Artificial Intelligence

Lecture 7, Chapter 13, 14
Quantifying Uncertainty
Probabilistic Reasoning

Supta Richard Philip

Department of CS
AIUB

AIUB, September 2025

Table of Contents

Table of Contents

Acting Under Uncertainty

- Agents may need to handle **uncertainty**, whether due to **partial observability, nondeterminism**, or a combination of the two.
- Agent knowledge can best provide only a **degree of belief**.
- Our main tool for dealing with **degrees of belief** is **probability theory**.
- An agent must first have **preferences** between the different possible **outcomes** of the various plans.
- We use utility theory to represent and reason with preferences.
- Preferences, as expressed by utilities, are combined with probabilities in the general theory of rational decisions called decision theory.
- Decision theory = probability theory + utility theory.

Table of Contents

Basic Probability Notation

- Random variable is some aspect of the world about which we (may) have uncertainty
- R = Is it raining?
- T = Is it hot?
- Capital letters: Random variables
- Lowercase letters: values that the Random variables can take
- $r \in +r, -r$
- $t \in +t, -t$

Single probability distribution

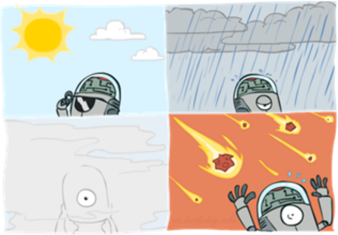
- Single probability distribution with binary Random variable Temperature P(T)

$$P(T)$$

T	P
hot	0.5
cold	0.5

Single probability distribution

- Single probability distribution with Random variable Weather P(W)



$$P(W)$$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Joint probability distribution

- Joint probability distribution with Random variable temperature and Weather P(T,W)
- A joint distribution over a set of random variables: specifies a real number for each assignment (or outcome):
  $P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$
  $P(x_1, x_2, ..., x_n)$
- $P(x_1, x_2, ..., x_n) >= 0$ (non-negativity)
- $\sum_{x_1, x_2, ..., x_n} P(x_1, x_2, ..., x_n) = 1$ (normalization).
- Size of distribution if n variables with domain sizes d?

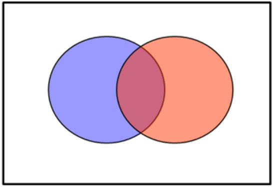
$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Conditional probability distribution

- Joint probability P(T,W) to conditional distribution.
- $P(X|Y) = P(X,Y)/P(Y) = P(X,Y)/\sum_X P(X,Y)$

$$P(a|b) = \frac{P(a,b)}{P(b)}$$



$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4$$

$$= P(W = s, T = c) + P(W = r, T = c)$$
$$= 0.5$$

Joint to singular probability distribution(Marginal)

- Joint probability P(X,Y) to singular probability distribution.

$P(X,Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$$P(x) = \sum_y P(x,y)$$

$$P(y) = \sum_x P(x,y)$$

$P(X)$

X	P
+x	.5
-x	.5

$P(Y)$

Y	P
+y	.6
-y	.4

Joint to conditional probability distribution(Marginal)

- Joint probability P(X,Y) to conditional probability distribution.

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W=s|T=c) = \frac{P(W=s,T=c)}{P(T=c)} = \frac{P(W=s,T=c)}{P(W=s,T=c)+P(W=r,T=c)} = \frac{0.2}{0.2+0.3} = 0.4$$

$$P(W=r|T=c) = \frac{P(W=r,T=c)}{P(T=c)} = \frac{P(W=r,T=c)}{P(W=s,T=c)+P(W=r,T=c)} = \frac{0.3}{0.2+0.3} = 0.6$$

$P(W|T=c)$

W	P
sun	0.4
rain	0.6

Joint to conditional probability distribution(Marginal)

- Joint probability P(X,Y) to conditional probability distribution.

Joint Distribution

$$P(T,W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Conditional Distributions

$P(W|T)$

$$P(W|T=hot)$$

W	P
sun	0.8
rain	0.2

$$P(W|T=cold)$$

W	P
sun	0.4
rain	0.6

Normalization Trick

- Normalization Trick.

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence →

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one) →

$P(W|T = c)$

W	P
sun	0.4
rain	0.6

- Step 1: Compute Z = sum over all entries
- Step 2: Divide every entry by Z

W	P
sun	0.2
rain	0.3

Normalize →
Z = 0.5

W	P
sun	0.4
rain	0.6

Table of Contents

Probabilistic Inference: Inference by Enumeration

- Compute a desired probability from other known probabilities (e.g. conditional from joint).

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

- P(W)?
- P(sun)=.3+.1+.1+.15=.65
- P(rain)=1-.65=.35

Inference by Enumeration

- P(W |winter, hot)?
- P(sun | winter,hot)=.1
- P(rain|winter,hot)=.05
- P(sun|winter,hot)=2/3
- P(rain|winter,hot)=1/3

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- P(W | winter)?
- P(sun|winter)∝.1+.15=.25
- P(sun,winter)=.25

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- $P(W \mid \text{winter})$?
- $P(\text{rain}|\text{winter})$:.05+.2=.25
- $P(\text{rain,winter})$=.25

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- $P(sun|winter) \approx .25$
- $P(rain|winter) \approx .25$
- $P(sun|winter) = .5$
- $P(rain|winter) = .5$

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

The Product Rule

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

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

$P(D,W)$

D	W	P
wet	sun	0.08
dry	sun	0.72
wet	rain	0.14
dry	rain	0.06

The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

Table of Contents

Bayes' Rule

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x)$$

$$P(x|y) = \frac{P(y|x)}{P(y)}P(x)$$

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

- The doctor knows that the disease meningitis causes the patient to have a stiff neck, say 70% most of the time. The doctor also knows some unconditional facts: the prior probability that a patient has meningitis is 1/50,000, and the prior probability that any patient has a stiff neck is 1%. Let s be the proposition that the patient has a stiff neck and m be the proposition that the patient has meningitis.

$$
\begin{aligned}
P(s \mid m) &= 0.7 \\
P(m) &= 1/50000 \\
P(s) &= 0.01 \\
P(m \mid s) &= \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \times 1/50000}{0.01} = 0.0014\,.
\end{aligned}
$$

Table of Contents

Independence

If two variables **X** and **Y** are independent($\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$), by definition the following are true:

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

This says that their joint distribution factors into a product of two simpler distributions

Conditional Independence

If two variables **X** and **Y** are are conditionally independent given **Z** ($\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$), by definition the following are true:

- $P(X, Y|Z) = P(X|Z)P(Y|Z)$
- $P(X|Y, Z) = P(X|Z)$
- $P(Y|X, Z) = P(Y|Z)$

$$P(x|z, y) = \frac{P(x, z, y)}{P(z, y)}$$

$$= \frac{P(x, y|z)P(z)}{P(y|z)P(z)}$$

$$= \frac{P(x|z)P(y|z)P(z)}{P(y|z)P(z)}$$

Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

$P(X_2)$

H	0.5
T	0.5

$\cdots$

$P(X_n)$

H	0.5
T	0.5

$P(h, h, t, h) =$ P(h)P(h)P(t)P(h)

Bayesian Networks

Two problems with using full joint distribution tables as our probabilistic models:

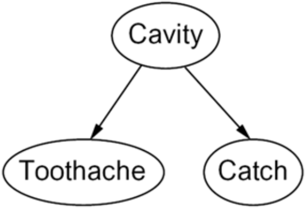
- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayesian Networks:

- a technique for describing complex joint distributions (models) using simple, local distributions (conditional probability tables, or CPTs)
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions

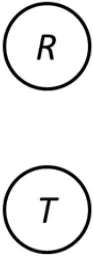
Graphical Model Notation

- Nodes: variables (with domains)
  Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  MAY indicate influence between variables
  Formally: encode conditional independence relationships (more later)
- For now: arrows mean that there may be a causal relationship between the two variables



Example: Traffic

- Variables:
  R: It rains
  T: There is traffic



Model 1: independence

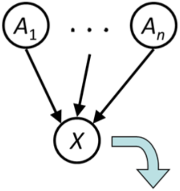
Model 2: rain may cause traffic

Example: Alarm Network

- Variables:
  B: Burglary
  A: Alarm goes off
  M: Mary calls
  J: John calls
  E: Earthquake!

Bayes Net Semantics

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
- A collection of distributions over X, one for each combination of parents' values
- CPT: conditional probability table
- **A Bayes net = Topology (graph) + Local Conditional Probabilities**

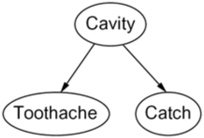


$P(X|A_1 \ldots A_n)$

Bayes Net Semantics

- Bayes nets implicitly encode joint distributions
- As a product of local conditional distributions
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together.
- $P(+cavity, +catch, -toothache) = P(-toothache| + cavity)P(+catch| + cavity)P(+cavity)$

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$



Bayes Net Example

$P(R)$

+r	1/4
-r	3/4

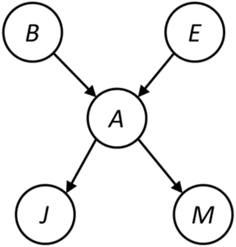
$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(+r, -t) =$ P(+r)P(-t|+r) = ¼* ¼

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

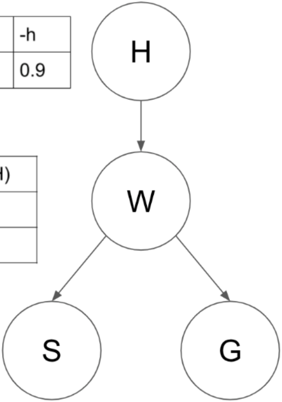
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b,-e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Example

	+h	-h
P(H)	0.1	0.9



	P(+w | H)	P(-w | H)
+h	0.667	0.333
-h	0.25	0.75

	P(+s | W)	P(-s | W)
+w	0.4	0.6
-w	0.2	0.8

	P(+g | W)	P(-g | W)
+w	0.25	0.75
-w	0.5	0.5

Query

$$P(+w \mid +s, -g)$$

$$= \frac{P(+w, +s, -g)}{P(+s, -g)} \; (chain\ rule)$$

$$= \frac{\sum_h P(h, +w, +s, -g)}{\sum_{h,w} P(h, w, +s, -g)} \; (marginalization)$$

$$= \frac{\sum_h P(h)P(+w \mid h)P(+s \mid +w)P(-g \mid +w)}{\sum_{h,w} P(h)P(w \mid h)P(+s \mid w)P(-g \mid w)} \; (joint = product\ of\ CPTs)$$

$$= \frac{0.1 * 0.667 * 0.4 * 0.75 + 0.9 * 0.25 * 0.4 * 0.75}{0.1 * 0.667 * 0.4 * 0.75 + 0.9 * 0.25 * 0.4 * 0.75 + 0.1 * 0.333 * 0.2 * 0.5 + 0.9 * 0.75 * 0.2 * 0.5}$$

$$= 0.553$$

Table of Contents

References

- Stuart Russell and Peter Norvig. 2009. Artificial Intelligence: A Modern Approach (3rd ed.). Prentice Hall Press, Upper Saddle River, NJ, USA.
- https://www.cs.cmu.edu/~15281/coursenotes/probability
- https://www.cs.cmu.edu/~15281/coursenotes/bayesnets
- https://www.cs.cmu.edu/~15281/coursenotes/hmm_pf