Lecture 6

Decision Theory

CS 3630

Lectures 3-5 Recap

Trash Sorting Robot

- States are uncertain:
  - Prior probability distribution on states
  - No dependence on past actions
- Three simple sensors
  - Discrete sensors, discrete conditional distributions
  - Continuous sensor, conditional distribution is Gaussian
- Perception using Bayes equation:
  - Bayes inversion equation to infer state from sensor measurements
  - Maximum likelihood estimation (MLE)
  - Maximum a posteriori estimate (MAP)
  - Sensors are conditionally independent, given state
- Deterministic actions make planning pretty easy:
  - Formulate decision making as an optimization problem
  - Minimize expected cost, minimize worst-case cost, etc.
- Learning prior distributions and sensor models
  - Counting outcomes and using proportions for discrete distributions
  - Parameter estimation for Gaussian distributions

A Vacuum Cleaning Robot

Chapter 3

Overview

- The state space is more interesting, but still discrete.
- Actions are not deterministic.
    - Uncertainty in the effects of actions.
    - Probability associated to an action’s effects depends on the current state.
- Very simple sensing system
- Perception includes both sensing and context
    - An individual measurement from a simple sensor doesn’t provide much information.
    - *History* of sensor observations affects current belief about the world.
- Planning is more complex in this scenario.
    - Because effects of actions depend on state, we need to think about more than one action, and about how the effects of actions propagate through time.
    - Because there is uncertainty, we plan to maximize expected reward, not deterministic outcomes or goals.
- Reinforcement Learning (RL) is appropriate when we don’t have access to large data sets, and when the robot operates in the same setting for a long period of time.

States

In this chapter, we consider the following scenario:

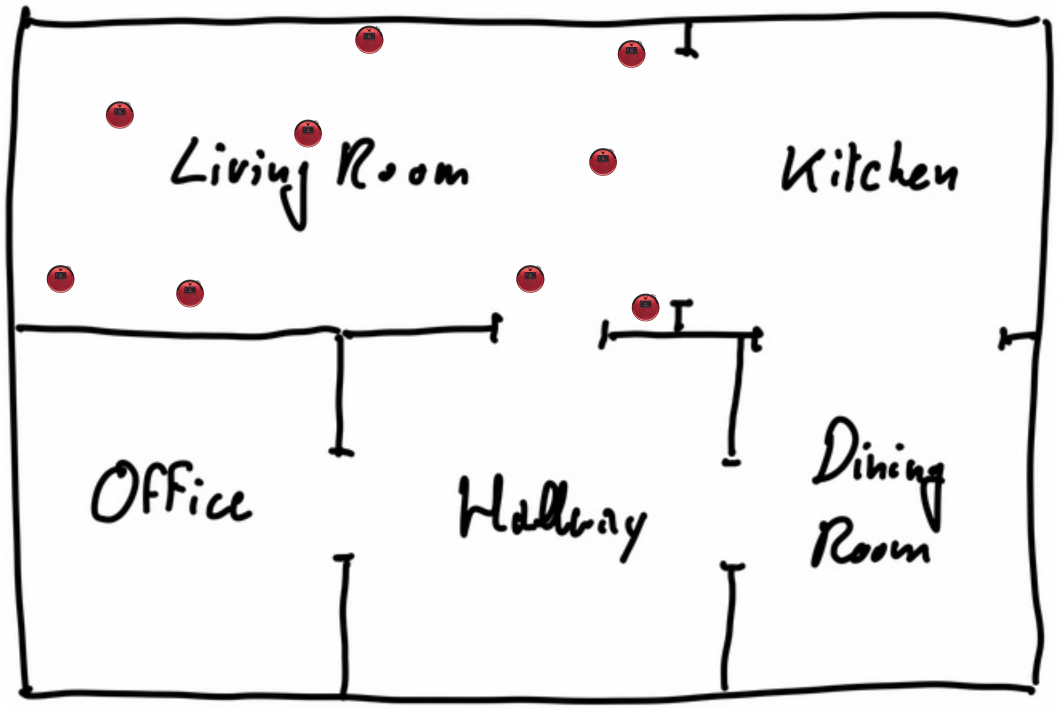
- The robot can move in any direction, so its orientation doesn't matter.
- The robot is equipped with navigation software (which is not perfect), so we won't worry about path planning from room to room.
- To clean a specific room, the robot can execute a preprogrammed motion (maybe boustrophedon, maybe random), so we don't need to worry about the exact position of the robot in a specific room.
- The robot has built-in collision avoidance, so no need to have a detailed map of object locations

➢ ***The room in which the robot is currently located is the only interesting piece of information for this robot.***

State Space

For this robot, the state, X, is defined as the room in which the robot is currently located:

$$X \in \{living\ room, kitchen, office, hallway, dining\ room\}$$



A typical vacuum cleaning robot.

For all of the robot locations shown here, we have:

$$X = living\ room$$

The exact location within the living room is not relevant for this robot.

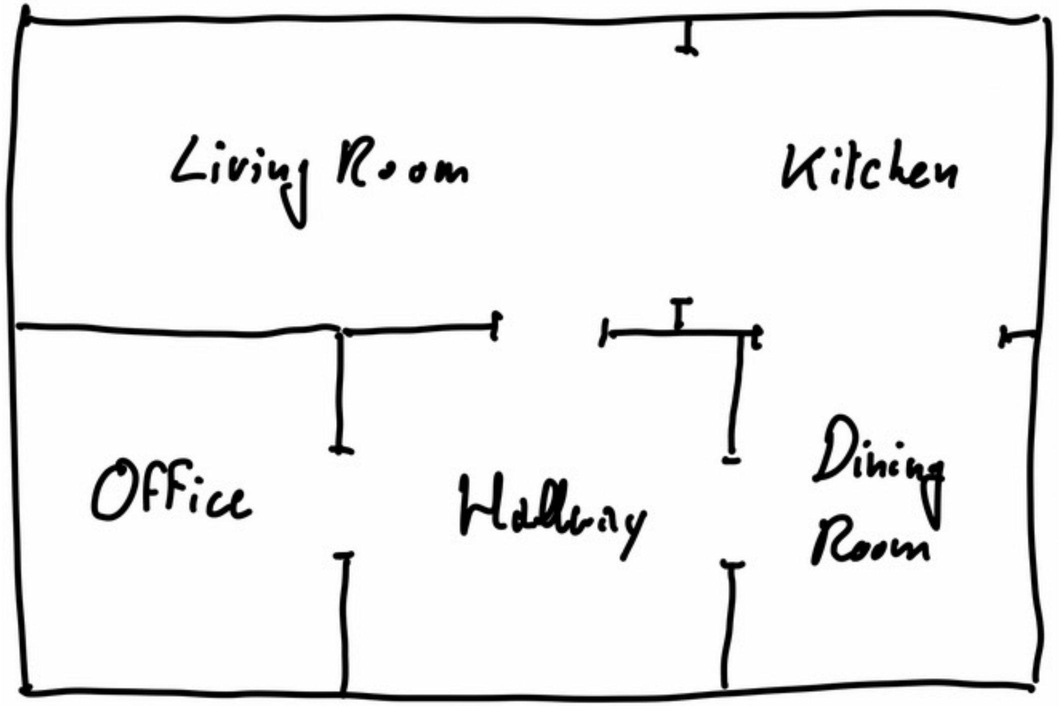
To simplify notation, we'll sometimes write $X \in \{L, K, O, H, D\}$.

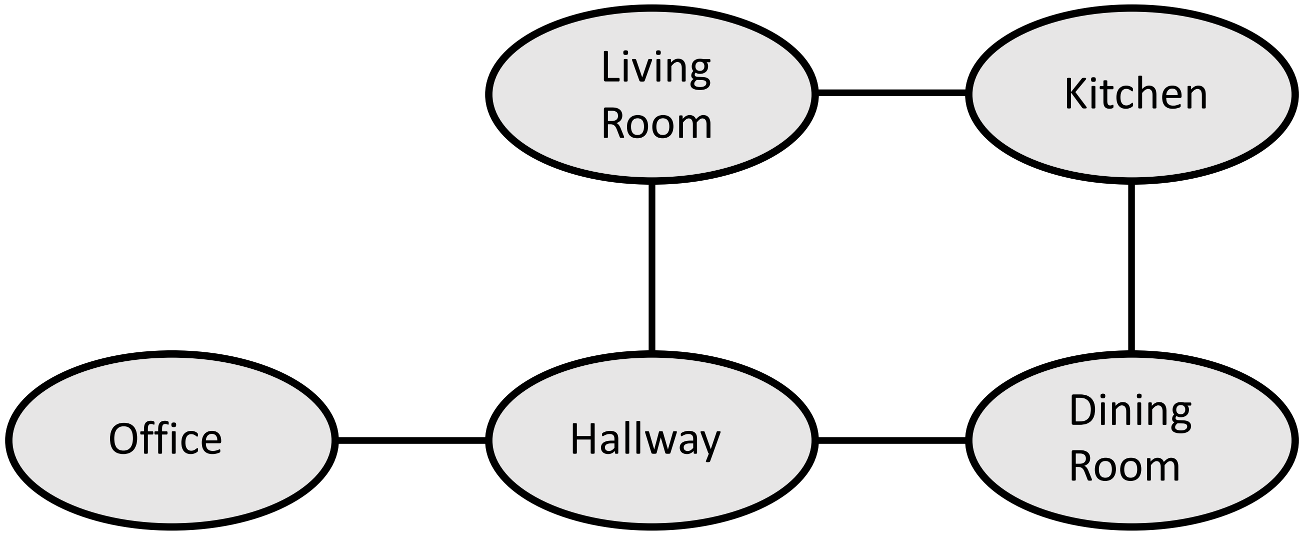
State Space

The state space is the set of all states, along with connectivity information (i.e., neighborhood relationships between states).

In this case, we can represent the state space by a simple undirected graph.

- Our robot can move directly from the *living room* to the *kitchen* or *hallway*, but cannot move directly from the *kitchen* to the *office*.
- This representation will be useful for both planning and for perception.





Prior probability distribution

- For our trash sorting robot, the prior probability distribution on state described our belief (before making any sensor measurements) about the type of object in the work cell.
- For our vacuum cleaning robot, the prior probability distribution on state describes our belief about the ***initial location*** of the robot: *When the robot "wakes up," where will it be?*
- In this chapter, we'll assume that the robot always returns to its charging station in the office after operation (perhaps with the help of a human, or a very smart dog).
- Therefore, our prior distribution on state at the start of the day is given by:

State, *x*	***P(X=x)***
Living room	0
Kitchen	0
Office	1
Hallway	0
Dining room	0

- $P(X = office) = 1$ implies that there is ***no uncertainty*** in our initial state.
- BUT… because there will be uncertainty associated to the effects of actions, this certainty will not long endure after the robot begins its daily activities.

Discrete time systems

- For our trash sorting robot, there was no need to consider the passing of time.
  - Past actions did not affect future performance,
  - Actions were executed in a single time step.
  - The state, X, denoted the state at the present time, and we never needed to represent the state at any other time (neither past nor present).
- For our vacuum cleaning robot, the passing of time is important.
  - We know the location of the robot at the start of the day, but after the robot executes its first actions, there will be uncertainty in the robot's state.
  - The state could change each time the robot executes an action.
  - Sensor measurements depend on state, and state depends on actions; therefore, the sequence in which sensor measurements occur will give us information about the world that can be used for perception.
- Most of the time, nothing interesting happens.
  - We don't need to keep track of the state for all $t \in \mathbb{R}_{\geq 0}$.
  - We only need to keep track of state at discrete time instants, $t \in \{t_0, t_1 \dots\}$, where $\{t_0, t_1 \dots\}$ is the set of times at which something "interesting" occurs.
- **We will represent the state at time $\boldsymbol{t}$ by $\boldsymbol{X_t}$, and we'll simplify notation by simply using $\boldsymbol{t \in \{0, 1, 2 \dots\}}$.**
- The initial state of the robot (i.e., when it wakes up in the morning) is therefore: $X_0 = office$.

Belief state

- It will sometimes be convenient to refer to the entire probability distribution at time t.
- We refer to this distribution as the belief state at time t, denoted by b_t.
- The belief state is a ***row vector*** whose elements correspond to the possible states.
- In our case, there are five possible states, so b_t has five elements.
- At $t = 0$, the belief state is merely our initial distribution:

$$b_0 = [P(X_0 = L), \quad P(X_0 = K) \quad P(X_0 = O) \quad P(X_0 = H) \quad P(X_0 = D)]$$

$$= [0 \quad 0 \quad 1 \quad 0 \quad 0]$$

- The belief state b_{t+1} is conditioned on the initial state x_0 and all actions taken until time t.

$$b_{t+1}^T = \begin{bmatrix} P(X_{t+1} = L \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = K \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = O \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = H \mid a_1 \dots a_t, x_0) \\ P(X_{t+1} = D \mid a_1 \dots a_t, x_0) \end{bmatrix}$$

We need to learn about actions…

➢ Note that we use b_{t+1}^T to denote the transpose of b_{t+1} (for formatting purposes).

Actions

- Our vacuum cleaning robot has four actions:
  - ➢ Move *left, right, up*, or *down* (relative to the map of the house)

- Effects of actions are probabilistic.
- Effects of actions depend on the current state.
  - ➢ ***Use conditional probabilities to model the effects of actions.***

- For a specific sequence of actions (e.g., *up, right, down, left*), computing probabilities for states in the distant future seems complicated.
  - ➢ ***Happily, thanks to the Markov property, these computations are not so difficult.***

Actions

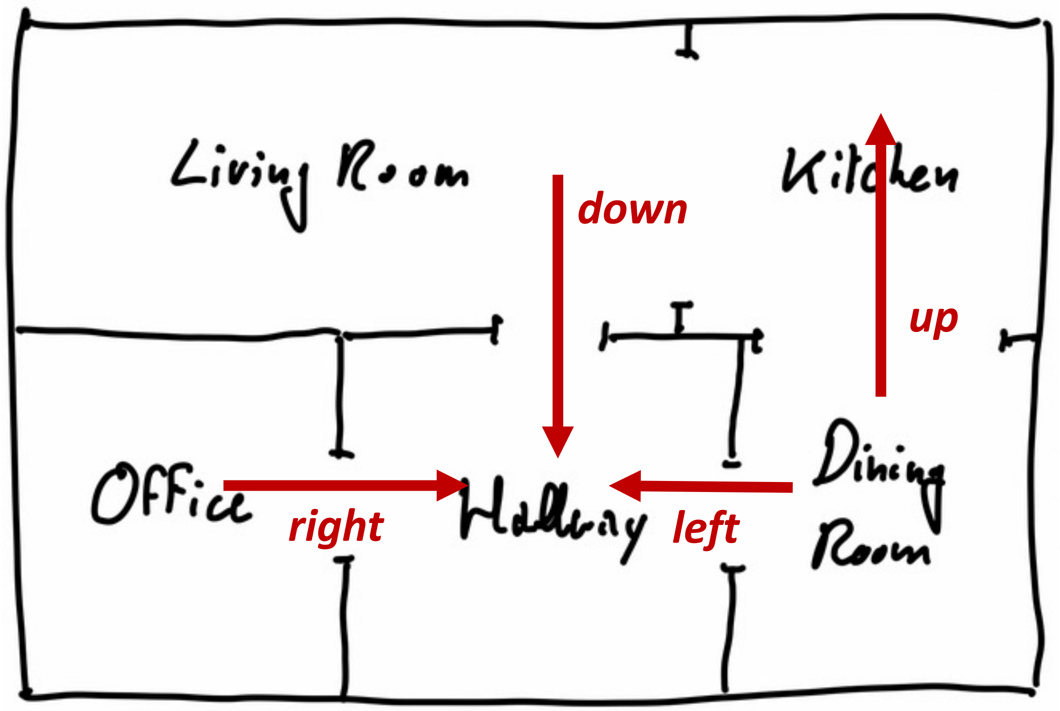
Our robot has four actions:

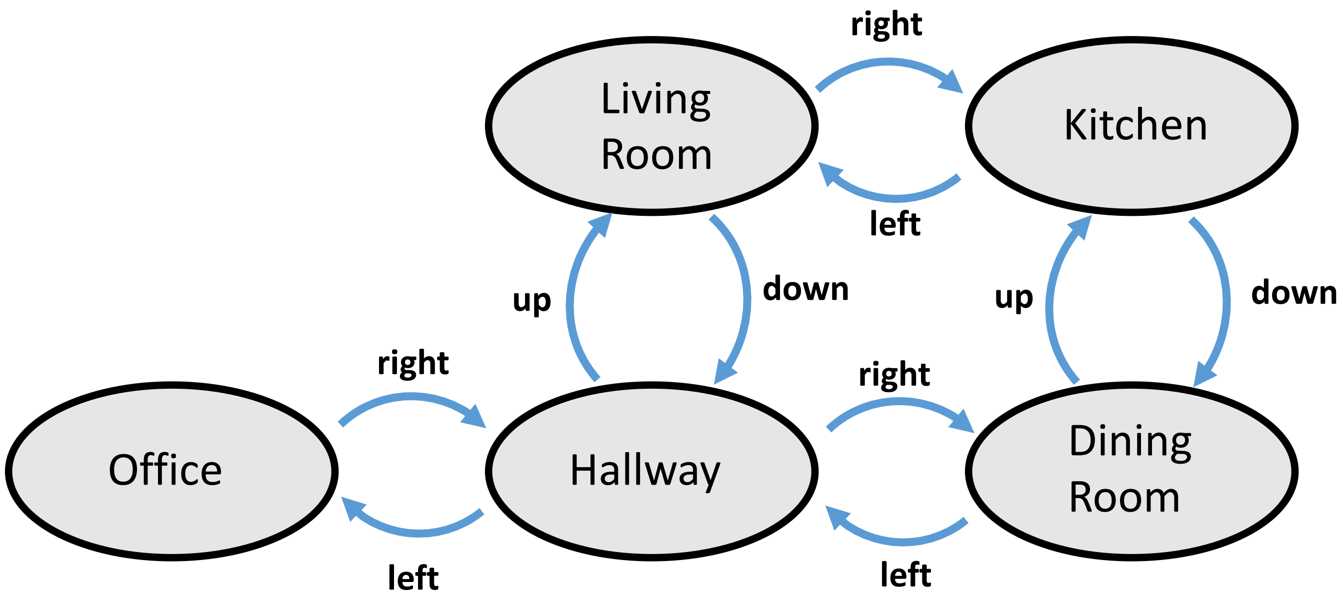
up, down, left, right.

- Effects of actions are context dependent.
- Actions potentially cause a change in state.
- Executing an action in state X_t produces state X_{t+1}

We can represent this by a slight modification to our state space:

- Instead of using an undirected graph, use a directed graph.
- Each edge (u, v) corresponds to an action meant to change the state from $x_t = u$ to $x_{t+1} = v$.
- Sadly, our actions are not deterministic, so we need to do a bit more work.





Uncertainty in the effects of actions

- We will model uncertainty in the effects of actions by using conditional probability distributions.
- In particular, we define the conditional probability distribution for the next state, X_{t+1}, given that the current state, X_t is room x_t, and that action a_t was executed at time t.

$$P(X_{t+1} = x_{t+1} \,|a_t, X_t = x_t)$$

Example: If we are in the *Office* at time *t* and execute the *move right* action, $P(X_{t+1} = H \mid right\ , X_t = O)$ denotes the conditional probability of arriving to the *Hallway*.

Uncertainty in the effects of actions

- We will model uncertainty in the effects of actions by using conditional probability distributions.
- In particular, we define the conditional probability distribution for the next state, X_{t+1}, given that the current state, X_t is room x_t, and that action a_t was executed at time t.

$$P(X_{t+1} = x_{t+1} | a_t, X_t = x_t)$$

Random variable that denotes the state at time $t+1$.

Value taken by X_{t+1} at time $t+1$.

Action taken at time t.

The random state X_t took the value x_t.

Example: If we are in the *Office* at time *t* and execute the *move right* action, $P(X_{t+1} = H \mid right, X_t = O)$ denotes the conditional probability of arriving to the *Hallway*.

The Markov property

- Suppose we start in state $X_0 = O$, execute two actions, *right, up*, and move through the states, $X_1 = H, X_2 = L$.
- What can we say about X_3 if we now execute the action *right*? Or, more formally, what can we say about the conditional probability

$$P(X_3 = x_3 \,|right, up, right, X_0 = O, X_1 = H, X_2 = L)$$

Key Observation:

➢ If we *know* that the robot is in the Living Room at time $t = 2$ and executes the action $a_2 = right$, our belief about X_3 is ***completely independent*** of where the robot may have been at times $t = 0{,}1$ or of the actions taken at times $t = 0{,}1$.

➢ *More generally, if we know the current room (aka , X_t) then the history of how the robot came to be in that room will not affect our belief about what happens when the robot executes its next action.*

➢ ***This is an example of a Markov property.***

The Markov property

- Using this Markov property, we can write

$$P(X_3 = x_3 \mid right, up, right, X_0 = O, X_1 = H, X_2 = L) = P(X_3 = x_3 \mid right, X_2 = L)$$

The Markov property

- Using this Markov property, we can write

$$P(X_3 = x_3 \mid right, up, right, X_0 = O, X_1 = H, X_2 = L) = P(X_3 = x_3 \mid right, X_2 = L)$$

What the robot has done before time t.

Where the robot has been before time t.

What the robot does now, at time t.

Where the robot is now, at time t.

Our Markov assumption:

$$\boldsymbol{P(X_{t+1} = x_{t+1} \mid a_0, \ldots a_t, X_0 = x_0, \ldots, X_t = x_t) = P(X_{t+1} = x_{t+1} \mid a_t, X_t = x_t)}$$

Actions: Part 2

- Our vacuum cleaning robot has four actions:
- ➢ Move *left, right, up*, or *down* (relative to the map of the house)

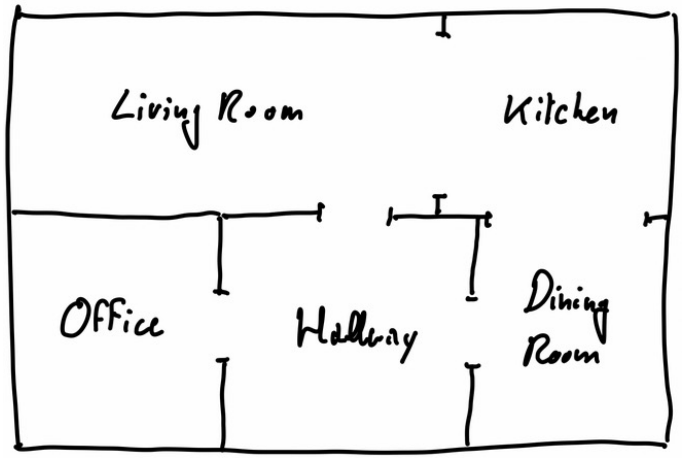
- Effects of actions are probabilistic.
- Effects of actions depend on the current state.
- ➢ ***Use conditional probabilities to model the effects of actions.***

- For a specific sequence of actions (e.g., *up, right, down, left*) Computing probabilities for states in the distant future seems complicated.
- ➢ ***Happily, thanks to the Markov property, these computations are not so difficult.***

Conditional probability distributions for actions

- Thanks to our Markov assumption, all necessary knowledge about the probabilistic effects of actions is included in our conditional probability tables.
- For example, if $X_t = L$, we can write conditional probability distributions for each of the four possible actions.
- In our example scenario, a reasonable distribution could be:

X_t	a_t	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	**L**	1	0	0	0	0
Living Room	**R**	0.2	0.8	0	0	0
Living Room	**U**	1	0	0	0	0
Living Room	**D**	0.2	0	0	0.8	0



Conditional probability distributions for actions

- Thanks to our Markov assumption, all necessary knowledge about the probabilistic effects of actions is included in our conditional probability tables.
- For example, if $X_t = L$, we can write conditional probability distributions for each of the four possible actions.
- In our example scenario, a reasonable distribution could be:

X_t	a_t	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	**L**	1	0	0	0	0
Living Room	**R**	0.2	0.8	0	0	0
Living Room	**U**	1	0	0	0	0
Living Room	**D**	0.2	0	0	0.8	0

$\mathrm{X_t}$ = Living Room
Action = *move right*.

➢ ***Regardless of how we came to be in the Living Room, if we now execute the action move right, we arrive to the Kitchen with probability 0.8, and stay in the Living Room with probability 0.2.***

Conditional probability distributions for actions

- Thanks to our Markov assumption, we can encapsulate all necessary knowledge about the probabilistic effects of actions using conditional probability tables.

- For example, if $X_t = L$, we can write conditional probability distributions for each of the four possible actions.
- In our example scenario, a reasonable distribution could be:

		X_{t+1}				
X_t	a_t	**Living Room**	**Kitchen**	**Office**	**Hallway**	**Dining Room**
Living Room	**L**	1	0	0	0	0
Living Room	**R**	0.2	0.8	0	0	0
Living Room	**U**	1	0	0	0	0
Living Room	**D**	0.2	0	0	0.8	0

Taken together, these four rows give the Conditional Probability Table for:

- Arriving to the each of the five possible rooms for X_{t+1}
- Given that

  $X_t = \text{Living Room}$
- For each possible action a_t
  - Left, Right, Up, Down

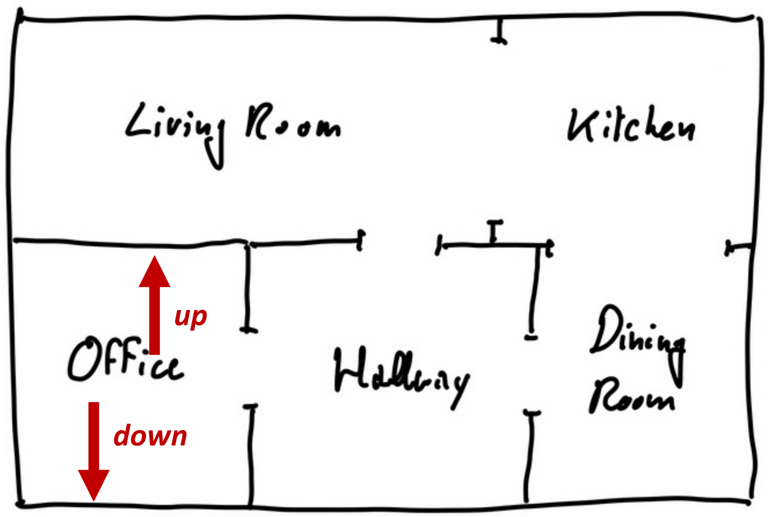
$$P(X_{t+1} = x_{t+1} \mid a_t, X_t = \textit{Living Room})$$

Conditional probability tables

In the book, you'll find the CPTs for the four actions collected into a very large table.

X1	A1	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	L	1	0	0	0	0
Living Room	R	0.2	0.8	0	0	0
Living Room	U	1	0	0	0	0
Living Room	D	0.2	0	0	0.8	0
Kitchen	L	0.8	0.2	0	0	0
Kitchen	R	0	1	0	0	0
Kitchen	U	0	1	0	0	0
Kitchen	D	0.2	0	0	0	0.8
Office	L	0	0	1	0	0
Office	R	0	0	0.2	0.8	0
Office	U	0	0	1	0	0
Office	D	0	0	1	0	0
Hallway	L	0	0	0.8	0.2	0
Hallway	R	0	0	0	0.2	0.8
Hallway	U	0.8	0	0	0.2	0
Hallway	D	0	0	0	1	0
Dining Room	L	0	0	0	0.8	0.2
Dining Room	R	0	0	0	0	1
Dining Room	U	0	0.8	0	0	0.2
Dining Room	D	0	0	0	0	1

This table was constructed by hand (*Thanks, Professor Dellaert!*), with intuitively reasonable probability values.



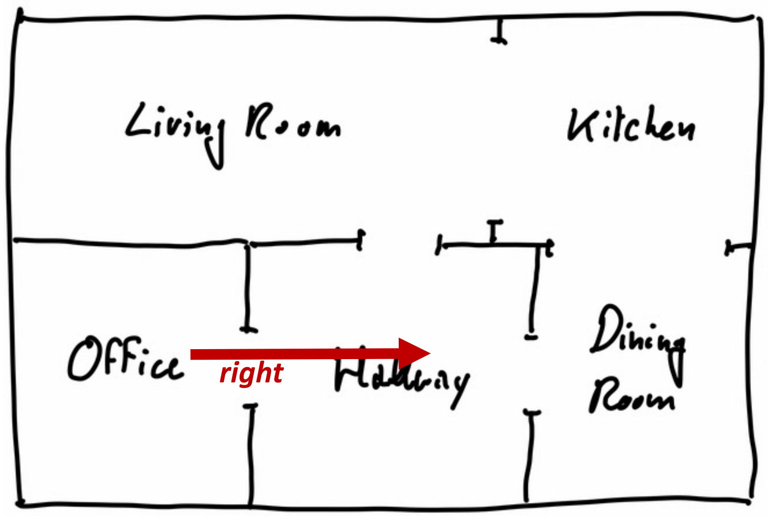
If the robot is in the ***Office***, then moving ***up*** or moving ***down*** will not allow the robot to change rooms.

Conditional probability tables

In the book, you'll find the CPTs for the four actions collected into a very large table.

X1	A1	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	L	1	0	0	0	0
Living Room	R	0.2	0.8	0	0	0
Living Room	U	1	0	0	0	0
Living Room	D	0.2	0	0	0.8	0
Kitchen	L	0.8	0.2	0	0	0
Kitchen	R	0	1	0	0	0
Kitchen	U	0	1	0	0	0
Kitchen	D	0.2	0	0	0	0.8
Office	L	0	0	1	0	0
Office	R	0	0	0.2	0.8	0
Office	U	0	0	1	0	0
Office	D	0	0	1	0	0
Hallway	L	0	0	0.8	0.2	0
Hallway	R	0	0	0	0.2	0.8
Hallway	U	0.8	0	0	0.2	0
Hallway	D	0	0	0	1	0
Dining Room	L	0	0	0	0.8	0.2
Dining Room	R	0	0	0	0	1
Dining Room	U	0	0.8	0	0	0.2
Dining Room	D	0	0	0	0	1

This table was constructed by hand (*Thanks, Professor Dellaert!*), with intuitively reasonable probability values.



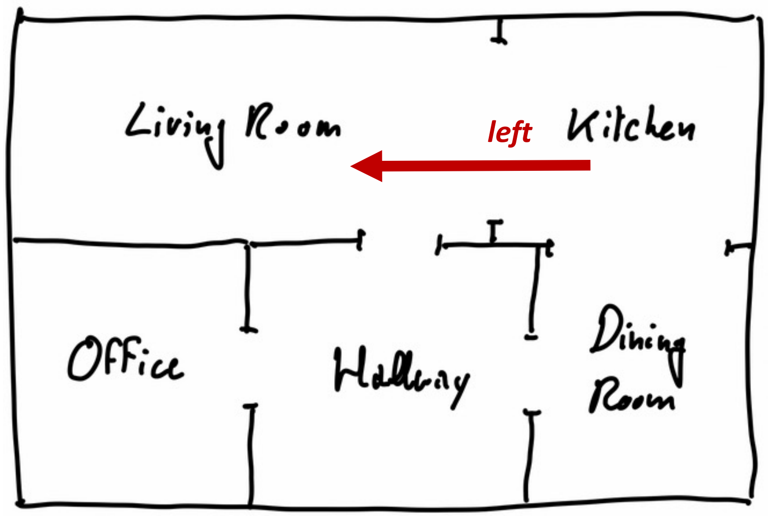
If the robot is in the ***Office*** and moves ***right***, it will stay in the ***Office*** (prob = 0.2) or arrive to the ***Hallway*** (prob = 0.8)

Conditional probability tables

In the book, you'll find the CPTs for the four actions collected into a very large table.

X1	A1	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	L	1	0	0	0	0
Living Room	R	0.2	0.8	0	0	0
Living Room	U	1	0	0	0	0
Living Room	D	0.2	0	0	0.8	0
Kitchen	L	0.8	0.2	0	0	0
Kitchen	R	0	1	0	0	0
Kitchen	U	0	1	0	0	0
Kitchen	D	0.2	0	0	0	0.8
Office	L	0	0	1	0	0
Office	R	0	0	0.2	0.8	0
Office	U	0	0	1	0	0
Office	D	0	0	1	0	0
Hallway	L	0	0	0.8	0.2	0
Hallway	R	0	0	0	0.2	0.8
Hallway	U	0.8	0	0	0.2	0
Hallway	D	0	0	0	1	0
Dining Room	L	0	0	0	0.8	0.2
Dining Room	R	0	0	0	0	1
Dining Room	U	0	0.8	0	0	0.2
Dining Room	D	0	0	0	0	1

This table was constructed by hand (*Thanks, Professor Dellaert!*), with intuitively reasonable probability values.



If the robot is in the ***Kitchen*** and moves ***left***, it will stay in the ***Kitchen*** (prob = 0.2) or arrive to the ***Living Room*** (prob = 0.8)

Conditional probability tables

We can construct a conditional probability table for each action using this large table.

X1	A1	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	L	1	0	0	0	0
Living Room	R	0.2	0.8	0	0	0
Living Room	U	1	0	0	0	0
Living Room	D	0.2	0	0	0.8	0
Kitchen	L	0.8	0.2	0	0	0
Kitchen	R	0	1	0	0	0
Kitchen	U	0	1	0	0	0
Kitchen	D	0.2	0	0	0	0.8
Office	L	0	0	1	0	0
Office	R	0	0	0.2	0.8	0
Office	U	0	0	1	0	0
Office	D	0	0	1	0	0
Hallway	L	0	0	0.8	0.2	0
Hallway	R	0	0	0	0.2	0.8
Hallway	U	0.8	0	0	0.2	0
Hallway	D	0	0	0	1	0
Dining Room	L	0	0	0	0.8	0.2
Dining Room	R	0	0	0	0	1
Dining Room	U	0	0.8	0	0	0.2
Dining Room	D	0	0	0	0	1

- Consider the action move right.
- We construct the conditional probability matrix for this action by collecting the move right rows from the table.

$$M_r = \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Conditional probability tables

We can construct a conditional probability table for each action using this large table.

X1	A1	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	L	1	0	0	0	0
Living Room	R	0.2	0.8	0	0	0
Living Room	U	1	0	0	0	0
Living Room	D	0.2	0	0	0.8	0
Kitchen	L	0.8	0.2	0	0	0
Kitchen	R	0	1	0	0	0
Kitchen	U	0	1	0	0	0
Kitchen	D	0.2	0	0	0	0.8
Office	L	0	0	1	0	0
Office	R	0	0	0.2	0.8	0
Office	U	0	0	1	0	0
Office	D	0	0	1	0	0
Hallway	L	0	0	0.8	0.2	0
Hallway	R	0	0	0	0.2	0.8
Hallway	U	0.8	0	0	0.2	0
Hallway	D	0	0	0	1	0
Dining Room	L	0	0	0	0.8	0.2
Dining Room	R	0	0	0	0	1
Dining Room	U	0	0.8	0	0	0.2
Dining Room	D	0	0	0	0	1

- Consider the action move right.
- We construct the conditional probability matrix for this action by collecting the move right rows from the table.

$$M_r = \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

Conditional probability tables

We can construct a conditional probability table for each action using this large table.

X1	A1	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	L	1	0	0	0	0
Living Room	R	0.2	0.8	0	0	0
Living Room	U	1	0	0	0	0
Living Room	D	0.2	0	0	0.8	0
Kitchen	L	0.8	0.2	0	0	0
Kitchen	R	0	1	0	0	0
Kitchen	U	0	1	0	0	0
Kitchen	D	0.2	0	0	0	0.8
Office	L	0	0	1	0	0
Office	R	0	0	0.2	0.8	0
Office	U	0	0	1	0	0
Office	D	0	0	1	0	0
Hallway	L	0	0	0.8	0.2	0
Hallway	R	0	0	0	0.2	0.8
Hallway	U	0.8	0	0	0.2	0
Hallway	D	0	0	0	1	0
Dining Room	L	0	0	0	0.8	0.2
Dining Room	R	0	0	0	0	1
Dining Room	U	0	0.8	0	0	0.2
Dining Room	D	0	0	0	0	1

- Consider the action move right.
- We construct the conditional probability matrix for this action by collecting the move right rows from the table.

$$M_r = \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

$$M_r = \begin{bmatrix} P(L|L,r) & P(K|L,r) & P(O|L,r) & P(H|L,r) & P(D|L,r) \\ P(L|K,r) & P(K|K,r) & P(O|K,r) & P(H|K,r) & P(D|K,r) \\ P(L|O,r) & P(K|O,r) & P(O|O,r) & P(H|O,r) & P(D|O,r) \\ P(L|H,r) & P(K|H,r) & P(O|H,r) & P(H|H,r) & P(D|H,r) \\ P(L|D,r) & P(K|D,r) & P(O|D,r) & P(H|D,r) & P(D|D,r) \end{bmatrix}$$

Posterior probabilities

- Suppose we start in the Office, and execute a sequence of commands $a_0, \dots a_t$.
- What should we believe about the state of the robot at time $t+1$?
- The belief state b_{t+1} represents our belief about the state of the robot at time $t+1$.
- The belief state is merely the conditional probability distribution for X_{t+1} given the initial state and all actions that have been executed.
- For each possible value, x_{t+1}, that can be assigned to X_{t+1} we want to determine:

$$P(X_{t+1} = x_{t+1} \mid a_0, \dots a_t, X_0 = x_0)$$

- How can we compute this?
- Is it necessary to do a long chain of reasoning all the way back to $t = 0$ every time we execute an action?

Posterior probabilities

- Remember the law of total probability:

$$P(A) = \sum P(A|B_t)P(B_t)$$

- We can condition everything on some context, K_t, to obtain:

$$P(A \mid K_t) = \sum P(A \mid B_t, K_t)P(B_t \mid K_t)$$

- Define the context at time t to be $K_t \triangleq a_0, \dots a_t = K_{t-1}, a_t$.
- Let B_t be the event $B_t \triangleq X_t = x_t$.
- Let A be the event $A \triangleq X_{t+1} = x_{t+1}$.
- Then we can write

$$P(X_{t+1} = x_{t+1} \mid K_t) = \sum_{x_t} P(X_{t+1} \mid X_t = x_t, K_t)P(X_t = x_t \mid K_t)$$

Posterior probabilities

- We can rewrite

$$P(X_{t+1} = x_{t+1} \mid K_t) = \sum P(X_{t+1} \mid X_t = x_t, K_t) P(X_t = x_t \mid K_t)$$

using the fact that $K_t = K_{t-1}, a_t$:

$$P(X_{t+1} = x_{t+1} \mid K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1} \mid X_t = x_t, K_{t-1}, a_t) P(X_t = x_t \mid K_{t-1}, a_t)$$

Sum over all possible values of X_t:

- Living Room, Kitchen, Office, Hallway, Dining Room

Complete history of the robot's actions to date:

- $\mathrm{K_t} = a_0, \dots a_t$

Posterior probabilities

- We can now apply our Markov assumption to the equation

$$P(X_{t+1} = x_{t+1}|K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1}|\, X_t = x_t, K_{t-1}, a_t\,)P(X_t = x_t|K_{t-1}, a_t)$$

- Given $X_t = x_t$, the next state X_{t+1} is conditionally independent of all past actions:

$$P(X_{t+1} = x_{t+1}|\, X_t = x_t, K_{t-1}, a_t\,) = P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)$$

- Furthermore, the action a_t does not affect the state X_t, and therefore

$$P(X_t = x_t|K_{t-1}, a_t) = P(X_t = x_t|K_{t-1})$$

Substitute these into the equation above, and we obtain:

$$\boldsymbol{P(X_{t+1} = x_{t+1}|K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)P(X_t = x_t|K_{t-1})}$$

Posterior probabilities

- Let's take a closer look at this result:

$$\boxed{P(X_{t+1} = x_{t+1} | K_t)} = \sum_{x_t} \boxed{P(X_{t+1} = x_{t+1} | X_t = x_t, a_t)} \boxed{P(X_t = x_t | K_{t-1})}$$

The posterior probability for state X_{t+1} given the actions $a_0, \ldots a_t$

Action model, given by a conditional probability table

Prior probability for X_t, given the actions $a_0, \ldots a_{t-1}$

How do we know $P(X_t = x_t | K_{t-1})$?

Posterior probabilities

- Let's take a closer look at this result:

$$\boldsymbol{P(X_{t+1} = x_{t+1}|K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)\boxed{P(X_t = x_t|K_{t-1})}}$$

For $t = 0$, the prior takes the form:

$$\boldsymbol{P(X_0 = x_0|K_{-1})}$$

But we never have $t = -1$, so for the base case at $t = 0$, we use the prior on initial state, $P(X_0 = x_0)$, which gives:

$$\boldsymbol{P(X_1 = x_1|K_0) = \sum_{x_0} P(X_1 = x_1|\, X_0 = x_0, a_1\,)P(X_0 = x_0)}$$

We now proceed iteratively to compute $P(X_{t+1} = x_{t+1}|K_t)$ for arbitrary t.

Posterior probabilities

- Let's take a closer look at this result:

$$\boldsymbol{P(X_{t+1} = x_{t+1}|K_t) = \sum_{\boxed{x_t}} P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)P(X_t = x_t|K_{t-1})}$$

The sum is taken over the set of all possible values for $\mathrm{x_t}$

$$\sum_{x_t \in \{L,K,O,H,D\}} P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)P(X_t = x_t|K_{t-1})$$

At time t, the state could be any of the rooms, $\{L, K, O, H, D\}$.

Posterior probabilities

- Let's take a closer look at this result:

Don't forget, a_t is hiding in $\boldsymbol{K_t}$.

$$\boldsymbol{P(X_{t+1} = \boxed{x_{t+1}} | K_t) = \sum_{x_t} P(X_{t+1} = \boxed{x_{t+1}} | X_t = x_t, \boxed{a_t}) P(X_t = x_t | K_{t-1})}$$

This equation applies to a specific action, a_t, e.g., *move up*.

If we want to know the probability distribution of X_{t+1} for a different action, e.g., *move right*, we need to use the equation again.

This equation tells us how to compute the probability that X_{t+1} is in the ***specific state,*** $\mathrm{x}_{t+1} \in \{L, K, O, H, D\}$.

To compute b_{t+1}, we would need to use this equation five time, once for each possible value for X_{t+1}.

Matrix form

- We can write the expression for $P(X_{t+1} = x_{t+1}\,|K_t)$ in a more compact form
- To keep things simple, let's use the action, *move right*, and compute the probability of arriving to the *Living Room*:

$$
\begin{aligned}
P(X_{t+1} = L|K_t) &= \textstyle\sum_{x_t} P(X_{t+1} = L|\, X_t = x_t, a_t = r\,)P(X_t = x_t|K_{t-1}) \\
&= P(X_{t+1} = L|\, X_t = L, r\,)P(X_t = L|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = K, r\,)P(X_t = K|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = O, r\,)P(X_t = O|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = H, r\,)P(X_t = H|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = D, r\,)P(X_t = D|K_{t-1})
\end{aligned}
$$

- We can write this as a simple matrix equation:

$$
P(X_{t+1} = L|K_t) = \begin{bmatrix} P(X_t = L\,|\,K_{t-1}) & P(X_t = K\,|\,K_{t-1}) & P(X_t = O\,|K_{t-1}) & P(X_t = H\,|K_{t-1}) & P(X_t = D\,|K_{t-1}) \end{bmatrix} \begin{bmatrix} P(L|L,r) \\ P(L|K,r) \\ P(L|O,r) \\ P(L|H,r) \\ P(L|D,r) \end{bmatrix}
$$

Matrix form

- We can write the expression for $P(X_{t+1} = x_{t+1} | K_t)$ in a more compact form
- To keep things simple, let's use the action, *move right*, and compute the probability of arriving to the *Living Room*:

$$
\begin{aligned}
P(X_{t+1} = L|K_t) &= \textstyle\sum_{x_t} P(X_{t+1} = L|\, X_t = x_t, a_t = r\,)P(X_t = x_t|K_{t-1}) \\
&= P(X_{t+1} = L|\, X_t = L, r\,)P(X_t = L|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = K, r\,)P(X_t = K|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = O, r\,)P(X_t = O|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = H, r\,)P(X_t = H|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = D, r\,)P(X_t = D|K_{t-1})
\end{aligned}
$$

This is merely one column from the conditional probability matrix for the action *move right*.

- We can write this as a simple matrix equation:

This row matrix is exactly the prior b_t

$$
P(X_{t+1} = L|K_t) = \begin{bmatrix} P(X_t = L \,|\, K_{t-1}) & P(X_t = K \,|\, K_{t-1}) & P(X_t = O \,|K_{t-1}) & P(X_t = H \,|K_{t-1}) & P(X_t = D \,|K_{t-1}) \end{bmatrix} \begin{bmatrix} P(L|L,r) \\ P(L|K,r) \\ P(L|O,r) \\ P(L|H,r) \\ P(L|D,r) \end{bmatrix}
$$

Matrix form

- We can write a similar expression for each state $X_{t+1} \in \{L, K, O, H, D\}$.
- We can then collect these five equations into a single matrix equation.
- Let $M_{\mathcal{A}}$ denote the conditional probability matrix for action $\mathcal{A}$.
- Recall, when $\mathcal{A}$ is *move right*, we have:

$$M_r = \begin{bmatrix} P(L|L,r) & P(K|L,r) & P(O|L,r) & P(H|L,r) & P(D|L,r) \\ P(L|K,r) & P(K|K,r) & P(O|K,r) & P(H|K,r) & P(D|K,r) \\ P(L|O,r) & P(K|O,r) & P(O|O,r) & P(H|O,r) & P(D|O,r) \\ P(L|H,r) & P(K|H,r) & P(O|H,r) & P(H|H,r) & P(D|H,r) \\ P(L|D,r) & P(K|D,r) & P(O|D,r) & P(H|D,r) & P(D|D,r) \end{bmatrix} = \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

- We can now compute $b_{t+1} = P(X_{t+1}|K_t)$ by combining these equations to obtain:

$$b_{t+1} = b_t M_r$$

Let’s take a closer look….

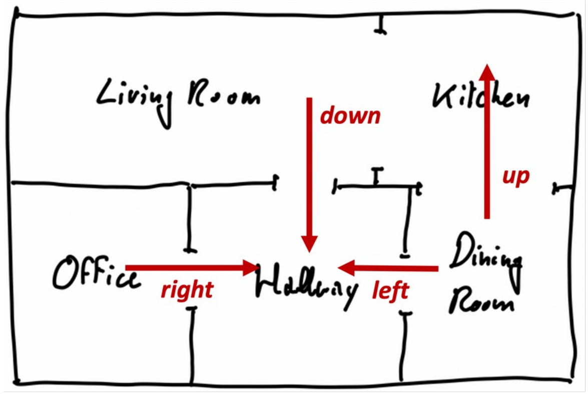
Lecture 6 ended here…

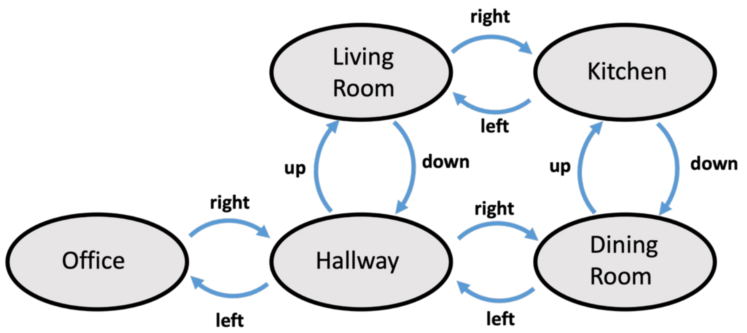
Lecture 6 recap…

Lecture 6 Recap

- Our vacuum cleaning robot has four actions:

➢ Move *left, right, up*, or *down* (relative to the map of the house)



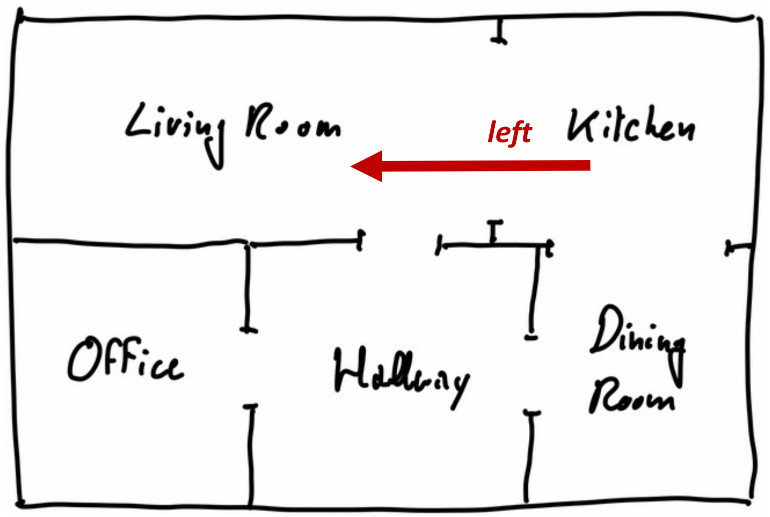


➢ We can use a large conditional probability table to track the effect of each action

X1	*A1*	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	**L**	1	0	0	0	0
Living Room	**R**	0.2	0.8	0	0	0
Living Room	**U**	1	0	0	0	0
Living Room	**D**	0.2	0	0	0.8	0
Kitchen	**L**	0.8	0.2	0	0	0
Kitchen	**R**	0	1	0	0	0
Kitchen	**U**	0	1	0	0	0
Kitchen	**D**	0.2	0	0	0	0.8
Office	**L**	0	0	1	0	0
Office	**R**	0	0	0.2	0.8	0
Office	**U**	0	0	1	0	0
Office	**D**	0	0	1	0	0
Hallway	**L**	0	0	0.8	0.2	0
Hallway	**R**	0	0	0	0.2	0.8
Hallway	**U**	0.8	0	0	0.2	0
Hallway	**D**	0	0	0	1	0
Dining Room	**L**	0	0	0	0.8	0.2
Dining Room	**R**	0	0	0	0	1
Dining Room	**U**	0	0.8	0	0	0.2
Dining Room	**D**	0	0	0	0	1

Conditional probability tables

X1	A1	Living Room	Kitchen	Office	Hallway	Dining Room
Living Room	L	1	0	0	0	0
Living Room	R	0.2	0.8	0	0	0
Living Room	U	1	0	0	0	0
Living Room	D	0.2	0	0	0.8	0
Kitchen	L	0.8	0.2	0	0	0
Kitchen	R	0	1	0	0	0
Kitchen	U	0	1	0	0	0
Kitchen	D	0.2	0	0	0	0.8
Office	L	0	0	1	0	0
Office	R	0	0	0.2	0.8	0
Office	U	0	0	1	0	0
Office	D	0	0	1	0	0
Hallway	L	0	0	0.8	0.2	0
Hallway	R	0	0	0	0.2	0.8
Hallway	U	0.8	0	0	0.2	0
Hallway	D	0	0	0	1	0
Dining Room	L	0	0	0	0.8	0.2
Dining Room	R	0	0	0	0	1
Dining Room	U	0	0.8	0	0	0.2
Dining Room	D	0	0	0	0	1



If the robot is in the ***Kitchen*** and moves ***left***, it will stay in the ***Kitchen*** (prob = 0.2) or arrive to the ***Living Room*** (prob = 0.8)

Posterior probabilities

- Given a known start location (the Office), and a sequence of commands $a_0, \dots a_t,$ we want to determine:

$$P(X_{t+1} = x_{t+1} \mid a_0, \dots a_t, X_0 = x_0)$$

- How can we compute this?

Posterior probabilities

- Given a known start location (the Office), and a sequence of commands $a_0, \dots a_t$, we want to determine:

$$P(X_{t+1} = x_{t+1} \mid \underbrace{a_0, \dots a_t}_{K_t \text{ or } K_{t-1}, a_t}, X_0 = x_0)$$

- How can we compute this?

We used various mathematical rules (slide 30) to rewrite the above equation as:

$$P(X_{t+1} = x_{t+1} | K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1} | X_t = x_t, K_{t-1}, a_t) P(X_t = x_t | K_{t-1}, a_t)$$

And then applied the Markov Property (slide 32) to further simplify the equation to:

$$\boldsymbol{P(X_{t+1} = x_{t+1} | K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1} | X_t = x_t, a_t) P(X_t = x_t | K_{t-1})}$$

Posterior probabilities

- Given a known start location (the Office), and a sequence of commands $a_0, \dots a_t,$ we want to determine:

$$P(X_{t+1} = x_{t+1} \mid \underbrace{a_0, \dots a_t}_{K_t \text{ or } K_{t-1}, a_t}, X_0 = x_0)$$

- How can we compute this?

We used various mathematical rules (slide 30) to rewrite the above equation as:

$$P(X_{t+1} = x_{t+1} | K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1} | \, X_t = x_t, \cancel{K_{t-1}}, a_t \,) P(X_t = x_t | K_{t-1}, \cancel{a_t})$$

And then applied the Markov Property (slide 32) to further simplify the equation to:

$$\boldsymbol{P(X_{t+1} = x_{t+1} | K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1} | \, X_t = x_t, a_t \,) P(X_t = x_t | K_{t-1})}$$

$$\boldsymbol{P(X_{t+1} = x_{t+1}|K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)P(X_t = x_t|K_{t-1})}$$

This equation has three components:

- $\boldsymbol{P(X_{t+1} = x_{t+1}|K_t)}$ -- the posterior (what we want to know)
- $\boldsymbol{P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)}$ – the conditional probability representing the effect of a given action $\boldsymbol{a_t}$ from a given state $\boldsymbol{x_t}$ (we have this)
- $\boldsymbol{P(X_t = x_t|K_{t-1})}$ – the prior (which is just our posterior from the previous timestep!)

$$P(X_{t+1} = x_{t+1}|K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1}|\, X_t = x_t, a_t\,)P(X_t = x_t|K_{t-1})$$

This equation has three components:

- $P(X_{t+1} = x_{t+1}|K_t)$ -- the posterior (what we want to know)
- $P(X_{t+1} = x_{t+1}|\, X_t$ … esenting the effect of a giv … e this)
- $P(X_t = x_t|K_{t-1})$ – the prior (which is just our posterior from the previous timestep!)

These two probabilities are the same, just at different timesteps

Calculating the posterior probability over time

Special case at t=0, no actions taken yet, so all we can use to infer the belief is the prior probability. In our example it's [0 0 1 0 0], that the robot begins in the office.

$$P(X_0 = x_0)$$

$$P(X_1 = x_1 | a_0) = \sum_{x_0} P(X_1 = x_1 | X_0 = x_0, a_0) P(X_0 = x_0)$$

$$P(X_2 = x_2 | a_0, a_1) = \sum_{x_1} P(X_2 = x_2 | X_1 = x_1, a_1) P(X_1 = x_1 | a_0)$$

$$P(X_3 = x_3 | a_0, a_1, a_2) = \sum_{x_2} P(X_3 = x_3 | X_2 = x_2, a_2) P(X_2 = x_2 | a_0, a_1)$$

…

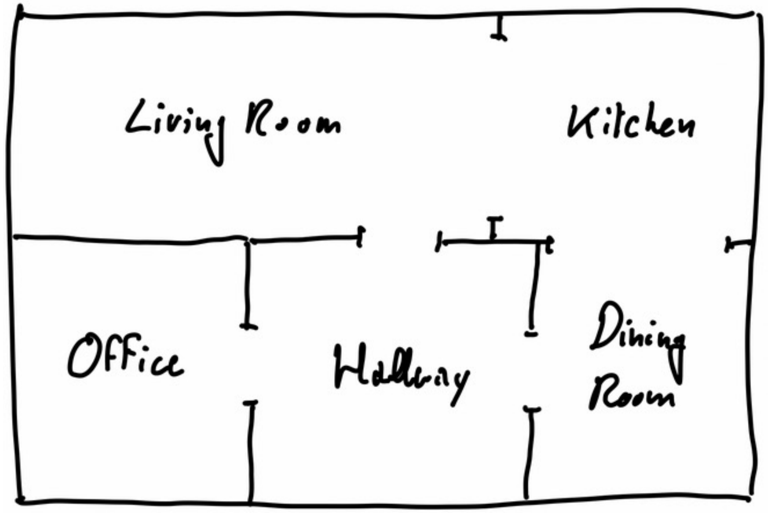
$$P(X_{t+1} = x_{t+1} | K_t) = \sum_{x_t} P(X_{t+1} = x_{t+1} | X_t = x_t, a_t) P(X_t = x_t | K_{t-1})$$

Calculating the posterior probability over time

$$P(X_0 = O) = 1$$

Special case at t=0, no actions taken yet, so all we can use to infer the belief is the prior probability. In our example it's [0 0 1 0 0], that the robot begins in the office.

Probability that the robot is in the Office (O) at timestep 0.



Full belief:

$$b_0 = [\ 0\ \ 0\ \ 1\ \ 0\ \ 0]$$

Calculating the posterior probability over time

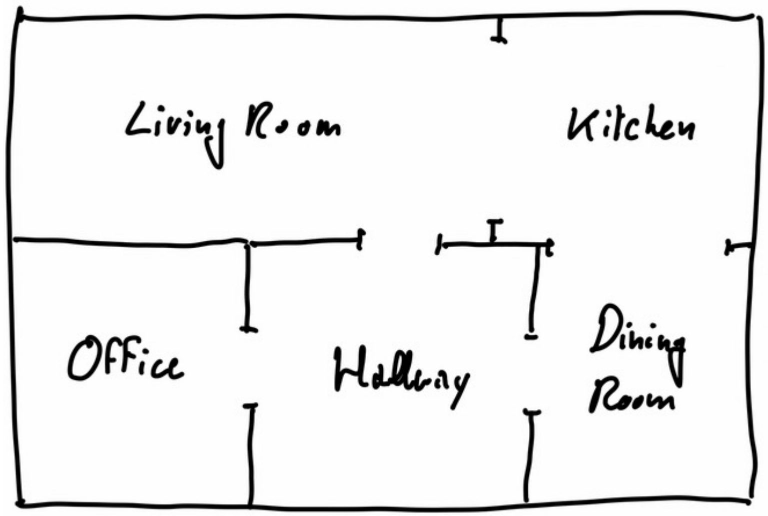
Calculations for timestep 1. What is the probability that state at timestep 1 (X_1) is "Hallway" (H) given that the action at timestep 0 was "Right" (R).

$$P(X_0 = O) = 1$$

$$P(X_1 = H|R) = \sum_{x_0} P(X_1 = H|\, X_0 = x_0, R\,)P(X_0 = x_0)$$

Unrolling the sigma...

$$= P(X_1 = H|\, X_0 = O, R\,)P(X_0 = O) + P(X_1 = H|\, X_0 = H, R\,)P(X_0 = H) +$$
$$P(X_1 = H|\, X_0 = D, R\,)P(X_0 = D) +$$
$$P(X_1 = H|\, X_0 = K, R\,)P(X_0 = K) + P(X_1 = H|\, X_0 = L, R\,)P(X_0 = L)$$



$$P(X_1 = H|R) = 0.32$$

If we do this calculation for all rooms, then the resulting belief is:

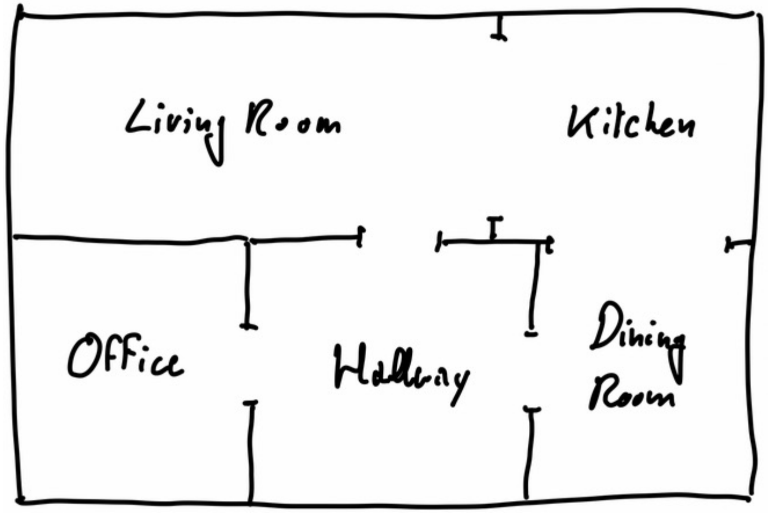
$$b_1 = [0 \quad 0 \quad 0.2 \;\; 0.8 \quad 0]$$

Calculating the posterior probability over time

Given $b_1 = [0 \quad 0 \quad 0.2 \;\; 0.8 \;\; 0]$, calculate $P(X_2 = H|R)$ given that the action at timestep 1 was again "Right" (R).

$$\boldsymbol{P(X_2 = H|R) = \sum_{x_1} P(X_2 = H|\, X_1 = x_1, R\,)P(X_1 = x_1)}$$

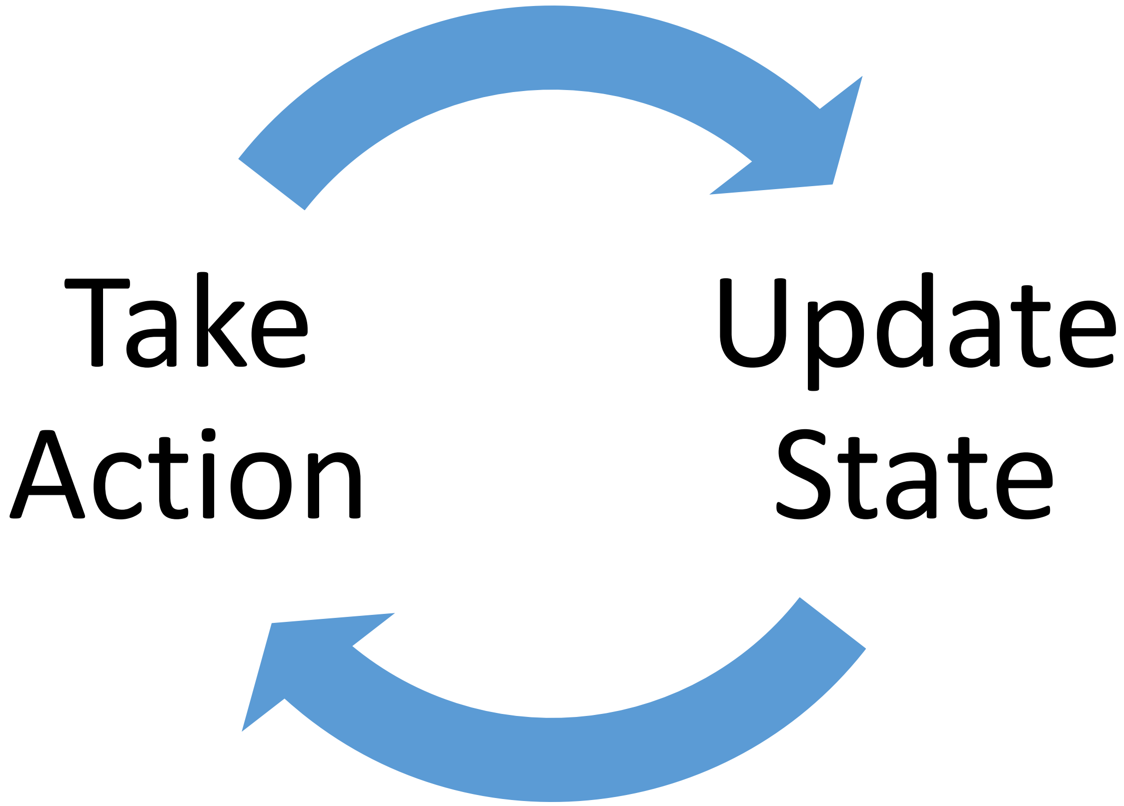
$$\boldsymbol{= P(X_2 = H|\, X_1 = O, R\,)P(X_1 = O) +}$$
$$\boldsymbol{P(X_2 = H|\, X_1 = H, R\,)P(X_1 = H) +}$$
$$\boldsymbol{P(X_2 = H|\, X_1 = D, R\,)P(X_1 = D) +}$$
$$\boldsymbol{P(X_2 = H|\, X_1 = K, R\,)P(X_1 = K) +}$$
$$\boldsymbol{P(X_2 = H|\, X_1 = L, R\,)P(X_1 = L)}$$



$P(X_2 = H|R) = 0.8$

If we do this calculation for all rooms, then the resulting belief is:

$b_1 = [0 \quad 0 \quad 0.04 \quad 0.32 \quad 0.64]$


Take
Action
Update
State

How do we compute this sum efficiently?

$$\boldsymbol{P}(\boldsymbol{X_{t+1}} = \boldsymbol{x_{t+1}} | \boldsymbol{K_t}) = \sum_{\boldsymbol{x_t}} \boldsymbol{P}(\boldsymbol{X_{t+1}} = \boldsymbol{x_{t+1}} | \, \boldsymbol{X_t} = \boldsymbol{x_t}, \boldsymbol{a_t}\,) \boldsymbol{P}(\boldsymbol{X_t} = \boldsymbol{x_t} | \boldsymbol{K_{t-1}})$$

Matrix form

- We can write the expression for $P(X_{t+1} = x_{t+1}\,|K_t)$ in a more compact form
- To keep things simple, let's use the action, *move right*, and compute the probability of arriving to the *Living Room*:

$$
\begin{aligned}
P(X_{t+1} = L|K_t) &= \sum_{x_t} P(X_{t+1} = L|\, X_t = x_t, a_t = r\,)P(X_t = x_t|K_{t-1}) \\
&= P(X_{t+1} = L|\, X_t = L, r\,)P(X_t = L|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = K, r\,)P(X_t = K|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = O, r\,)P(X_t = O|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = H, r\,)P(X_t = H|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = D, r\,)P(X_t = D|K_{t-1})
\end{aligned}
$$

- We can write this as a simple matrix equation:

$$
P(X_{t+1} = L|K_t) = \begin{bmatrix} P(X_t = L\,|\,K_{t-1}) & P(X_t = K\,|\,K_{t-1}) & P(X_t = O\,|K_{t-1}) & P(X_t = H\,|K_{t-1}) & P(X_t = D\,|K_{t-1}) \end{bmatrix} \begin{bmatrix} P(L|L,r) \\ P(L|K,r) \\ P(L|O,r) \\ P(L|H,r) \\ P(L|D,r) \end{bmatrix}
$$

Matrix form

- We can write the expression for $P(X_{t+1} = x_{t+1} \,|K_t)$ in a more compact form
- To keep things simple, let's use the action, *move right*, and compute the probability of arriving to the *Living Room*:

$$
\begin{aligned}
P(X_{t+1} = L|K_t) &= \textstyle\sum_{x_t} P(X_{t+1} = L|\, X_t = x_t, a_t = r\,)P(X_t = x_t|K_{t-1}) \\
&= P(X_{t+1} = L|\, X_t = L, r\,)P(X_t = L|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = K, r\,)P(X_t = K|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = O, r\,)P(X_t = O|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = H, r\,)P(X_t = H|K_{t-1}) + \\
&\quad P(X_{t+1} = L|\, X_t = D, r\,)P(X_t = D|K_{t-1})
\end{aligned}
$$

This is merely one column from the conditional probability matrix for the action *move right*.

- We can write this as a simple matrix equation:

This row matrix is exactly the prior b_t

$$
P(X_{t+1} = L|K_t) = \begin{bmatrix} P(X_t = L\,|\,K_{t-1}) & P(X_t = K\,|\,K_{t-1}) & P(X_t = O\,|K_{t-1}) & P(X_t = H\,|K_{t-1}) & P(X_t = D\,|K_{t-1}) \end{bmatrix} \begin{bmatrix} P(L|L,r) \\ P(L|K,r) \\ P(L|O,r) \\ P(L|H,r) \\ P(L|D,r) \end{bmatrix}
$$

Matrix form

- We can write a similar expression for each state $X_{t+1} \in \{L, K, O, H, D\}$.
- We can then collect these five equations into a single matrix equation.
- Let $M_{\mathcal{A}}$ denote the conditional probability matrix for action $\mathcal{A}$.
- Recall, when $\mathcal{A}$ is *move right*, we have:

$$M_r = \begin{bmatrix} P(L|L,r) & P(K|L,r) & P(O|L,r) & P(H|L,r) & P(D|L,r) \\ P(L|K,r) & P(K|K,r) & P(O|K,r) & P(H|K,r) & P(D|K,r) \\ P(L|O,r) & P(K|O,r) & P(O|O,r) & P(H|O,r) & P(D|O,r) \\ P(L|H,r) & P(K|H,r) & P(O|H,r) & P(H|H,r) & P(D|H,r) \\ P(L|D,r) & P(K|D,r) & P(O|D,r) & P(H|D,r) & P(D|D,r) \end{bmatrix} = \begin{bmatrix} 0.2 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

- We can now compute $b_{t+1} = P(X_{t+1}|K_t)$ by combining these equations to obtain:

$$b_{t+1} = b_t M_r$$

Let's take a closer look….

Calculating the belief state

$b_{t+1} = b_t M_r =$

$$[P(X_t = L \mid K_{t-1}) \quad P(X_t = K \mid K_{t-1}) \quad P(X_t = O \mid K_{t-1}) \quad P(X_t = H \mid K_{t-1}) \quad P(X_t = D \mid K_{t-1})] \begin{bmatrix} P(L|L,r) & P(K|L,r) & P(O|L,r) & P(H|L,r) & P(D|L,r) \\ P(L|K,r) & P(K|K,r) & P(O|K,r) & P(H|K,r) & P(D|K,r) \\ P(L|O,r) & P(K|O,r) & P(O|O,r) & P(H|O,r) & P(D|O,r) \\ P(L|H,r) & P(K|H,r) & P(O|H,r) & P(H|H,r) & P(D|H,r) \\ P(L|D,r) & P(K|D,r) & P(O|D,r) & P(H|D,r) & P(D|D,r) \end{bmatrix}$$

Let's look at the first entry of b_{t+1} --- the product of b_t and the first column of M_r.

Calculating the belief state

$b_{t+1} = b_t M_r =$

$$[P(X_t = L \mid K_{t-1}) \quad P(X_t = K \mid K_{t-1}) \quad P(X_t = O \mid K_{t-1}) \quad P(X_t = H \mid K_{t-1}) \quad P(X_t = D \mid K_{t-1})] \begin{bmatrix} P(L|L,r) & P(K|L,r) & P(O|L,r) & P(H|L,r) & P(D|L,r) \\ P(L|K,r) & P(K|K,r) & P(O|K,r) & P(H|K,r) & P(D|K,r) \\ P(L|O,r) & P(K|O,r) & P(O|O,r) & P(H|O,r) & P(D|O,r) \\ P(L|H,r) & P(K|H,r) & P(O|H,r) & P(H|H,r) & P(D|H,r) \\ P(L|D,r) & P(K|D,r) & P(O|D,r) & P(H|D,r) & P(D|D,r) \end{bmatrix}$$

$$\begin{aligned} \rightarrow\ & P(X_{t+1} = L \mid X_t = L, r)P(X_t = L \mid K_{t-1}) + \\ & P(X_{t+1} = L \mid X_t = K, r)P(X_t = K \mid K_{t-1}) + \\ & P(X_{t+1} = L \mid X_t = O, r)P(X_t = O \mid K_{t-1}) + \\ & P(X_{t+1} = L \mid X_t = H, r)P(X_t = H \mid K_{t-1}) + \\ & P(X_{t+1} = L \mid X_t = D, r)P(X_t = D \mid K_{t-1}) \end{aligned}$$

- ***This is exactly the computation we performed above!***
- ***This works for each entry of b_{t+1}.***

Calculating the belief state

If we execute action $\mathcal{A}$ at time t, the belief state, $\boldsymbol{b_{t+1} = P(X_{t+1} = x_{t+1} \mid K_t)}$, is given by

$$\boldsymbol{b_{t+1} = b_t M_{\mathcal{A}}}$$

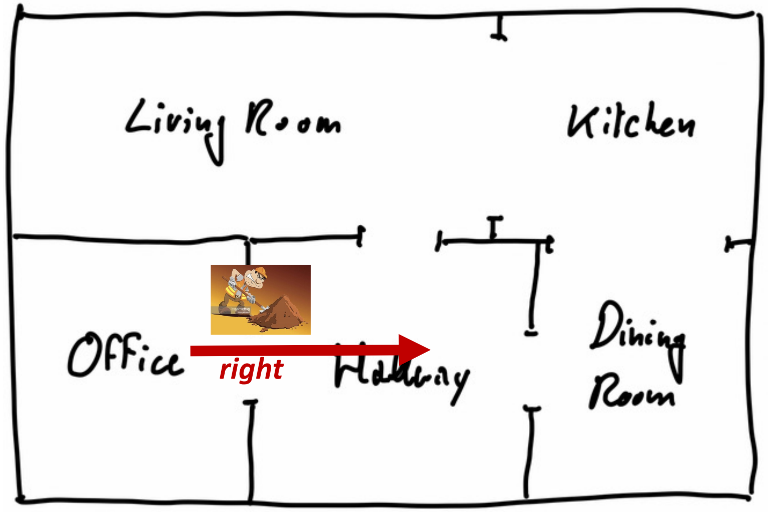
in which $\boldsymbol{M_{\mathcal{A}}}$ is the conditional probability matrix for action $\mathcal{A}$ and $\boldsymbol{b_t}$ is the belief state at time $\boldsymbol{t}$.

Example: Move Right

As we have seen above, if we execute the command move right from the initial state, $x_0 = Office$, we obtain

$$b_1 = b_0 M_r = [0\ 0\ 1\ 0\ 0]\begin{bmatrix} 0.2 & 0.80 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} = [0.0 \quad 0.0 \quad 0.2 \quad 0.8 \quad 0.0]$$

$$= [P(X_1 = L \mid r, X_0 = O), \quad P(X_1 = K \mid r, X_0 = O) \quad P(X_1 = O \mid r, X_0 = O), \quad P(X_1 = H \mid r, X_0 = O) \quad P(X_1 = D \mid r, X_0 = O)]$$

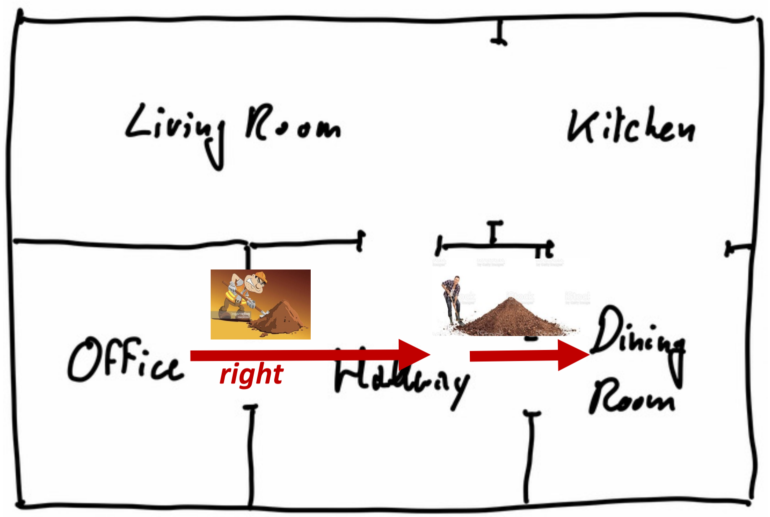


- You can imagine probability mass being pushed to the right by a sloppy worker.
- Only 80% of the probability arrives to the Hallway.

Move right multiple times

If we now again execute the action move right at time $t = 1$, we obtain

$$b_2 = b_1 M_r = [0.0 \quad 0.0 \quad 0.2 \quad 0.8 \quad 0.0] \begin{bmatrix} 0.2 & 0.80 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} = [0.0 \quad 0.0 \quad 0.04 \quad 0.32 \quad 0.64]$$



Move right multiple times

If we now again execute the action move right at time $t = 1$, we obtain

$$b_2 = b_1 M_r = [0.0 \quad 0.0 \quad 0.2 \quad 0.8 \quad 0.0] \begin{bmatrix} 0.2 & 0.80 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} = [0.0 \quad 0.0 \quad 0.04 \quad 0.32 \quad 0.64]$$

If we execute the action *move right* n times in succession, we obtain

$$b_n = b_0 M_r^n$$

- ***As you can imagine, there's a very beautiful theory to systems like this – a combination of linear algebra and probability theory.***

Markov chains

- Suppose we have chosen a specific sequence of actions: $a_0, \dots a_n$
- At stage $t+1$, we compute the belief b_{t+1} using conditional probability matrix M_{a_t} and the prior belief b_t:

$$b_{t+1} = b_t M_{a_t} = \underbrace{\underbrace{\underbrace{\underbrace{b_0 M_{a_0}}_{b_1} M_{a_1}}_{b_2} M_{a_2}}_{b_3} \dots M_{a_{t-1}}}_{b_t} M_{a_t}$$

Recall that b_0 is the initial distribution for state, in our example scenario:

$$b_0 = [0\ 0\ 1\ 0\ 0]$$

At any time t, all of the available information about the history of the robot (where it has been, what it has done) is contained in the belief state b_t.

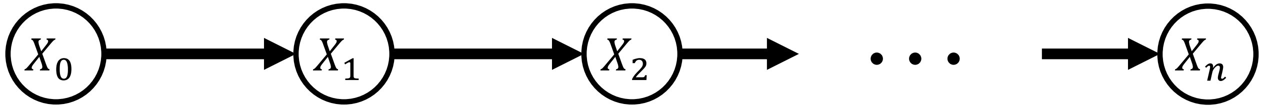
If you know b_t, learning specific previous actions does not add information.

Markov chains

- A sequence of random variables $X_0, \dots X_n$ is a ***Markov chain*** if the Markov property holds

$$P(X_{t+1} | X_t, X_{t-1}, \dots X_0) = P(X_{t+1} | X_t)$$

- In our case, we have a fixed action sequence $a_0 \dots a_t$, which defines the distributions for each of the X_i.
- For a fixed sequence of actions, the state of our vacuum cleaning robot forms a Markov chain.
- A Markov chain has a simple graphical representation:



$$P(X_{t+1} | a_0 \dots a_t, x_0) \rightarrow b_{t+1} = b_t M_{a_t}$$

Each node includes the distribution b_t and each arc corresponds to the computation $b_{t-1} M_{a_t}$

Controlled Markov chains

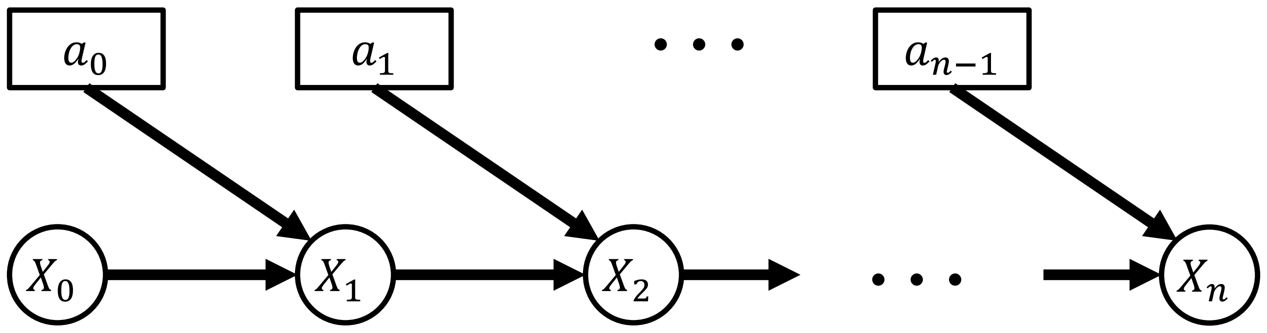
- So far, in our discussions about the Markov chain $X_0, \dots X_t$, we have been careful to always add the phrase *"for a fixed action sequence $a_0 \dots a_t$."*
- We can think of the actions, $a_0 \dots a_t$, as ***control inputs*** to the system.

➢ Our choice of $a_0 \dots a_t$ controls how the system evolves.

➢ We don't control the actual state X_t, but we do control which conditional probability matrix is used to update the belief state.

❑ We call this kind of process a ***controlled Markov chain***.

- A controlled Markov chain also has a nice graphical representation:



Note:
States are ***random*** – circles.
Actions are ***deterministic*** – boxes.