FIT5047 Probability

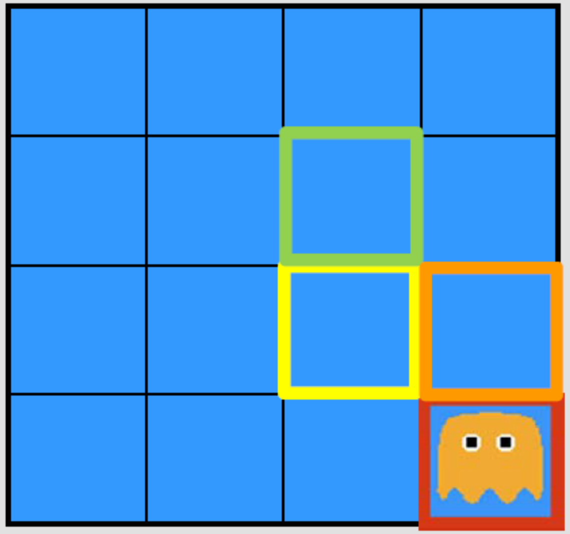
授课老师：Joe

FIT5047平时班 – Probability

- Probability fundamentals
- Inference by enumeration
- Product Rule, Chain Rule, Bayes’ Rule
- Independence and conditional independence

FIT5047平时班 – Probability

- A ghost is somewhere in the grid
- Sensor readings tell how close a tile is to the ghost
  - On the ghost: red
  - 1 away: orange
  - 2 away: yellow
  - 3+ away: green
- Sensors are noisy, but we know Pr(Color|Distance)

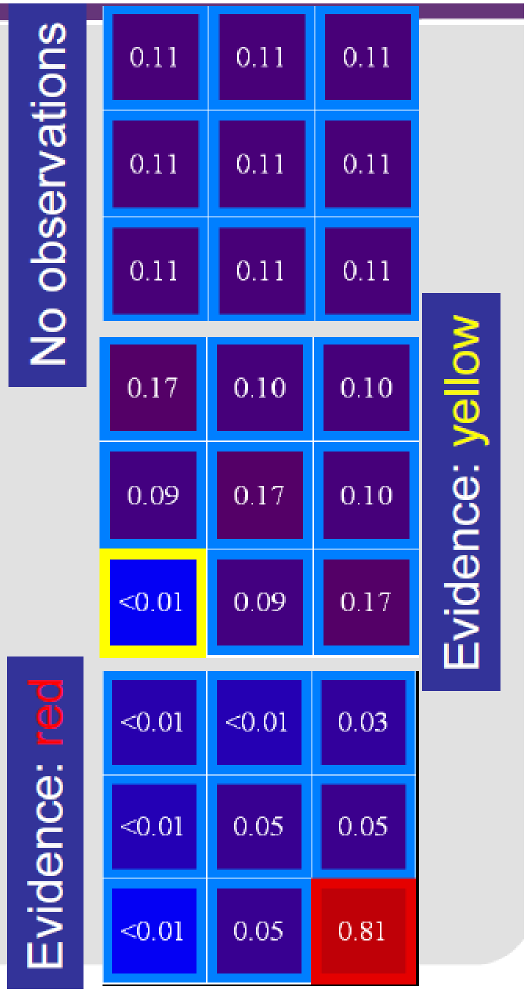


Pr(red|2)	Pr(orange|2)	Pr(yellow|2)	Pr(green|2)	TOTAL
0.05	0.17	0.46	0.32	1

We want to know: Pr(Location | Color)

FIT5047平时班 – Probability

- **General situation:**
  - **Evidence**: Agent knows certain things about the state of the world
  - **Hidden variables**: Agent needs to reason about other aspects
  - **Model**: Agent knows something about how the known variables relate to the unknown variables
- **Probabilistic reasoning gives us a framework for managing our beliefs and knowledge**



FIT5047平时班 – Probability

- **Unobserved random variables have distributions that represent probabilities of value assignments**

Pr(Temp)

Temp	Pr
warm	0.5
cold	0.5

Pr(Weather)

Weather	Pr
sunny	0.6
rain	0.1
fog	0.3

- **A probability is a single number**

$\Pr(\text{Weather}=\text{rain}) = 0.1$ or $\Pr(\text{rain}) = 0.1$

FIT5047平时班 – Probability

Kolmogorov's axioms for finite discrete random variables – where $e_1, \ldots, e_n$ are the possible distinct values of random variable E

$$\Pr(e_i) \geq 0 \quad \forall i = 1, \ldots, n$$

$$\Pr(e_i) \leq 1 \quad \forall i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \Pr(e_i) = 1$$

$$\forall e_i, e_j \subseteq E$$

$$\text{if } e_i \cap e_j = \varnothing \text{ then } \Pr(e_i \vee e_j) = \Pr(e_i) + \Pr(e_j)$$

FIT5047平时班 – Probability

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

  $\Pr(X_1{=}x_1, \ldots, X_n{=}x_n)$ or $\Pr(x_1, \ldots, x_n)$

  – Size of distribution of n variables with domain sizes d?

- Must obey:

$$\forall x_i \quad \Pr(x_1, \ldots, x_n) \geq 0$$

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

- For all but small distributions, impractical to write out

$\Pr(W,T)$

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

FIT5047平时班 – Probability

- **From a joint distribution, we can calculate the probability of any event**
  - Probability that it is hot AND sunny
  - Probability that it is hot
  - Probability that it is hot OR sunny
- **Typically, the events we care about are *partial assignments*, like Pr(T=hot)**

$$\mathbf{Pr}(W, T)$$

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

Marginal distributions **are sub-tables that eliminate variables**

Marginalization **(summing out): Combine collapsed rows by adding**

$\mathbf{Pr}(W,T)$

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

$$\Pr(t) = \sum_{w\in\{sun,rain\}} \Pr(t,w)$$

$\Pr(T)$

T	Pr
hot	0.5
cold	0.5

$$\Pr(w) = \sum_{t\in\{hot,cold\}} \Pr(t,w)$$

$\Pr(W)$

W	Pr
sun	0.6
rain	0.4

FIT5047平时班 – Probability

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

Joint Distribution

$$\Pr(W, T)$$

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

Conditional Distributions

$\Pr(W|T)$

$$\Pr(W|T = hot)$$

W	Pr
sun	0.8
rain	0.2

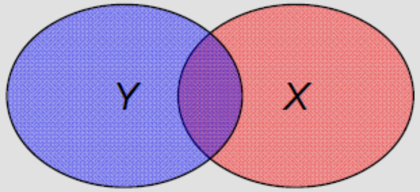
$$\Pr(W|T = cold)$$

W	Pr
sun	0.4
rain	0.6

FIT5047平时班 – Probability

$$\Pr(X \mid Y) = \frac{\Pr(X \wedge Y)}{\Pr(Y)}$$

$\Pr(X \cap Y)$



$\Pr(Y)$ $\Pr(X)$

$\mathbf{Pr}(\boldsymbol{W}, \boldsymbol{T})$

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

$$\Pr(W = rain | T = cold) = ?$$

FIT5047平时班 – Probability

- ***Conditional or posterior probabilities:***
  - E.g., Pr(*cavity* | *toothache*)=0.8, given that *toothache* is all I know
- **Notation for conditional distributions:**
  - Pr(*cavity* | *toothache*) = a single number
  - Pr(Cavity, Toothache) = 2x2 table sums to 1
  - Pr(Cavity | Toothache) = Two 2-element vectors, each sums to 1
- **If we know more:**
  - Pr(*cavity* | *toothache*, *catch*) = 0.9
  - Pr(*cavity* | *toothache*, *cavity*) = 1
- **Less specific beliefs remain *valid* after more evidence arrives, but are not always *useful***
- **New evidence may be irrelevant, allowing simplification:**
  - Pr(*cavity* | *toothache*, *traffic*) = Pr(*cavity* | *toothache*) = 0.8

- **A trick to get a whole conditional distribution at once:**
  - Select the joint probabilities matching the evidence
  - ***Normalize*** the selection (make it sum to 1)

$\Pr(W,T)$

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

$\xrightarrow{\text{Select}}$

$\Pr(T, rain)$

T	R	Pr
hot	rain	0.1
cold	rain	0.3

$\xrightarrow{\text{Normalize}}$

$\Pr(T \mid rain)$

T	Pr
hot	0.25
cold	0.75

FIT5047平时班 – Probability

Pr(sun)?

Pr(sun | summer)?

Pr(sun | winter, hot)?

S	T	W	Pr
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

- **Sometimes we have conditional distributions but want the joint distribution**

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

- **Example:**

$\mathbf{Pr}(\boldsymbol{W})$

W	Pr
sun	0.8
rain	0.2

$\mathbf{Pr}(\boldsymbol{T}|\boldsymbol{W})$

T	W	Pr
cold	sun	0.1
hot	sun	0.9
cold	rain	0.7
hot	rain	0.3

$\Longleftrightarrow$

$\mathbf{Pr}(\boldsymbol{T},\boldsymbol{W})$

T	W	Pr
cold	sun	0.08
hot	sun	0.72
cold	rain	0.14
hot	rain	0.06

The Chain Rule

- **We can always write a joint distribution as an incremental product of conditional distributions**

$$\Pr(x_1, \ldots, x_n) = \prod_{i=1}^{n} \Pr(x_i | x_1, \ldots, x_{i-1})$$

- **Example:**
  Pr(Traffic,Umbrella,Rain)=
  Pr(Umbrella|Rain,Traffic) x Pr(Traffic|Rain) x Pr(Rain)
- **Why is this true?**

$$\Pr(x_1, \ldots, x_n) = \Pr(x_n | x_1, \ldots, x_{n-1}) \Pr(x_1, \ldots, x_{n-1})$$
$$= \Pr(x_n | x_1, \ldots, x_{n-1}) \Pr(x_{n-1} | x_1, \ldots, x_{n-2}) \Pr(x_1, \ldots, x_{n-2})$$

FIT5047平时班 – Probability

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

$$\Pr(x, y) = \Pr(x|y)\Pr(y) = \Pr(y|x)\Pr(x)$$

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

FIT5047平时班 – Probability

- **Attributed to Rev. Thomas Bayes**

$$\Pr(h \mid e) = \frac{\Pr(e \mid h)\Pr(h)}{\Pr(e)}$$

- **Also called *Conditionalization*:**

$$\Pr'(h) = \Pr(h \mid e)$$

- **Also read as**

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Prob of evidence}}$$

- **Assumptions:**
  - Joint priors over $\{h_i\}$ and e exist
  - Total evidence: e is observed

FIT5047平时班 – Probability

Diagnosis of breast cancer (hypothesis), given xray (evidence)

- Let $Pr(h)=0.01$, $Pr(e|h)=0.8$ and $Pr(e|{\sim}h)=0.1$
- Bayes theorem yields

$$
\begin{aligned}
\Pr(h \mid e) &= \frac{\Pr(e \mid h)\Pr(h)}{\Pr(e)} \\
&= \frac{\Pr(e \mid h)\Pr(h)}{\Pr(e \mid h)\Pr(h) + \Pr(e \mid \sim h)\Pr(\sim h)} \\
&= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times 0.99} \\
&= \frac{0.008}{0.008 + 0.099} = \frac{0.008}{0.107} \approx 0.075
\end{aligned}
$$

Ghostbusters Revisited (I)

- **We have two distributions:**
  - **Prior distribution** over ghost location: $\Pr(L)$
  - **Sensor model**: $\Pr(\underbrace{R}_{\text{reading}} \mid \underbrace{D}_{\text{distance}})$
    - > Given by some "black box" process
    - > Assume reading is at the lower left corner
    - > E.g., $\Pr(yellow|D\geq3)=0.27$
      $\Pr(yellow|D=2)=0.46$
      $\Pr(yellow|D=1)=0.25$
      $\Pr(yellow|D=0)=0.03$

Should these probabilities sum to 1?

- **The posterior distribution Pr(L|R) over ghost locations given a reading**

$$\Pr(l=(3,1)|yellow)$$
$$\propto \Pr(yellow|l=(3,1))\Pr(l=(3,1))$$
$$\propto 0.03*0.11=0.0033$$

	1	2	3
1	0.11	0.11	0.11
2	0.11	0.11	0.11
3	0.11	0.11	0.11

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

Example Problems

- Suppose a murder occurs in a town of population 10,000 (10,001 before the murder). A suspect is brought in and DNA tested. The probability that there is a DNA match given that a person is innocent is 1/100,000; the probability of a match on a guilty person is 1. What is the probability he is guilty given a DNA match?

FIT5047平时班 – Probability

- Doctors have found that people with Creutzfeldt–Jakob disease (CJ) almost invariably ate lots of hamburgers, thus Pr(HamburgerEater|CJ) = 0.9. CJ is a rare disease: about 1 in 100,000 people get it. Eating hamburgers is widespread: Pr(HamburgerEater) = 0.5. What is the probability that a regular hamburger eater will have CJ disease?

FIT5047平时班 – Probability

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

$$\Pr(X,Y) = \Pr(X)\Pr(Y)$$

$$\forall x,y \ \ \Pr(x,y) = \Pr(x)\Pr(y) \ \ \text{or} \ \ \Pr(x|y)=\Pr(x)$$

$$X \perp\!\!\!\perp Y$$

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

Which Variables are Independent?

$\mathbf{Pr}(T)$

T	Pr
warm	0.5
cold	0.5

$\mathbf{Pr}_1(T, W)$

T	W	Pr
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

$\mathbf{Pr}_2(T, W)$

T	W	Pr
warm	sun	0.3
warm	rain	0.2
cold	sun	0.3
cold	rain	0.2

$\mathbf{Pr}(W)$

W	Pr
sun	0.6
rain	0.4

FIT5047平时班 – Probability

- Unconditional (absolute) independence is rare
- Conditional independence is our most basic and robust form of knowledge about uncertain environments:

$$\forall x, y, z \qquad \Pr(x, y|z) = \Pr(x|z)\Pr(y|z) \text{ or}$$

Extra practice questions on probability

FIT5047平时班 – Probability

Consider a fair, seven sided dice; a roll X from a such a dice follows the probability distribution:

$$\mathbb{P}(X = x) = \frac{1}{7}, x \in \{1, 2, 3, 4, 5, 6, 7\}$$

1. If we roll one of our seven sided dice 3 times and record the results, what is the probability that all three rolls will be be greater than 5?

2. If we roll one of our seven sided dice five times. Find the probability that exactly three rolls show the same number, (i.e., three of a kind), and the remaining two rolls show the same number different from the other number.

3. If we roll one of our seven sided dice five times, what is the probability that three or more rolls show the same Number?

Hard question

FIT5047平时班 – Probability

What is the probability of getting a royal flush but where the cards ordered by rank have alternate color? That is, order the cards as 10,J,Q,K,A and then check to see they have alternate colour. Note in a proper royal flush, it is all the one suit, but we have changed that to alternate colour. So, for example “red 10, black J, red Q, black K, red A” is OK but “red J, black 10, red Q, black K, red A” is not OK because once reordered in rank the alternating colour no longer holds. Note the order in which they are drawn from the pack is not considered.

FIT5047平时班 – Probability

What is the probability that in the sequence of cards, as they are drawn, no rank occurs twice in a row? So ignoring the suit, the following are allowed: A, 10, 4, J, 10 or A, 10, A, 4, A, but the following are not allowed: A, A, 10, 4, A (A repeated in positions 1 and 2), A, 4, 10, 10, J (10 repeated in positions 3 and 4).

Hard question

FIT5047平时班 – Probability

A car manufacturing company has two two lines of production, A and B. Line A manufactures 65% of cars and line B manufactures 35%. 70% of the cars from line A and 85% of the cars from line B are rated standard quality.

1.A car is chosen at random and is found to be of standard quality. What is the probability that it has come from line A?

2. If two cars are randomly chosen from this factory, and one of them is found to be of standard quality. What is the probability that the other one is defective?

A red die, a blue die, and a yellow die (all six-sided) are rolled. We are interested in the probability that the number appearing on the blue die is less than that appearing on the yellow die which is less than that appearing on the red die. (That is, if B (R) [Y] is the number appearing on the blue (red) [yellow] die, then we are interested in $P(B < Y < R)$.)