Machine Intelligence

Lecture 4: Reasoning under uncertainty - probabilities

Thomas Dyhre Nielsen

Aalborg University

Tentative course overview

Topics:

- Introduction
- Search-based methods
- Constrained satisfaction problems
- **Representing domains endowed with uncertainty**
- Bayesian networks
- Machine learning
- Planning
- Multi-agent systems

Certainty in Search

Assumptions for using search for solving planning problems:

- Current state is fully known
- Actions have deterministic effects

Certainty in CSP and Logic

- Possible worlds are possible or impossible (according to given constraints/propositions)
- Propositions are fully known/believed, or unknown
- Generalization: soft constraints – possible worlds are more or less desirable

More realistic scenarios

- Agents do not observe the world perfectly
- Actions have uncertain effects
- Propositions are believed only with a certain confidence

Degrees of Belief for Propositions

In reality, states of knowledge may better be represented by degrees of belief:

$$\begin{aligned}
&\mathit{Bel}(\mathit{light_on} \leftarrow \ \mathit{switch_on} \wedge \mathit{breaker_up}) = 0.7 \\
&\mathit{Bel}(\neg \mathit{umbrella} \rightarrow \mathit{rain}) = 1.5 \\
&\mathit{Bel}(\mathit{umbrella} \rightarrow \mathit{rain}) = 0.2 \\
&\mathit{Bel}(\mathit{global_warming}) = 0.8
\end{aligned}$$

Question: what rules must (rational) degrees of belief obey?

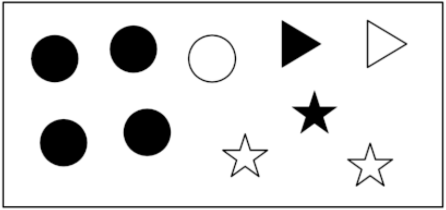
Basic Probability Calculus

Probability theory is built on the foundation of variables and worlds.

Worlds described by the variables:

- Filled: {true, false}
- Shape: {circle, triangle, star}

as well as position.



Probability measures

Ω: set of all possible worlds (for a given, fixed set of variables). A **probability measure over** Ω, is a function P, that assigns **probability values**

$$P(\Omega') \in [0, 1]$$

to subsets $\Omega' \subseteq \Omega$, such that

Axiom 1: $P(\Omega) = 1$.

Axiom 2: if $\Omega_1 \cap \Omega_2 = \emptyset$, then $P(\Omega_1 \cup \Omega_2) = P(\Omega_1) + P(\Omega_2)$.

Simplification for finite Ω

If all variables have a finite domain, then

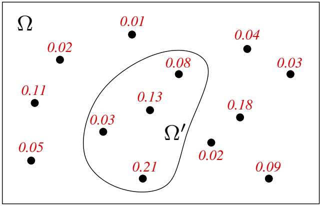
- Ω is finite, and
- a probability distribution is defined by assigning a probability value

$$P(\omega)$$

to each individual possible world $\omega \in \Omega$.

For any $\Omega' \subseteq \Omega$ then

$$P(\Omega') = \sum_{\omega \in \Omega'} P(\omega)$$



$$P(\Omega') = 0.08 + 0.13 + 0.03 + 0.21 = 0.45$$

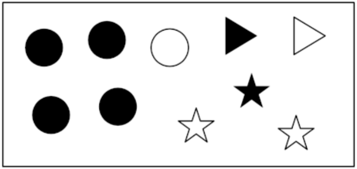
From now on, we will only consider variables with finite domains.

Probabilities of Propositions

A probability distribution over possible worlds defines probabilities for propositions α:

$$\begin{aligned} P(\alpha) &= P(\{\omega \in \Omega \mid \omega \models \alpha\}) \\ &= \sum_{\omega:\alpha \text{ is true in } \omega} P(\omega) \end{aligned}$$

Example



Assume probability for each world is 0.1:

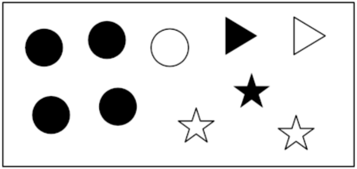
- $P(\textit{Shape} = \textit{circle}) =$

Probabilities of Propositions

A probability distribution over possible worlds defines probabilities for propositions α:

$$P(\alpha) = P(\{\omega \in \Omega \mid \omega \models \alpha\}) = \sum_{\omega:\alpha \text{ is true in } \omega} P(\omega)$$

Example



Assume probability for each world is 0.1:

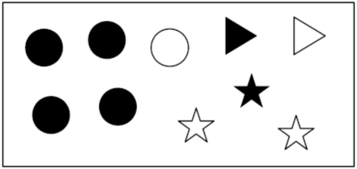
- $P(\mathit{Shape} = \mathit{circle}) = 0.5$
- $P(\mathit{Filled} = \mathit{false}) =$

Probabilities of Propositions

A probability distribution over possible worlds defines probabilities for propositions α:

$$
\begin{aligned}
P(\alpha) &= P(\{\omega \in \Omega \mid \omega \models \alpha\}) \\
&= \sum_{\omega:\alpha \text{ is true in } \omega} P(\omega)
\end{aligned}
$$

Example



Assume probability for each world is 0.1:

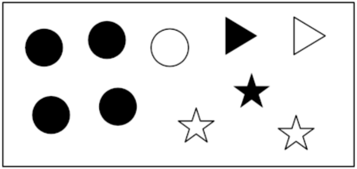
- $P(\mathit{Shape} = \mathit{circle}) = 0.5$
- $P(\mathit{Filled} = \mathit{false}) = 0.4$
- $P(\mathit{Shape} = c \wedge \mathit{Filled} = f) =$

Probabilities of Propositions

A probability distribution over possible worlds defines probabilities for propositions α:

$$\begin{aligned} P(\alpha) &= P(\{\omega \in \Omega \mid \omega \models \alpha\}) \\ &= \sum_{\omega : \alpha \text{ is true in } \omega} P(\omega) \end{aligned}$$

Example



Assume probability for each world is 0.1:

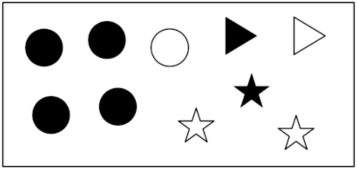
- $P(\textit{Shape} = \textit{circle}) = 0.5$
- $P(\textit{Filled} = \textit{false}) = 0.4$
- $P(\textit{Shape} = \textit{c} \land \textit{Filled} = \textit{f}) = 0.1$

Probabilities of Propositions

A probability distribution over possible worlds defines probabilities for propositions α:

$$P(\alpha) = P(\{\omega \in \Omega \mid \omega \models \alpha\}) = \sum_{\omega:\alpha \text{ is true in } \omega} P(\omega)$$

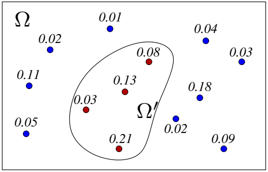
Example



Assume probability for each world is 0.1:

- $P(\textit{Shape} = \textit{circle}) = 0.5$
- $P(\textit{Filled} = \textit{false}) = 0.4$
- $P(\textit{Shape} = \textit{c} \wedge \textit{Filled} = \textit{f}) = 0.1$

Another example



$$P(\textit{Color} = \textit{red}) = 0.08 + 0.13 + 0.03 + 0.21 = 0.45$$

Axiom

If $\mathcal{A}$ and $\mathcal{B}$ are disjoint, then $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$.

Example

Consider a deck with 52 cards. If $\mathcal{A} = \{2, 3, 4, 5\}$ and $\mathcal{B} = \{7, 8\}$, then

$$P(\mathcal{A} \cup \mathcal{B}) =$$

Basic probability axioms

Axiom

If $\mathcal{A}$ and $\mathcal{B}$ are disjoint, then $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$.

Example

Consider a deck with 52 cards. If $\mathcal{A} = \{2, 3, 4, 5\}$ and $\mathcal{B} = \{7, 8\}$, then

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) = \frac{4}{13} + \frac{2}{13} = \frac{6}{13}.$$

Axiom

If $\mathcal{A}$ and $\mathcal{B}$ are disjoint, then $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$.

Example

Consider a deck with 52 cards. If $\mathcal{A} = \{2, 3, 4, 5\}$ and $\mathcal{B} = \{7, 8\}$, then

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) = \frac{4}{13} + \frac{2}{13} = \frac{6}{13}.$$

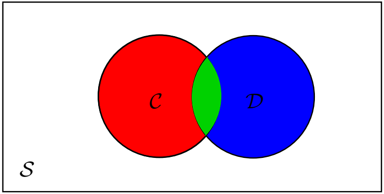
More generally

If $\mathcal{C}$ and $\mathcal{D}$ are not disjoint, then $P(\mathcal{C} \cup \mathcal{D}) = P(\mathcal{C}) + P(\mathcal{D}) - P(\mathcal{C} \cap \mathcal{D})$.

Example

If $\mathcal{C} = \{2, 3, 4, 5\}$ and $\mathcal{D} = \{\spadesuit\}$, then

$$P(\mathcal{C} \cup \mathcal{D}) =$$



Axiom

If $\mathcal{A}$ and $\mathcal{B}$ are disjoint, then $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$.

Example

Consider a deck with 52 cards. If $\mathcal{A} = \{2, 3, 4, 5\}$ and $\mathcal{B} = \{7, 8\}$, then

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) = \frac{4}{13} + \frac{2}{13} = \frac{6}{13}.$$

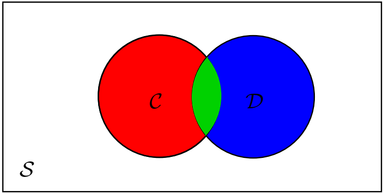
More generally

If $\mathcal{C}$ and $\mathcal{D}$ are not disjoint, then $P(\mathcal{C} \cup \mathcal{D}) = P(\mathcal{C}) + P(\mathcal{D}) - P(\mathcal{C} \cap \mathcal{D})$.

Example

If $\mathcal{C} = \{2, 3, 4, 5\}$ and $\mathcal{D} = \{\spadesuit\}$, then

$$P(\mathcal{C} \cup \mathcal{D}) = \frac{4}{13} + \frac{1}{4} - \frac{4}{52} = \frac{25}{52}.$$



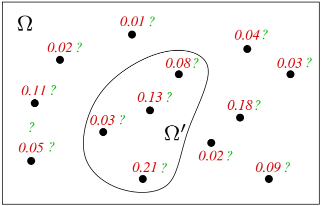
Probability Updating

- Given new information (**evidence**), degrees of belief change.
- Evidence can consist of:
  - learning that a certain proposition p is true ("*switch_up*")
  - measuring the value of some variable ("room_ai = 0.2.90")
  - obtaining partial information on the value of some variable ("room_ai $\neq$ 0.2.90")
  - ...
- In all cases: evidence can be represented as the set of possible world Ω' not ruled out by the observation.

How should the probabilities be updated, when I observe Ω'?:

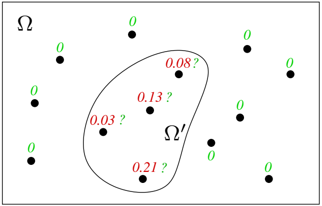
- Given new information (**evidence**), degrees of belief change.
- Evidence can consist of:
  - learning that a certain proposition p is true ("*switch_up*")
  - measuring the value of some variable ("room_ai = 0.2.90")
  - obtaining partial information on the value of some variable ("room_ai $\neq$ 0.2.90")
  - ...
- In all cases: evidence can be represented as the set of possible world Ω' not ruled out by the observation.

How should the probabilities be updated, when I observe Ω'?:



- Given new information (**evidence**), degrees of belief change.
- Evidence can consist of:
  - learning that a certain proposition p is true ("*switch_up*")
  - measuring the value of some variable ("room_ai = 0.2.90")
  - obtaining partial information on the value of some variable ("room_ai $\neq$ 0.2.90")
  - ...
- In all cases: evidence can be represented as the set of possible world Ω' not ruled out by the observation.

How should the probabilities be updated, when I observe Ω'?:

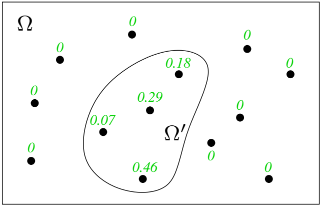


- worlds that are not consistent with evidence have probability 0

Probability Updating

- Given new information (**evidence**), degrees of belief change.
- Evidence can consist of:
  - learning that a certain proposition p is true ("*switch_up*")
  - measuring the value of some variable ("room_ai = 0.2.90")
  - obtaining partial information on the value of some variable ("room_ai $\neq$ 0.2.90")
  - ...
- In all cases: evidence can be represented as the set of possible world Ω' not ruled out by the observation.

How should the probabilities be updated, when I observe Ω'?:



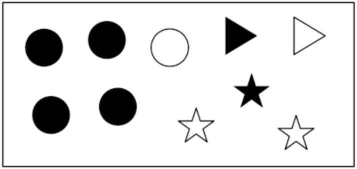
- worlds that are not consistent with evidence have probability 0
- the probabilities of worlds consistent with evidence are proportional to their probability before observation, and they must sum to 1

Definition

The **conditional probability** of proposition p given e is

$$P(p \mid e) = \frac{P(p \wedge e)}{P(e)}$$

Example



(probability for each world is 0.1)

$$P(\textit{S=circ.} \mid \textit{Fill} = f) = \frac{P(\textit{S=circ.} \wedge \textit{Fill} = f)}{P(\textit{Fill} = f)}$$

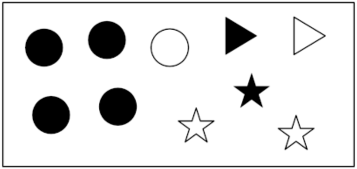
$$= \frac{0.1}{0.4} = 0.25$$

Definition

The **conditional probability** of proposition p given e is

$$P(p \mid e) = \frac{P(p \wedge e)}{P(e)}$$

Example



(probability for each world is 0.1)

$$P(\textit{S=circ.} \mid \textit{Fill} = f) = \frac{P(\textit{S=circ.} \wedge \textit{Fill} = f)}{P(\textit{Fill} = f)}$$

$$= \frac{0.1}{0.4} = 0.25$$

What is the probability of $P(\textit{S=star} \mid \textit{Fill} = f)$?

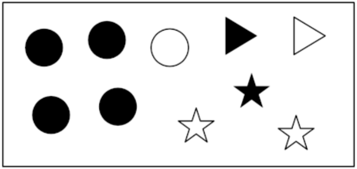
Conditional probability

Definition

The **conditional probability** of proposition p given e is

$$P(p \mid e) = \frac{P(p \wedge e)}{P(e)}$$

Example

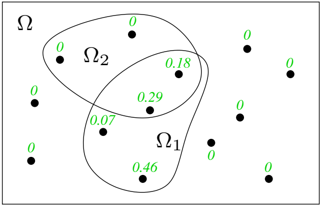


(probability for each world is 0.1)

$$P(\textit{S=circ.} \mid \textit{Fill} = f) = \frac{P(\textit{S=circ.} \wedge \textit{Fill} = f)}{P(\textit{Fill} = f)} = \frac{0.1}{0.4} = 0.25$$

What is the probability of $P(\textit{S=star} \mid \textit{Fill} = f)$?

Another example



- e and p are represented by possible worlds Ω_1 and Ω_2
- division by $P(\Omega_1)$ already in green numbers

$$P(\Omega_2 \mid \Omega_1) = 0.18 + 0.29$$

Bayes rule

For propositions p, e:

$$P(p \mid e) = \frac{P(e \wedge p)}{P(e)} = \frac{P(e \mid p)P(p)}{P(e)}$$

Two important rules

Bayes rule

For propositions p, e:

$$P(p \mid e) = \frac{P(e \wedge p)}{P(e)} = \frac{P(e \mid p)P(p)}{P(e)}$$

Example

A doctor observes symptoms and wishes to find the probability of a disease:

$$P(\textit{disease} \mid \textit{symp.}) = \frac{P(\textit{symp.} \mid \textit{disease})P(\textit{disease})}{P(\textit{symp.})}$$

Two important rules

Bayes rule

For propositions p, e:

$$P(p \mid e) = \frac{P(e \wedge p)}{P(e)} = \frac{P(e \mid p)P(p)}{P(e)} = \frac{P(e \mid p)P(p)}{P(e \wedge p) + P(e \wedge \neg p)}$$

Example

A doctor observes symptoms and wishes to find the probability of a disease:

$$P(\textit{disease} \mid \textit{symp.}) = \frac{P(\textit{symp.} \mid \textit{disease})P(\textit{disease})}{P(\textit{symp.})}$$

Bayes rule

For propositions p, e:

$$P(p \mid e) = \frac{P(e \wedge p)}{P(e)} = \frac{P(e \mid p)P(p)}{P(e)} = \frac{P(e \mid p)P(p)}{P(e \wedge p) + P(e \wedge \neg p)}$$

Chain rule

For propositions $p_1, \ldots, p_n$:

$$P(p_1 \wedge \ldots \wedge p_n) = P(p_1)P(p_2 \mid p_1) \cdots P(p_i \mid p_1 \wedge \ldots \wedge p_{i-1}) \cdots P(p_n \mid p_1 \wedge \ldots \wedge p_{n-1})$$

Both rules are immediate consequences of the definition of conditional probability!

Random Variables

Variables defining possible worlds on which probabilities are defined are called **random variables**.

Distributions

For a random variable A, and $a \in D_A$ we have the probability

$$P(A = a) = P(\{\omega \in \Omega \mid A = a \text{ in } \omega\})$$

The **probability distribution of** A is the function on D_A that maps a to $P(A = a)$. The distribution of A is denoted

$$P(A)$$

Joint Distributions

Extension to several random variables:

$$P(A_1, \ldots, A_k)$$

is the **joint distribution of** $A_1, \ldots, A_k$. The joint distribution maps tuples $(a_1, \ldots, a_k)$ with $a_i \in D_{A_i}$ to the probability

$$P(A_1 = a_1, \ldots, A_k = a_k)$$

With random variables instead of propositions, the chain rule becomes:

$$P(A_1, \ldots, A_n) = P(A_1)P(A_2 \mid A_1) \cdots P(A_i \mid A_1, \ldots, A_{i-1}) \cdots P(A_n \mid A_1, \ldots, A_{n-1})$$

Note: each $P(p_i \mid p_1 \wedge \ldots \wedge p_{i-1})$ was a *number*. Each $P(A_i \mid A_1, \ldots, A_{i-1})$ is a *function* on tuples $(a_1, \ldots, a_i)$.

Bayes' rule for variables

Bayes' rule can also be written for variables:

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

Bayes' rule for variables

Bayes' rule can also be written for variables:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(A,B)}$$

Bayes' rule for variables

Bayes' rule can also be written for variables:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(A,B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

Bayes' rule for variables

Bayes' rule can also be written for variables:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(A,B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

Example

Consider the variables

- Temp : $\mathrm{sp}(\text{Temp}) = \{l, m, h\}$
- Sensor : $\mathrm{sp}(\text{Sensor}) = \{l, m, h\}$

$P(\text{Temp}) = (0.1, 0.6, 0.3)$

$P(\text{Sensor}|\text{Temp}) =$

		Temp		
		l	m	h
Sensor	l	0.8	0.1	0.05
	m	0.15	0.8	0.1
	h	0.05	0.1	0.85

Bayes' rule for variables

Bayes' rule can also be written for variables:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(A,B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

Example

Consider the variables

- Temp : $\mathrm{sp}(\text{Temp}) = \{l, m, h\}$
- Sensor : $\mathrm{sp}(\text{Sensor}) = \{l, m, h\}$

$P(\text{Temp}) = (0.1, 0.6, 0.3)$

$P(\text{Sensor}|\text{Temp}) =$

		Temp		
		l	m	h
Sensor	l	0.8	0.1	0.05
	~~m~~	~~0.15~~	~~0.8~~	~~0.1~~
	~~h~~	~~0.05~~	~~0.1~~	~~0.85~~

Assume we observe $S =$ low:

$P(T|\text{low}) = \frac{P(\text{low}|T)P(T)}{\sum_T P(\text{low}|T)P(T)} =$

Bayes' rule can also be written for variables:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(A,B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

Example

Consider the variables

- Temp : $\mathrm{sp}(\text{Temp}) = \{l, m, h\}$
- Sensor : $\mathrm{sp}(\text{Sensor}) = \{l, m, h\}$

$P(\text{Temp}) = (0.1, 0.6, 0.3)$

$P(\text{Sensor}|\text{Temp}) =$

		Temp		
		l	m	h
Sensor	l	0.8	0.1	0.05
	m	0.15	0.8	0.1
	h	0.05	0.1	0.85

Assume we observe $S = \text{low}$:

$P(T|\text{low}) = \frac{P(\text{low}|T)P(T)}{\sum_T P(\text{low}|T)P(T)} =$

$P(\text{low}|T)P(T) =$

	Temp				
	l	m	h		$P(S = \text{low})$
$S = \text{low}$	0.08	0.06	0.015	$\rightarrow$	0.155

Bayes' rule for variables

Bayes' rule can also be written for variables:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(A,B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

Example

Consider the variables

- Temp : $\mathrm{sp}(\text{Temp}) = \{l, m, h\}$
- Sensor : $\mathrm{sp}(\text{Sensor}) = \{l, m, h\}$

$P(\text{Temp}) = (0.1, 0.6, 0.3)$

$P(\text{Sensor}|\text{Temp}) =$

		Temp		
		l	m	h
Sensor	l	0.8	0.1	0.05
	~~m~~	$\cancel{0.15}$	$\cancel{0.8}$	$\cancel{0.1}$
	~~h~~	$\cancel{0.05}$	$\cancel{0.1}$	$\cancel{0.85}$

Assume we observe $S =$ low:

$P(T|\text{low}) = \frac{P(\text{low}|T)P(T)}{\sum_T P(\text{low}|T)P(T)} =$

	Temp		
	l	m	h
$S =$ low	0.08 / 0.155	0.06 / 0.155	0.015 / 0.155

$P(\text{low}|T)P(T) =$

	Temp				$P(S = \text{low})$
	l	m	h		
$S =$ low	0.08	0.06	0.015	$\rightarrow$	0.155

Bayes' rule can also be written for variables:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(A,B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

Example

Consider the variables

- Temp : $\mathrm{sp}(\text{Temp}) = \{l, m, h\}$
- Sensor : $\mathrm{sp}(\text{Sensor}) = \{l, m, h\}$

$P(\text{Temp}) = (0.1, 0.6, 0.3)$

$P(\text{Sensor}|\text{Temp}) =$

		Temp		
		l	m	h
Sensor	l	0.8	0.1	0.05
	$\cancel{m}$	$\cancel{0.15}$	$\cancel{0.8}$	$\cancel{0.1}$
	$\cancel{h}$	$\cancel{0.05}$	$\cancel{0.1}$	$\cancel{0.85}$

Assume we observe $S = \text{low}$:

$P(T|\text{low}) = \frac{P(\text{low}|T)P(T)}{\sum_T P(\text{low}|T)P(T)} =$

	Temp		
	l	m	h
$S = \text{low}$	0.52	0.39	0.09

$P(\text{low}|T)P(T) =$

	Temp				
	l	m	h		$P(S = \text{low})$
$S = \text{low}$	0.08	0.06	0.015	$\rightarrow$	0.155

Independence

Results for Bayern Munich and SC Freiburg in seasons 2001/02 and 2003/04. (Not counting the matches Munich vs. Freiburg):

$$D_{\textit{Munich})} = D_{\textit{Freiburg}} = \{\textit{Win, Draw, Loss}\}$$

	2001/02
Munich:	`LWDWWWWWWWWLDLDLDLWLDWWWDWDDWWWW`
Freiburg:	`WLLDDWLDWDWLLLDDLWDDLLDLLLLLLWLW`
	2003/04
Munich:	`WDWWLDWWDWLWWDDWDWLWWWDDWWWLWWLL`
Freiburg:	`LDDWDWLWLLLWWLWLWLLDWLDDWDLLLWLD`

Summary:

	Freiburg			
Munich	W	D	L	
W	12	9	15	36
D	3	4	9	16
L	6	4	2	12
	21	17	26	

The joint distribution of *Munich* and *Freiburg*:

P(*Munich*,*Freiburg*):

Munich	Freiburg W	D	L	P(*Munich*)
W	.1875	.1406	.2344	.5625
D	.0468	.0625	.1406	.25
L	.0937	.0625	.0312	.1875
P(*Freiburg*)	.3281	.2656	.4062	

The joint distribution of *Munich* and *Freiburg*:

$P($*Munich,Freiburg*$)$:

Munich	*Freiburg* W	D	L	$P($*Munich*$)$
W	.1875 .571	.1406 .529	.2344 .577	.5625
D	.0468 .143	.0625 .235	.1406 .346	.25
L	.0937 .285	.0625 .235	.0312 .077	.1875
$P($*Freiburg*$)$	.3281	.2656	.4062	

Conditional distribution: $P($*Munich* $\mid$ *Freiburg*$)$

The joint distribution of *Munich* and *Freiburg*:

P(*Munich,Freiburg*):

Munich	*Freiburg* W	D	L	P(*Munich*)
W	.1875	.1406	.2344	.5625
	.571	.529	.577	
D	.0468	.0625	.1406	.25
	.143	.235	.346	
L	.0937	.0625	.0312	.1875
	.285	.235	.077	
P(*Freiburg*)	.3281	.2656	.4062	

Conditional distribution: P(*Munich* | *Freiburg*)

We have (almost):

$$P(\textit{Munich} \mid \textit{Freiburg}) = P(\textit{Munich})$$

The variables *Munich* and *Freiburg* are **independent**.

Definition of Independence

The variables $A_1, \ldots, A_k$ and $B_1, \ldots, B_m$ are **independent** if

$$P(A_1, \ldots, A_k \mid B_1, \ldots, B_m) = P(A_1, \ldots, A_k)$$

This is equivalent to:

$$P(B_1, \ldots, B_m \mid A_1, \ldots, A_k) = P(B_1, \ldots, B_m)$$

and also to:

$$P(A_1, \ldots, A_k, B_1, \ldots, B_m) = P(A_1, \ldots, A_k) \cdot P(B_1, \ldots, B_m)$$

Compact Specifications by Independence

Independence properties can greatly simplify the specification of a joint distribution:

	$F =$			
$M =$	W	D	L	$P(M)$
W				.5625
D	M and F are independent			.25
L				.1875
$P(F)$	.3281	.2656	.4062	

The probability for each possible world then is defined, e.g.

$$P(M = D, F = L) = 0.25 \cdot 0.4062 = 0.10155$$

Example

Joint distribution for variables

Sex :	$D_{\textit{Sex}} = \{$*male,female*$\}$
Hair length :	$D_{\textit{Hair length}} = \{$*long, short*$\}$
Height :	$D_{\textit{Height}} = \{$*tall, medium*$\}$

	Sex			
	male		*female*	
	Hair length		*Hair length*	
Height	*long*	*short*	*long*	*short*
tall	0.06	0.24	0.07	0.03
medium	0.04	0.16	0.28	0.12

P(*Hair length*, *Height*) P(*Height*), P(*Height* | *Hair length*):

	Hair length		
Height	*long*	*short*	
tall	0.13	0.27	0.4
	0.289	0.49	
medium	0.32	0.28	0.6
	0.711	0.51	

$\rightsquigarrow$ *Hair length* and *Height* are not independent.

Example Continued

$P($*Hair length*, *Height* $\mid$ *Sex* $=$ *female*$)$, $P($*Height* $\mid$ *Sex* $=$ *female*$)$,
$P($*Height* $\mid$ *Hair length*, *Sex* $=$ *female*$)$:

Height	*Hair length* long	*Hair length* short	
tall	0.14	0.06	0.2
	0.2	0.2	
medium	0.56	0.24	0.8
	0.8	0.8	

$\rightsquigarrow$ *Hair length* and *Height* are independent given *Sex*=*female*.

Also: *Hair length* and *Height* are independent given *Sex*=*male*.

$\rightsquigarrow$ *Hair length* and *Height* are independent given *Sex*.

Definition of Conditional Independence

The variables $A_1, \ldots, A_n$ are **conditionally independent** of the variables $B_1, \ldots, B_m$ **given** $C_1, \ldots, C_k$, if

$$P(A_1, \ldots, A_n \mid B_1, \ldots, B_m, C_1, \ldots, C_k) = P(A_1, \ldots, A_n \mid C_1, \ldots, C_k)$$

Agenda: use conditional independence to facillitate specification of probability distributions on complex state spaces