P(X)\,P(Y/X) \rightarrow P(X)P(Y)$$

- $X \perp Y$ implies $Y \perp X$
- Eg: Weather is independent of dental problems.

Conditional Independence of Events

$A = 2X + 3Y$
$B = 5Z + 8Y$
specify Y

$P(A, B/Y) = P(A/Y)P(B/Y)$

- An event α is conditionally independent of event β given event γ in P

$$P(\alpha \perp \beta \mid \gamma) \quad \text{if} \quad \begin{cases} P(\alpha \mid \beta \wedge \gamma) = P(\alpha \mid \gamma) \\ P(\alpha \mid \beta \wedge \gamma) = P(\beta \mid \gamma) = 0 \\ P(\alpha \wedge \beta \mid \gamma) = P(\alpha \mid \gamma)P(\beta \mid \gamma) \end{cases} \tag{15}$$

- Eg: Getting Admission in MIT is independent of getting admission in Stanford, given the student has scored Grade A.

Conditional Independence of Random Variables

- A random variable *X* is conditionally independent of random variable *Y* given random variable *Z*

$$P(X \perp Y \mid Z) \quad \text{if} \quad \begin{cases} P(X \mid Y, Z) = P(X \mid Z) \\ P(X \mid Y, Z) = P(Y \mid Z) = 0 \\ P(X, Y \mid Z) = P(X \mid Z) P(Y \mid Z) \end{cases} \tag{16}$$

- Eg: The random variables Toothache and Catch are independent given the presence or absence of Cavity.

Student Example

- Model the difficulty of a course, intelligence of students, Grade the students score in a particular course.
- Let D represent the difficulty of a course.

$$\begin{aligned} \textit{Domain of } \mathbf{D} &= \{easy, hard\} \\ \mathbf{P}(\mathbf{D}) &= \{d^0, d^1\} \\ &= \{0.6, 0.4\} \end{aligned}$$

- Let I represent the intelligence of a student.

$$\begin{aligned} \textit{Domain of } \mathbf{I} &= \{low, high\} \\ \mathbf{P}(\mathbf{I}) &= \{i^0, i^1\} \\ &= \{0.7, 0.3\} \end{aligned}$$

Student Example

- Let G represent the grade a student gets for a course.

$$\textit{Domain of } \mathbf{G} = \{A, B, C\}$$
$$\mathbf{P}(\mathbf{G}) = \{g^1, g^2, g^3\}$$

- $\mathbf{P}(\mathbf{D}, \mathbf{I}, \mathbf{G})$ denotes the probabilities of all combinations of the values of the 3 random variables.
- These $2 * \overset{2}{\cancel{3}} * 3 = 12$ values can be represented using a Joint Distribution Table.

Joint Probability Distribution

$P(\text{high}, \text{high}, \text{easy})$ $P(\text{low}, \text{low}, \text{difficult})$

- Joint Probability Distribution completely represents the joint distribution for all random variables.
- In the students example, the **P(D, I, G)**, the 12 parameters cannot be determined by the value of the other parameters. Hence called Independent parameters.
- Independent parameters are parameters whose values are not completely determined by the values of the other parameters.

Student Example - Joint Distribution

I	D	G	P
i^0	d^0	g^1	0.126
		g^2	0.168
		g^3	0.126
i^0	d^1	g^1	0.009
		g^2	0.045
		g^3	0.126
i^1	d^0	g^1	0.052
		g^2	0.0224
		g^3	0.0056
i^1	d^1	g^1	0.069
		g^2	0.036
		g^3	0.024

$P(g^1, i^0, d^0)$

$P(g^2, i^0, d^0)$

$$\sum \mathbf{P(D, I, G)} = 1$$

$P(g^3, i^0, d^0)$

$$P(X=x, Y=y \mid Z=z) = P(X=x \mid Z=z)\, P(Y=y \mid Z=z)$$

Conditional Independence

Let X,Y , Z be sets of random variables. We say that X is conditionally independent of Y given conditional independence Z in a distribution P if P satisfies (X = x ⊥ Y = y | Z = z) for all values x ∈ Val(X), y ∈ Val(Y), and z ∈ Val(Z). The variables in the set Z are often said to be observed. If the set observed variable Z is empty, then instead of writing (X ⊥ Y | ∅), we write (X ⊥ Y) and say that X and Y are marginally independent

(Definition 2.4 from Daphne Koller's book)

Conditional Independence

Proposition 2.3 The distribution P satisfies $(X \perp Y \mid Z)$ if and only if $P(X,Y \mid Z) = P(X \mid Z)P(Y \mid Z)$. Suppose we learn about a conditional independence.

Can we conclude other independence properties that must hold in the distribution?

Symmetry: $(X \perp Y \mid Z) \Rightarrow (Y \perp X \mid Z)$.

innovate achieve lead

Properties that hold

(2.7) Decomposition: $(X \perp Y, W \mid Z) \Rightarrow (X \perp Y \mid Z)$.

(2.8) Weak union: $(X \perp Y, W \mid Z) \Rightarrow (X \perp Y \mid Z, W)$.

(2.9) Contraction: $(X \perp W \mid Z, Y) \;\&\; (X \perp Y \mid Z) \Rightarrow (X \perp Y, W \mid Z)$.

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

Why is decomposition true?

$$P(X, Y \mid Z) = \sum_w P(X, Y, w \mid Z) \quad \text{(summing over all } w\text{)}$$

$$= \sum_w \boxed{P(X \mid Z)} P(Y, w \mid Z) \quad \text{because } (X \perp Y, W \mid Z)$$

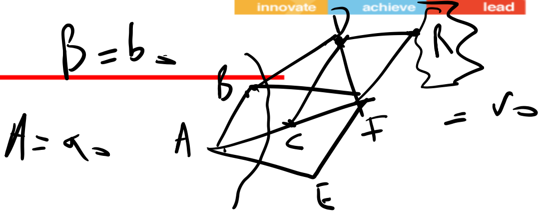
$$= P(X \mid Z)\left(\sum_w P(Y, w \mid Z)\right).$$

But $\sum_w P(Y, w/z) = \underline{P(Y/z)}$

marginal

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

Querying a Distribution



Compute $\underline{P(Y / E = e)}$

E = evidence random variables instantiated to e.

Y → query variables, a subset of random variables in the network

We want the posterior probability distribution over the values y given that E = e

MAP Query

Find a high probability joint assignment to some subset of variables

$$MAP(W \mid e) = \arg\max_{W} P(W, e)$$

How is a MAP query different from a probability query?

MAP Query

To find the most likely assignment to a single variable A, we could simply compute P(A | e) and then pick the most likely value.

The assignment where each variable individually picks its most likely value can be quite different from the most likely joint assignment to all variables simultaneously.

[Joint Distribution Matters]

Example

Example 2.4 from the textbook [Daphne Koller]

Let random variables A and B be both binary valued

a^0	a^1
0.4	0.6

A	b^0	b^1
a^0	0.1	0.9
a^1	0.5	0.5

these are conditional probabilities

$P(B = b^i / A = a^j)$

Example

Now $MAP(A) = a^1$ since $P(a^1) > P(a^0)$

What is $MAP(A, B)$?

It is (a^0, b^1) since $P(a^0, b^1) = P(a^0)\,P(b^1/a^0)$
$= 0.4 \times 0.9 = 0.36$ is greater than for any other combination.

good or bad?

$$\boxed{\arg\max_{a,b} P(a,b) \neq \left(\arg\max_{a} P(a), \arg\max_{b} P(b)\right)}$$

Continuous Spaces

$$\int p(x)\,dx = 1$$

$Val(X)$

$$P(X \le a) = \int_{-\infty}^{a} p(x)\,dx$$

$$P(a \le X \le b) = \int_{a}^{b} p(x)\,dx$$

Continuous space

$$P(a_1 \le X_1 \le b_1,\ a_2 \le X_2 \le b_2, \ldots a_n \le X_n \le b_n)$$

$$= \int_{a_1}^{b_1}\int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} p(x_1, x_2, \ldots x_n)\, dx_n\, dx_{n-1} \ldots dx_1$$

p is integrable joint density function.

$p(x_1, x_2 \ldots x_n) \ge 0$ for all values $x_1, x_2, \ldots x_n$

Conditional Density Functions

We cannot write $P(Y/X=x)$ as $\frac{P(X=x, Y)}{P(X=x)}$ since both these probabilities are equal to 0.

Define $P(Y/x)$ as $\lim_{\varepsilon \to 0} P(Y / x-\varepsilon \leq X \leq x+\varepsilon)$

$$P(a \leq Y \leq b \mid x-\varepsilon \leq X \leq x+\varepsilon) = \frac{P(a \leq Y \leq b, x-\varepsilon \leq X \leq x+\varepsilon)}{P(x-\varepsilon \leq X \leq x+\varepsilon)}$$

Conditional Density Functions

$$\frac{\int_a^b \int_{x-\varepsilon}^{x+\varepsilon} p(x', y)\, dx\, dy}{\int_{x-\varepsilon}^{x+\varepsilon} p(x')\, dx} = \frac{\int_a^b 2\varepsilon\, p(x, y)\, dy}{2\varepsilon\, p(x)}$$

Conditional Density Functions

$$P\left(a \le Y \le b \,/\, x-\varepsilon \le X \le x+\varepsilon\right)$$

$$\simeq \frac{\int_a^b 2\varepsilon\, p(x,y)\,dy}{2\varepsilon\, p(x)} = \int_a^b \frac{p(x,y)}{p(x)}\,dy$$

$$\boxed{\therefore \text{Density of } p(Y/X=x) = p(x,y)/p(x)}$$

Table of Contents

Graph

- Data structure used to represent the probability distribution of data.
- A graph is a data structure $\mathcal{K}$ consisting of a set of nodes and a set of edges.

$$\text{Graph} \qquad \mathcal{K} = (\mathcal{X}, \mathcal{E}) \tag{17}$$

- The set of nodes denote each random variable.

$$\text{Set of Nodes} \qquad \mathcal{X} = \{X_1 \ldots X_n\} \tag{18}$$

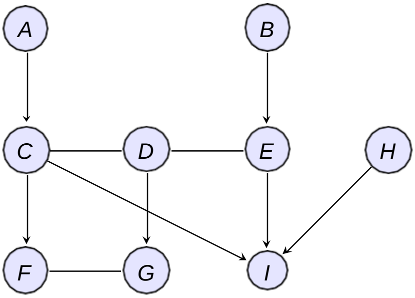
- A pair of nodes X_i, X_j can be connected by a directed edge $X_i \rightarrow X_j$ or an undirected edge $X_i - X_j$.

$$\text{Set of Edges} \qquad \mathcal{E} = X_i \rightarrow X_j \quad \text{or} \quad X_i - X_j \tag{19}$$

Directed Graph

- A graph is directed if all edges are directed. $X_i \rightarrow X_j$.

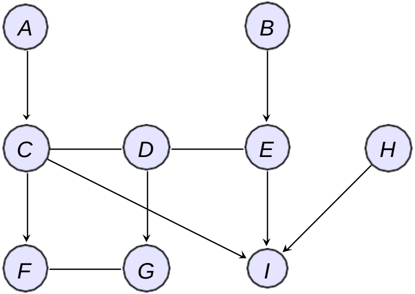
$$\mathrm{G} = (\mathrm{X}, \mathrm{E}) \quad \textit{where} \quad \mathrm{E} = \{X_i \rightarrow X_j\} \tag{20}$$



Undirected Graph

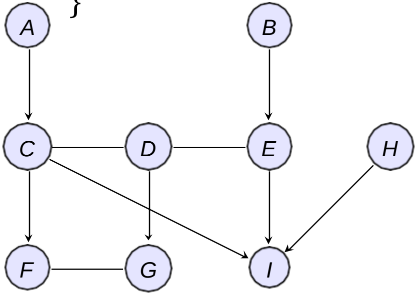
- A graph is undirected if all edges are undirected. $X_i - X_j$.

$$H = (X, E') \quad \textit{where} \quad E' = \{X_i - X_j\} \tag{21}$$



Parent and Child

Graph $\mathcal{K} = (\mathcal{X}, \mathcal{E})$ *where* $\mathcal{E} = \{X \rightleftharpoons Y\}$



- Parent
  - X is called the parent of Y.
  - Pa_X denote parents of X.
- Child
  - Y is called the child of X.
  - Ch_X denote children of X.
- Example: Identify the parents and children of Node E.

  Ans: $Pa_E = B$ $\quad Ch_E = I$

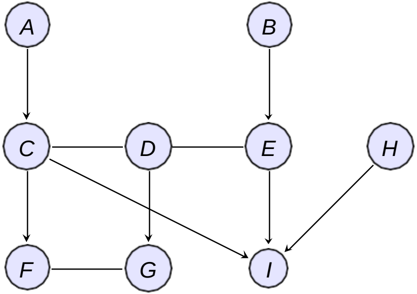
Neighbor and Boundary

- Neighbor
  - Whenever $X - Y \in \mathcal{E}$, X and Y are adjacent.
  - Nb_X denote neighbors of X.
- Boundary
  - In Directed graph, $Boundary_X = Pa_X$.
  - In Undirected graph, $Boundary_X = Nb_X$.

$$Boundary_X = Pa_X \cup Nb_X$$

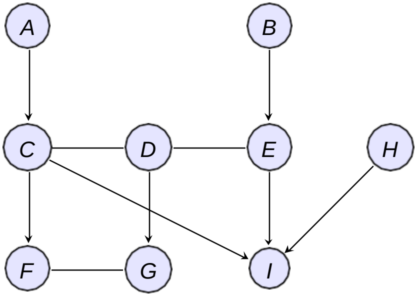
- Example: Identify the neighbors and boundary of Node C.

  Ans: $Nb_C = \cancel{F}, D, \cancel{I}$ $\quad B_C = A, D, \cancel{F, I}$



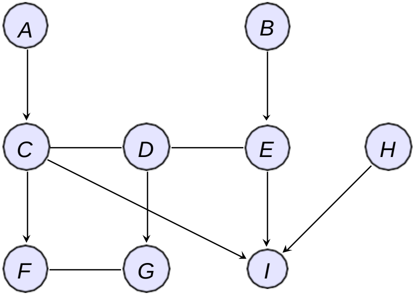
Degree of a Graph

- The degree of a node X is the number of edges in which it participates.
- The in-degree of a node is the number of directed edges $Y \rightarrow X$.
- The degree of a graph is the maximal degree of a node in the graph.
- Example: Identify the degree Node I. Ans=3



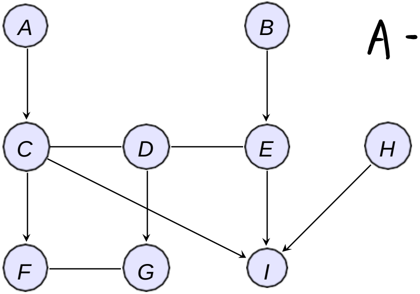
Path

- $X_1, \ldots, X_k$ form a path in the graph $\mathcal{K} = (\mathcal{X}, \mathcal{E})$, if for every $i = 1, \ldots, k-1$ we have either $X_i \rightarrow X_{i+1}$ or $X_i - X_{i+1}$.
- Example: Identify a path. Ans: $A \rightarrow C \rightarrow I$



Trail

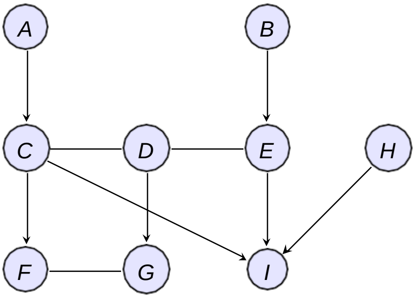
- $X_1, \ldots, X_k$ form a trail in the graph $\mathcal{K} = (\mathcal{X}, \mathcal{E})$, if for every $i = 1, \ldots, k-1$ we have either ~~X_i � X_{i+1}~~ $X_i \rightleftarrows X_{i+1}$ (any sort of edge)
- A graph is connected, if there is a trail between X_i and X_j.
- Example: Identify a trail. Ans: $A \to C - D - E \to I$ (both path & trail)



A - C - F - G - D
(trail, but not a path)

Ancestor and Descendant

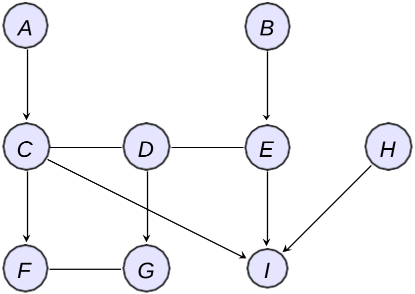
- X is an ancestor of Y in a graph $\mathcal{K}$ if there is a directed path $X_1, \ldots, X_k$ with $X_1 = X$ and $X_k = Y$.
- Y is the descendant of X.
- $Ancestor_X$ and $Descendant_X$



Cycle

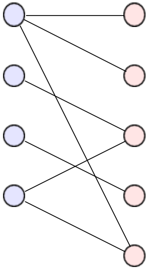
- A cycle in graph $\mathcal{K}$ is a directed path $X_1, \ldots, X_k$ with $X_1 = X_k$.
- Example: Identify a cycle.

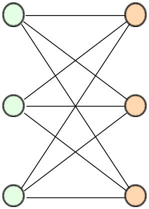
  Ans: No cycle



Bipartite Graphs

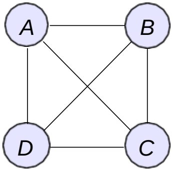
- If the vertex-set of a graph G can be split into two disjoint sets, V_1 and V_2, in such a way that each edge in the graph joins a vertex in V_1 to a vertex in V_2, and there are no edges in G that connect two vertices in V_1 or two vertices in V_2, then the graph G is called a bipartite graph.
- Document-Terms, Student-Class, Movie preference of viewers

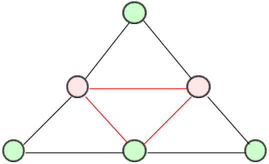




Clique

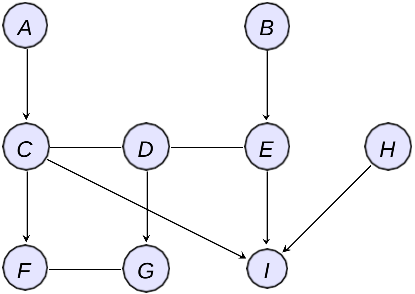
- Clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent.
- A ~~maximum~~ maximal clique of a graph, is a clique, such that there is no clique with more vertices.





Directed Acyclic Graph (DAG)

- A graph is acyclic if it contains no cycles.
- A directed acyclic graph is a graph that has directed edges but no cycles.
- DAG is the basic graphical representation of Bayesian Networks.



References

1. Probabilistic Graphical Models: Principles and Techniques by Daphne Koller and Nir Friedman. MIT Press. 2009
2. Artificial Intelligence: A Modern Approach (3rd Edition) by Stuart Russell, Peter Norvig
3. Mastering Probabilistic Graphical Models using Python by Ankur Ankan, Abhinash Panda. Packt Publishing 2015.
4. Learning in Graphical Models by Michael I. Jordan. MIT Press. 1999

Thank You !!!