Probability

Forest Agostinelli
University of South Carolina

Topics Covered in This Class

- **Part 1: Search**
  - Pathfinding
    - Uninformed search
    - Informed search
  - Adversarial search
  - Optimization
    - Local search
    - Constraint satisfaction
- **Part 2: Knowledge Representation and Reasoning**
  - Propositional logic
  - First-order logic
  - Prolog
- **Part 3: Knowledge Representation and Reasoning Under Uncertainty**
  - Probability
  - Bayesian networks
- **Part 4: Machine Learning**
  - Supervised learning
    - Inductive logic programming
    - Linear models
    - Deep neural networks
    - PyTorch
  - Reinforcement learning
    - Markov decision processes
    - Dynamic programming
    - Model-free RL
  - Unsupervised learning
    - Clustering
    - Autoencoders

Outline

- Motivation
- Probability
- Joint distribution
- Conditional distribution
- Inference
- Independence
- Baye’s rule

Uncertainty

- We cannot predict everything with absolute certainty due to immense computational complexity, simply not knowing enough, or inherit randomness
- “If a fruit is a tomato and it is red, then it is ripe”
  - Sometimes, a tomato can be red but might not be ready
  - Other indicators: tomato texture, weather, etc.
  - We may not know everything that makes a tomato ripe
- “ If I drive to the airport 90 minutes before my flight, I will get there in time”
  - Must be perfect road conditions, no accidents, no earthquakes, etc.
  - It would be impractical to model every possibility
- Apparent/Inherent randomness
  - Chaotic systems
  - Quantum physics

How to Handle Uncertainty?

- Uncertainty comes from limits on computation, lack of knowledge, and inherent randomness
- We cannot say something is true or false with absolute certainty, however, we can provide a degree of belief
- Probability theory gives us the tools to deal with degrees of belief

Toothache Example

- A toothache may be a sign of a cavity
- However, a toothache may also be a sign of a gum problem, abscess, etc.
- Before running a battery of tests on a patient, we can assess the probability that they have a cavity given that they have a toothache and other relevant factors (age, diet, medical history, etc.)
- If we say that they have a cavity with probability 0.8, this means that, out of all the situations that are indistinguishable from the current situation, as far as our knowledge goes, the patient will have a cavity in 80% of them

Epistemological Commitments

Language	Ontological Commitment (What exists in the world)	Epistemological Commitment (What an agent believes about facts)
Propositional logic	facts	true/false/unknown
First-order logic	facts, objects, relations	true/false/unknown
Temporal logic	facts, objects, relations, times	true/false/unknown
Probability theory	facts	degree of belief $\in [0, 1]$
Fuzzy logic	facts with degree of truth $\in [0, 1]$	known interval value

- Propositional and first-order logic
  - This can either be true, false, or unknown
- Probability
  - There are degrees of belief

Random Variables

- A variable can take on a value from its domain
  - Variable: Temperature
  - Domains could be
    - {Hot, Cold}
    - {0deg, 0.1deg, … 100deg}
- Types of random variables
  - Boolean: {True, False}
  - Discrete
  - Continuous
    - We will not focus on these in this class
- Domain values must be exhaustive and mutually exclusive
  - Exhaustive: The domain must have every possible value for the variable
  - Mutually exclusive: Two or more values cannot be assigned at the same time (i.e. cannot be both hot and cold)

Probability Model

- Associates a numerical probability $P(w)$ for every possible world w
- Axioms:
  - $0 \leq P(w) \leq 1$
  - $\sum_{w \in \Omega} P(w) = 1$ (Ω is the set of all possible worlds)
- Probability of a specific event ϕ
  - $\sum_{w \in \phi} P(w)$
  - For example, if we have two six sided die, there are 36 possible worlds
  - If we want to know the probability that the die sum to 11, we sum the probability of rolling a 5 and 6 as well as rolling a 6 and 5.

Notation

- The probability of variable A taking on value a can be written as
  - $P(A = a)$ or $P(a)$
- The probability of A=a AND B=b is written as
  - $P(a, b)$
- The probability of A=a OR B=b is written as
  - $P(a \vee b)$

Probability Distributions

- Unobserved random variables have distributions
- A distribution is a table of probabilities of values
- A probability is a single number
  - $P(W = rain) = 0.1$
  - Sometimes written $P(rain) = 0.1$

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Joint Distributions

- A joint distribution over a set of random variables specifies a probability for every possible assignment to those variables
  - $P(T, W)$
  - In this example, the temperature can only be hot or cold and the weather can only be sun or rain
- Probabilities must still be between 0 and 1 and probabilities must sum to 1

$$P(T, W)$$

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

Quick Quiz

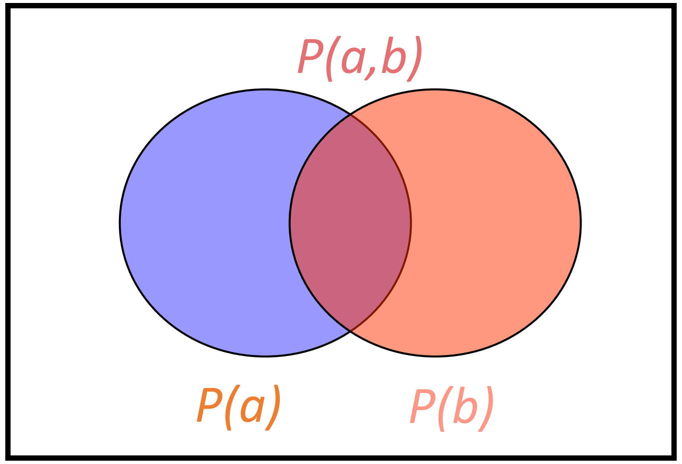
- Probability for a specific event is the sum of the probabilities of the worlds in which it holds
- P(+x, +y) ?
  - 0.2
- P(+x) ?
  - 0.2+0.3=0.5
- P(-y V +x) ?
  - 0.1 + 0.3 + 0.2 = 0.6

$$P(X, Y)$$

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

Probability Space

- Probability of two binary variables
  - $A \in \{a, \neg a\}$
  - $B \in \{b, \neg b\}$
- $P(a) + P(\neg a) = 1$
  - $P(\neg a) = 1 - P(a)$
- $P(a, b) = P(a|b)P(b)$ <-Product rule
- $P(a \vee b) = P(a) + P(b) - P(a, b)$



Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$P(T, W)$

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

$$P(t) = \sum_s P(t, s)$$

$$P(s) = \sum_t P(t, s)$$

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.4

Marginal Distributions

$P(X,Y)$

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	0.5
-x	0.5

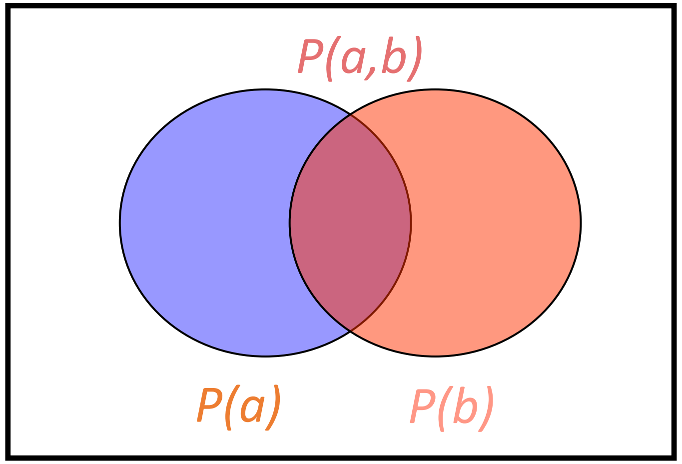
$P(Y)$

Y	P
+y	0.6
-y	0.4

Conditional Probabilities

- The probability of an event given the value of one or more variables have been observed
- For example, the probability of a cavity given a toothache
- Often said to be conditioned on the evidence
- Can obtain from joint probability

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



$$P(T,W)$$

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

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

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

$$= 0.2 + 0.3 \quad = 0.5$$

Quick Quiz

$$P(X, Y)$$

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

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

- P(+x | +y) ?
  - .2/.6=1/3
- P(-x | +y) ?
  - .4/.6=2/3
- P(-y | +x) ?
  - .3/.5=.6

Conditional Probability Distributions

- Conditional probability distributions are probability distributions over some variables given fixed values of others

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

Joint Distribution

$P(T, W)$

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

Normalization Trick

$$P(T, W)$$

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

Normalization Trick

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

$P(T,W)$

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 →

$P(c,W)$

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

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

Probabilistic Inference

- Compute a desired probability from other known probabilities
- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's *beliefs* given the evidence
- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes *beliefs to be updated*

Inference by Enumeration

- P(W)?

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)?
  - P(sun)=.3+.1+.1+.15=.65
  - P(rain)=1- P(sun)=1-0.65=0.35

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, hot)?

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

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

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

- P(W|winter, hot)?
  - P(W, winter, hot)/P(winter, hot)
  - P(sun|winter,hot)=0.1/(0.1 + 0.05)=2/3
  - P(rain|winter,hot)=0.05/(0.1 + 0.05)=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

$$P(a|b) = \frac{P(a,b)}{P(b)} \qquad P(t) = \sum_s P(t,s)$$

- P(W|winter)?

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

$$P(a|b) = \frac{P(a,b)}{P(b)} \qquad P(t) = \sum_s P(t,s)$$

- P(W|winter)?
  - P(W|winter)=P(W,winter)/P(winter)
  - P(sun|winter)=(0.1 + 0.15)/0.5=0.5
  - P(rain|winter)=(0.05 + 0.2)/0.5=0.5

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

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \ldots H_r$

$\left.\right\}$ $X_1, X_2, \ldots X_n$ *All variables*

- We want:

** Works fine with multiple query variables, too*

$$P(Q|e_1 \ldots e_k)$$

- Step 1: Select the entries consistent with the evidence
- Step 2: Sum out H to get joint of Query and evidence
- Step 3: Normalize

$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(\underbrace{Q, h_1 \ldots h_r, e_1 \ldots e_k}_{X_1, X_2, \ldots X_n})$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Inference by Enumeration Analysis

- Worst case time complexity $O(d^n)$
- Space complexity $O(d^n)$
- Where there are n variables that can take on d values

Independence

- Independence
  - $P(a|b) = P(a)$
  - $P(sunny|cavity) = P(sunny)$
  - $P(a, b) = P(a)P(b)$
  - Derived from the product rule
    - $P(a, b) = P(a|b)P(b) = P(a)P(b)$
- Independence between two variables can be written as
  - $P(A, B) = P(A)P(B)$
  - $P(A|B) = P(A)$
  - $P(B|A) = P(B)$

Independence

- P(A,B,C)=P(A)P(B)P(C)
- Greatly reduces the representation size

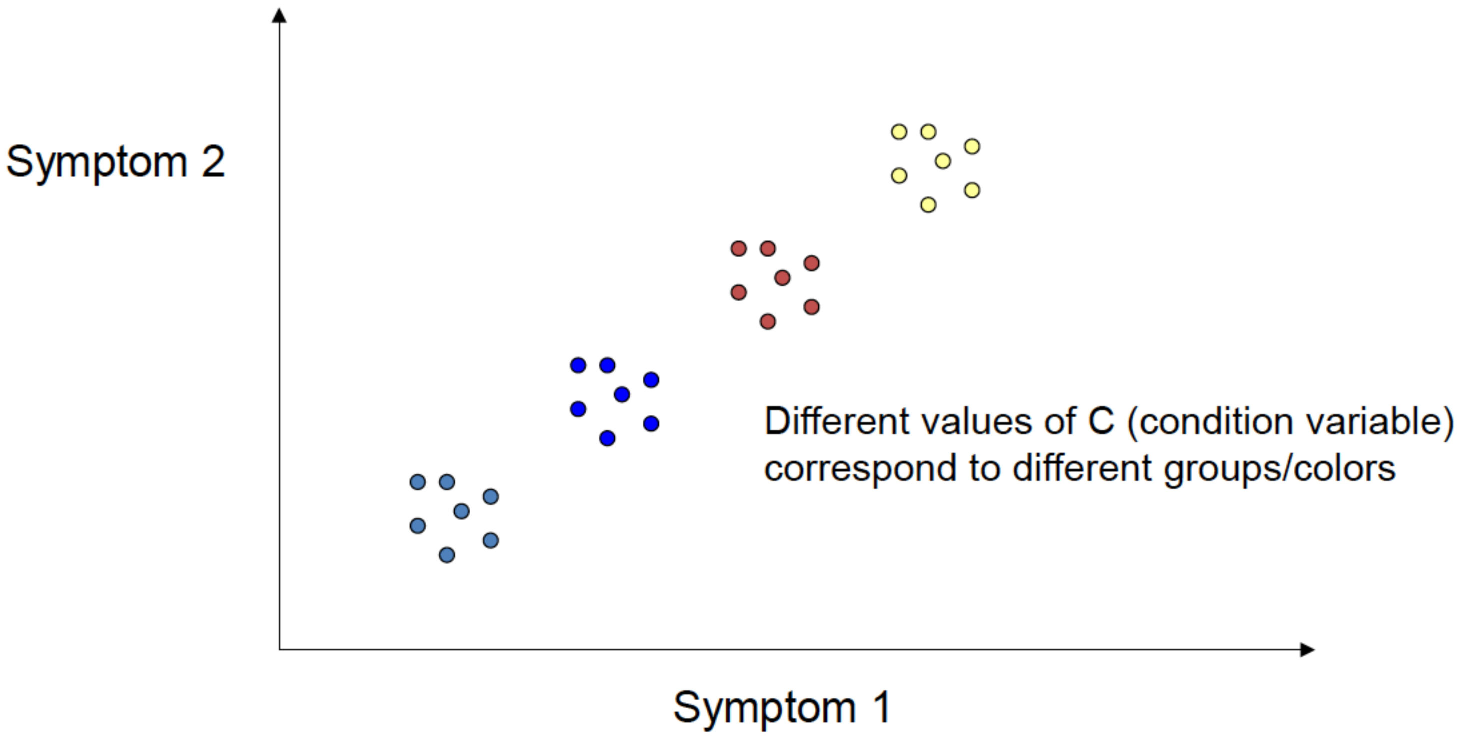
A	P(A)
0	0.4
1	0.6

B	P(B)
0	0.7
1	0.3

C	P(C)
0	0.1
1	0.9

A	B	C	P(A,B,C)
0	0	0	.4 * .7 * .1 = .028
0	0	1	.4 * .7 * .9 = .252
0	1	0	.4 * .3 * .1 = .012
0	1	1	.4 * .3 * .9 = .108
1	0	0	.6 * .7 * .1 = .042
1	0	1	.6 * .7 * .9 = .378
1	1	0	.6 * .3 * .1 = .018
1	1	1	.6 * .3 * .9 = .162

Conditional Independence



- Symptom 1 and symptom 2 are conditionally independent given the group
- But, symptom 1 and symptom 2 are dependent when the group has not been given

Conditional Independence

- $P(A,B|C)=P(A|C)P(B|C)$
- Rain causes both increased umbrella usage and worsened road conditions. These events are not independent because seeing lots of umbrellas makes worsened road conditions more likely
- However, given the condition that it is raining makes the events conditionally independent. Once you know it is raining, seeing umbrellas tells you nothing more about road conditions

Conditional Independence

- $P(T,D|C)=P(T|C)P(D|C)$
- Like independence, conditional independence can greatly reduce the representation size

T	C	P(T|C)
0	0	.9
0	1	.4
1	0	.1
1	1	.6

D	C	P(D|C)
0	0	.8
0	1	.1
1	0	.2
1	1	.9

Joint:

Conditional probabilities:

T	D	C	P(T,D|C)
0	0	0	.9 * .8 = .72
0	0	1	.4 * .1 = .04
0	1	0	.9 * .2 = .18
0	1	1	.4 * .9 = .36
1	0	0	.1 * .8 = .08
1	0	1	.6 * .1 = .06
1	1	0	.1 * .2 = .02
1	1	1	.6 * .9 = .54

Bayes' Rule

- Product Rule
  - $P(a, b) = P(a|b)P(b)$
  - $P(a, b) = P(b|a)P(a)$
- Bayes' Rule
  - $P(b|a) = \frac{P(a|b)P(b)}{P(a)}$
- Often, we perceive evidence as the effect of some unknown cause
  - We perceive toothache, which may be due to a cavity
- It may be a lot easier to model the probability of the effect given the cause
  - I.e. P(symptoms|disease) may be known but P(disease|symptoms) may be unknown
- $P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$

Quick Quiz

2. (10 pts total, 5 pts each) We have a database describing 100 examples of printer failures. Of these, 75 examples are hardware failures, and 25 examples are driver failures. Of the hardware failures, 15 had Windows operating system. Of the driver failures, 15 had Windows operating system. Show your work.

a. (5 pts) Calculate P(windows | hardware) using the information in the problem.

b. (5 pts) Calculate P(driver | windows) using Bayes' rule and the information in the problem.

Quick Quiz

- P(windows|hardware)=15/75
- P(driver|windows) = p(windows|driver)p(driver)/p(windows)=0.5
  - p(windows|driver)p(driver) = (15/25) * (25/100)
  - P(windows)=P(windows,hardware) + P(windows,driver)
  - P(windows)=P(windows|hardware)P(hardware) + P(windows|driver)P(driver)
  - P(windows)=(15/75) * (75/100) + (15/25) * (25/100)
  - P(driver|windows)= $\frac{\frac{15}{25}\frac{25}{100}}{\frac{15}{75}\frac{75}{100}+\frac{15}{25}\frac{25}{100}} = \frac{\frac{15}{100}}{\frac{15}{100}+\frac{15}{100}} = 0.5$

Summary

- Probabilistic inference by enumeration
  - Start with a table that represents the joint distribution
  - Select entries consistent with the evidence
  - Obtain the marginal distribution by summing out all hidden variables (not evidence or query variables)
  - Obtain the conditional distribution by normalizing
  - Space and time complexity is $O(d^n)$
- Exploiting independence can lead to a more compact representation of the joint distribution
- Using Bayes' rule, we can do probabilistic inference with known conditional and marginal distributions