USC Viterbi School of Engineering

Probability and Bayesian Inference

[Resources: 6.825 Techniques in Artificial Intelligence, MIT OCW]

CSCI 545 Introduction to Robotics
Instructor: Stefanos Nikolaidis

Motivation

- Why do we need probability?
  - To understand how likely an outcome is!

Motivation

Motivation

Is it deterministic or non-deterministic?

Motivation

Is it deterministic or non-deterministic?

Probability Theory

- It is a logic; a language that speaks about likelihood of events.

- We can start with a universe of atomic events: things that could happen or ways that the world could be

- Then, a probability distribution is a function that maps events into a range between zero and one: P: events→[0, 1]

Probability Theory Axioms

- P(True) = 1 = P(U)

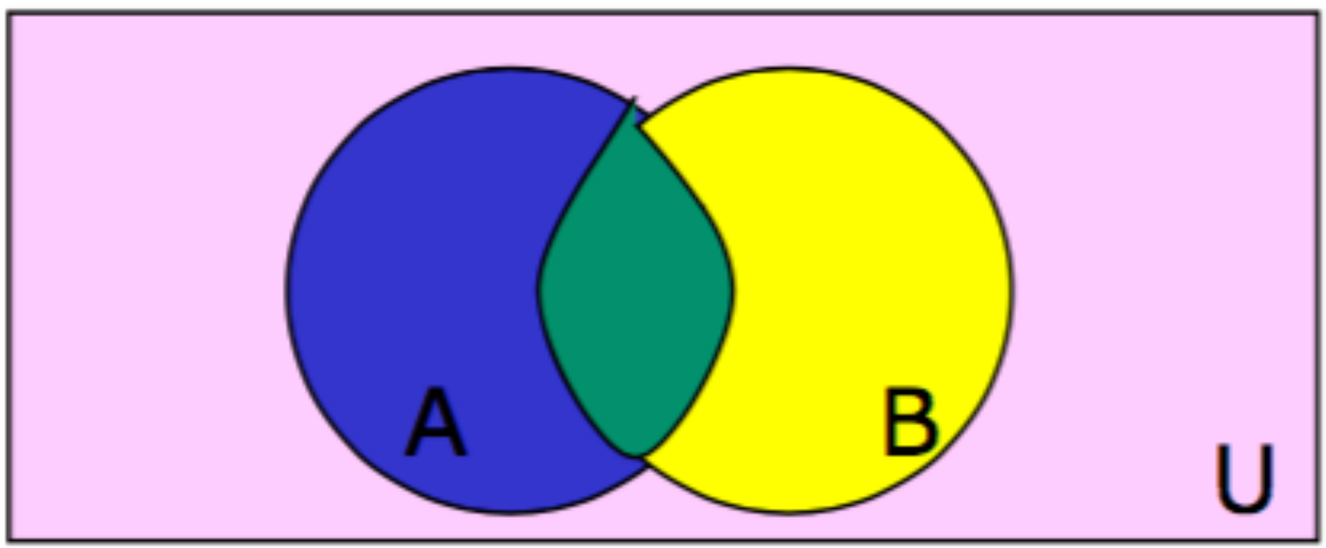
The probability of True is one. It is the probability of the universe, the probability that something in this realm of discussion that we have available to use is one.

- P(False) = 0 = P(null)

The probability of False is 0. In terms of atomic events, False is the empty set.

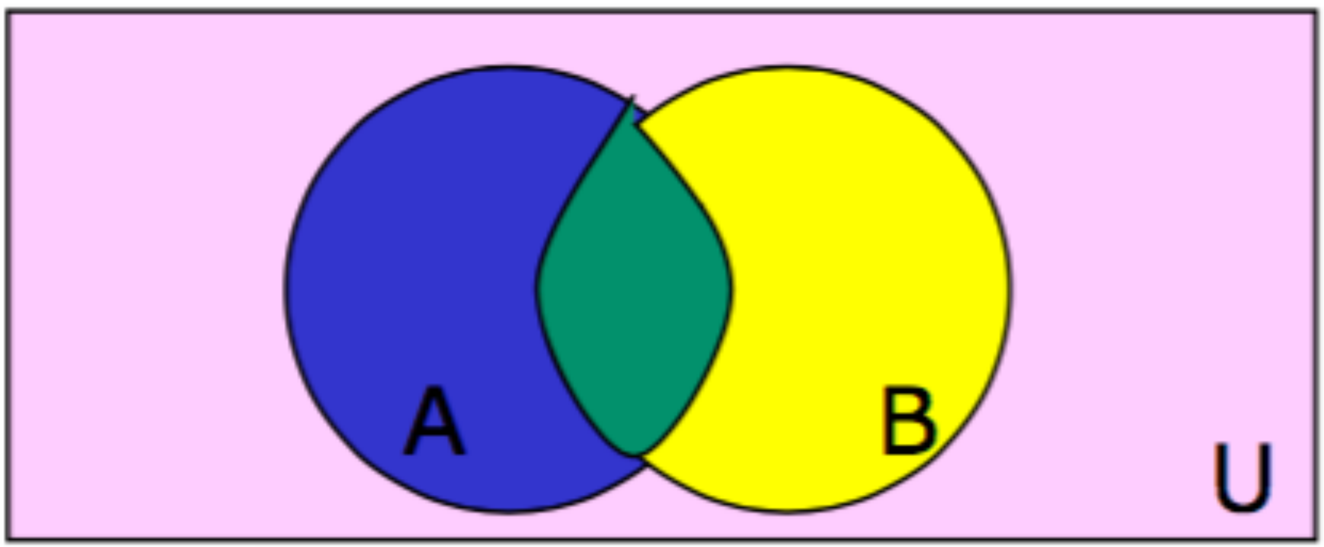
Probability Theory Axioms

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Probability Theory Axioms

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Random Variables

- A random variable: it can be considered as a mapping from a discrete domain to a range from 0 to 1.

- It sums to 1 over the domain

- Example RV: Raining
  - Domain: True / False
  - Raining (True) = 0.2

Random Variables

- Raining (True) = 0.2
- P(Raining = True) = 0.2
- P(Raining = False) = ?

Random Variables

- Raining (True) = 0.2
- P(Raining = True) = 0.2
- P(Raining = False) = **0.8**


Ocean
Ocean
Ocean
MoonStar
SARDINES
kinova

Multiple Random Variables

	Not intervene	**Intervene**
Trusting	0.4	0.1
Not trusting	0.2	0.3

Multiple Random Variables

	Not intervene	**Intervene**
Trusting	0.4	0.1
Not trusting	0.2	0.3

P(Trusting) = 0.5
P(Not Trusting) = 0.5

Multiple Random Variables

	Not intervene	Intervene
Trusting	0.4	0.1
Not trusting	0.2	0.3

P(Not intervene) = ?

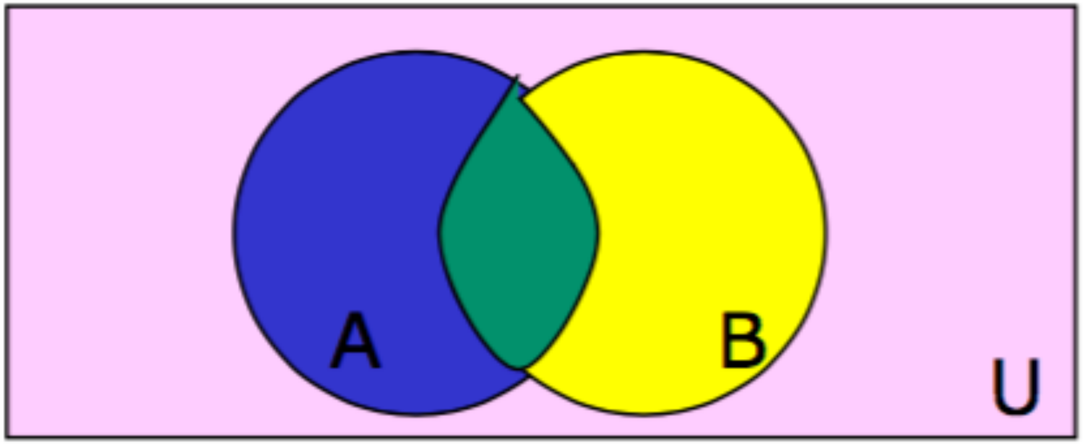
Multiple Random Variables

	Not intervene	Intervene
Trusting	0.4	0.1
Not trusting	0.2	0.3

P(Not intervene) = 0.6

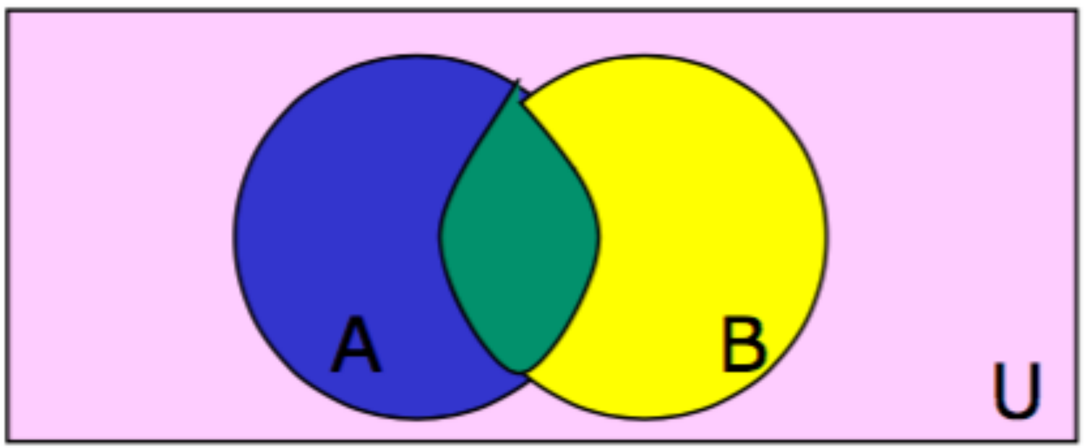
Conditional Probability

- $P(A \mid B)$?



Conditional Probability

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



We only consider the part of the world, where B is True.

Conditional Probability

	Not intervene	**Intervene**
Trusting	0.4	0.1
Not trusting	0.2	0.3

$$P(\text{Not-Intervene}|\text{Trusting}) =$$

Conditional Probability

	Not intervene	**Intervene**
Trusting	0.4	0.1
Not trusting	0.2	0.3

$$P(\text{Not-Intervene}|\text{Trusting}) = \frac{0.4}{0.5} = 0.8$$

Posterior probability

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

$$P(\mathrm{State}|\mathrm{Observation}) = \frac{P(\mathrm{Observation}|\mathrm{State})P(\mathrm{State})}{P(\mathrm{Observation})}$$

Posterior probability

$$P(A) = P(A|B)P(B) + P(A|-B)P(-B)$$

$$P(A) = P(A \cap B) + P(A \cap -B)$$

Posterior probability

	Not intervene	Intervene
Trusting	0.4	0.1
Not trusting	0.2	0.3

$$P(Trusting|Intervene) = \frac{P(Intervene|Trusting)P(Trusting)}{P(Intervene)}$$

Posterior probability

	Not intervene	Intervene
Trusting	0.4	0.1
Not trusting	0.2	0.3

$$P(Trusting|Intervene) = \frac{0.2 * 0.5}{0.4} = 0.25$$

Common Reasoning

- Start with a prior about whether human is trusting or not

- Observe evidence, e.g., human intervening

- Update belief on human trust

Computing Values

- How can we compute P(Intervene | Trusting)?

Computing Values

- How can we compute P(Intervene | Trusting)?

- Ask people about their trust in the robot then examine their behavior .

- Learning conditional probabilities with causal structure is easier.

Learning Causal Structures

- We let running a user study with graduate students
- We then run a second study with a different age group
- Will P(trusting | intervene) be the same?

Learning Causal Structures

- We let running a user study with graduate students
- We then run a second study with a different age group
- Will P(trusting | intervene) be the same?
- Will P(trusting) be the same?

Learning Causal Structures

- We let running a user study with graduate students
- We then run a second study with a different age group
- Will P(trusting | intervene) be the same?
- Will P(trusting) be the same?
- P(intervene | trusting) does not depend on the prior trust in the robot!

Independent Events

- Knowing that B is true doesn't give us any more information about A.

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

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

Independent Events

- Are temperature and time of the day independent?
- Are the events “intervene” and “trusting” independent?

Independent Events

- Are temperature and time of the day independent?
- Are the events “intervene” and “trusting” independent?
- What about “intervene” and “sad”?

Modeling Physical Phenomena (and human behaviors)

- We want the simplest model that explains the observations

Conditional Independence

- A and B are conditionally independent given C, iff the probability of A given B and C is equal to the probability of A given C:

$$P(A|B,C) = P(A|C)$$

Conditional Independence

- Learning about one's sadness does not say much about whether they will intervene

$$P(intervene|trusting, sad) = P(intervene|trusting)$$

State

- A *state* captures all the information that the robot needs o make decisions. It may change over time such as its location.

- Example: what is the state of an object dropping?

State in Human-Robot Interaction

- Example: Say that before going for the glass, the robot goes for the can the human intervenes. Would this provide us with information about whether the person will intervene when the robot goes for the glass?

- What if we can measure their trust right before the robot goes for the glass?

Multiple Observations

- The robot attempts to pick up the can, then the glass, the person intervenes twice.

$$P(trusting|I_c, I_g) = \frac{P(I_c, I_g|trusting)P(trusting)}{P(I_c, I_g)}$$

- How can we compute $P(I_c, I_g|trusting)$?

Multiple Observations

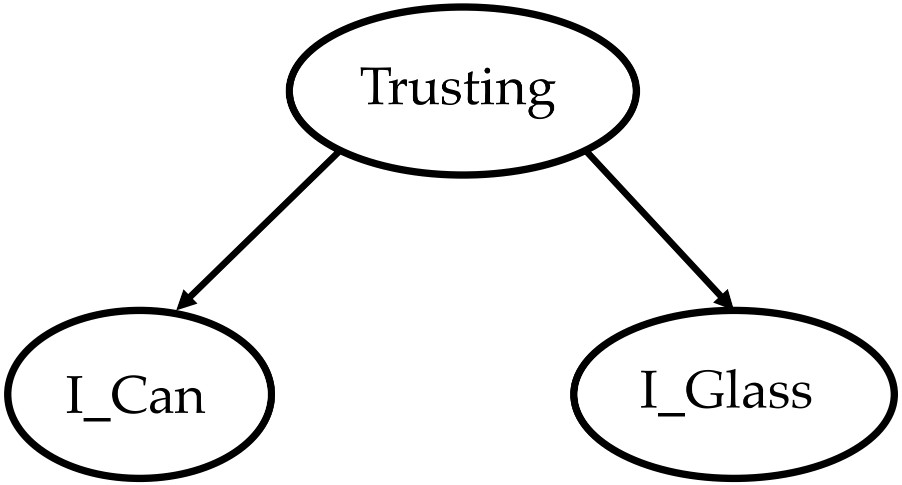
- The robot attempts to pick up the can, then the glass, the person intervenes twice.

$$P(trusting|I_c, I_g) = \frac{P(I_c, I_g|trusting)P(trusting)}{P(I_c, I_g)}$$

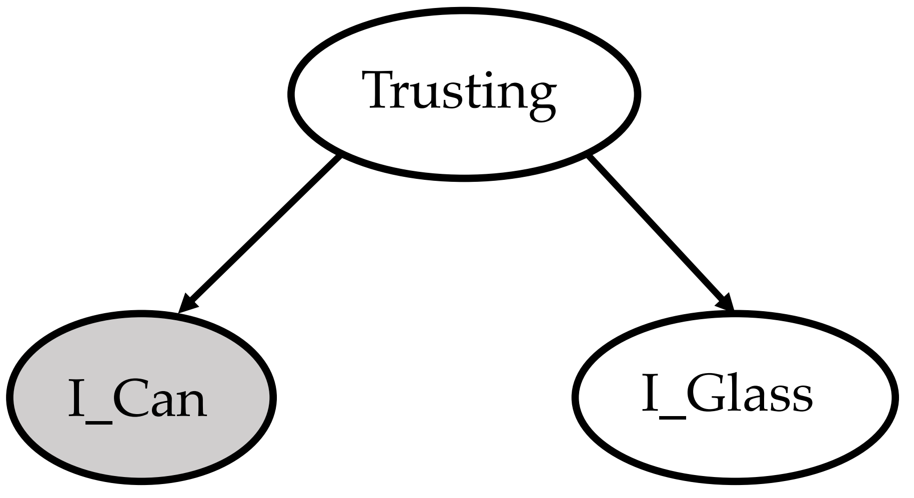
- How can we compute $P(I_c, I_g|trusting)$?

$$P(I_c, I_g|trusting) = P(I_c|trusting)P(I_g|trusting)$$

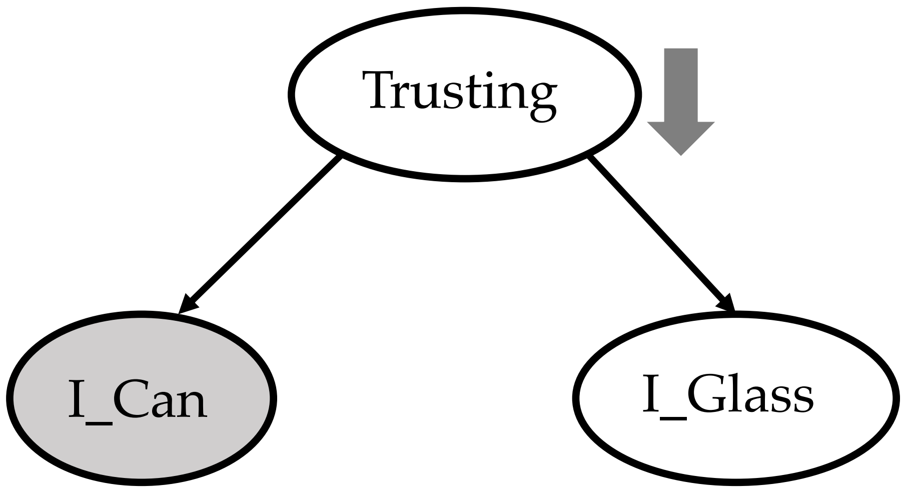
Bayesian Networks



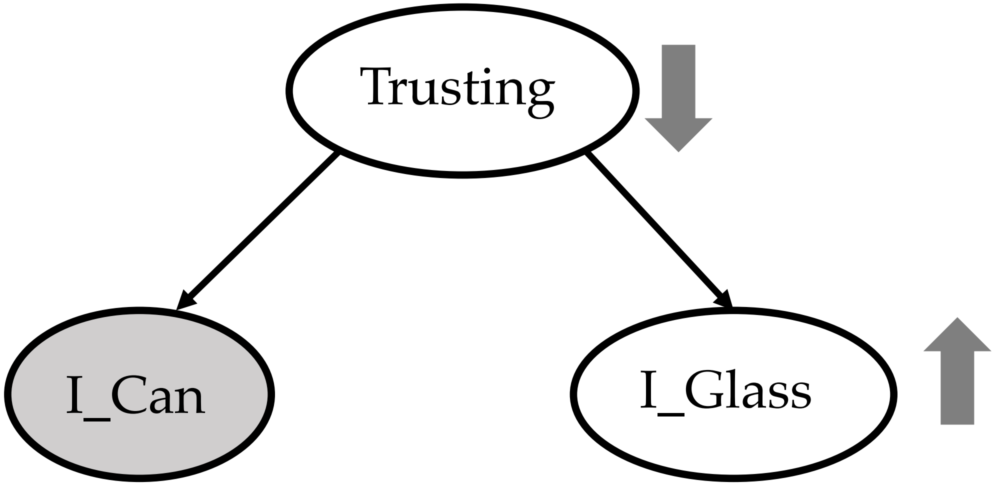
Bayesian Networks



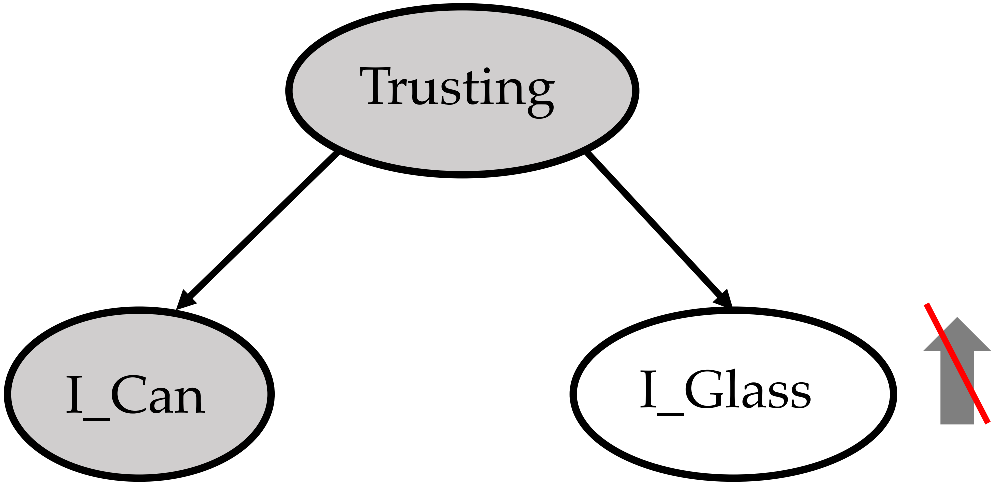
Bayesian Networks



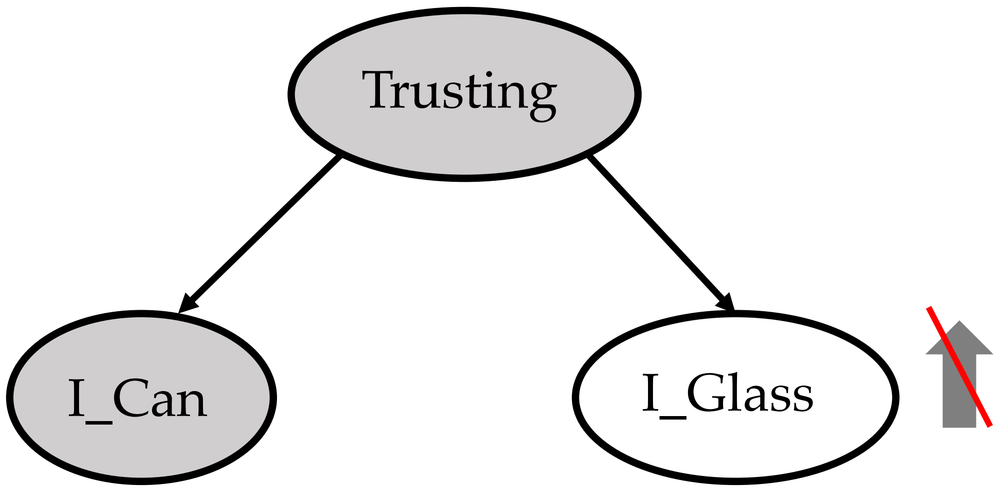
Bayesian Networks



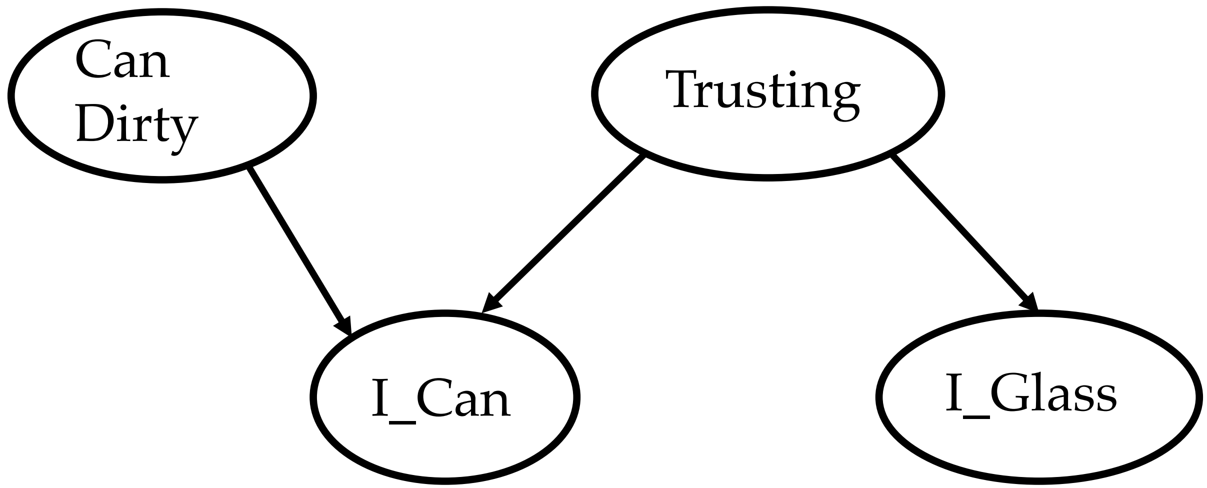
Bayesian Networks



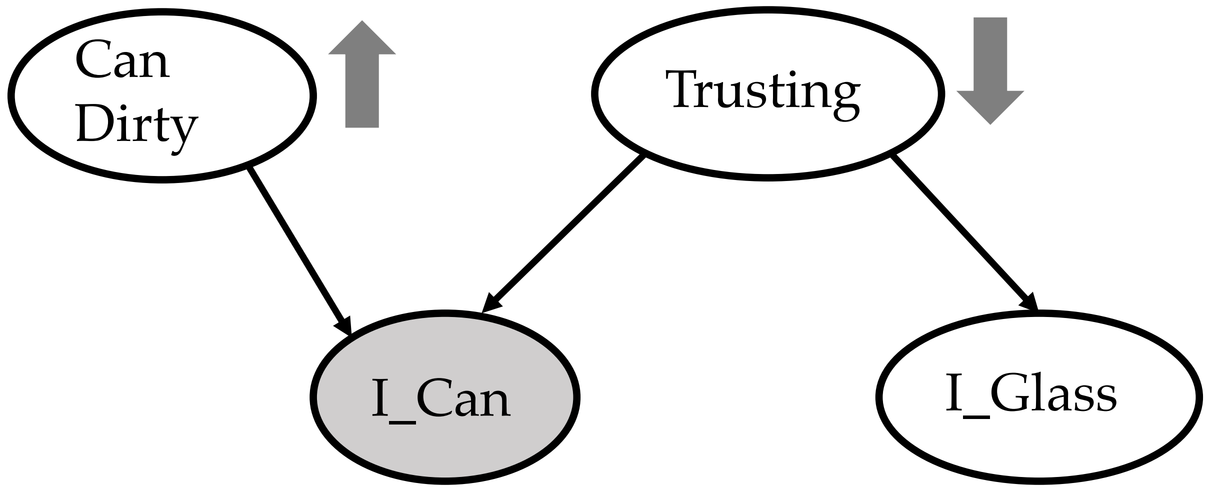
Bayesian Networks



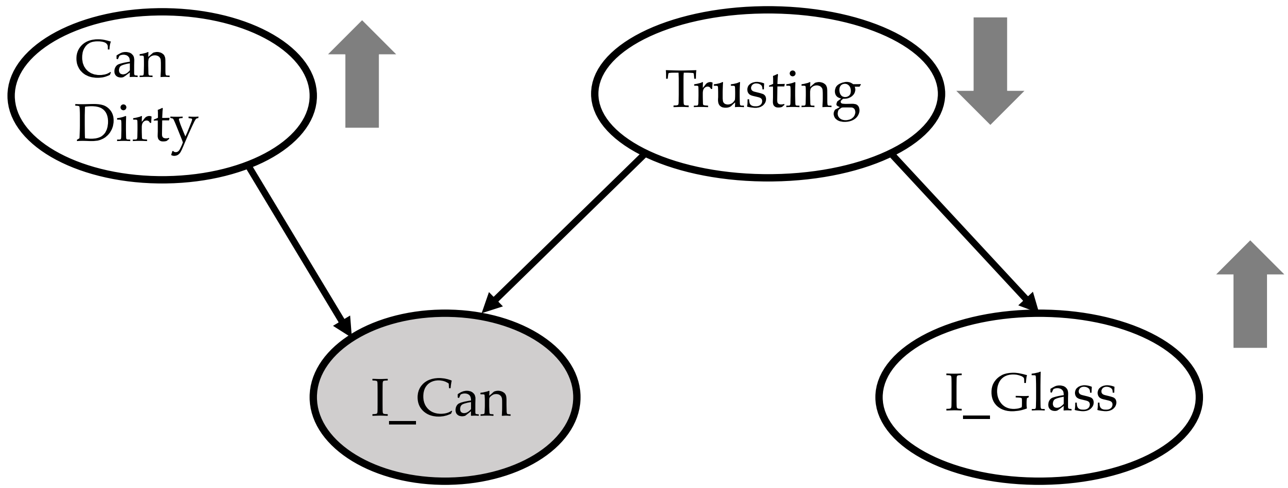
Bayesian Networks



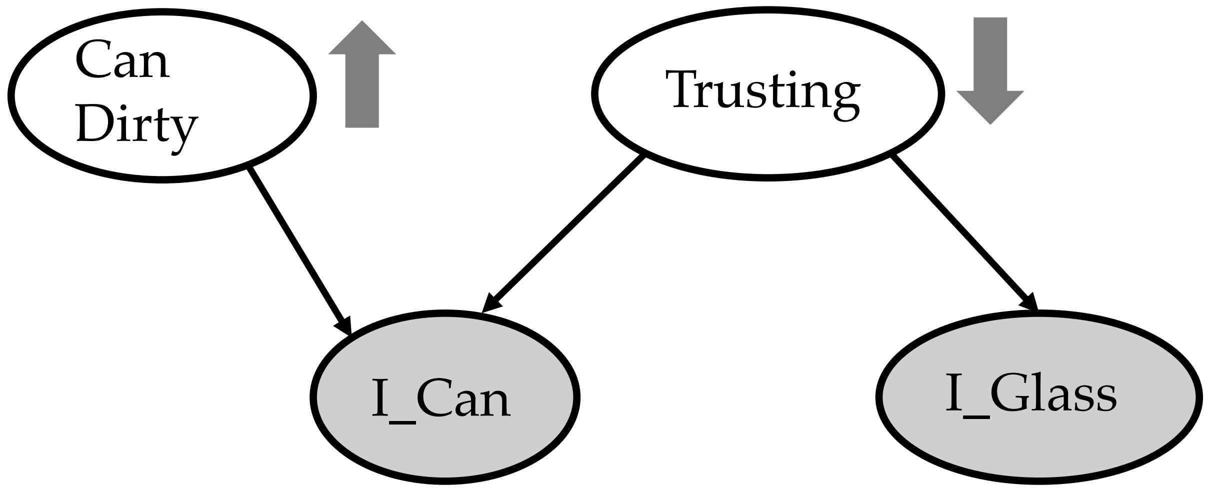
Bayesian Networks



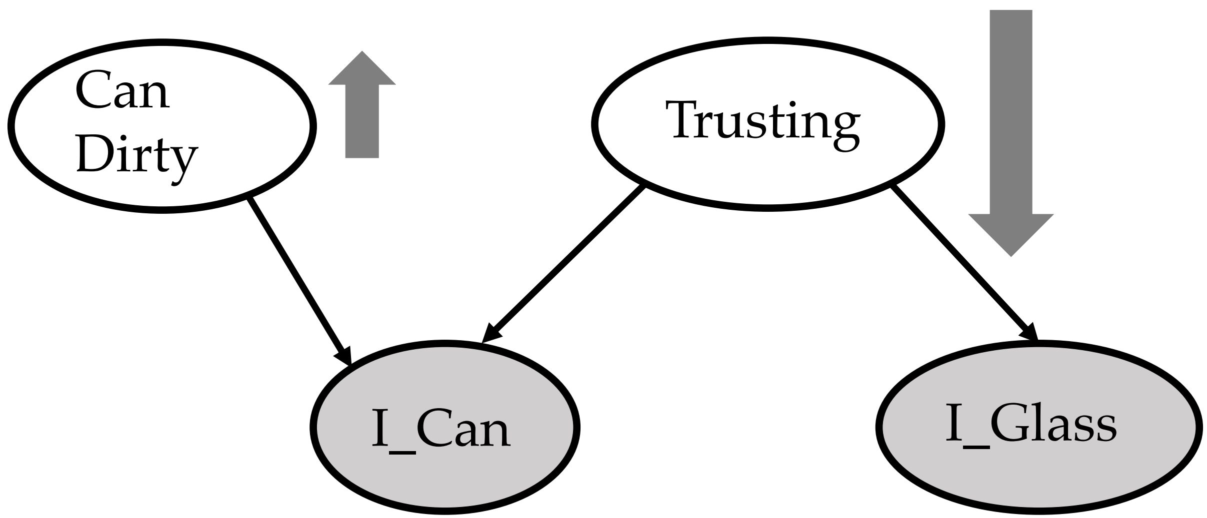
Bayesian Networks



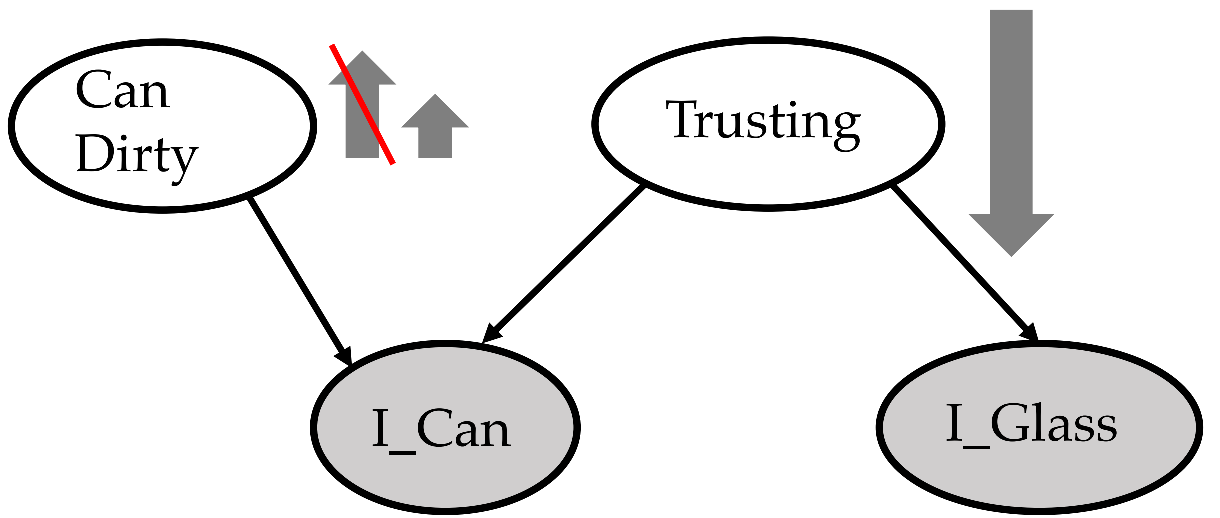
Bayesian Networks



Bayesian Networks



Bayesian Networks



Bayesian Networks

- Each *node* corresponds to a random variable, which is either discrete or continuous.

- Directed links connect pairs of nodes; $X \rightarrow Y$ means X is a parent of Y.

- Each node has a conditional probability distribution: $P(X|parents(X))$.

- Each node is **conditionally independent** of its non-descendents given its parents

Bayesian Networks

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

Bayesian Networks

$$\begin{aligned} P(X_1, \ldots, X_n) &= P(X_1 | X_2, \ldots, X_n) P(X_2, \ldots, X_n) \\ &= P(X_1 | X_2, \ldots, X_n) P(X_2 | X_3, \ldots, X_n) P(X_3, \ldots, X_n) \end{aligned}$$

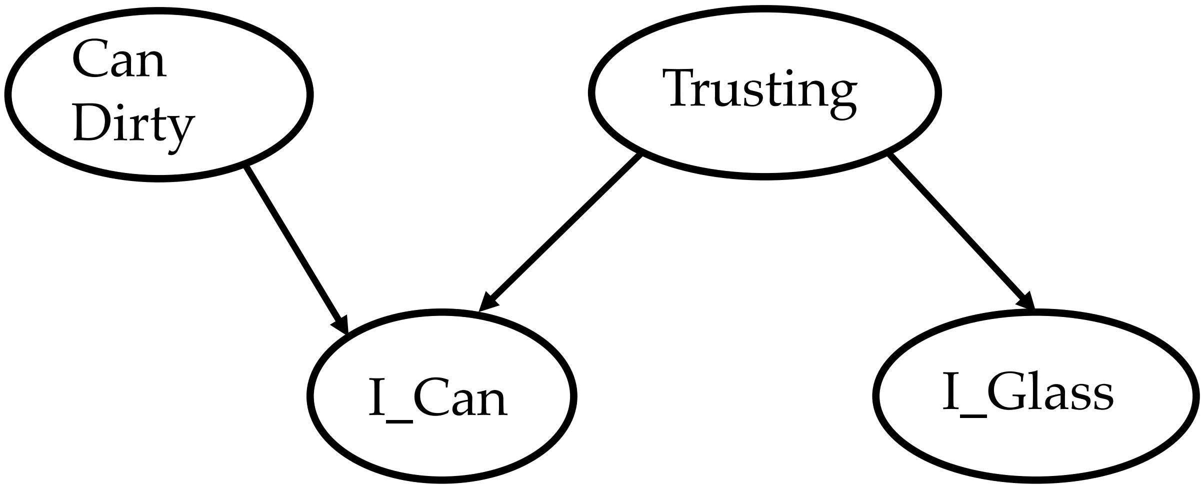
Bayesian Networks

$$
\begin{aligned}
P(X_1, \ldots, X_n) &= P(X_1 | X_2, \ldots, X_n) P(X_2, \ldots, X_n) \\
&= P(X_1 | X_2, \ldots, X_n) P(X_2 | X_3, \ldots, X_n) P(X_3, \ldots, X_n) \\
&\vdots \\
&= \Pi_{i=1}^{n} P(X_i | parents(X_i))
\end{aligned}
$$

Bayesian Network

D	P(D)
t	0.3

T	P(T)
t	0.5



T	P(I_G|T)
t	0.2
f	0.7

D	T	P(I_C|D,T)
f	f	0.7
f	t	0.05
t	f	0.9
t	t	0.2

Bayesian Network

$$P(I_C, I_G, T, D) = P(I_C, I_G | T, D) P(T, D)$$

Bayesian Network

$$
\begin{aligned}
P(I_C, I_G, T, D) &= P(I_C, I_G | T, D) P(T, D) \\
&= P(I_C | T, D) P(I_G | T, D) P(T) P(D)
\end{aligned}
$$

Bayesian Network

$$
\begin{aligned}
P(I_C, I_G, T, D) &= P(I_C, I_G | T, D) P(T, D) \\
&= P(I_C | T, D) P(I_G | T, D) P(T) P(D) \\
&= P(I_C | T, D) P(I_G | T) P(T) P(D)
\end{aligned}
$$

Example

- What is the probability of the human intervening in the can?

Example

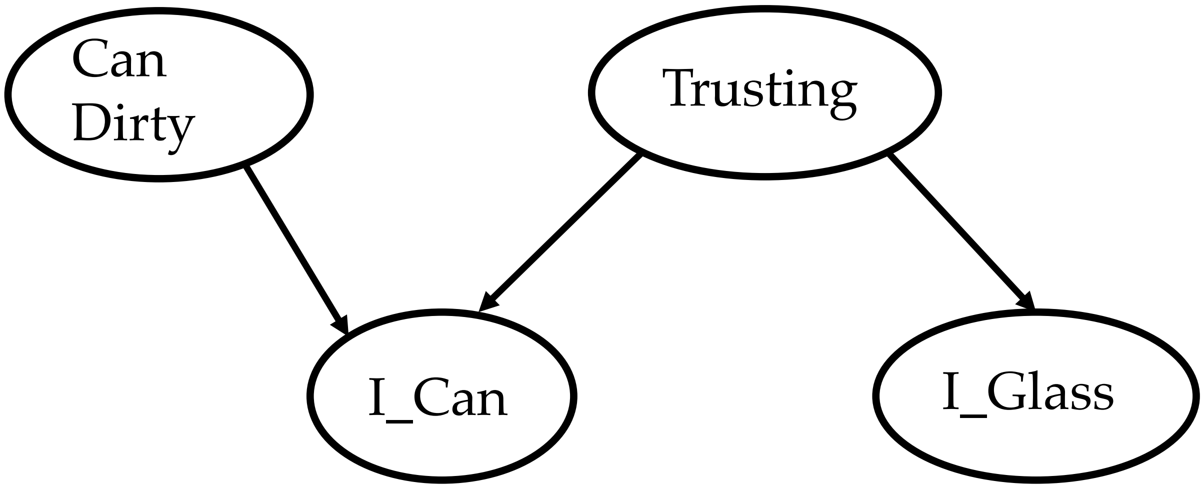
- What is the probability of the human intervening in the can?

$$
\begin{aligned}
P(I_C) &= P(I_C|D,T)P(D,T) + P(I_C|-D,-T)P(-D,-T) + P(C|D,-T)P(D,-T) + P(C|-D,T)P(-D,T \\
&= 0.2 * 0.3 * 0.5 + 0.7 * 0.7 * 0.5 + 0.9 * 0.3 * 0.5 + 0.05 * 0.7 * 0.5 \\
&= 0.428
\end{aligned}
$$

Bayesian Network

D	P(D)
t	0.3

T	P(T)
t	0.5

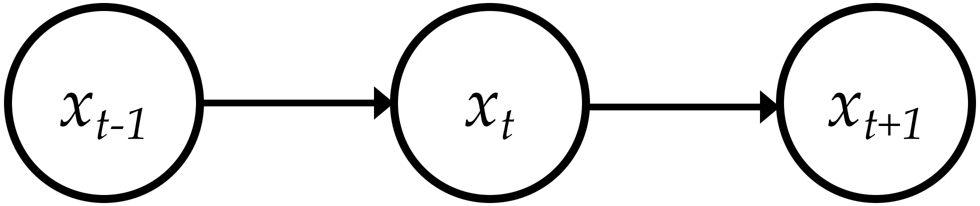


T	P(I_G|T)
t	0.2
f	0.7

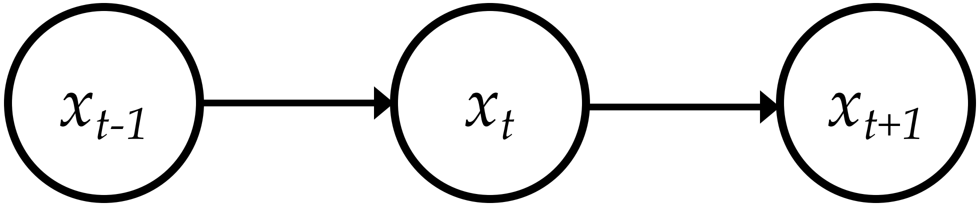
D	T	P(I_C|D,T)
f	f	0.7
f	t	0.05
t	f	0.9
t	t	0.2

Probabilistic Reasoning over Time

State Estimation Over Time



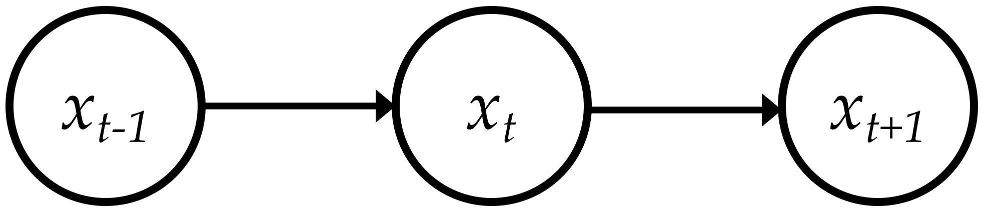
State Estimation Over Time



State Transitions:

$$P(X_t|X_{0:t-1}) =$$

State Estimation Over Time

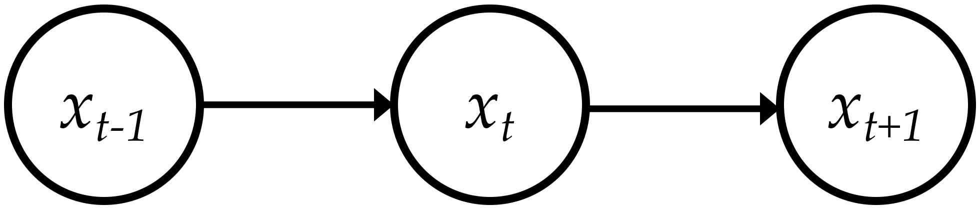


State Transitions:

$$P(X_t|X_{0:t-1}) = P(X_t|X_{t-1})$$

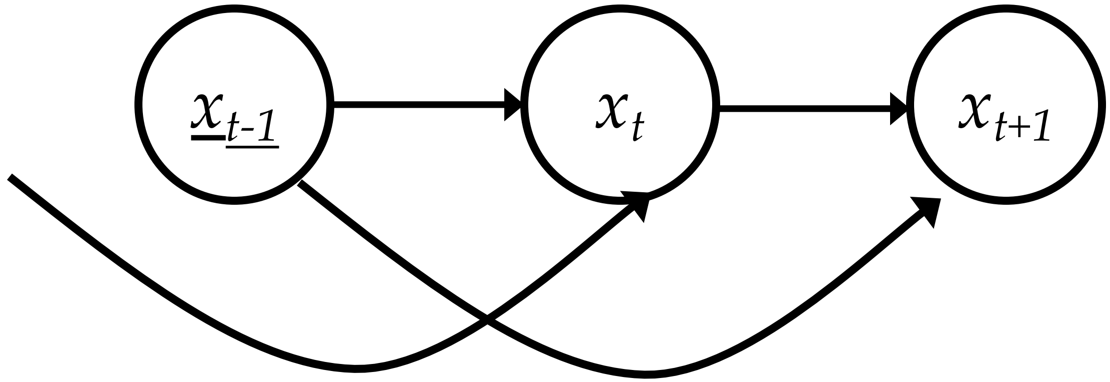
Markov Assumption

Second Order Model



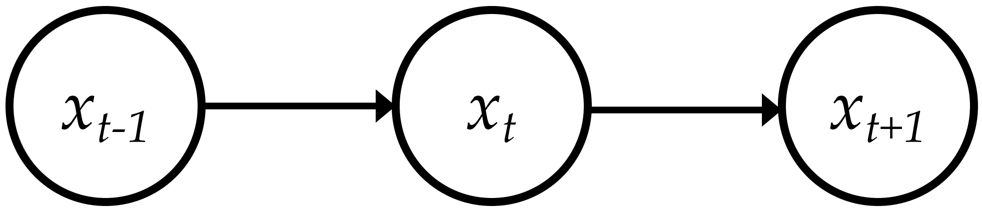
$$P(X_t|X_{0:t-1}) = P(X_t|X_{t-1}, X_{t-2})$$

Second Order Model



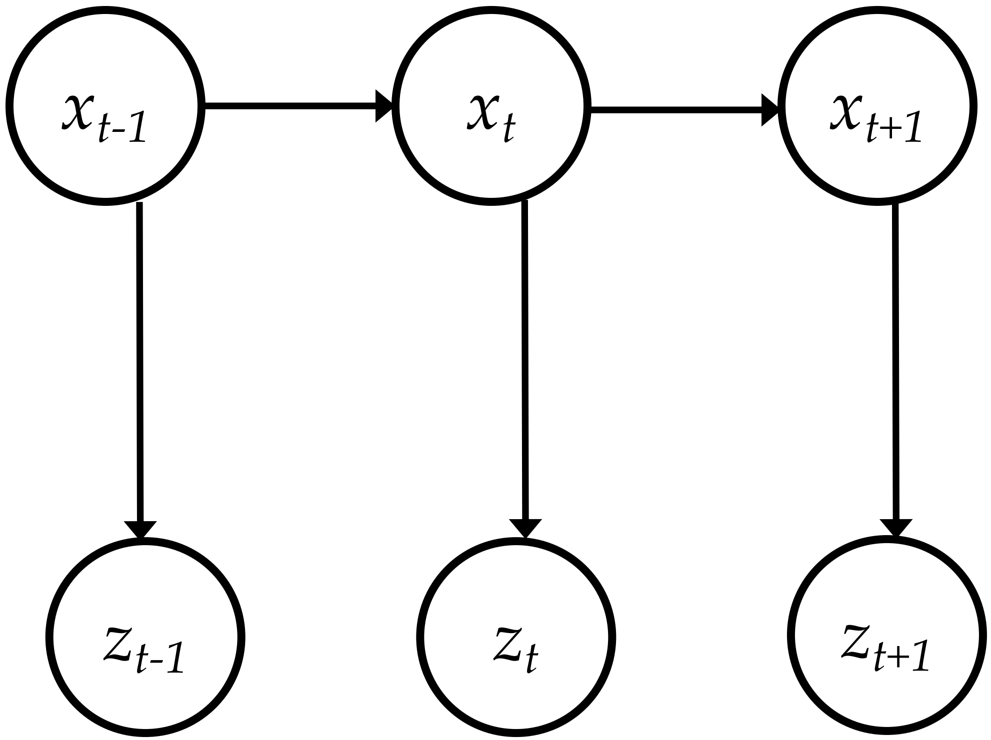
$$P(X_t|X_{0:t-1}) = P(X_t|X_{t-1}, X_{t-2})$$

Assumption of Stationary Dynamics



$$P(X_t|X_{t-1}) = P(X_{t-1}|X_{t-2})$$

Sensor Model



$$P(Z_t|X_{0:t}) = P(Z_t|X_t)$$

Bayesian Filtering

$$P(x_t|z_{1:t}) =$$

Bayesian Filtering

$$P(x_t|z_{1:t}) = P(x_t|z_t, z_{1:t-1})$$

Bayesian Filtering

$$
\begin{aligned}
P(x_t|z_{1:t}) &= P(x_t|z_t, z_{1:t-1}) \\
&= \frac{P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1})}{P(z_t|z_{1:t-1})}
\end{aligned}
$$

Bayesian Filtering

$$\begin{aligned}
P(x_t|z_{1:t}) &= P(x_t|z_t, z_{1:t-1}) \\
&= \frac{P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1})}{P(z_t|z_{1:t-1})} \\
&= \eta P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1})
\end{aligned}$$

Bayesian Filtering

$$
\begin{aligned}
P(x_t|z_{1:t}) &= P(x_t|z_t, z_{1:t-1}) \\
&= \frac{P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1})}{P(z_t|z_{1:t-1})} \\
&= \eta P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t, x_{t-1}|z_{1:t-1})
\end{aligned}
$$

Bayesian Filtering

$$
\begin{aligned}
P(x_t|z_{1:t}) &= P(x_t|z_t, z_{1:t-1}) \\
&= \frac{P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1})}{P(z_t|z_{1:t-1})} \\
&= \eta P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t, x_{t-1}|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}, z_{1:t-1})P(x_{t-1}|z_{1:t-1})
\end{aligned}
$$

Bayesian Filtering

$$
\begin{aligned}
P(x_t|z_{1:t}) &= P(x_t|z_t, z_{1:t-1}) \\
&= \frac{P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1})}{P(z_t|z_{1:t-1})} \\
&= \eta P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t, x_{t-1}|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}, z_{1:t-1})P(x_{t-1}|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1}|z_{1:t-1})
\end{aligned}
$$

Bayesian Filtering

$$
\begin{aligned}
\underline{P(x_t|z_{1:t})} &= P(x_t|z_t, z_{1:t-1}) \\
&= \frac{P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1})}{P(z_t|z_{1:t-1})} \\
&= \eta P(z_t|x_t, z_{1:t-1})P(x_t|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t, x_{t-1}|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}, z_{1:t-1})P(x_{t-1}|z_{1:t-1}) \\
&= \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1})\underline{P(x_{t-1}|z_{1:t-1})}
\end{aligned}
$$

Bayesian Filtering

$$P(x_t|z_{1:t}) = \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}) P(x_{t-1}|z_{1:t-1})$$

$$b(x_t) = \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}) b(x_{t-1})$$

Bayesian Filtering

$$b(x_t) = \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}, z_{1:t-1}, u_{1:t}) b(x_{t-1})$$

Algorithm **Bayes Filter**(*$b(x_{t-1})$, u_t , z_t*)

For all x_t do:

$$\bar{b}(x_t) = \sum_{x_{t-1}} P(x_t|x_{t-1}, u_t) b(x_{t-1})$$

$$b(x_t) = \eta P(z_t|x_t) \bar{b}(x_t)$$

Endfor

Return $b(x_t)$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$
\begin{array}{cc}
 & X_{t+1} \\
X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array}
\end{array}
$$

Sensor:

$$
\begin{array}{cc}
 & Z_t \\
X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array}
\end{array}
$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0) P(x_0)$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0)P(x_0)$$

$$P(desk|sense - desk) = \eta P(sense - desk|desk)[P(desk|desk)P(desk) + P(desk|whiteboard)P(whiteboard) + P(desk|door)P(door)]$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0)P(x_0)$$

$$\begin{aligned} P(desk|sense-desk) &= \eta P(sense-desk|desk)[P(desk|desk)P(desk) \\ &+ P(desk|whiteboard)P(whiteboard) + P(desk|door)P(door)] \\ &= \eta 0.8(0.4 * 0.333 + 0.4 * 0.333 + 0 * 0.333) \end{aligned}$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0)P(x_0)$$

$$\begin{aligned} P(desk|sense-desk) &= \eta P(sense-desk|desk)[P(desk|desk)P(desk) \\ &+ P(desk|whiteboard)P(whiteboard) + P(desk|door)P(door)] \\ &= \eta 0.8(0.4 * 0.333 + 0.4 * 0.333 + 0 * 0.333) \\ &= \eta 0.8 * 0.8 * 0.333 \end{aligned}$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0)P(x_0)$$

$$\begin{aligned} P(desk|sense-desk) &= \eta P(sense-desk|desk)[P(desk|desk)P(desk) \\ &+ P(desk|whiteboard)P(whiteboard) + P(desk|door)P(door)] \\ &= \eta 0.8(0.4 * 0.333 + 0.4 * 0.333 + 0 * 0.333) \\ &= \eta 0.8 * 0.8 * 0.333 \\ &= \eta 0.213 \end{aligned}$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0) P(x_0)$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0)P(x_0)$$

$$P(whiteboard|sense-desk)$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array} \end{array}$$

Assume that we observe: "sense-desk", "sense-door", "sense-desk"

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0) P(x_0)$$

$$\begin{aligned} P(whiteboard|sense-desk) &= \eta 0.1(0.4 * 0.333 + 0.4 * 0.333 + 0 * 0.333) \\ &= \eta 0.1 * 0.8 * 0.333 \\ &= \eta 0.027 \end{aligned}$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_1|z_1) = \eta P(z_1|x_1) \sum_{x_0} P(x_1|x_0)P(x_0)$$

$$P(door|sense - desk) = \eta 0.1(0.2 * 0.333 + 0.2 * 0.333 + 1.0 * 0.333)$$

$$= \eta 0.1 * 1.4 * 0.333$$

$$= \eta 0.047$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$\eta 0.213 + \eta 0.027 + \eta 0.047 = 1$$

$$\eta = 1/0.287$$

$$\eta = 3.48$$

$$P(x_1) = (0.744, 0.093, 0.163)$$

Next, we see “door”

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$P(x_1) = (0.744, 0.093, 0.163)$ Next, we see “door”

$$P(x_2|z_1, z_2) = \eta P(z_2|x_2) \sum_{x_1} P(x_2|x_1) P(x_1)$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$P(x_1) = (0.744, 0.093, 0.163)$ Next, we see “door”

$$P(x_2|z_1, z_2) = \eta P(z_2|x_2) \sum_{x_1} P(x_2|x_1)P(x_1)$$

$$\begin{aligned} P(x_2 = desk|o_1 = desk, o_2 = door) &= \eta 0.1 * (0.4 * 0.742 + 0.4 * 0.094 + 0 * 0.163) \\ &= \eta 0.034 \end{aligned}$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$P(x_1) = (0.744, 0.093, 0.163)$ Next, we see “door”

$$P(x_2|z_1, z_2) = \eta P(z_2|x_2) \sum_{x_1} P(x_2|x_1) P(x_1)$$

$$\begin{aligned} P(x_2 = whiteboard|o_1 = desk, o_2 = door) &= \eta 0.1 * (0.4 * 0.742 + 0.4 * 0.094 + 0 * 0.164) \\ &= 0.034 \\ &= \eta 0.034 \end{aligned}$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{c} \\ X_t = \end{array} \begin{array}{ccc} & X_{t+1} & \\ 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array}$$

Sensor:

$$\begin{array}{c} \\ X_t = \end{array} \begin{array}{ccc} & Z_t & \\ 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$P(x_1) = (0.744, 0.093, 0.163)$ Next, we see “door”

$$P(x_2|z_1, z_2) = \eta P(z_2|x_2) \sum_{x_1} P(x_2|x_1)P(x_1)$$

$$\begin{aligned} P(x_2 = whiteboard|o_1 = desk, o_2 = door) &= \eta 0.1 * (0.4 * 0.742 + 0.4 * 0.094 + 0 * 0.164) \\ &= 0.034 \\ &= \eta 0.034 \end{aligned}$$

Why P(whiteboard) and P(desk) become equal?

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{array}{ccc} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{array} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{array}{ccc} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{array} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$P(x_1) = (0.744, 0.093, 0.163)$ Next, we see “door”

$$P(x_2|z_1, z_2) = \eta P(z_2|x_2) \sum_{x_1} P(x_2|x_1) P(x_1)$$

$$P(x_2 = door|o_1 = desk, o_2 = door) = \eta 0.8 * (0.2 * 0.744 + 0.4 * 0.093 + 0 * 0.163)$$

$$= \eta 0.2643$$

$$P(x_2) = [0.1, 0.1, 0.80]$$

Example

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense - desk, sense - whiteboard, sense - door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

$$P(x_2) = [0.1, 0.1, 0.80]$$

Finally, we observe “desk”

$$P(x) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0.744 & 0.093 & 0.163 \\ 0.1 & 0.1 & 0.8 \\ 0.41 & 0.05 & 0.54 \end{bmatrix}$$

Predicting the Future

- Bayesian Filtering: $P(x_t | z_{1:t})$
- Prediction:

Predicting the Future

- Bayesian Filtering: $P(x_t | z_{1:t})$

- Prediction: $P(x_{t+k} | z_{1:t})$

Predicting the Future

- Bayesian Filtering: $P(x_t|z_{1:t})$

- Prediction: $P(x_{t+k}|z_{1:t})$

$$P(x_{t+k}|z_{1:t}) = \sum_{x_{t+k-1}} P(x_{t+k}, x_{t+k-1}|z_{1:t})$$

Predicting the Future

- Bayesian Filtering: $P(x_t|z_{1:t})$

- Prediction: $P(x_{t+k}|z_{1:t})$

$$P(x_{t+k}|z_{1:t}) = \sum_{x_{t+k-1}} P(x_{t+k}, x_{t+k-1}|z_{1:t})$$

$$= \sum_{x_{t+k-1}} P(x_{t+k}|z_{1:t}, x_{t+k-1})P(x_{t+k-1}|z_{1:t})$$

Predicting the Future

- Bayesian Filtering: $P(x_t | z_{1:t})$

- Prediction: $P(x_{t+k} | z_{1:t})$

$$
\begin{aligned}
P(x_{t+k} | z_{1:t}) &= \sum_{x_{t+k-1}} P(x_{t+k}, x_{t+k-1} | z_{1:t}) \\
&= \sum_{x_{t+k-1}} P(x_{t+k} | z_{1:t}, x_{t+k-1}) P(x_{t+k-1} | z_{1:t}) \\
&= \sum_{x_{t+k-1}} P(x_{t+k} | x_{t+k-1}) P(x_{t+k-1} | z_{t:t})
\end{aligned}
$$

Predicting the Future

FIltering
1/3, 1/3, 1/3
0.744, 0.093, 0.163
0.101, 0.101,0.798
0.413,0.052, 0.536

How will the probabilities change in subsequent timesteps?

States: $x = \{Desk, Whiteboard, Door\}$

Observations: $z = \{sense-desk, sense-whiteboard, sense-door\}$

Transitions:

$$\begin{array}{cc} & X_{t+1} \\ X_t = & \begin{matrix} 0.4 & 0.4 & 0.2 \\ 0.4 & 0.4 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{matrix} \end{array}$$

Sensor:

$$\begin{array}{cc} & Z_t \\ X_t = & \begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{matrix} \end{array}$$

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

Predicting the Future

- Bayesian Filtering

FIltering
1/3, 1/3, 1/3
0.744, 0.093, 0.163
0.101, 0.101,0.798
0.413,0.052, 0.536
0.1860, 0.1860, 0.6280
0.1490, 0.1490, 0.7020
0.1190 , 0.1190, 0.7620
0.0950, 0.0950, 0.8100
0.0760, 0.0760, 0.8480

Probabilities move to a stationary distribution

Predicting the Future

- Bayesian Filtering: $P(x_t | z_{1:t})$

- Prediction: $P(x_{t+k} | z_{1:t})$

- Smoothing: $P(x_k | z_{1:t}), k < t$

Predicting the Future

- Bayesian Filtering: $P(x_t|z_{1:t})$

- Prediction: $P(x_{t+k}|z_{1:t})$

- Smoothing: $P(x_k|z_{1:t}), k < t$

Predicting the Future

- Bayesian Filtering: $P(x_t|z_{1:t})$

$$P(x_t|z_{1:t}) = \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}) P(x_{t-1}|z_{1:t-1})$$

Predicting the Future

- Bayesian Filtering: $P(x_t|z_{1:t})$

$$P(x_t|z_{1:t}) = \eta P(z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}) P(x_{t-1}|z_{1:t-1})$$

$$f_{1:t} = \eta FORWARD(f_{1:t-1}, z_t)$$

Smoothing

$$P(x_k|z_{1:t}), k < t$$

$$P(x_k|z_{1:t}) = P(x_k|z_{1:k}, z_{k+1:t})$$

Smoothing

$$P(x_k|z_{1:t}), k < t$$

$$\begin{aligned} P(x_k|z_{1:t}) &= P(x_k|z_{1:k}, z_{k+1:t}) \\ &= \eta P(z_{k+1:t}|x_k, z_{1:k}) P(x_k|z_{1:k}) \end{aligned}$$

Smoothing

$P(x_k|z_{1:t}), k < t$

$$
\begin{aligned}
P(x_k|z_{1:t}) &= P(x_k|z_{1:k}, z_{k+1:t}) \\
&= \eta P(z_{k+1:t}|x_k, z_{1:k})P(x_k|z_{1:k}) \\
&= \eta P(z_{k+1:t}|x_k)P(x_k|z_{1:k})
\end{aligned}
$$

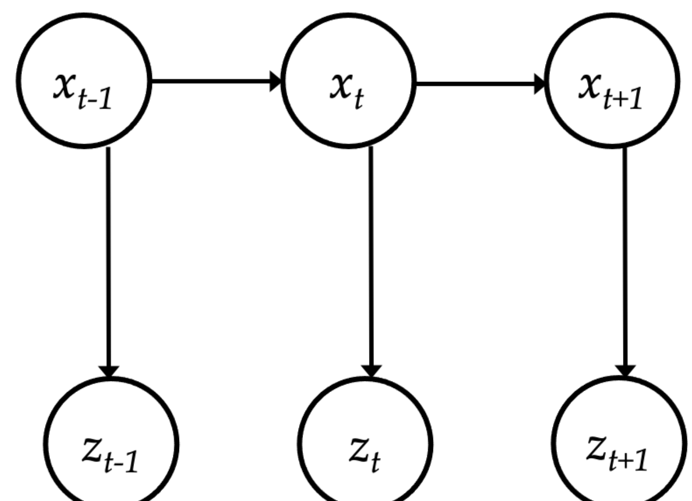
Smoothing

$$P(x_k|z_{1:t}), k < t$$

$$\begin{aligned} P(x_k|z_{1:t}) &= P(x_k|z_{1:k}, z_{k+1:t}) \\ &= \eta P(z_{k+1:t}|x_k, z_{1:k}) P(x_k|z_{1:k}) \\ &= \eta P(z_{k+1:t}|x_k) P(x_k|z_{1:k}) \\ &= \eta b_{k+1:t} \times f_{1:k} \end{aligned}$$

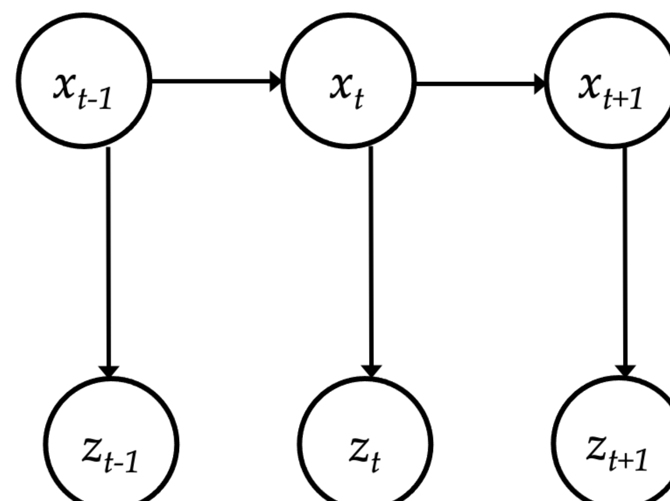
Backward Pass

$$b_{k+1:t} = P(z_{k+1:t} | x_k)$$



Backward Pass

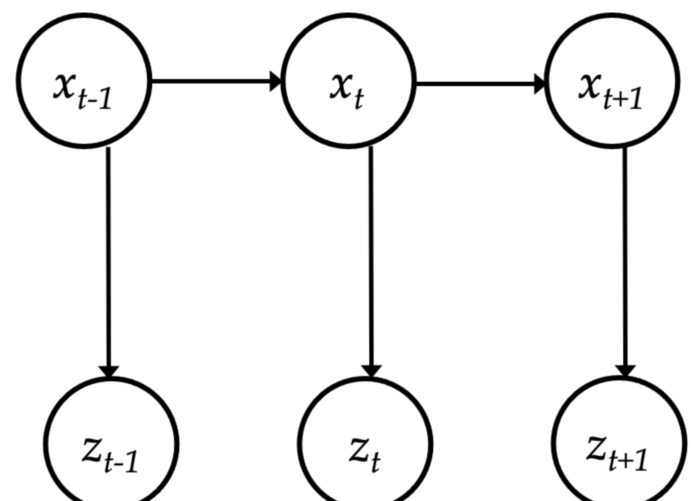
$$b_{k+1:t} = P(z_{k+1:t}|x_k) = \sum_{x_{k+1}} P(z_{k+1:t}, x_{k+1}|x_k)$$



Backward Pass

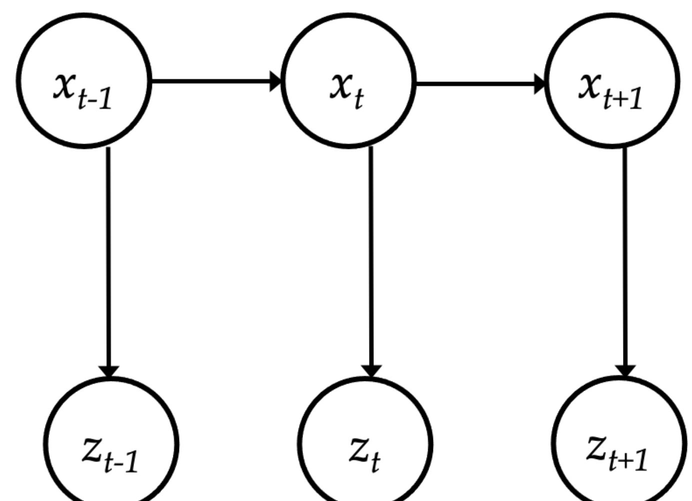
$$b_{k+1:t} = P(z_{k+1:t}|x_k) = \sum_{x_{k+1}} P(z_{k+1:t}, x_{k+1}|x_k)$$

$$= \sum_{x_{k+1}} P(z_{k+1:t}|x_{k+1}, x_k)P(x_{k+1}|x_k)$$



Backward Pass

$$
\begin{aligned}
b_{k+1:t} = P(z_{k+1:t}|x_k) &= \sum_{x_{k+1}} P(z_{k+1:t}, x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1:t}|x_{k+1}, x_k)P(x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1:t}|x_{k+1})P(x_{k+1}|x_k)
\end{aligned}
$$



Backward Pass

$$
\begin{aligned}
b_{k+1:t} = P(z_{k+1:t}|x_k) &= \sum_{x_{k+1}} P(z_{k+1:t}, x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1:t}|x_{k+1}, x_k) P(x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1:t}|x_{k+1}) P(x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1}, z_{k+2:t}|x_{k+1}) P(x_{k+1}|x_k)
\end{aligned}
$$

Backward Pass

$$\begin{aligned}
b_{k+1:t} = \underline{P(z_{k+1:t}|x_k)} &= \sum_{x_{k+1}} P(z_{k+1:t}, x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1:t}|x_{k+1}, x_k) P(x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1:t}|x_{k+1}) P(x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1}, z_{k+2:t}|x_{k+1}) P(x_{k+1}|x_k) \\
&= \sum_{x_{k+1}} P(z_{k+1}|x_{k+1}) \underline{P(z_{k+2:t}|x_{k+1})} P(x_{k+1}|x_k)
\end{aligned}$$

Smoothing

$$P(x_k|z_{1:t}), k < t$$

$$\begin{aligned}
P(x_k|z_{1:t}) &= P(x_k|z_{1:k}, z_{k+1:t}) \\
&= \eta P(z_{k+1:t}|x_k, z_{1:k}) P(x_k|z_{1:k}) \\
&= \eta P(z_{k+1:t}|x_k) P(x_k|z_{1:k}) \\
&= \eta b_{k+1:t} \times f_{1:k}
\end{aligned}$$

Filtering vs Smoothing

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

Filtering	**Smoothing**
1/3, 1/3, 1/3	1/3, 1/3, 1/3
0.744, 0.093, 0.163	
0.101, 0.101,0.798	
0.413,0.052, 0.536	0.413,0.052, 0.536

How do you expect the values to look like?

Filtering vs Smoothing

Assume that we observe: ”sense-desk”, “sense-door”, “sense-desk”

Filtering	Smoothing
1/3, 1/3, 1/3	1/3, 1/3, 1/3
0.744, 0.093, 0.163	0.666, 0.083, 0.251
0.101, 0.101,0.798	0.245, 0.245, 0.51
0.413,0.052, 0.536	0.413,0.052, 0.536

Why did the P(desk) decrease?