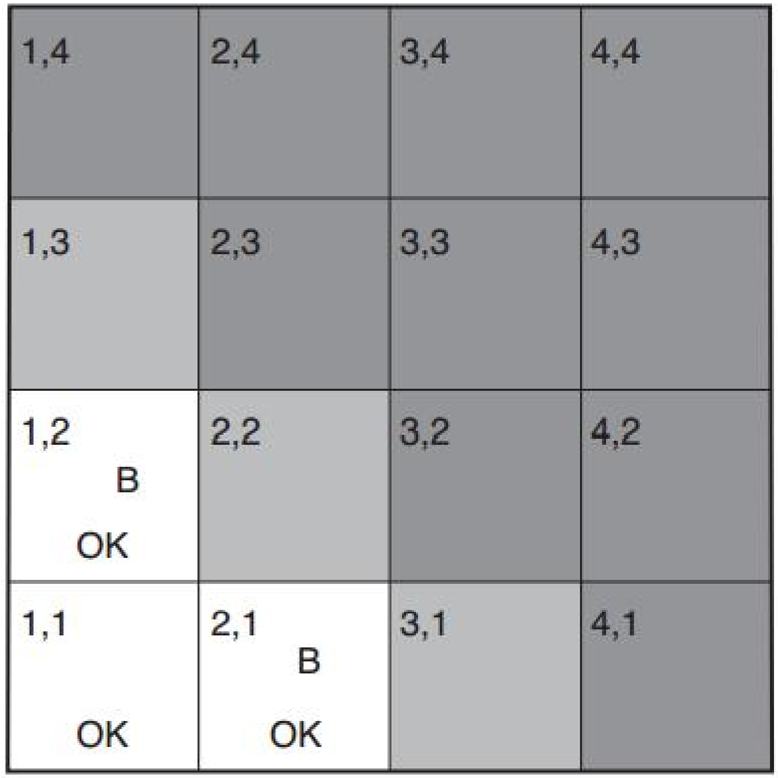
Probabilistic Agents

Quantifying Uncertainty

Why Deal with Uncertainty???



- How can a logical agent decide its next move rationally ???
- Can probabilities help ?

How does uncertainty arise ?

- Partial Observability
- Non-determinism
- A combination of both

Belief State

- Set of all possible worlds that the agent might be in

Uncertainty

Let action A_t = leave for airport t minutes before flight
Will A_t get me there on time?

Problems:
1) partial observability (road state, other drivers' plans, etc.)
2) noisy sensors (traffic reports)
3) uncertainty in action outcomes (flat tire, etc.)
4) immense complexity of modeling and predicting traffic

- Is A_{90} a good plan

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport ...)

Rational Decisions

A rational decision must consider:

- The relative importance of the sub-goals
- Utility theory
- The degree of belief that the sub-goals will be achieved
- Probability theory

Decision theory = probability theory + utility theory :
the agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all possible outcomes of the action"

Probability

Compare the following:

1) Propositional logic:
   "The patient has a cavity"
2) Probabilistic:
   "The probability that the patient has a cavity is 0.8"

1) Is either valid or not, depending on the state of the world
2) Validity depends on the agents perception history, the evidence

Probability - Things you already know

- Sample Space
  - Set of all possible worlds
  - The Greek letter Ω (uppercase omega) is used to refer to the sample space, and ω (lowercase omega) refers to elements of the space, that is, particular possible worlds.
- Probability Model
  - Associates a numerical value for each possible world

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1$$

- Events

Probability

Probabilities are either:

Prior probability (unconditional)

Before any evidence is obtained

Posterior probability (conditional)

After evidence is obtained

Evidence: Already known information

Probability

Notation for unconditional probability for a proposition A:
$P(A)$

Ex: $P(cavity)=0.2$ means:
"the degree of belief for "Cavity" given no extra evidence is 0.2"

Axioms for probabilities:

1. $0 \leq P(A) \leq 1$
2. $P(True)=1, P(False)=0$
3. $P(A \vee B)=P(A)+P(B)-P(A \wedge B)$

Probability

The **axioms** of probability constrain the possible assignments of probabilities to propositions.

An **agent that violates** the axioms will **behave irrationally** in some circumstances

- De Finetti's Theorem

Conditional Probability

The Posterior prob. (conditional prob.) after obtaining evidence:
Notation:
$P(A|B)$ means:
"The probability of A given that all we know is B". Example:
$P(\text{Sunny} \mid \text{Summer}) = 0.65$

Is defined as:

$$\boxed{P(A \mid B) = \frac{P(A \wedge B)}{P(B)}} \quad \text{if } P(B) \neq 0$$

Can be rewritten as the product rule:

$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$

"For A and B to be true,
B has to be true, and A has to be true given B"

Conditional Probability

For the entire random variables:

$$\boxed{\mathbf{P}(A,B) = \mathbf{P}(B)\mathbf{P}(A \mid B)}$$

should be interpreted as a set of equations for all possible values on the random variables A and B.

Example:

$$\boxed{\mathbf{P}(Weather\,, Season\,) = \mathbf{P}(Season\,)\mathbf{P}(Weather \mid Season\,)}$$

Conditional Probability

A general version holds for whole distributions, e.g.,

$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\,\mathbf{P}(Cavity)$

(View as a 4 x 2 set of equations)

Random variable

A random variable has a *domain* of possible values

Each value has a assigned probability between 0 and 1

The values are :

- **Mutually exclusive** (disjoint): (only one of them are true)
- **Complete** (there is always one that is true)

Example: The random variable Weather:

P(Weather=sunny) = 0.7

P(Weather=rain) = 0.2

P(Weather=cloudy) = 0.08

P(Weather=snow) = 0.02

Random Variable

The random variable Weather as a whole is said to have a probability distribution which is a vector (in the discrete case):

P(Weather) = [0.7 0.2 0.08 0.02]

(Notice the bold **P** which is used to denote the prob.distribution)

Random variable

The random variable Weather:

P(Weather=sunny) = 0.7

P(Weather=rain) = 0.2

P(Weather=cloudy) = 0.08

P(Weather=snow) = 0.02

Variable names starts with Uppercase letter

Values are all lowercase

Probability Model

Begin with a set Ω - the sample space
e.g., 6 possible rolls of a die.
$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$

$0 \leq P(\omega) \leq 1$

$\Sigma_{\omega} P(\omega) = 1$

e.g., $P(1) \leq P(2) \leq P(3) \leq P(4) \leq P(5) \leq P(6) \leq 1/6$.

An event A is any subset of Ω

$P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$

E.g., $P(\text{die roll} < 4) = 1/6 + 1/6 + 1/6 = 1/2$

Are mutually exclusive events independent ?

Inference using Full Joint Distributions

The Joint Probability Distribution

Assume that an agent describing the world using the random variables $X_1, X_2, ..X_n$.

The joint probability distribution (or "joint") assigns values for <u>all</u> combinations of values on $X_1, X_2, ..X_n$.
Notation: $\mathbf{P}(X_1, X_2, ..X_n)$ (i.e. **P** bold)

Example of ”Joint” P(Season, Weather)

	Weather			
Season	Sun	Rain	Cloud	Snow
Spring	0.07	0.03	0.10	0.06
Summer	0.13	0.01	0.05	0.01
Autumn	0.05	0.05	0.15	0.03
Winter	0.05	0.01	0.10	0.10

Example: $P(\text{Weather=Sun} \wedge \text{Season=Summer}) = 0.13$

The Joint Probability Distribution – A dentist world

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

The Joint Probability Distribution – A dentist world

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the $Toothache$, $Cavity$, $Catch$ world.

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

The Joint Probability Distribution – A dentist world

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the $Toothache$, $Cavity$, $Catch$ world.

What is P(cavity) ???

Marginalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

This technique is called Marginalization

Marginalization

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z})$$

This technique is called Marginalization

Conditioning

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z})$$

This technique is called Conditioning

Normalization

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

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

$$\begin{aligned}
&\mathbf{P}(Cavity \mid toothache) = \alpha \, \mathbf{P}(Cavity, toothache) \\
&\quad = \alpha \left[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch) \right] \\
&\quad = \alpha \left[\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle \right] = \alpha \, \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle .
\end{aligned}$$

Normalization

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

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

$$\begin{aligned}
&\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity, toothache) \\
&\quad = \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
&\quad = \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] = \alpha\, \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle \,.
\end{aligned}$$

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Inference using Full Joint Probability Distribution

- can answer probabilistic queries for discrete variables
- We can "learn" the cell entries from "examples"

What about the time and space needed to do the inference?

Exponential

Inference using Full Joint Probability Distribution

- can answer probabilistic queries for discrete variables
- We can "learn" the cell entries from "examples"

What about the time and space needed to do the inference?

We now see some techniques with which we can reduce these requirements

Independence

- Let us expand the full joint distribution by adding a fourth variable, Weather. The full joint distribution then becomes **P**(Toothache,Catch,Cavity,Weather)
- Weather has four values
- Entry count increases from 8 to 32
- 4 different tables like this, for each value of Weather

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

Independence

$$P(toothache, catch, cavity, cloudy)$$
$$= P(cloudy \mid toothache, catch, cavity) P(toothache, catch, cavity)$$

Clouds in the sky don't care about your toothache and other dental problems

$$P(cloudy \mid toothache, catch, cavity) = P(cloudy)$$

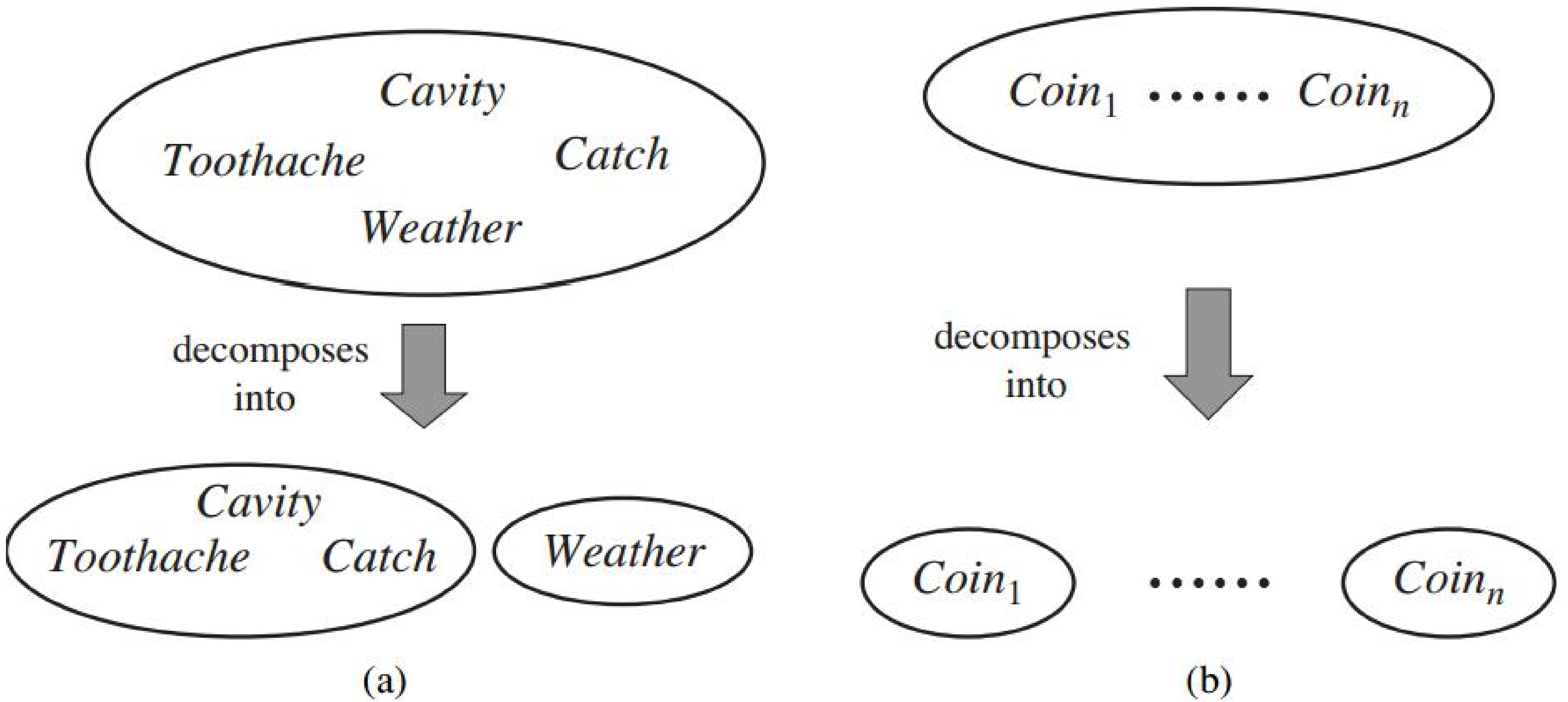
$$P(toothache, catch, cavity, cloudy) = P(cloudy) P(toothache, catch, cavity)$$

Inependence

$$P(toothache, catch, cavity, cloudy) = P(cloudy)P(toothache, catch, cavity)$$

- The 32 element table is now reduced to two tables
  - One containing 8 entries
  - One containing 2 entries

Independence



Inependence

$$\mathbf{P}(X \mid Y) = \mathbf{P}(X) \quad \text{or} \quad \mathbf{P}(Y \mid X) = \mathbf{P}(Y) \quad \text{or} \quad \mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$$

$$P(a \mid b) = P(a) \quad \text{or} \quad P(b \mid a) = P(b) \quad \text{or} \quad P(a \wedge b) = P(a)P(b)$$

Independence

- Independence assertions are usually based on knowledge of the domain
- Can we guess independence between two variables by observations ???

Independence can reduce time and space complexities! Still we can do better!!

Bayes' Rule

The left side of the product rule is symmetric w.r.t B and A:

$$P(A \wedge B) = P(A)P(B \mid A)$$
$$P(A \wedge B) = P(B)P(A \mid B)$$

Equating the two right-hand sides yields Bayes' rule:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

Bayes' Rule

Useful for assessing diagnostic probability from causal probability:

$$P(\text{Cause}|\text{Effect}) = \frac{P(\text{Effect}|\text{Cause})P(\text{Cause})}{P(\text{Effect})}$$

Two distinct relationships

- Causal Relationship
- Diagnostic Relationship

Example of Medical Diagnosis using Bayes' rule

Known facts:

Meningitis causes stiff neck 50% of the time.

The probability of a patient having meningitis (m) is 1/50.000.

The probability of a patient having stiff neck (s) is 1/20.

Query:

What is the probability of meningitis given stiff neck ?

Example of Medical Diagnosis using Bayes' rule

Known facts:

Meningitis causes stiff neck 50% of the time.

The probability of a patient having meningitis (m) is 1/50.000.

The probability of a patient having stiff neck (s) is 1/20.

Query:

What is the probability of meningitis given stiff neck ?

Solution:

$P(s|m)=0.5$

$P(m) = 1/50.000$

$P(s) = 1/20$

Example of Medical Diagnosis using Bayes' rule

Known facts:

Meningitis causes stiff neck 50% of the time.

The probability of a patient having meningitis (m) is 1/50.000.

The probability of a patient having stiff neck (s) is 1/20.

Query:

What is the probability of meningitis given stiff neck ?

Solution:

P(s|m)=0.5

P(m) = 1/50.000

P(s) = 1/20

$$\boxed{P(m \mid s) = \frac{P(s \mid m) \quad P(m)}{P(s)} = \frac{0.5 \bullet 1/50000}{1/20} = 0.0002}$$

Bayes' Rule

In distribution form

$$P(Y|X) = \frac{\mathbf{P}(X|Y)\,\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha \mathbf{P}(X|Y)\,\mathbf{P}(Y)$$

Why prefer causal relationships to diagnostic relationships??

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

- compute the single term P(b | a) in terms of three terms: P(a | b), P(b), and P(a)

- The conditional probability P(*effect* | cause) quantifies the relationship in the **causal** direction, whereas P(cause | *effect*) describes the **diagnostic** direction

- Estimating the probabilities from the evidence is more reliable in this direction

Why prefer causal relationships to diagnostic relationships??

- The conditional probability P(*effect* | cause) quantifies the relationship in the **causal** direction, whereas P(cause | *effect*) describes the **diagnostic** direction

- Estimating the probabilities from the evidence is more reliable in this direction

$$\mathrm{P(m\,|\,s)} = \frac{\mathrm{P(s\,|\,m)} \quad \mathrm{P(m)}}{\mathrm{P(s)}} = \frac{0.5 \bullet 1/50000}{1/20} = 0.0002$$

Combining Evidence

Task: Compute $P(Cavity | Toothache \wedge Catch)$

1. Rewrite using the definition and use the joint. With N evidence variables, the “joint” will be an N dimensional table. It is often impossible to compute probabilities for all entries in the table.

2. Rewrite using Bayes’ rule. This also requires a lot of condprob. to be estimated. Other methods are to prefer.

Combining Evidence

Task: Compute P(Cavity|Toothache ∧ Catch)

2. Rewrite using Bayes' rule. This also requires a lot of condprob. to be estimated. What to do ?

$$\mathbf{P}(Cavity \mid toothache \wedge catch) \\ = \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$

Conditional Independence

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \alpha\,\mathbf{P}(toothache \wedge catch \mid Cavity)\,\mathbf{P}(Cavity)$$

- Assume that, given the value of cavity, toothache and catch are independent of each other

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \alpha\,\mathbf{P}(toothache \mid Cavity)\,\mathbf{P}(catch \mid Cavity)\,\mathbf{P}(Cavity)$$

- The assumption might be true, and can also be false
- However, can lead to very good AI systems

Bayes' Rule and Conditional Independence

P(Cavity|toothache $\wedge$ catch)

= α P(toothache $\wedge$ catch|Cavity) P(Cavity)

= α P(toothache|Cavity) P(catch|Cavity) P(Cavity)

This is an example of a "**naïve" Bayes model**:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$$

Total number of parameters is linear in n

Summary

Probability can be used to reason about uncertainty

Uncertainty arises because of both **laziness** and **ignorance** and it is inescapable in **complex**, **dynamic**, or **inaccessible worlds**.

Probabilities **summarize** the agent's **beliefs**.

Summary

Basic probability statements include **prior probabilities** and **conditional probabilities** over simple and complex propositions

The **full joint probability distribution** specifies the **probability of each** complete as assignment of **values to random variables**. It is usually too large to create or use in its explicit form

The **axioms** of probability constrain the possible assignments of probabilities to propositions. An **agent that violates** the axioms will **behave irrationally** in some circumstances

Summary

Bayes' rule allows unknown **probabilities** to be **computed** from known **conditional probabilities**, usually in the causal direction.

With many pieces of evidence it will in general run into the same scaling problems as does the full joint distribution

Conditional independence brought about by direct causal relationships in the domain might allow the full joint distribution to be factored into smaller, conditional distributions

The **naive Bayes model** assumes the conditional independence of all effect variables, given a single cause variable, and grows linearly with the number of effects

Reference

- 13.1 – 13.6