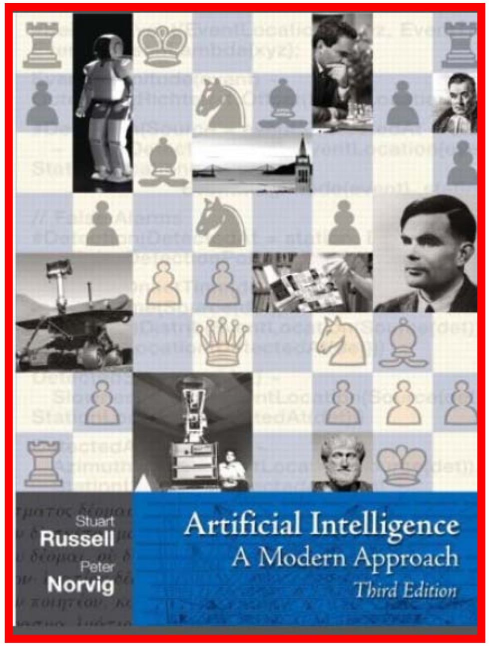


(Chapter-13)

PROBABILITY

Yanmei Zheng

Outline

- Probability
  - ✓ Random Variables
  - ✓ Joint and Marginal Distributions
  - ✓ Conditional Distribution
  - ✓ Product Rule, Chain Rule, Bayes' Rule
  - ✓ Inference
  - ✓ Independence
- You'll need all this stuff A LOT for the next few weeks, so make sure you go over it now!

Random Variables

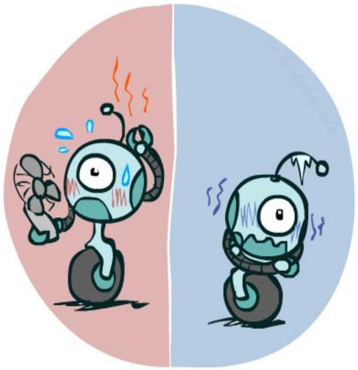
- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - T = Is it hot or cold?
  - D = How long will it take to drive to work?
  - L = Where is the ghost?
- We denote random variables with capital letters
- Like variables in a CSP, random variables have domains
  - R in {true, false}   (often write as {+r, -r})
  - T in {hot, cold}
  - D in $[0, \infty)$
  - L in possible locations, maybe {(0,0), (0,1), …}

Probability Distributions

- Associate a probability with each value

  - ✓ **Temperature:**

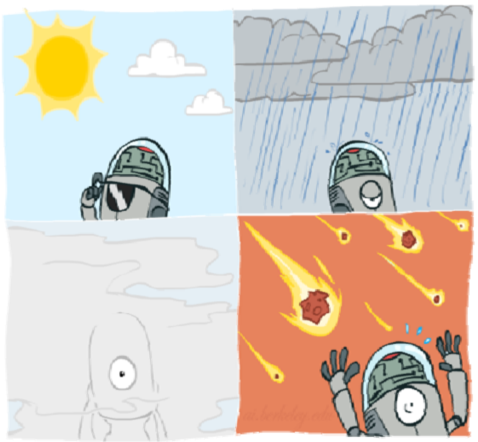
$P(T)$



T	P
hot	0.5
cold	0.5

  - **Weather**:

$P(W)$



W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Dr.Yanmei Zheng

Probability Distributions

- Unobserved random variables have distributions

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Shorthand notation:

OK *if* all domain entries are unique

$$P(hot) = P(T = hot),$$

$$P(cold) = P(T = cold),$$

$$P(rain) = P(W = rain),$$

- A distribution is a TABLE of probabilities of values
- A probability (lower case value) is a single number $P(W = rain) = 0.1$
- Must have: $\forall x \;\; P(X = x) \geq 0$ and $\sum_x P(X = x) = 1$

Joint Distributions

- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots X_n$
  specifies a real number 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 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

Probabilistic Models

- A probabilistic model is a joint distribution over a set of random variables
- Probabilistic models:
  - ✓ (Random) variables with domains
  - ✓ Assignments are called *outcomes*
  - ✓ Joint distributions: say whether assignments (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

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?

$P(T, W)$

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

- Typically, the events we care about are *partial assignments*, like P(T=hot)

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

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

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

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

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

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

$P(X)$

X	P
+x	
-x	

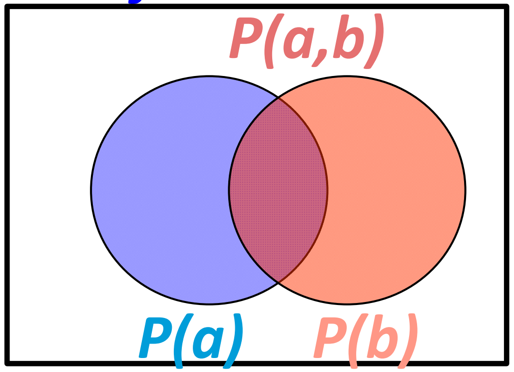
$P(Y)$

Y	P
+y	
-y	

Conditional Probabilities

- A simple relation between joint and conditional 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

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

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

Quiz: Conditional Probabilities

- $P(+x \mid +y)$?

$P(X, Y)$

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

- $P(-x \mid +y)$?

- $P(-y \mid +x)$?

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

Normalization Trick

$P(T, W)$

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

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

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

$P(W|T = c)$

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

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

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

$P(c, W)$

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

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}$$

Dr.Yanmei Zheng

Quiz: Normalization Trick

- P(X | Y=-y) ?

$P(X, Y)$

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

SELECT the joint probabilities matching the evidence →

NORMALIZE the selection (make it sum to one) →

To Normalize

- (Dictionary) To bring or restore to a normal condition

All entries sum to ONE

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

- Example 1

W	P
sun	0.2
rain	0.3

Normalize → Z = 0.5

W	P
sun	0.4
rain	0.6

- Example 2

T	W	P
hot	sun	20
hot	rain	5
cold	sun	10
cold	rain	15

Normalize → Z = 50

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

Normalization Trick

- A trick to get a whole conditional distribution at once:
  - ✓ Select the joint probabilities matching the evidence
  - ✓ Normalize the selection (make it sum to one)

$P(T, W)$

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

→ Select

$P(T, r)$

T	R	P
hot	rain	0.1
cold	rain	0.3

→ Normalize

$P(T|r)$

T	P
hot	0.25
cold	0.75

- ✓ Why does this work? Sum of selection is P(evidence)! (P(r), here)

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}$$

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 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

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

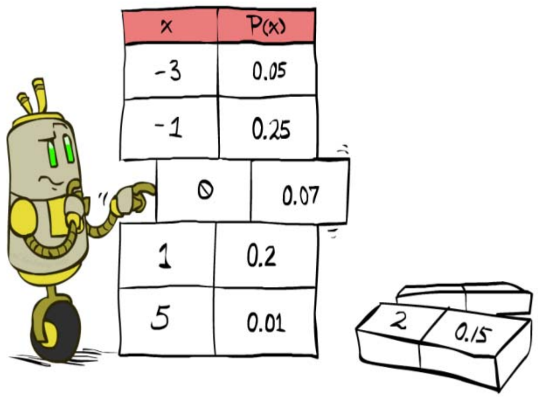
$\left.\begin{matrix} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{matrix}\right\} X_1, X_2, \dots X_n$

All variables

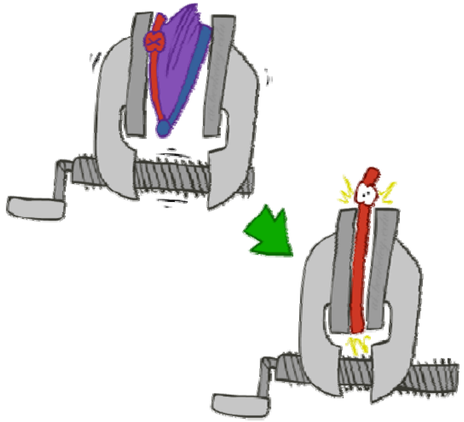
- **We want:** * **Works fine with multiple query variables, too**

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

- **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**

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

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

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

Inference by Enumeration

- **Start with the joint distribution:**

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

- **For any proposition *φ*, sum the atomic events where it is true:**

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

(catch = dentist's steel probe gets caught in cavity)

Inference by Enumeration

- Start with the joint distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

$$\boldsymbol{P}\ (\boldsymbol{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inference by Enumeration

- Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega:\omega \models \phi} P(\omega)$$

$$\boldsymbol{P}\,(\boldsymbol{cavity} \vee \boldsymbol{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

Inference by Enumeration

- **Start with the joint distribution:**

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- **Can also compute conditional probabilities:**

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- **Denominator can be viewed as a normalization constant α**

$$\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)$$
$$= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$
$$= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$
$$= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$$

- **General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables**

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

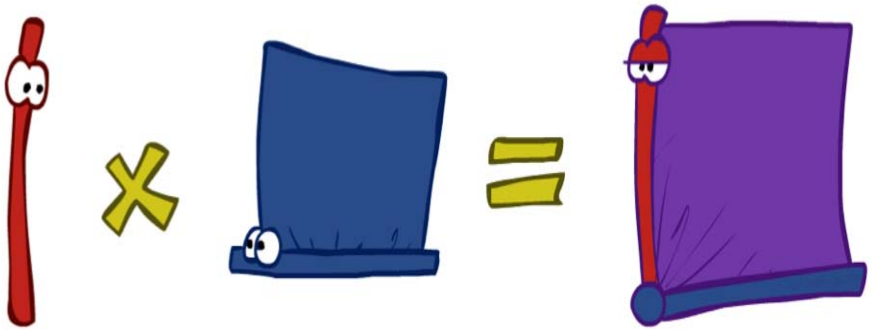
Inference by Enumeration

- **Obvious 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

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



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

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

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

- Why is this always true?

Independence

Independence

- Two variables are ***independent*** if:

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

  - ✓ This says that their joint distribution *factors* into a product two simpler distributions
  - ✓ Another form:

$$\forall x, y : P(x|y) = P(x)$$

  - ✓ We write: $X \perp\!\!\!\perp Y$

- Independence is a simplifying *modeling assumption*
  - ✓ *Empirical* joint distributions: at best "close" to independent
  - ✓ What could we assume for {Weather, Traffic, Cavity, Toothache}?

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

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

$P(W)$

W	P
sun	0.6
rain	0.4

$P_2(T, W)$

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

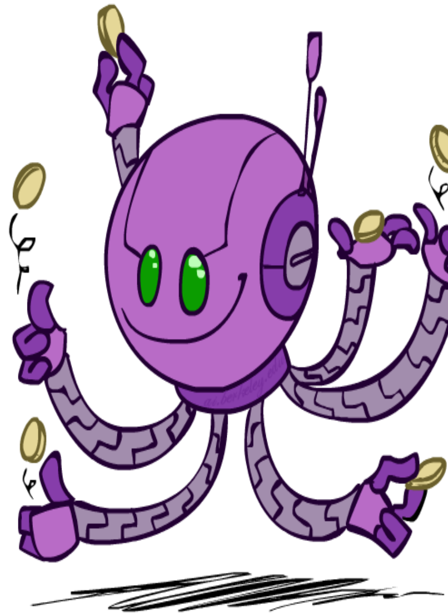
$P(X_2)$

H	0.5
T	0.5

$\cdots$

$P(X_n)$

H	0.5
T	0.5

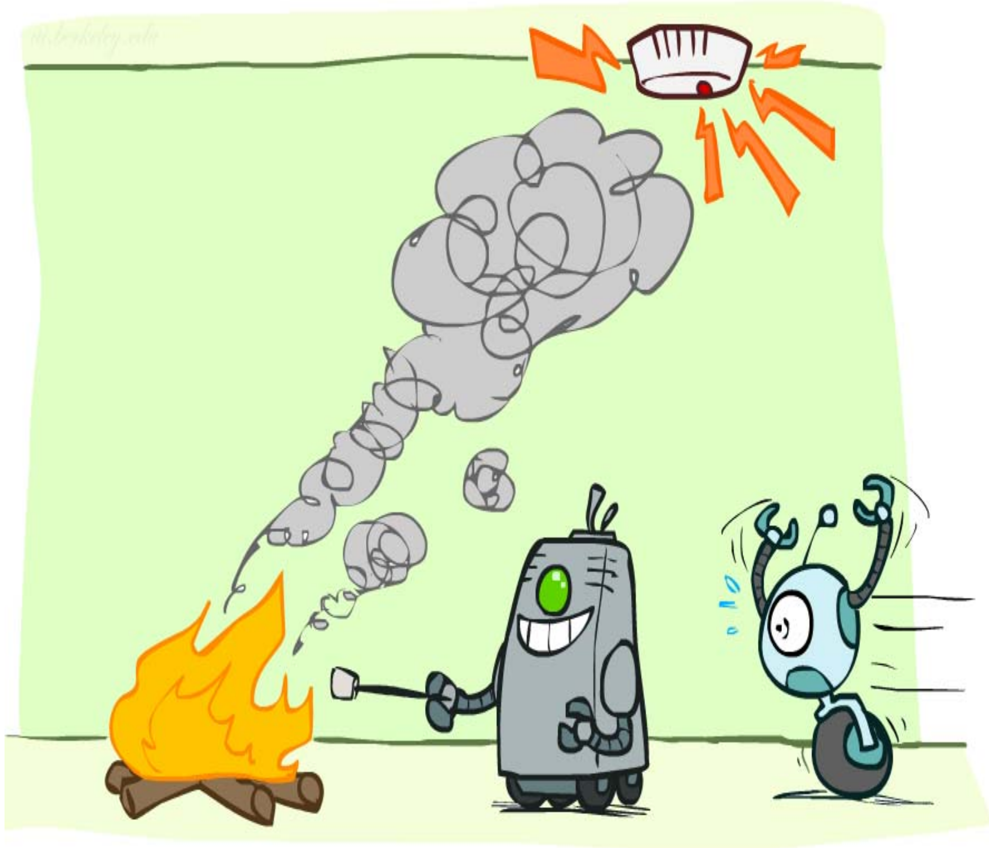


$P(X_1, X_2, \ldots X_n)$

2^n

Dr.Yanmei Zheng

Conditional Independence



Conditional Independence

- P(Toothache, Cavity, Catch)
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - ✓ P(+catch | +toothache, +cavity) = P(+catch | +cavity)
- The same independence holds if I don’t have a cavity:
  - ✓ P(+catch | +toothache, -cavity) = P(+catch| -cavity)
- Catch is *conditionally independent* of Toothache given Cavity:
  - ✓ P(Catch | Toothache, Cavity) = P(Catch | Cavity)

- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily

Conditional Independence

- **Unconditional (absolute) independence very rare (why?)**
- ***Conditional independence* is our most basic and robust form of knowledge about uncertain environments.**
- **X is conditionally independent of Y given Z** $X \perp\!\!\!\perp Y | Z$

  if and only if: $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$

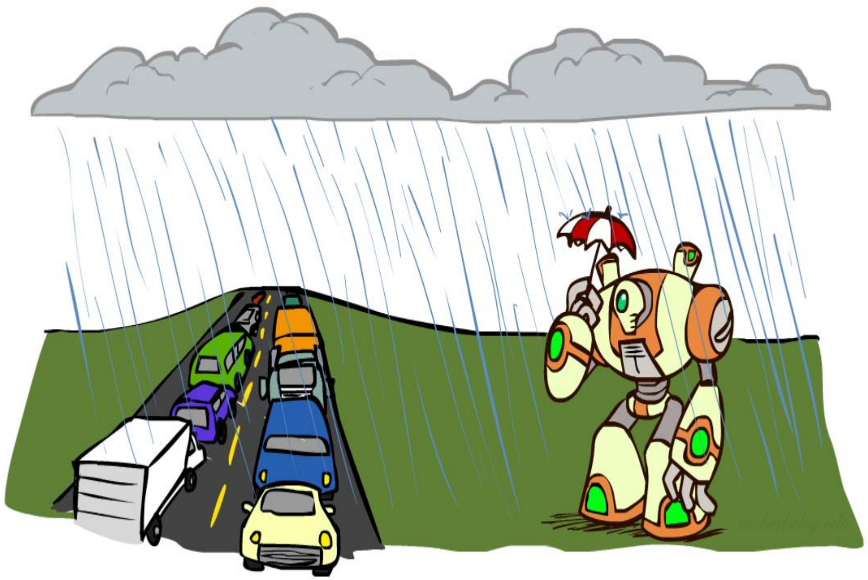
  or, equivalently, if and only if

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

Conditional Independence

- What about this domain:
  - ✓ **Traffic**
  - ✓ **Umbrella**
  - ✓ **Raining**

P(Umbrella | Traffic, Rain)=P(Umbrella | Rain)

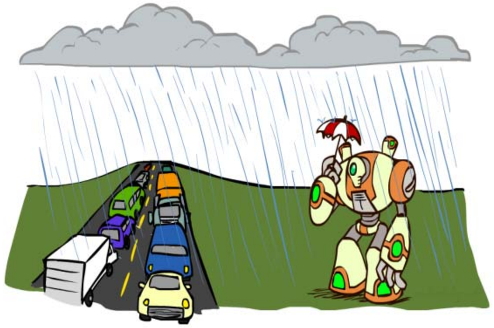


Dr.Yanmei Zheng

Conditional Independence and the Chain Rule

- **Chain rule:** $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- **Trivial decomposition:**



$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- **With assumption of conditional independence:**

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

- **Bayes' nets / graphical models help us express conditional independence assumptions**

Bayes Rule

Bayes' Rule

- Two ways to factor a joint distribution over two variables:

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

- Dividing, we get:

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

- Why is this at all helpful?
  - ✓ Lets us build one conditional from its reverse
  - ✓ Often one conditional is tricky but the other one is simple
  - ✓ Foundation of many systems we'll see later (e.g. ASR, MT)
- In the running for most important AI equation!

Inference with Bayes' Rule

- **Example: Diagnostic probability from causal probability:**

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

- **Example:**
  - ✓ **M: meningitis, S: stiff neck**

$$\left.\begin{aligned} P(+m) &= 0.0001 \\ P(+s|+m) &= 0.8 \\ P(+s|-m) &= 0.01 \end{aligned}\right\} \text{Example givens}$$

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}$$

  - ✓ **Note: posterior probability of meningitis still very small**
  - ✓ **Note: you should still get stiff necks checked out! Why?**

Quiz: Bayes' Rule

- Given:

$P(W)$

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

- What is P(W | dry) ?

Quiz: Bayes' Rule

- **Given:**

$P(W)$

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

- **What is P(W | dry) ?**

$$P(sun \mid dry) = \frac{P(dry \mid sun)P(sun)}{P(dry)} = \frac{0.9*0.8}{0.78}$$

$$P(dry) = \frac{P(dry, sun) + P(dry, rain)}{0.08 + 0.72 + 0.14 + 0.06}$$

$$= \frac{P(dry \mid sun)P(sun) + P(dry \mid rain)P(rain)}{0.08 + 0.72 + 0.14 + 0.06}$$

$$= \frac{0.9*0.8 + 0.3*0.2}{1} = 0.78$$

$$P(rain \mid dry) = \frac{P(dry \mid rain)P(rain)}{P(dry)} = \frac{0.3*0.2}{0.78}$$

$$P(dry) = \frac{P(dry, sun) + P(dry, rain)}{0.08 + 0.72 + 0.14 + 0.06}$$

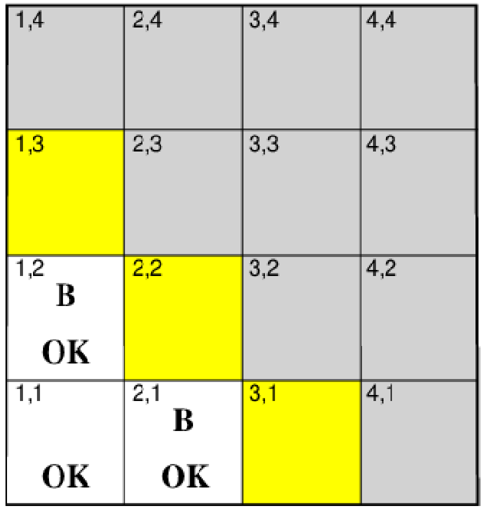
$$= \frac{P(dry \mid sun)P(sun) + P(dry \mid rain)P(rain)}{0.08 + 0.72 + 0.14 + 0.06}$$

$$= \frac{0.9*0.8 + 0.3*0.2}{1} = 0.78$$

Wampus World

Dr.Yanmei Zheng

Wumpus World



- P_{ij} *=true* iff [*i, j*] contains a pit

- B_{ij} *=true* iff [*i, j*] is breezy

  Include only $B_{1,1}$, $B_{1,2}$, $B_{2,1}$ in the probability model

Specifying the Probability Model

- The full joint distribution is $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

- Apply product rule: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$

  This gives us: $P(Effect|Cause)$

- First term: 1 if pits are adjacent to breezes, 0 otherwise

- Second term: pits are placed randomly, probability 0.2 per square:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \prod_{i,j=1,1}^{4,4} \mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

for n pits.

Observations and Query

- We know the following facts:

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$
$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

- Query is $\mathbf{P}(P_{1,3} | known, b)$

- Define $Unknown$ = P_{ij}s other than $P_{1,3}$ and $Known$

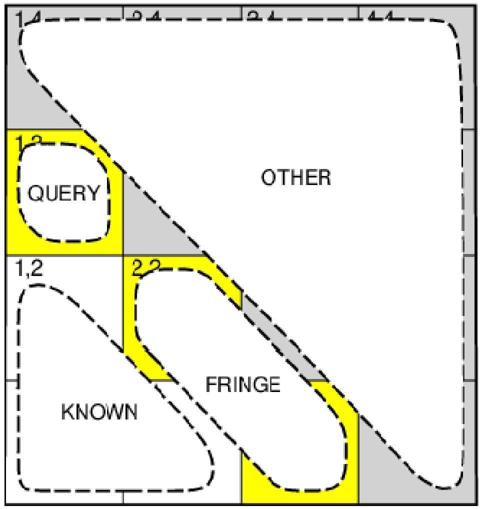
- For inference by enumeration, we have

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

- Grows exponentially with number of squares!

Using Conditional Independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares



- **Define** $Unknown = Fringe \cup Other$

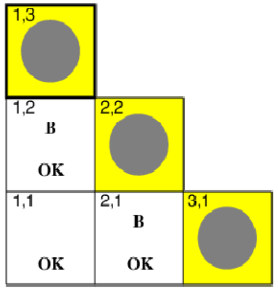
  $P(b \mid P_{1,3}, Known, Unknown) = P(b \mid P_{1,3}, Known, Fringe)$

- **Manipulate query into a form where we can use this!**

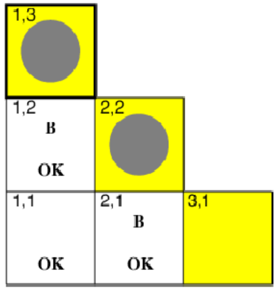
Using Conditional Independence

$$
\begin{aligned}
\mathbf{P}(P_{1,3}|known, b) &= \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b) \\
&= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\mathbf{P}(P_{1,3}, known, unknown) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other) \\
&= \alpha\, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other) \\
&= \alpha' \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)
\end{aligned}
$$

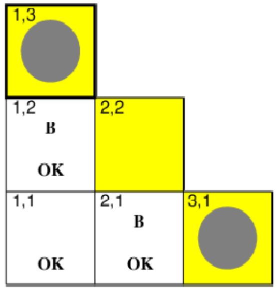
Using Conditional Independence



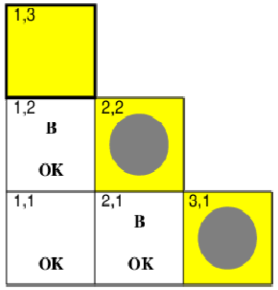
$0.2 \times 0.2 = 0.04$



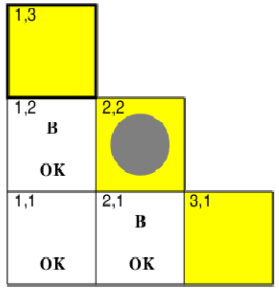
$0.2 \times 0.8 = 0.16$



$0.8 \times 0.2 = 0.16$



$0.2 \times 0.2 = 0.04$



$0.2 \times 0.8 = 0.16$

$$\mathbf{P}(P_{1,3} \mid known, b) = \alpha' \left(0.2(0.04 + 0.16 + 0.16),\ 0.8(0.04 + 0.16)\right) \approx (0.31, 0.69)$$

$$\mathbf{P}(P_{2,2} \mid known, b) \approx (0.86, 0.14)$$

Thank you

End of

Chapter 13