Un
Uncertainty

Probabilistic Reasoning

CSE 415: Introduction to Artificial Intelligence
University of Washington
Spring 2019

Un
Uncertainty

Outline

- Motivation
- Definitions and Laws of Probability
- Random Variables
- Joint and Marginal Distributions
- Conditional Distribution
- Product Rule, Chain Rule, Bayes' Rule
- Inference
- Independence

Un
Uncertainty

Motivation

In the real world:
- data is often noisy,
- many processes cannot be modeled deterministically in a reliable and practical way;
  - parts may be insufficiently understood, other
  - parts too complex for efficient computation.
- logical reasoning does not match the available information,
- adversarial agents may not always obey rationality assumptions.

Un
Uncertainty

Motivation (continued)

Therefore, we need a means to
- represent the certainty/uncertainty of information,
- compute certainty/uncertainty of related information.

The mathematics of probability is our tool of choice.

Un
Uncertainty

Definitions

A *random variable* X is a symbol that represents a class of events that may occur any number of times, and which take on values in a given set D called a *domain*.

Example: Let C be a coin-toss random variable with domain D={H, T}.

A *probability distribution* P for a random variable is a function that assigns to each domain element d_i a value p_i in the range 0 to 1. If D is finite then P is often given as a table.

d_i	p_i
H	0.5
T	0.5

Un
Uncertainty

Definitions

A *joint distribution* over a set of random variables $X_1, X_2, \ldots, X_n$ is a function that assigns to each n-tuple of domain elements $(d_{1,i_1}, d_{2,i_2}, \ldots, d_{n,i_n})$ a value $p_{i_1,i_2,\ldots,i_n}$ in the range 0 to 1. If all the domains D_j are finite then P is often given as a table.

Example with n=2:

$d_{1,i_1} \in D_1$	$d_{2,i_2} \in D_2$	$P(d_{1,i_1}, d_{2,i_2})$
rain	no crash	1/4
rain	crash	1/8
clear	no crash	5/16
clear	crash	1/16
snow	no crash	1/8
snow	crash	1/8

$P(X_1=d_{1,i_1}, X_2=d_{2,i_2})$

Definitions

The *conditional probability* of $P(X{=}d_i \mid Y{=}d_j)$ is the probability of $X{=}d_i$, given $Y{=}d_j$.

Example:
Let D={0, 1, 2, 3, 4, 5, 6, 7, 8, 9}.

Let Odd be the set of outcomes X=1, X=3, ..., X=9.
Let Even " " X=0, X=2, ..., X=8.
Let Prime " " X=2, X=3, X=5, X=7.
P(Odd) = 5/10 = 0.5
P(Prime) = 4/10 = 0.4
P(Odd | Prime) = |odd and prime | / | prime | = 3/4 = 0.75.
P(Prime | Odd) = |odd and prime | / | odd | = 3/5 = 0.6.

Two Laws of Probability

The sum rule: $\Sigma_{1<i<m} P(X = d_i) = 1$
Adding the probabilities of all the possible outcomes for a random variable must give a total of 1.0.

The product rule: $P(X{=}x, Y{=}y) = P(X{=}x)\, P(Y{=}y \mid X{=}x)$.
The joint probability of two random variables is equal to the product of the marginal of one times the conditional of the other given the one.

Logic vs Probability: Generalizing Modus Ponens

Modus Ponens:

$P \rightarrow Q$	If it's raining then I do my homework.
P	It's raining.
Q	I do my homework.

Bayes' Rule: (general idea)

If P then sometimes Q	If it's raining then I might do my homework.
P	It's raining.
Maybe Q	I might do my homework.

(Bayes' rule lets us calculate the probability of Q, taking P into account.)

Bayes' Rule

E: Some evidence exists, i.e., a particular condition is true
H: some hypothesis is true.

P(E|H) = probability of E given H.
P(E|~H) = probability of E given not H.
P(H) = probability of H, independent of E.

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)}$$

P(E) = P(E|H) P(H) + P(E|~H)(1 - P(H))

Applying Bayes' Rule

E: The patient's white blood cell count exceeds 110% of average.

H: The patient is infected with tetanus.

P(E|H) = 0.8 class-conditional probability
P(E|~H) = 0.3 "
P(H) = 0.01 prior probability

posterior probability:

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)} = \frac{(0.8)\ (0.01)}{(0.8)\ (0.01) + (0.3)(0.99)} = \frac{0.008}{0.305} = 0.0262$$

P(E) = P(E|H) P(H) + P(E|~H)(1 - P(H))

What is....?

Value

Random Variable

$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Probability Distribution

Un
Uncertainty

Joint Distributions

- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots X_n$ specifies a probability for each assignment (or *outcome*):

$$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$$
$$P(x_1, x_2, \ldots x_n)$$

  – Must obey: $P(x_1, x_2, \ldots x_n) \geq 0$

$$\sum_{(x_1, x_2, \ldots x_n)} P(x_1, x_2, \ldots x_n) = 1$$

- Size of joint distribution if n variables with domain sizes d?
  – For all but the smallest distributions, impractical to write out!

$P(T, W)$

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

Un
Uncertainty

Probabilistic Models

- A *probabilistic model* is a joint distribution over a set of random variables
- Probabilistic models:
  – (Random) variables with domains
  – Joint distributions: say whether assignments (called "*outcomes*") are likely
  – Normalized: sum to 1.0
  – Ideally: only certain variables directly interact
- Constraint satisfaction problems:
  – Variables with domains
  – Constraints: state whether assignments are possible
  – Ideally: only certain variables directly interact

Distribution over T,W

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

Constraint over T,W

T	W	P
hot	sun	T
hot	rain	F
cold	sun	F
cold	rain	T

Un
Uncertainty

Events

- An *event* is a set E of outcomes

$$P(E) = \sum_{(x_1 \ldots x_n) \in E} P(x_1 \ldots x_n)$$

- From a joint distribution, we can calculate the probability of any event
  – Probability that it's hot AND sunny?
  – Probability that it's hot?
  – Probability that it's hot OR sunny?
- Typically, the events we care about are *partial assignments*, like P(T=hot)

$P(T, W)$

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

Un
Uncertainty

Quiz: Events

- P(+x, +y) ?
- P(+x) ?
- P(-y OR +x) ?

$P(X, Y)$

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

Un
Uncertainty

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

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

$P(T)$

T	P
hot	0.5
cold	0.5

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

$P(W)$

W	P
sun	0.6
rain	0.4

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

Un
Uncertainty

Quiz: Marginal Distributions

$P(X, Y)$

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

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

$P(X)$

X	P
+x	
-x	

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

$P(Y)$

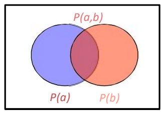
Y	P
+y	
-y	

Un
Uncertainty

Conditional Probabilities

- A simple relation between joint and marginal probabilities
  - In fact, this is taken as the **definition** of a conditional 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

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

$$= P(W=s,T=c) + P(W=r,T=c)$$
$$= 0.2 + 0.3 = 0.5$$

Un
Uncertainty

Quiz: Conditional Probabilities

$P(X, Y)$

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

- P(+x | +y) ?
- P(-x | +y) ?
- P(-y | +x) ?

Un
Uncertainty

Conditional Distributions

- Conditional 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

Un
Uncertainty

Conditional Distribs - The Slow Way…

$P(T, W)$

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

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

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

$P(W|T = c)$

W	P
sun	0.4
rain	0.6

Un
Uncertainty

Probabilistic Inference

- Probabilistic inference = *"compute a desired probability from other known probabilities (e.g. conditional from joint)"*
- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's *beliefs* given the eviden
- 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 update*

Un
Uncertainty

Inference by Enumeration

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

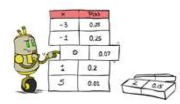
$\left.\right\}$ $X_1, X_2, \dots X_n$ *All variables*

- We want:

** Works fine with multiple query variables, too*

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

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence

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

- Step 3: Normalize

$$\times \frac{1}{Z}$$

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

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

Inference by Enumeration

- P(W)?
- P(W | winter)?
- P(W | winter, hot)?

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

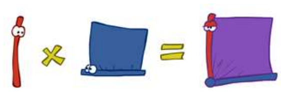
Inference by Enumeration

- Computational problems?
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

The Product Rule

- Sometimes have conditional distributions but want the joint

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



The Product Rule

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

- Example:

$P(W)$

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	
dry	sun	
wet	rain	
dry	rain	

The Chain Rule

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

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

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

Independence

- Two variables are *independent* in a joint distribution if:

$$P(X,Y) = P(X)P(Y) \qquad X \perp\!\!\!\perp Y$$

$$\forall x, y \; P(x,y) = P(x)P(y)$$

  - Says the joint distribution *factors* into a product of two simple ones
  - Usually variables aren't independent!
- Can use independence as a *modeling assumption*
  - Independence can be a simplifying assumption
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity}?
- Independence is like something from CSPs: what?

Example: Independence?

$P_1(T, W)$

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

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.4

$P_2(T, W) = P(T)P(W)$

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

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

$P(X_2)$

H	0.5
T	0.5

$\ldots$

$P(X_n)$

H	0.5
T	0.5

$P(X_1, X_2, \ldots X_n)$

2^n