Inference with Uncertain Knowledge

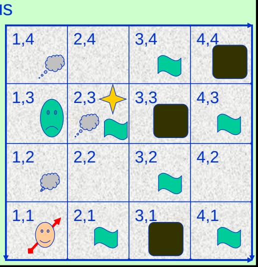
- Objectives
  - Why must agent use uncertain knowledge?
  - Fundamentals of Bayesian probability
  - Inference with full joint distributions
  - Inference with Bayes' rule
  - Bayesian networks (BNs)
  - Exact inference in BNs
- Reference
  - Russell & Norvig: Chapter 13 & 14

Limitation of Logic Inference

- When a logic agent knows enough about its env, it can derive plans that guarantee to work.
  - What if it does not have all necessary facts?
- Ex Wumpus World has at most 2 pits. Starting from [1,1], agent senses breeze at [1,2] and [2,1].
  - What should agent do?
- A logic agent may not be able to act under uncertainty.

Review of Wumpus World

- Gold is located in a room.
  - Gold glitters.
- Wumpus eats anyone in its room.
  - Stench in room next to wumpus
- Pits trap explorers.
  - Breeze in room next to a pit
- Goal of agent
  - Find gold without being eaten or falling into a pit.



Uncertainty and Decision Making

- The presence of uncertainty in env radically changes how agent should make decisions.
  - A logic agent can select or reject an action on whether it achieves the goal, regardless of what other actions might achieve.
- Ex Flight will departure 9am at Pearson airport.
  - Action a_1: Drive from Guelph at 6:30am.
  - Action a_2: Drive from Guelph at 5:30am.
  - Action a_3: Drive from Guelph at 1:30am.
  - Which action should be selected?

Preference and Utility

- Making decision under uncertainty requires agent to have preferences over env states.
  - The degree of desirability of a state is specified by a numerical utility.
  - The degrees of desirability over all relevant states is specified by a utility function.
  - How to specify preference using utility functions is investigated in utility theory.
- Ex Catching 9am flight

Decision Theory

- Making decision under uncertainty also requires agent to estimate likelihood of env states resultant from alternative actions.
  - Likelihood of the unobservable given observations is investigated by probability theory, among others.
  - Ex Catching 9am flight
- Decision theory = probability theory + utility theory
- Let utility of state s be U(s) and probability of s from action a be P(s|a), then expected utility of action a is $EU(a) = \Sigma_s U(s)P(s|a)$.
  - What is $EU(a_1)$, where a_1 is leaving Guelph at 6:30am?

Maximum Expected Utility Principle

- Maximum expected utility (MEU) principle
  - Rational agent should select action a* with max EU(a).
  - Ex Catching 9am flight: a_1, a_2 or a_3?
- Can a probabilistic WW agent do better?
  - Which room should be entered, [1,3], [2,2] or [3,1]?
- Focus
  - Likelihood estimation by probability theory
- Approach
  - Extend propositional logic to Bayesian probability

Variables and Their Domains

- Describe env state by a set $V = \{x_1,\ldots,x_n\}$ of discrete variables.
  - Ex Health state of a cough patient
- Associate each variable x_i in V with a finite domain $D_i = \{x_{i_1},\ldots,x_{i_k}\}$ of possible values.
  - Values in each domain must be mutually exclusive and exhaustive.
  - Ex An athlete's highest achievement in an Olympic game

Propositions

- A proposition is an assertion about an env state.
- A simple proposition asserts an assignment over $x \in V$.
- A complex proposition is formed by combining simple propositions with logical connectives.
  - A complex proposition asserts an assignment over $X \subseteq V$.
- Intuitively, a proposition asserts an event in env.

Atomic Event

- An atomic event is a conjunction of simple propositions one over each $x \in V$.
  - It corresponds to a complete assignment.
- Properties of atomic events
  - They are mutually exclusive.
  - They are collectively exhaustive.
  - An atomic event entails the truth value of every proposition.
  - Any proposition q is logically equivalent to disjunction of all atomic events that entail the truth of q.
- An atomic event is also referred to as a model.

Express Uncertain Knowledge Probabilistically

- Represent an env state s by a proposition q.
- Express an agent's uncertain knowledge about s by assigning q a probability p.
  - p is degree of belief of the agent over q or s.
- Agent's belief state is the collection of probabilities it assigns to all env states.
- Frequentist versus Bayesian probability

Prior Probability

- Prior probability P(q) for proposition q is an agent's degree of belief over q at initial belief state.
- Prior probability distribution P(x) for $x \in V$ is a set of probabilities one for each value of x.
- Joint probability distribution P(X) for $X \subseteq V$ is a set of probabilities one for each assignment over X.
- Full joint probability distribution P(V) is a set of probabilities one for each atom event.

Conditional Probability

- Conditional or posterior probability P(q|e), where q and e are propositions, is agent's degree of belief over q, after knowing e beyond initial belief state.
- Conditional probability is related to prior probability according to product rule:
  - $P(q|e) = P(q \wedge e) / P(e)$, whenever $P(e)>0$
  - Alternatively: $P(q \wedge e) = P(q|e) P(e)$
- Conditional probability distribution P(X|y) for assignment y over Y is the set of all conditional probabilities P(x|y), one for each assignment x of X.
- P(q) is a special case of conditional probability.

Conditional Probability Table

- Conditional probability table (CPT) P(X|Y) is a set of conditional probability distributions (CPDs), one for each assignment y over Y.
- Ex CPD P(lung_cancer|pos_x_ray)

  = { P(early_lung_cancer|pos_x_ray),
  P(late_lung_cancer|pos_x_ray),
  P(absent_lung_cancer|pos_x_ray) }
- Ex CPT P(lung_cancer|x_ray)

  = { P(lung_cancer|pos_x_ray),
  P(lung_cancer|neg_x_ray) }

Axioms of Probability

- Since degree of belief is subjective, can agent assign numerical values to propositions arbitrarily?
- Probability assignment must satisfy a set of axioms, where q, r and e are propositions:

1. [range] For any proposition q and r, $0 \leq P(q|r) \leq 1$.
2. [certainty] $P(q|q) = 1$.
3. [sum] If q and r are mutually exclusive,
   $P(q \vee r|e) = P(q|e) + P(r|e)$.
4. [product] $P(q \wedge r|e) = P(q|r \wedge e) P(r|e)$.

Properties Derived from Axioms

- Rest of probability theory is derivable from axioms.
- In the following, h, e and q are propositions.

- [Bayes rule] $P(h|e \wedge q) = P(e|h \wedge q)P(h|q)/P(e|q)$
- [Negation] $P(\neg q|e) = 1 - P(q|e)$
- [Marginalization] Given P(X,Z), $P(X) = \sum_{\underline{z}} P(X, \underline{z})$, where $\underline{z}$ is an assignment over Z
  - Marginal distribution

Properties Derived from Axioms (Cont)

- [Sum to 1] For any $Z \subseteq V$, the sum $\sum_{\underline{z}} P(\underline{z}) = 1$, where $\underline{z}$ is an assignment over Z.
- Marginalization and Sum to 1 hold for conditional probability as well.
- [Sum from atomic events]

  For any proposition q, it holds $P(q) = \sum_{\underline{z}} P(\underline{z})$, where $\underline{z}$ is an atomic event that entails q.

Why Should Agent Belief be Probability?

- Degree of belief of agent Ag in proposition e being p implies the following:
  - When offered alternative lotteries $L_1 = \{\$1|e, \$0|\neg e\}$ and $L_2 = \{\$p|e \vee \neg e\}$, Ag is indifferent btw them.
  - Ag is indifferent among lotteries $L_3 = \{\$(p-1)|e, \$p|\neg e\}$, $L_4 = \{\$(1-p)|e, -\$p|\neg e\}$, and $L_5 = \{\$0|e \vee \neg e\}$.
- Ag's degree of belief either follows axioms of probability (hence it is probability), or it does not.
- There exists a combination lottery that guarantees Ag's loss iff its degree of belief is not probability.

When Agent Belief is Not Probability

Proposition	Belief	Lottery	$a \wedge b$	$a \wedge \neg b$	$\neg a \wedge b$	$\neg a \wedge \neg b$
a	0.1	{-\$0.9|a, \$0.1|¬a}	-\$0.9	-\$0.9	\$0.1	\$0.1
b	0.6	{-\$0.4|b, \$0.6|¬b}	-\$0.4	\$0.6	-\$0.4	\$0.6
a ∨ b	0.8	{\$0.2|a∨b, -\$0.8|¬(a∨b)}	\$0.2	\$0.2	\$0.2	-\$0.8
		Combo Lottery:	-\$1.1	-\$0.1	-\$0.1	-\$0.1

Reflection

- In stochastic and partially observable envs, agent frequently faces decision situations like lotteries.
- If agent does not base its belief on probability, it will encounter situations (by nature or by design) where its performance is guaranteed sub-optimal.
- The only way that agent can always avoid such undesirable position is to adopt probabilistic belief.

Probabilistic Inference

- To act effectively, agent needs to determine what is the state of env given perception (observation).
- In stochastic and partially observable env, this task becomes computation of posterior distribution P(X|e) from prior distribution P(X) and observation e, where $X \subseteq V$.
  - This task is referred to as probabilistic inference.
- Ex Given P(hygiene, cavity, toothache) and observation toothache=yes, what is the posterior P(hygiene | toothache=yes)?

Ex Dental Hygiene

- V = {hygiene, cavity, toothache}, where hygiene $\in$ {good, bad}, cavity $\in$ {yes, no}, and toothache $\in$ {yes, no}

h	c	t	P(h,c,t)
good	yes	yes	0.0595
good	yes	no	0.0105
good	no	yes	0.0315
good	no	no	0.5985
bad	yes	yes	0.2040
bad	yes	no	0.0360
bad	no	yes	0.0030
bad	no	no	0.0570

Inference Using Full Joint Distribution

- Query: What is P(hygiene | toothache=yes)?
- Algorithm
  1. Constrain P(h,c,t) to P(h,c,t=y)
  2. Compute P(t=y)
  3. Condition P(h,c,t=y) into P(h,c,t|t=y)
  4. Marginalize out c and t to get P(h|t=y)

- Normalization: Steps 2 and 3 above
  - Normalization constant: α = 1/P(t=y)
- Hence, P(h|t=y) = $\alpha \Sigma_{c,t}$ P(h,c,t=y)

Inference by Bayes Rule

- [Bayes rule] P(h|e) = P(e|h)P(h)/P(e)
  - Compute probability of hypothesis h given observation e.

- Ex Meningitis patients often have stiff neck.
  - P(s|m): knowledge on causal mechanism
  - P(m): disease incidence rate in the population
  - P(s): symptom statistics from the population

Complexity of Inference Using Full Joint

- Suppose $|V|=n$, $|D_x|\leq k$ for $x \in V$, and $V'=V\setminus\{x\}$.
- How many independent probability parameters are needed to specify P(V)?
- How may parameters in P(V) make up $P(V'|x=x_0)$?
  - These parameters must be processed to compute $P(y|x=x_0)$ for $y \in V'$.
- Inference using full joint distributions is intractable!
  - Idea for efficiency improvement:
    Explore independence using graphical models

Explore Independence

- Inference by full joint assumes that every variable is directly dependent on every other.
  - This is often not the case in an application env.
- Ex Knowing child's dental hygiene affects belief on suffering of toothache.
  - But P(t|c,h) = P(t|c)
  - This allows full joint over V={h,c,t} to be specified with a less number of (independent) parameters.
- Variables x and y are conditionally independent given variable z, iff P(x'|y',z') = P(x'|z') holds for every value x', y', z' of x, y, z, respectively.

Conditional Independence

- When x and y are conditionally independent given z, we write P(x|y,z) = P(x|z) and I(x,z,y).
- Subsets of variables X and Y are conditionally independent given Z, where X, Y and Z are disjoint, iff $P(\underline{x}|\underline{y},\underline{z}) = P(\underline{x}|\underline{z})$ holds for every assignment $\underline{x}$, $\underline{y}$, $\underline{z}$ of X, Y, Z, respectively.
  - We write P(X|Y,Z) = P(X|Z) and I(X,Z,Y).
- Ex $V = \{x_0,\ldots,x_9\}$ and $|D_i|=2$, where x_{k+1} (k=1,…,8) is conditionally independent of $x_0, \ldots, x_{k-1}$ given x_k.
  - How many parameters are needed to specify P(V)?
- Symmetry of I(X,Z,Y) & unconditional independence

Encode Conditional Independence Concisely

- When |V| is large, how can agent encode knowledge on conditional independence concisely?
- Many dependence and independence relations can be encoded in a graph.
  1. Direct dependence
  2. Indirect dependence
  3. Causal dependence
  4. Unconditional independence
  5. Conditional independence

Bayesian Network (BN)

- A BN is an acyclic, directed graph G, where each node is associated with a CPT.
  - ❑V is a set of discrete env variables.
  - ❑Each node in G is labeled by a variable $v \in V$.
  - ❑Each node v with parent set $\pi(v)$ is associated with conditional probability table $P(v|\pi(v))$.
- Ex Burglar-quake
  - ❑A home burglar alarm is reliable, but responds to quakes on occasion. Neighbor John calls if alarm is heard, but may be confused with phone ringing. Mary also calls, but may miss the alarm due to loud music.
  - ❑Number of independent probability parameters

Semantics of BNs

1. A variable v in BN is conditionally independent of its non-descendants given its parents $\pi(v)$.

- [Chain rule] Given a BN over V, the full joint distribution is $P(V) = \prod_{v \in V} P(v|\pi(v))$.
  - ❑Ex Reduced BN for burglar alarm.
- Markov blanket of a variable in BN includes its parents, children, and children's parents.

2. A variable in BN is conditionally independent of all other variables given its Markov blanket.

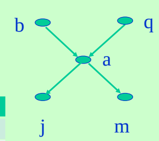
Inference in BNs

- BNs avoid specification of intractable number of probability parameters through chain rule.
- Ex Burglar-quake
  - ❑Compute P(b | j=t, m=t) through chain rule.
- Key operations
  - ❑Product: multiply($B_1(X_1)$, …, $B_m(X_m)$)
    - Multiply a set of potentials.
  - ❑Marginalization: marginOut(B(X), Z)
    - Marginalize out variables in Z from potential B(X).
- A potential is a non-negative real function with at least one positive value.

Ex Burglary-Quake

- Compute P(b | j=t, m=t)

b	P(b)
t	0.001
f	0.999



q	P(q)
t	0.002
f	0.998

b	q	a	P(a|b,q)
t	t	t	0.95
t	t	f	0.05
t	f	t	0.94
t	f	f	0.06
f	t	t	0.29
f	t	f	0.71
f	f	t	0.001
f	f	f	0.999

a	j	P(j|a)
t	t	0.90
t	f	0.10
f	t	0.05
f	f	0.95

a	m	P(m|a)
t	t	0.70
t	f	0.30
f	t	0.01
f	f	0.99

Multiply Potentials

multiply($B_1(X_1)$, …, $B_m(X_m)$) {
  $Z = \cup_{i=1,\dots,m} X_i$;
  init F(Z) s.t. F($\underline{z}$) = 1 for each assignment $\underline{z}$ of Z;
  for each assignment $\underline{z}$ of Z,
    for i=1 to m,
      $\underline{x}_i$ = assignment of X_i consistent with $\underline{z}$;
      F($\underline{z}$) *= $B_i(\underline{x}_i)$;
  return F(Z);
}

Marginalize Out Variables

marginOut(B(X), Z) {
  Y = X\Z;
  init F(Y) s.t. F($\underline{y}$) = 0 for each assignment $\underline{y}$ of Y;
  for each assignment $\underline{y}$ of Y,
    for each assignment $\underline{z}$ of Z,
      F($\underline{y}$) += B($\underline{y} \bowtie \underline{z}$);
  return F(Y);
}

BN Inference By Variable Elimination

- When BN inference is based on full joint, it still processes an intractable number of parameters.
- Ideas to improve inference efficiency
  - Avoid processing full joint directly.
  - Reduce size of intermediate result (potential or factor).
    - Marginalize as soon as possible.
    - Multiply as late as possible.
- Algorithm varElim(S, x, $\underline{y}$)
  - Given a BN S, a query variable x, and observation $\underline{y}$ over Y, compute posterior P(x|$\underline{y}$).

varElim(S, x, $\underline{y}$) { // S: over V; $\underline{y}$ over $Y \subset V$; $x \in V \setminus Y$;
  T = set of all CPTs in S;
  for each $v \in Y$ with v=u in $\underline{y}$,
    pick potential B over v in T; constrain B by v=u;
  for i=1 to |V|-1,
    select variable $z \neq x$ covered by a potential in T;
    Q = set of potentials over z in T;
    B_1 = multiply(potentials in Q);
    B_2 = marginOut(B_1, z);
    T = $(T \setminus Q) \cup \{B_2\}$;
  F = multiply(potentials in T);
  normalize F and return result;
}

Select Variable to Eliminate

- Heuristic: Select the variable whose elimination produces the smallest potential.
- Algorithm

  for each variable v to be eliminated,
    $U_v = \{\}$;
    for each potential B(X) s.t. $v \in X$,
      $U_v = U_v \cup X$;
  return variable y s.t. $|U_y|$ is minimal;

Complexity of VE

- What is the complexity if BN has a chain topology?
- If each variable in BN is directly connected to each other variable, what is the complexity?
- The sparser the BN topology, the more efficient the inference computation.
- VE can be extended to query about a set $X \subset V$ of variables directly.

Remarks

- Logic reasoning was once considered as sufficient for AI, but that view was soon found to be flawed.
- BNs and related graphical models are the dominant paradigm for uncertain reasoning in AI today.
- More advanced topics
  - Acquisition of BNs by learning or elicitation
  - Alternative BN inference paradigms
  - BNs with mixed variables or under dynamic envs
  - Multi-agent BNs
  - Decision making with graphical models
  - Uncertain reasoning in practical applications