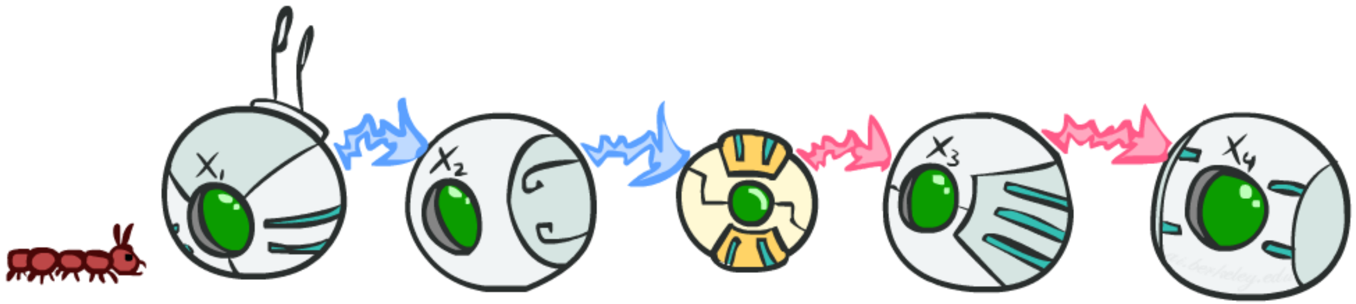
CS 5522: Artificial Intelligence II

Markov Models



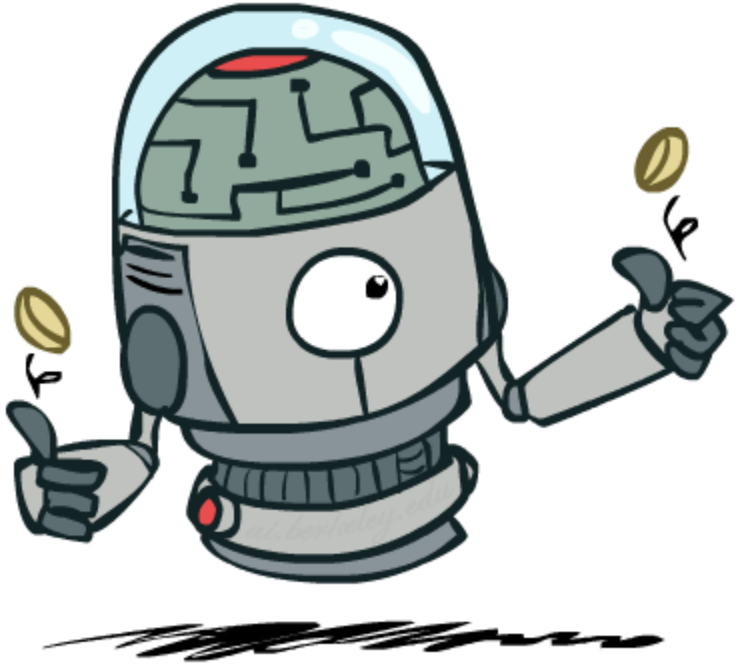
Instructor: Alan Ritter

Ohio State University

[These slides were adapted from CS188 Intro to AI at UC Berkeley. All materials available at http://ai.berkeley.edu.]

Independence

- Two variables are *independent* in a joint distribution if:

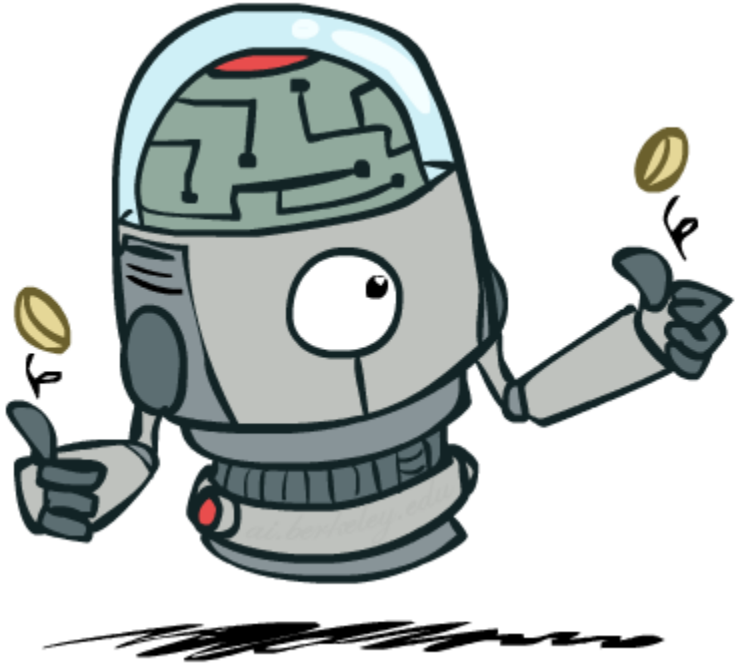


Independence

- Two variables are *independent* in a joint distribution if:

$$P(X,Y) = P(X)P(Y)$$

$$\forall x, y\, P(x,y) = P(x)P(y)$$



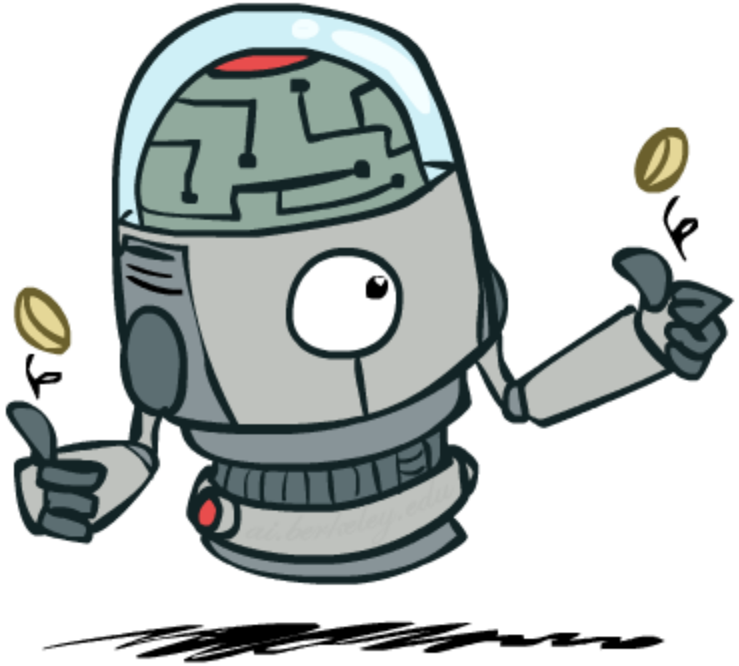
Independence

- Two variables are *independent* in a joint distribution if:

$$P(X, Y) = P(X)P(Y)$$

$$\forall x, y\, P(x, y) = P(x)P(y)$$

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



Independence

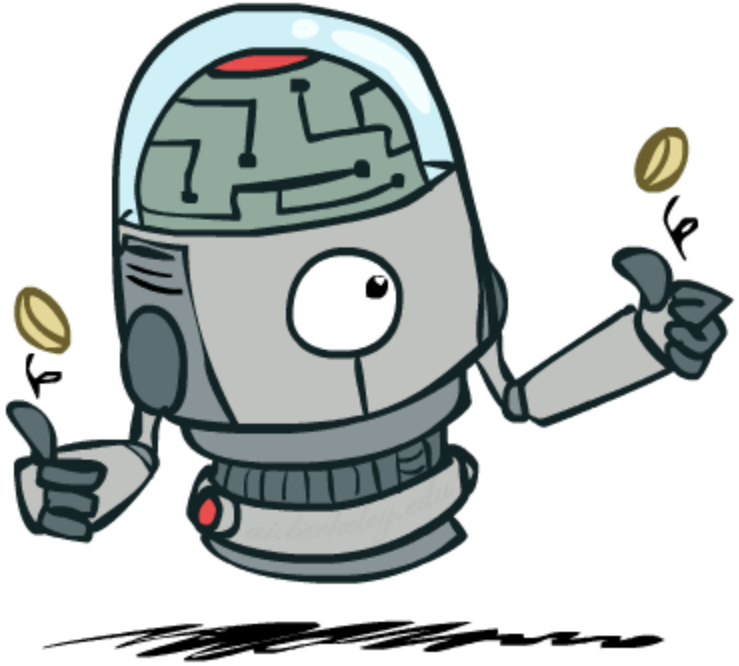
- Two variables are *independent* in a joint distribution if:

$$P(X,Y) = P(X)P(Y)$$

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

$$\forall x, y\, P(x,y) = P(x)P(y)$$

  - Says the joint distribution *factors* into a product of two simple ones
  - Usually variables aren't independent!



Independence

- Two variables are *independent* in a joint distribution if:

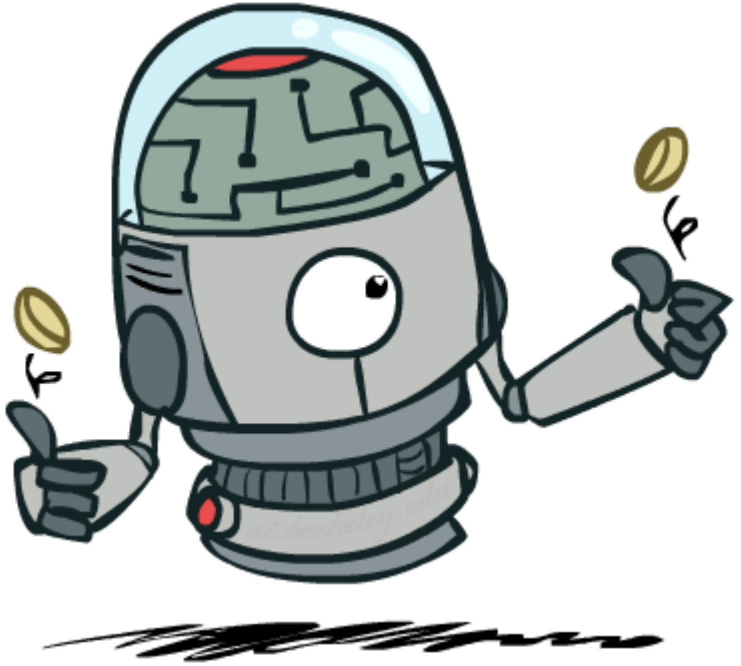
$$P(X,Y) = P(X)P(Y)$$

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

$$\forall x,y\, P(x,y) = P(x)P(y)$$

  - Says the joint distribution *factors* into a product of two simple ones
  - Usually variables aren't independent!

- Can use independence as a *modeling assumption*



Independence

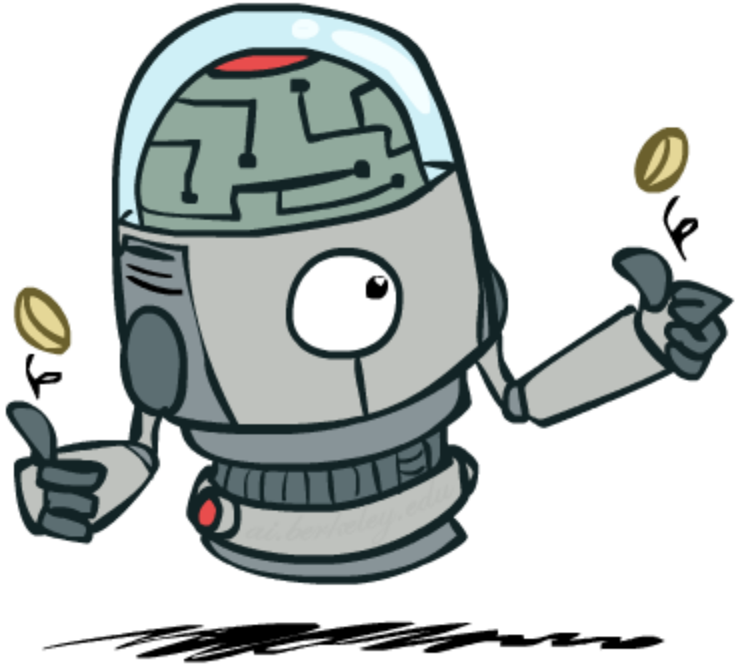
- Two variables are *independent* in a joint distribution if:

$$P(X,Y) = P(X)P(Y) \qquad X \perp\!\!\!\perp Y$$

$$\forall x, y \, P(x,y) = P(x)P(y)$$

  - Says the joint distribution *factors* into a product of two simple ones
  - Usually variables aren't independent!

- Can use independence as a *modeling assumption*
  - Independence can be a simplifying assumption
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity}?



Example: Independence?

$$P_1(T, W)$$

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

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

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

Example: Independence?

$P_1(T, W)$

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

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

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

$P_2(T, W) = P(T)P(W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

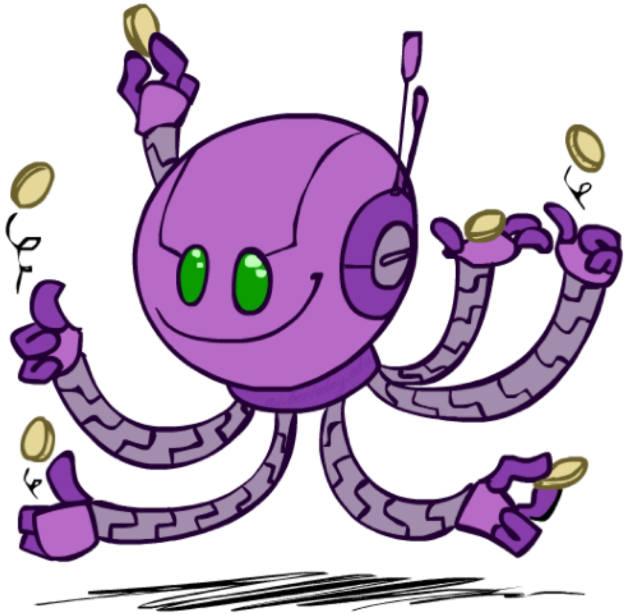
$P(X_2)$

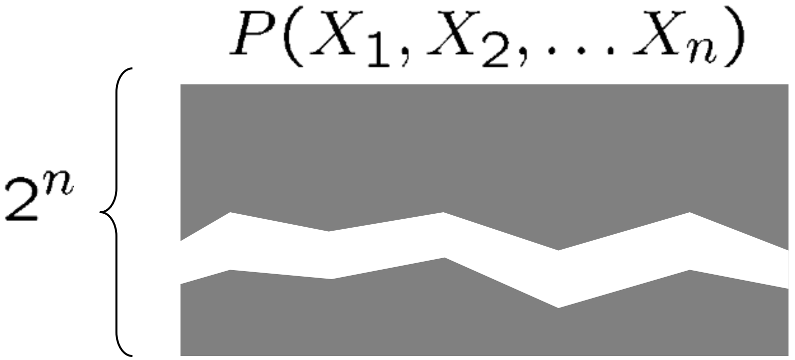
H	0.5
T	0.5

$\ldots$

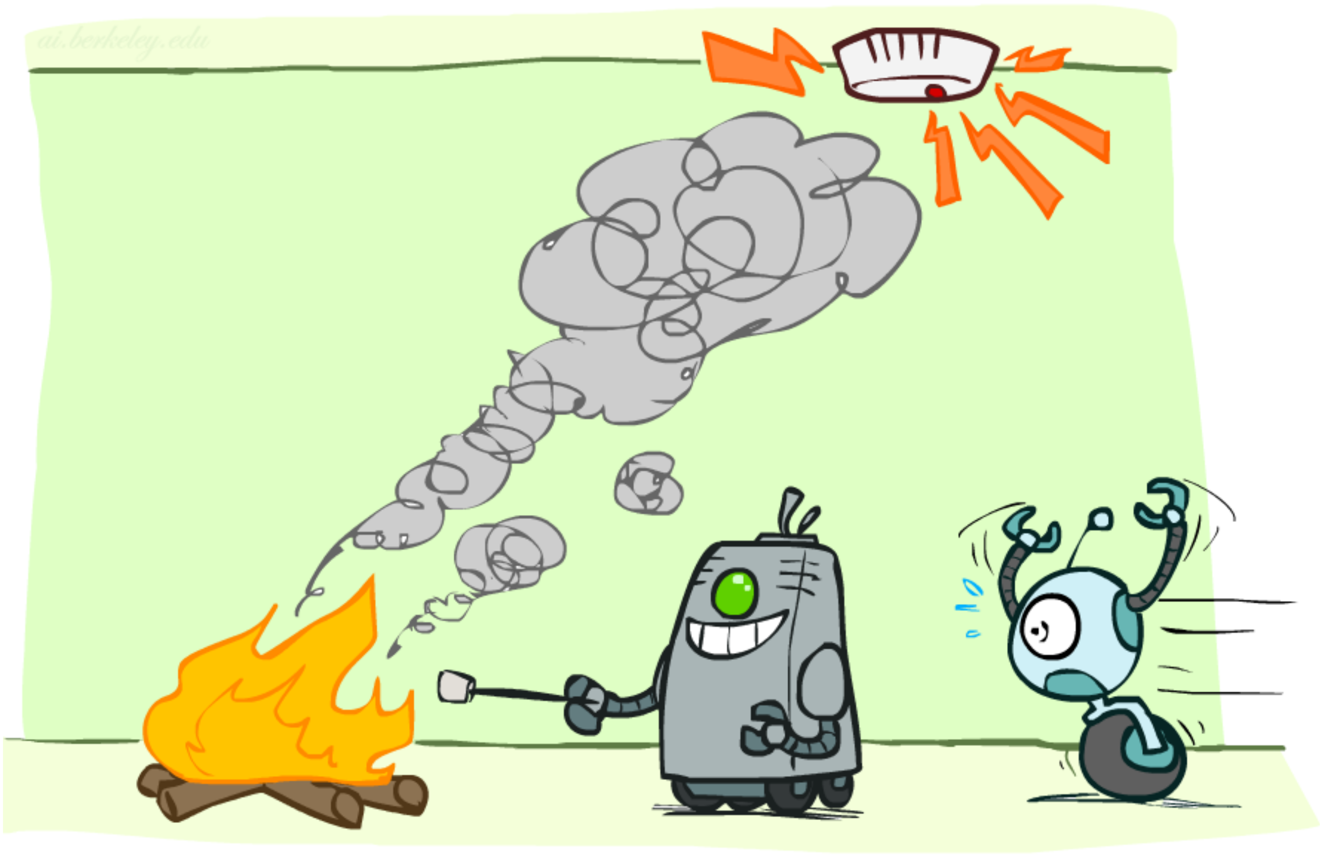
$P(X_n)$

H	0.5
T	0.5



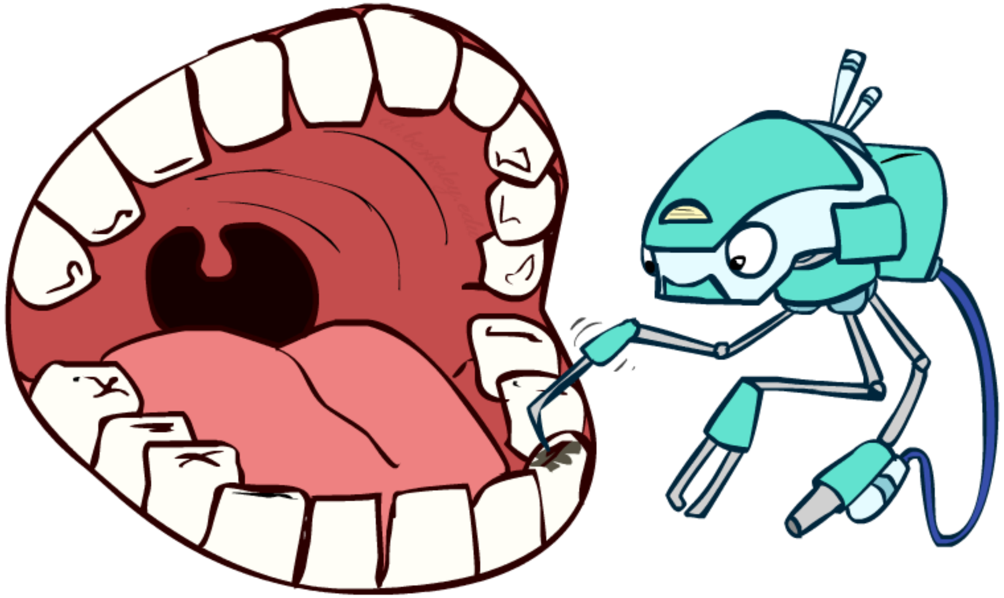


Conditional Independence



Conditional Independence

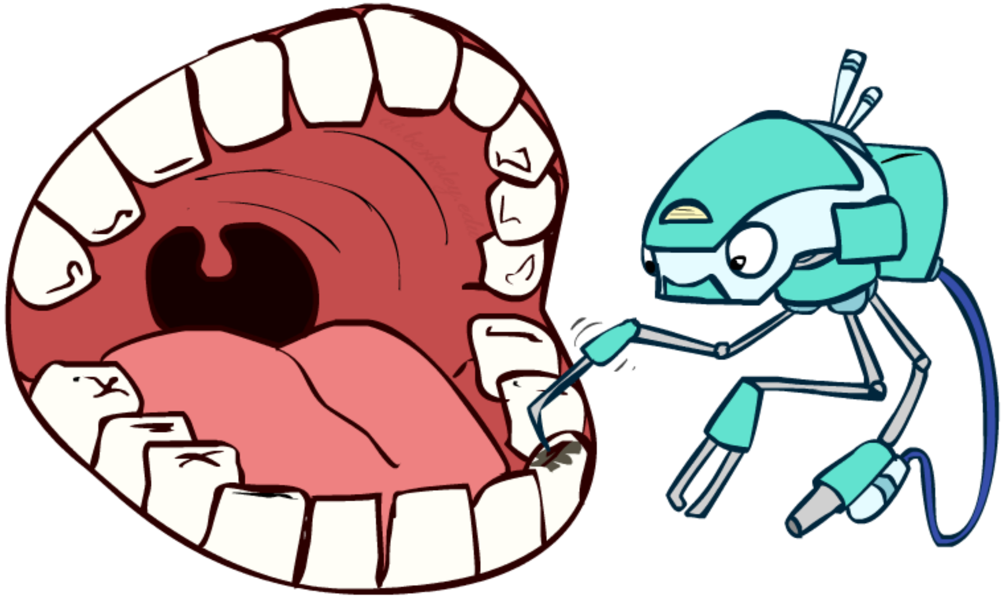
- P(Toothache, Cavity, Catch)



Conditional Independence

- P(Toothache, Cavity, Catch)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

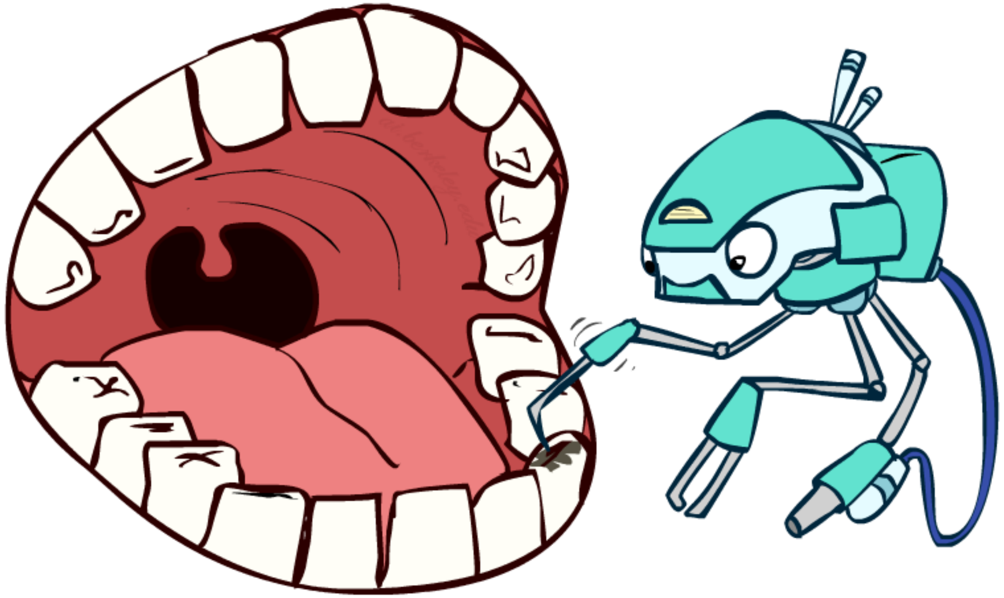


Conditional Independence

- P(Toothache, Cavity, Catch)

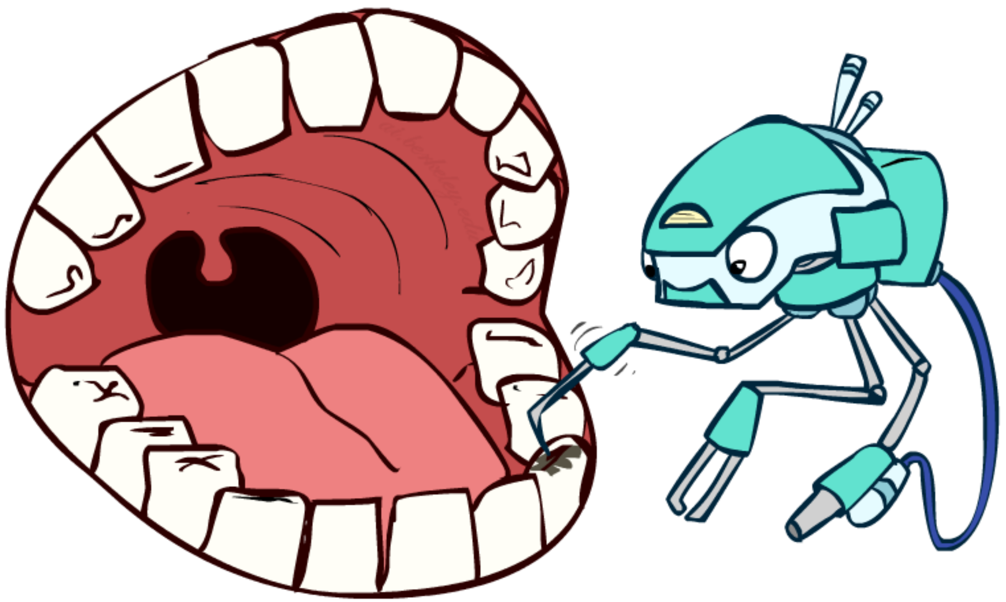
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

- The same independence holds if I don’t have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)



Conditional Independence

- P(Toothache, Cavity, Catch)
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)
- The same independence holds if I don’t have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)
- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)



Conditional Independence

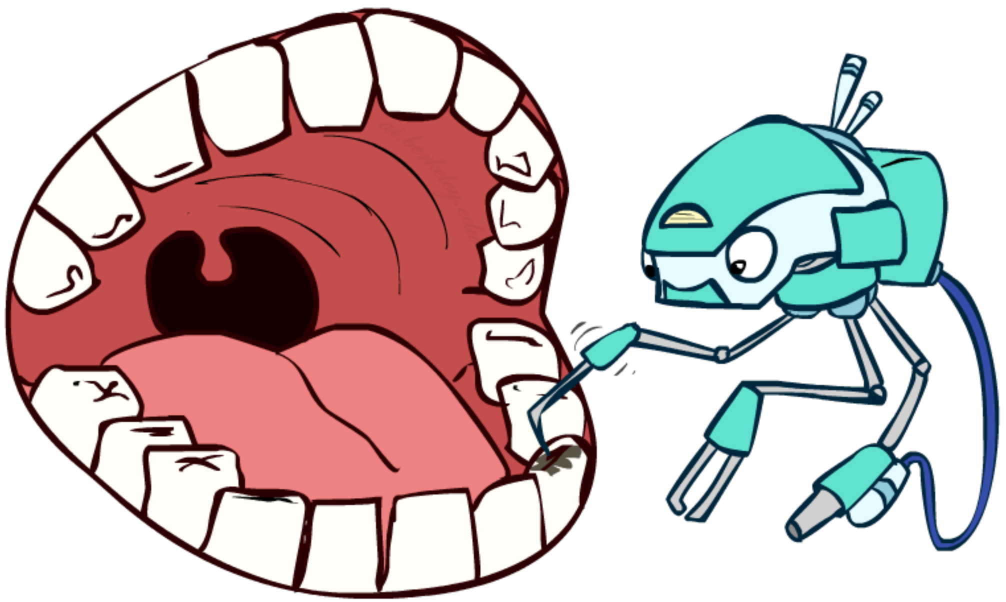
- P(Toothache, Cavity, Catch)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)

- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)

- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily



Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.
- X is conditionally independent of Y given Z $X \perp\!\!\!\perp Y | Z$

  if and only if:

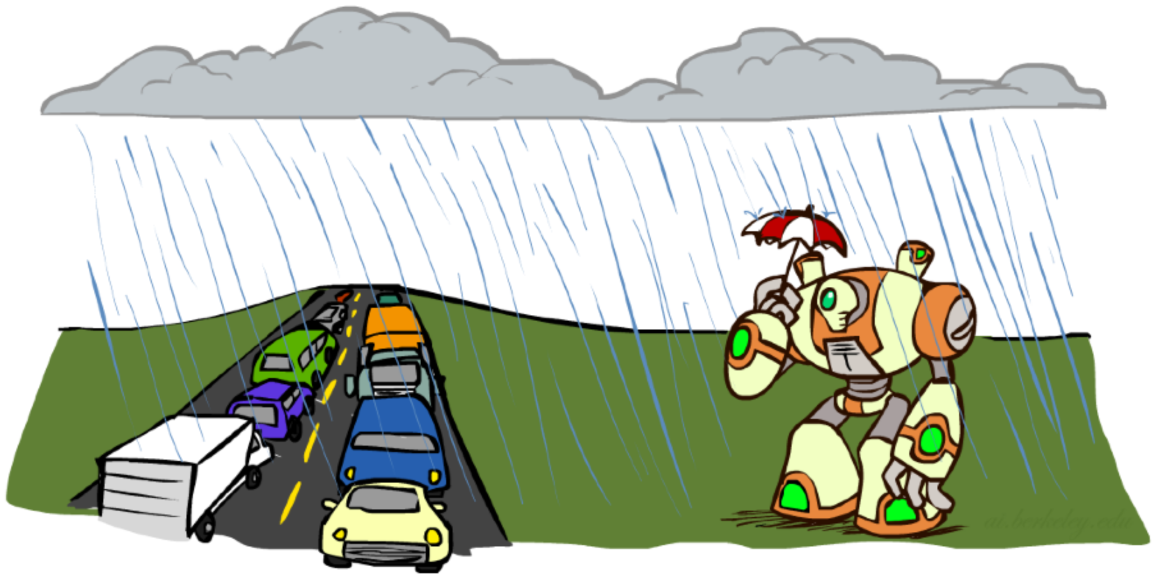
  $$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

  or, equivalently, if and only if

  $$\forall x, y, z : P(x|z, y) = P(x|z)$$

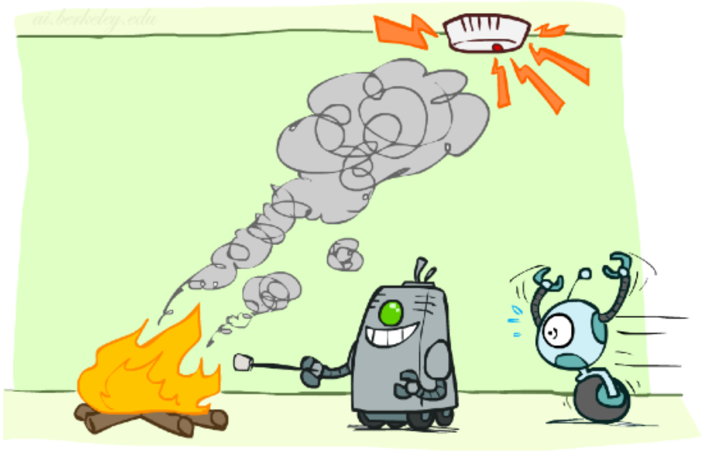
Conditional Independence

- What about this domain:
  - Traffic
  - Umbrella
  - Raining



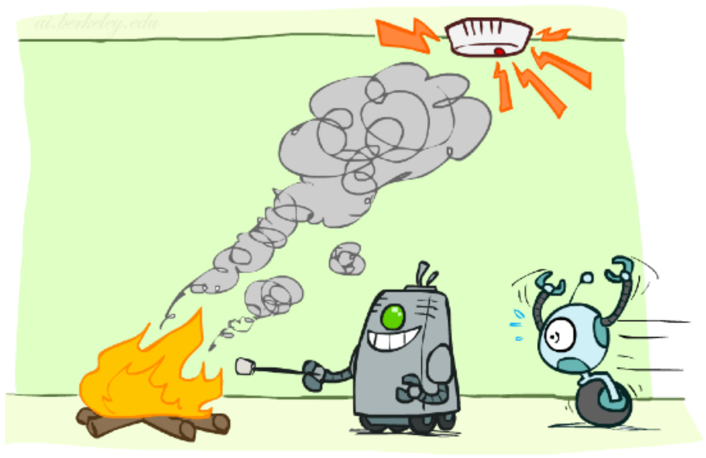
Conditional Independence

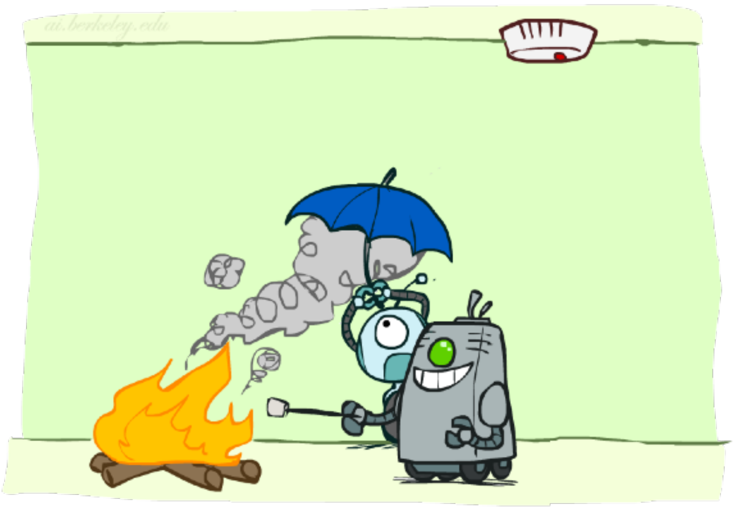
- What about this domain:
  - Fire
  - Smoke
  - Alarm



Conditional Independence

- What about this domain:
  - Fire
  - Smoke
  - Alarm





Probability Recap

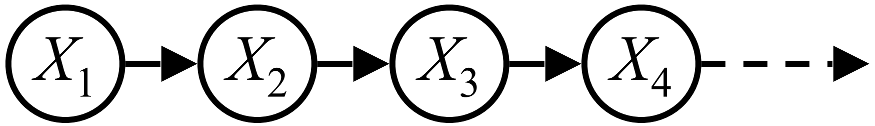
- Conditional probability $$P(x|y) = \frac{P(x,y)}{P(y)}$$
- Product rule $$P(x,y) = P(x|y)P(y)$$
- Chain rule $$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$
- X, Y independent if and only if: $\forall x, y : P(x,y) = P(x)P(y)$
- X and Y are conditionally independent given Z if and only if: $X \perp\!\!\!\perp Y | Z$
  $$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

Reasoning over Time or Space

- Often, we want to reason about a sequence of observations
  - Speech recognition
  - Robot localization
  - User attention
  - Medical monitoring
- Need to introduce time (or space) into our models

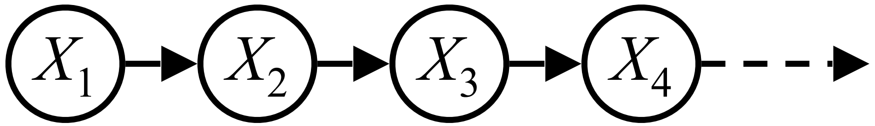
Markov Models

- Value of X at a given time is called the state



Markov Models

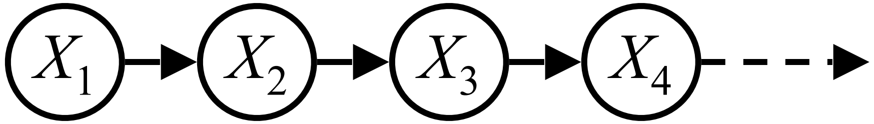
- Value of X at a given time is called the state



$P(X_1)$ $\quad$ $P(X_t|X_{t-1})$

Markov Models

- Value of X at a given time is called the state

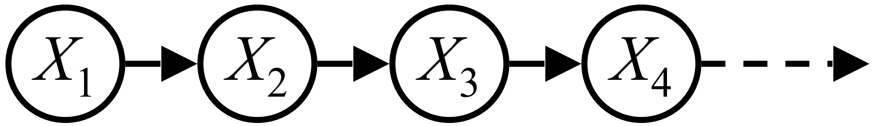


$$P(X_1) \qquad P(X_t|X_{t-1})$$

- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial state probabilities)

Markov Models

- Value of X at a given time is called the state

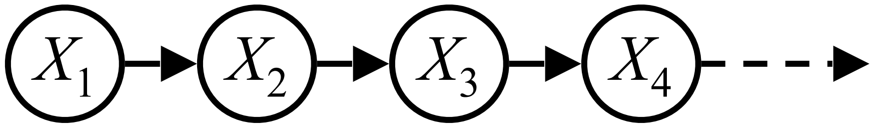


$P(X_1)$ $\qquad$ $P(X_t|X_{t-1})$

- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial state probabilities)
- Stationarity assumption: transition probabilities the same at all times

Markov Models

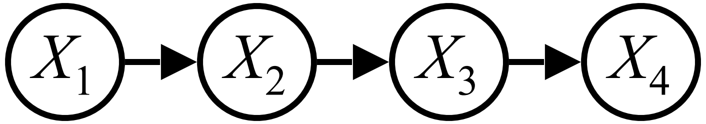
- Value of X at a given time is called the state



$$P(X_1) \qquad P(X_t|X_{t-1})$$

- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial state probabilities)
- Stationarity assumption: transition probabilities the same at all times
- Same as MDP transition model, but no choice of action

Joint Distribution of a Markov Model



$P(X_1)$ $\qquad$ $P(X_t|X_{t-1})$

- Joint distribution:

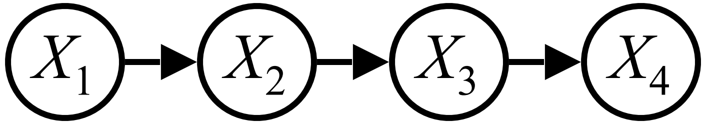
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

- More generally:

$$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1})$$

$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

Joint Distribution of a Markov Model



$P(X_1)$ $\qquad$ $P(X_t|X_{t-1})$

- Joint distribution:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

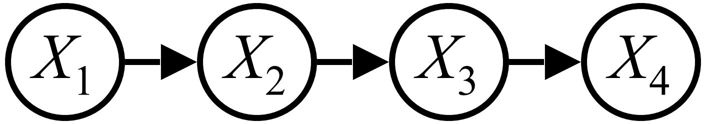
- More generally:

$$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1})$$

$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

- Questions to be resolved:

Joint Distribution of a Markov Model



$P(X_1)$ $\qquad$ $P(X_t|X_{t-1})$

- Joint distribution:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

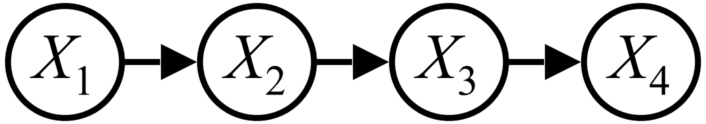
- More generally:

$$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1})$$

$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

- Questions to be resolved:
  - Does this indeed define a joint distribution?

Joint Distribution of a Markov Model



$P(X_1)$ $\qquad$ $P(X_t|X_{t-1})$

- Joint distribution:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

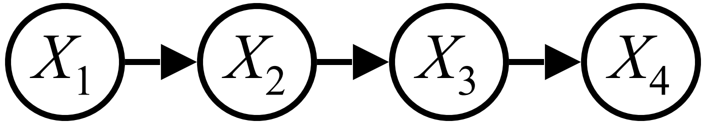
- More generally:

$$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1})$$

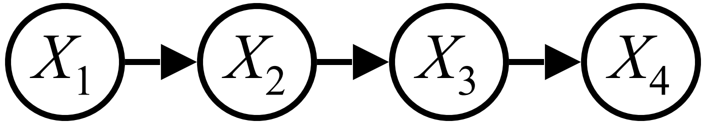
$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

- Questions to be resolved:
  - Does this indeed define a joint distribution?
  - Can every joint distribution be factored this way, or are we making some assumptions about the joint distribution by using this factorization?

Chain Rule and Markov Models



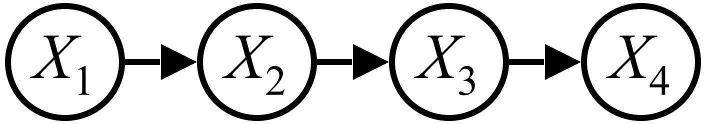
Chain Rule and Markov Models



- From the chain rule, every joint distribution over X_1, X_2, X_3, X_4 can be written as:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

Chain Rule and Markov Models

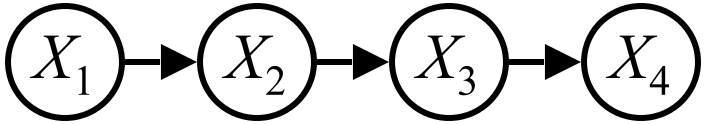


- From the chain rule, every joint distribution over X_1, X_2, X_3, X_4 can be written as:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

- Assuming that $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

Chain Rule and Markov Models

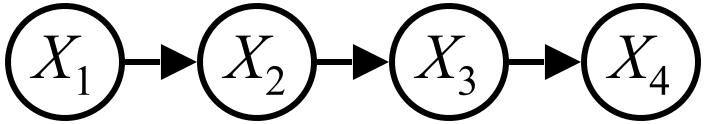


- From the chain rule, every joint distribution over X_1, X_2, X_3, X_4 can be written as:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

- Assuming that $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

Chain Rule and Markov Models



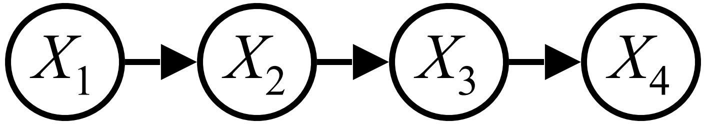
- From the chain rule, every joint distribution over X_1, X_2, X_3, X_4 can be written as:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

- Assuming that $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

Chain Rule and Markov Models



- From the chain rule, every joint distribution over X_1, X_2, X_3, X_4 can be written as:

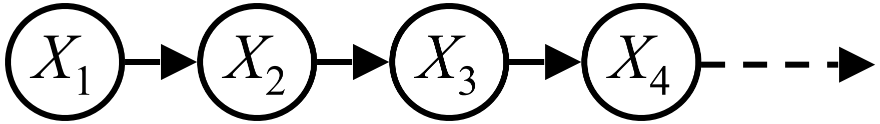
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

- Assuming that $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

  results in the expression posited on the previous slide:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

Chain Rule and Markov Models



- From the chain rule, every joint distribution over $X_1, X_2, \ldots, X_T$ can be written as:

$$P(X_1, X_2, \ldots, X_T) = P(X_1) \prod_{t=2}^{T} P(X_t | X_1, X_2, \ldots, X_{t-1})$$

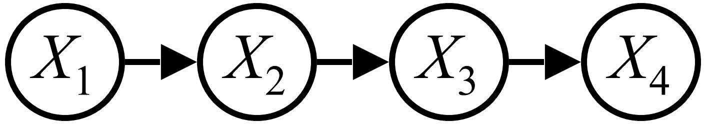
- Assuming that for all *t*:

$$X_t \perp\!\!\!\perp X_1, \ldots, X_{t-2} \mid X_{t-1}$$

gives us the expression posited on the earlier slide:

$$P(X_1, X_2, \ldots, X_T) = P(X_1) \prod_{t=2}^{T} P(X_t | X_{t-1})$$

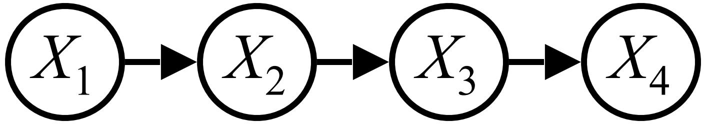
Implied Conditional Independencies



- We assumed: $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

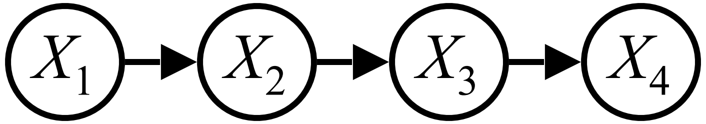
- Do we also have $X_1 \perp\!\!\!\perp X_3, X_4 \mid X_2$?

Implied Conditional Independencies



- We assumed: $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$
- Do we also have $X_1 \perp\!\!\!\perp X_3, X_4 \mid X_2$?
  - Yes!

Implied Conditional Independencies



- We assumed: $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

- Do we also have $X_1 \perp\!\!\!\perp X_3, X_4 \mid X_2$?
  - Yes!
  - Proof:

$$\begin{aligned} P(X_1 \mid X_2, X_3, X_4) &= \frac{P(X_1, X_2, X_3, X_4)}{P(X_2, X_3, X_4)} \\ &= \frac{P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_2)P(X_4 \mid X_3)}{\sum_{x_1} P(x_1)P(X_2 \mid x_1)P(X_3 \mid X_2)P(X_4 \mid X_3)} \\ &= \frac{P(X_1, X_2)}{P(X_2)} \\ &= P(X_1 \mid X_2) \end{aligned}$$

Markov Models Recap

- Explicit assumption for all *t* : $X_t \perp\!\!\!\perp X_1, \ldots, X_{t-2} \mid X_{t-1}$
- Consequence, joint distribution can be written as:

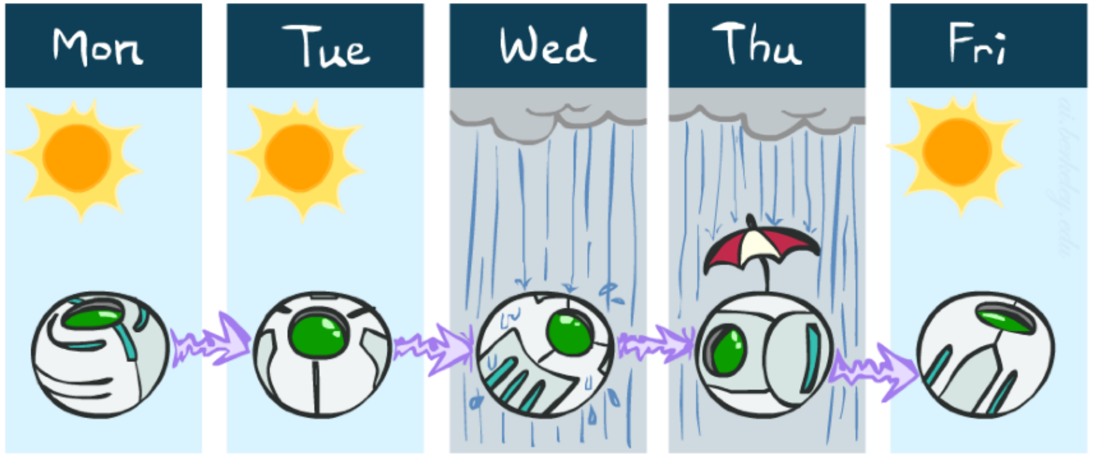
$$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1})$$

$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

- Implied conditional independencies: (try to prove this!)
  - Past variables independent of future variables given the present

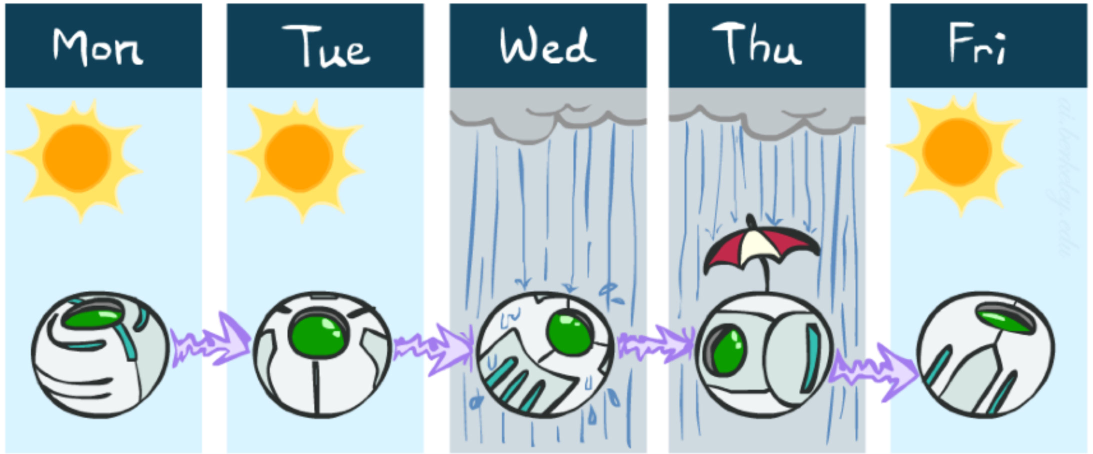
  i.e., if $t_1 < t_2 < t_3$ or $t_1 > t_2 > t_3$ then: $X_{t_1} \perp\!\!\!\perp X_{t_3} \mid X_{t_2}$
- Additional explicit assumption: $P(X_t \mid X_{t-1})$ is the same for all *t*

Example Markov Chain: Weather



- States: X = {rain, sun}
- Initial distribution: 1.0 sun
- CPT $P(X_t \mid X_{t-1})$:

Example Markov Chain: Weather

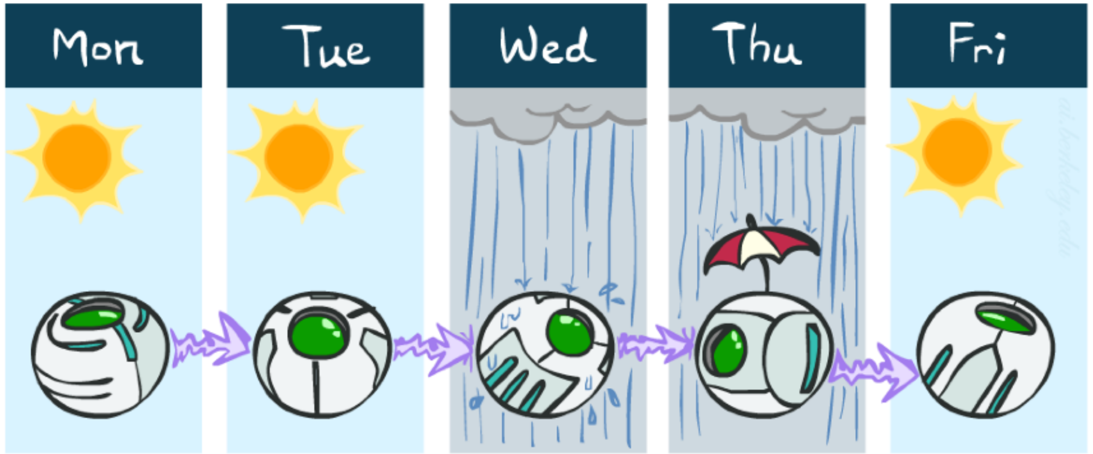


- States: X = {rain, sun}
- Initial distribution: 1.0 sun
- CPT $P(X_t \mid X_{t-1})$:

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Example Markov Chain: Weather

- States: X = {rain, sun}



- Initial distribution: 1.0 sun

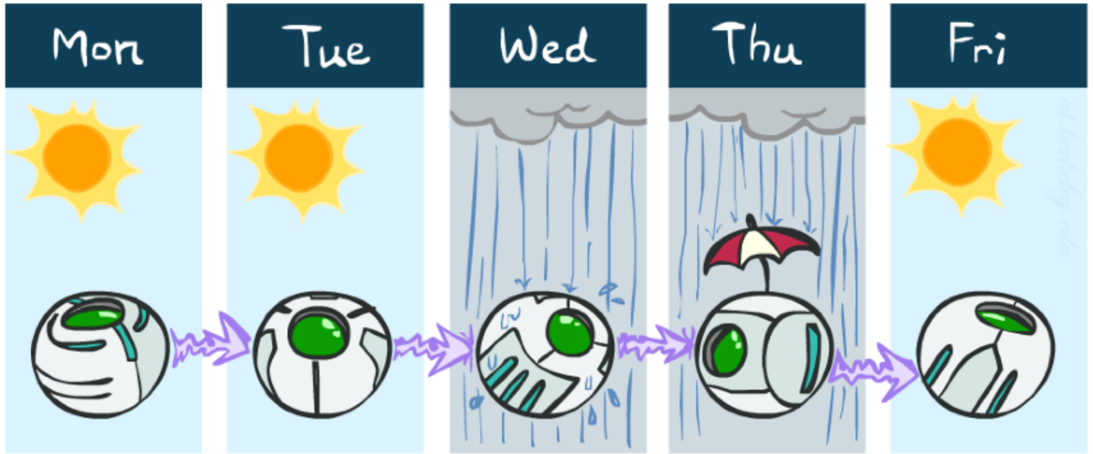
- CPT $P(X_t \mid X_{t-1})$:

Two new ways of representing the same CPT

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Example Markov Chain: Weather

- States: X = {rain, sun}

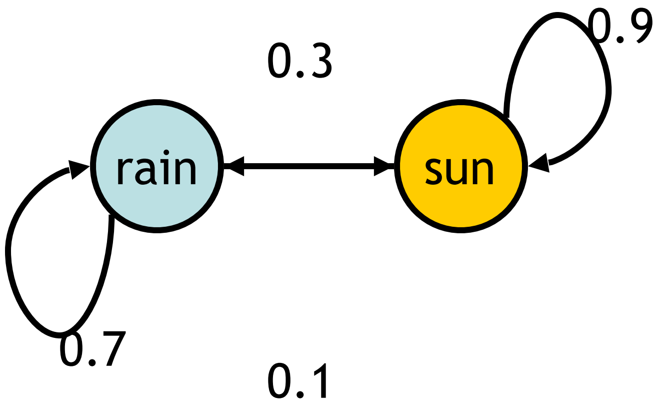


- Initial distribution: 1.0 sun

- CPT $P(X_t \mid X_{t-1})$:

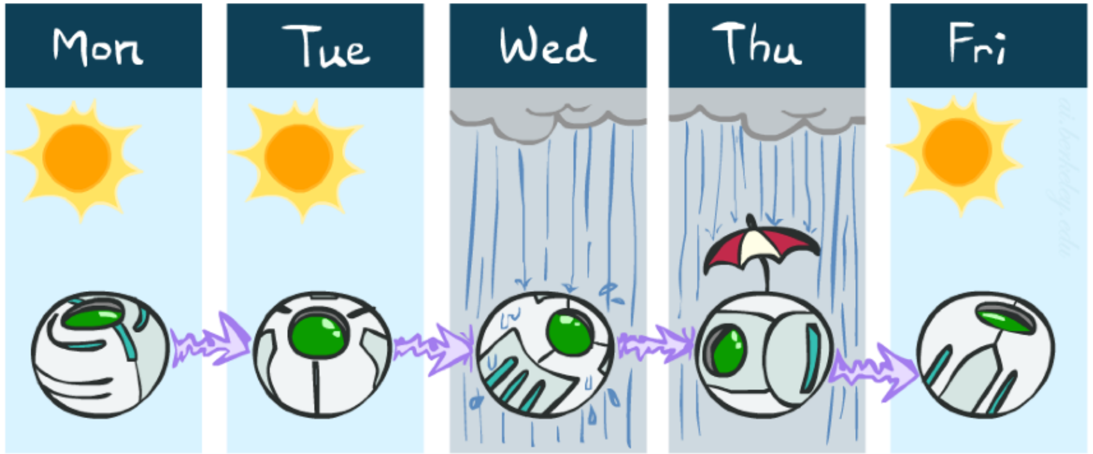
Two new ways of representing the same CPT

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7



Example Markov Chain: Weather

- States: X = {rain, sun}

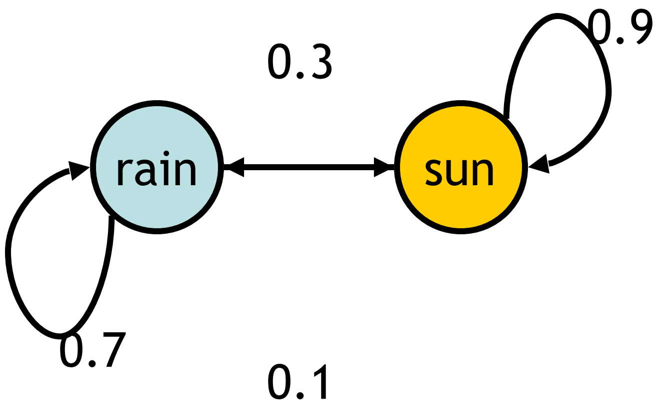


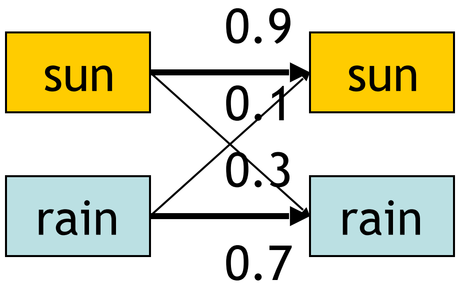
- Initial distribution: 1.0 sun

- CPT $P(X_t \mid X_{t-1})$:

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

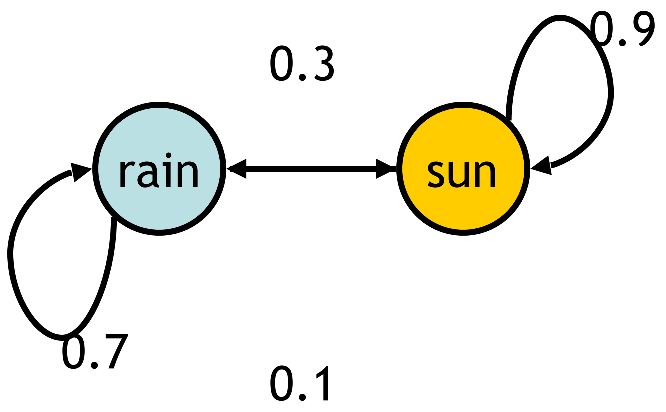
Two new ways of representing the same CPT





Example Markov Chain: Weather

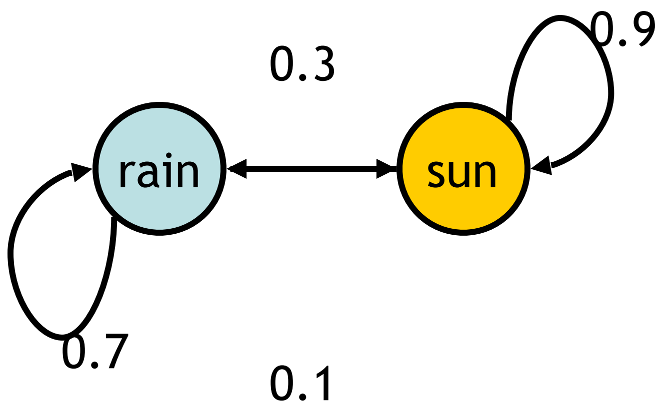
- Initial distribution: 1.0 sun



- What is the probability distribution after one step?

Example Markov Chain: Weather

- Initial distribution: 1.0 sun

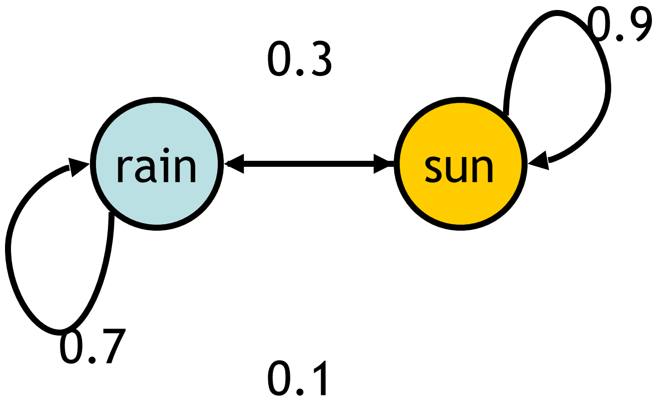


- What is the probability distribution after one step?

$$P(X_2 = \text{sun}) = P(X_2 = \text{sun} | X_1 = \text{sun}) P(X_1 = \text{sun}) + P(X_2 = \text{sun} | X_1 = \text{rain}) P(X_1 = \text{rain})$$

Example Markov Chain: Weather

- Initial distribution: 1.0 sun



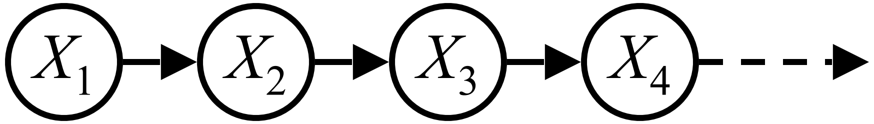
- What is the probability distribution after one step?

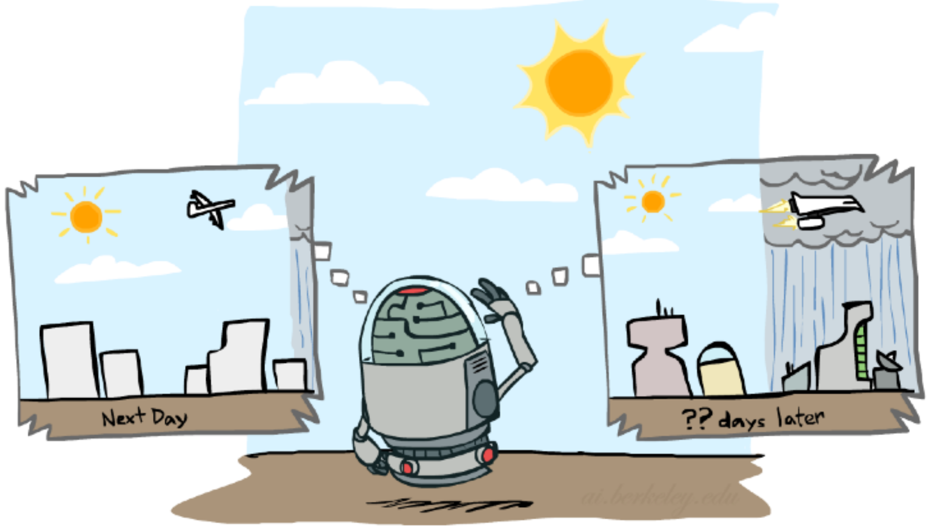
$$P(X_2 = \text{sun}) = P(X_2 = \text{sun} | X_1 = \text{sun}) P(X_1 = \text{sun}) + P(X_2 = \text{sun} | X_1 = \text{rain}) P(X_1 = \text{rain})$$

$$0.9 \cdot 1.0 + 0.3 \cdot 0.0 = 0.9$$

Mini-Forward Algorithm

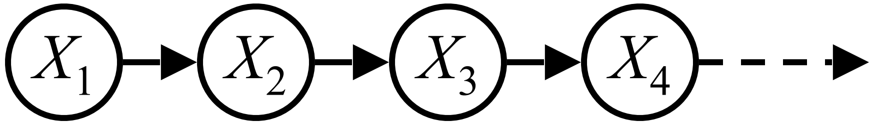
- Question: What's P(X) on some day t?

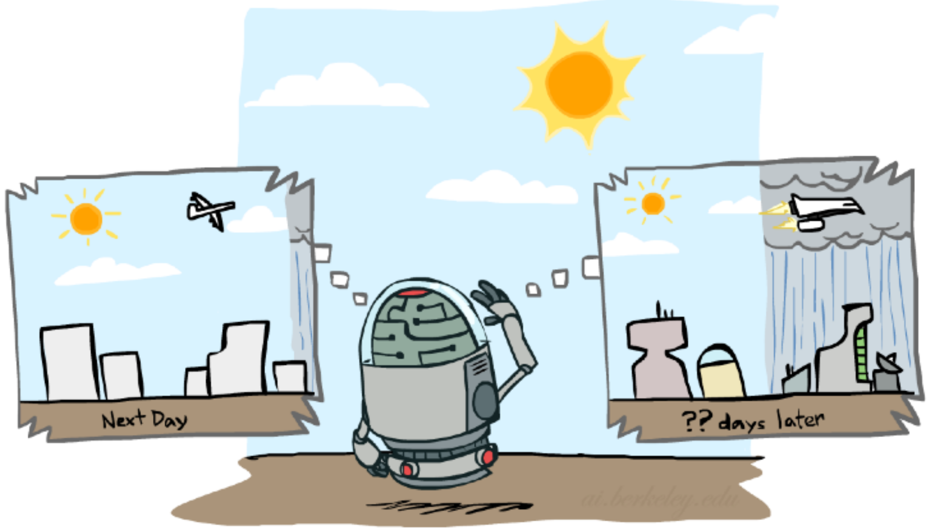




Mini-Forward Algorithm

- Question: What's P(X) on some day t?

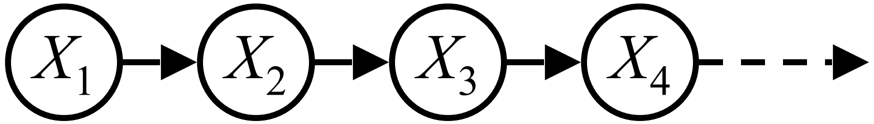


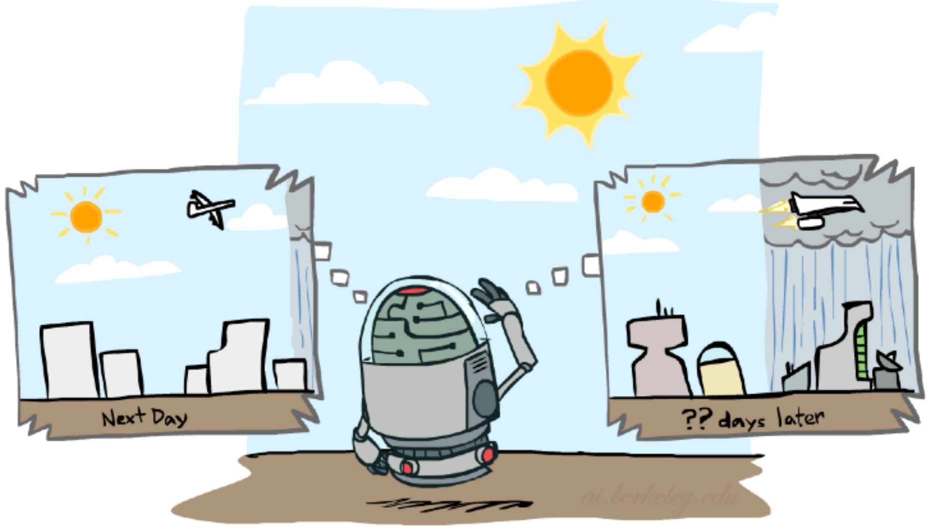


$$P(x_1) = \text{known}$$

Mini-Forward Algorithm

- Question: What's P(X) on some day t?



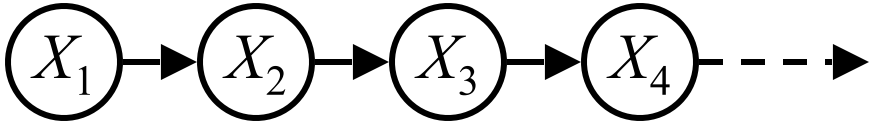


$P(x_1) = \text{known}$

$P(x_t) =$

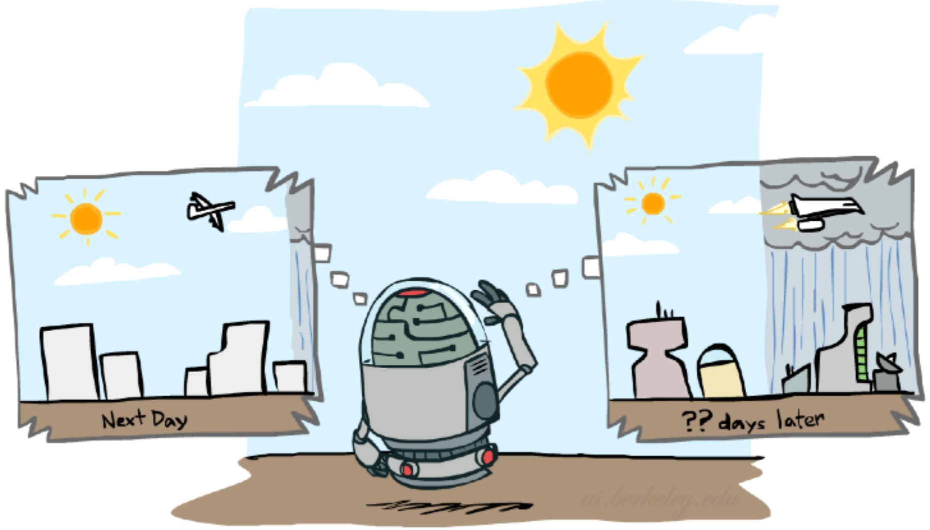
Mini-Forward Algorithm

- Question: What's P(X) on some day t?



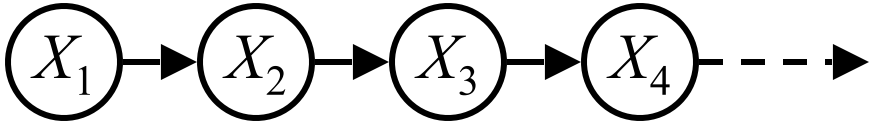
$$\color{red}{P(x_1)} = \text{known}$$

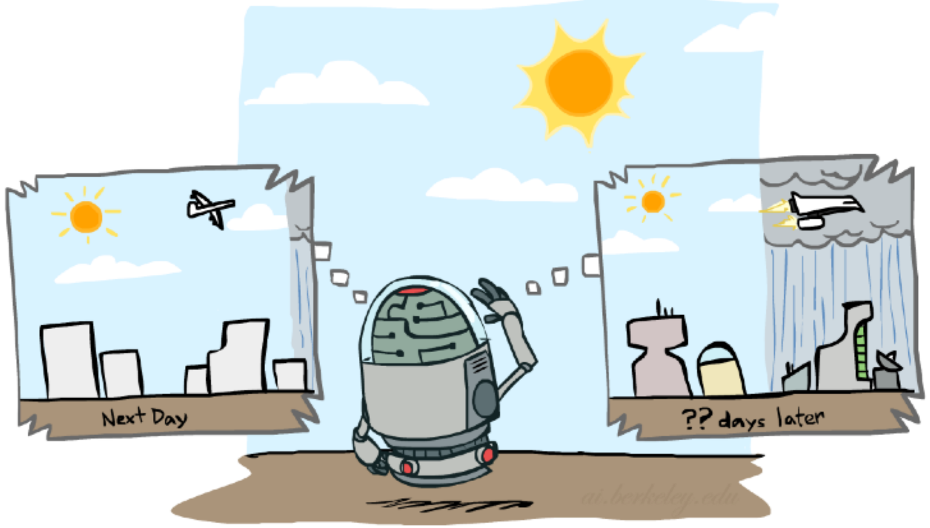
$$P(x_t) = \sum_{x_{t-1}} P(x_{t-1}, x_t)$$



Mini-Forward Algorithm

- Question: What's P(X) on some day t?





$$P(x_1) = \text{known}$$

$$P(x_t) = \sum_{x_{t-1}} P(x_{t-1}, x_t)$$

$$= \sum_{x_{t-1}} P(x_t \mid x_{t-1}) P(x_{t-1})$$

Forward simulation

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle$$

$P(X_1)$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle$$

$P(X_1)$ $P(X_2)$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle$$

$$P(X_1) \qquad P(X_2) \qquad P(X_3)$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle$$

$$P(X_1) \qquad P(X_2) \qquad P(X_3) \qquad P(X_4)$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{array}{c} 1.0 \\ 0.0 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.9 \\ 0.1 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.84 \\ 0.16 \end{array} \right\rangle \quad \left\langle \begin{array}{c} 0.804 \\ 0.196 \end{array} \right\rangle \Rightarrow \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \quad P(X_\infty)$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle \quad \Rightarrow \quad \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$\mathrm{P}(X_1) \quad \mathrm{P}(X_2) \quad \mathrm{P}(X_3) \quad \mathrm{P}(X_4) \quad \mathrm{P}(X_\infty)$$

- From initial observation of rain

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$P(X_1) \qquad P(X_2) \qquad P(X_3) \qquad P(X_4) \qquad\qquad P(X_\infty)$$

- From initial observation of rain

$$\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle$$

$$P(X_1)$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \quad P(X_\infty)$$

- From initial observation of rain

$$\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2)$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \quad P(X_\infty)$$

- From initial observation of rain

$$\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3)$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\underset{P(X_1)}{\begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix}} \quad \underset{P(X_2)}{\begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix}} \quad \underset{P(X_3)}{\begin{pmatrix} 0.84 \\ 0.16 \end{pmatrix}} \quad \underset{P(X_4)}{\begin{pmatrix} 0.804 \\ 0.196 \end{pmatrix}} \Rightarrow \underset{P(X_\infty)}{\begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}}$$

- From initial observation of rain

$$\underset{P(X_1)}{\begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix}} \quad \underset{P(X_2)}{\begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix}} \quad \underset{P(X_3)}{\begin{pmatrix} 0.48 \\ 0.52 \end{pmatrix}} \quad \underset{P(X_4)}{\begin{pmatrix} 0.588 \\ 0.412 \end{pmatrix}}$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \quad P(X_\infty)$$

- From initial observation of rain

$$\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.588 \\ 0.412 \end{matrix} \right\rangle \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \quad P(X_\infty)$$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle \quad \Rightarrow \quad \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \quad P(X_\infty)$$

- From initial observation of rain

$$\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.588 \\ 0.412 \end{matrix} \right\rangle \quad \Rightarrow \quad \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \quad P(X_\infty)$$

- From yet another initial distribution $P(X_1)$:

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix} \quad \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix} \quad \begin{pmatrix} 0.84 \\ 0.16 \end{pmatrix} \quad \begin{pmatrix} 0.804 \\ 0.196 \end{pmatrix} \Rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

$P(X_1)$ $P(X_2)$ $P(X_3)$ $P(X_4)$ $P(X_\infty)$

- From initial observation of rain

$$\begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix} \quad \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} \quad \begin{pmatrix} 0.48 \\ 0.52 \end{pmatrix} \quad \begin{pmatrix} 0.588 \\ 0.412 \end{pmatrix} \Rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

$P(X_1)$ $P(X_2)$ $P(X_3)$ $P(X_4)$ $P(X_\infty)$

- From yet another initial distribution $P(X_1)$:

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} \quad \ldots \quad \Rightarrow$$

$P(X_1)$

[Demo: L13D1,2,3

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\begin{pmatrix} 1.0 \\ 0.0 \end{pmatrix} \quad \begin{pmatrix} 0.9 \\ 0.1 \end{pmatrix} \quad \begin{pmatrix} 0.84 \\ 0.16 \end{pmatrix} \quad \begin{pmatrix} 0.804 \\ 0.196 \end{pmatrix} \rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \qquad P(X_\infty)$$

- From initial observation of rain

$$\begin{pmatrix} 0.0 \\ 1.0 \end{pmatrix} \quad \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} \quad \begin{pmatrix} 0.48 \\ 0.52 \end{pmatrix} \quad \begin{pmatrix} 0.588 \\ 0.412 \end{pmatrix} \rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

$$P(X_1) \quad P(X_2) \quad P(X_3) \quad P(X_4) \qquad P(X_\infty)$$

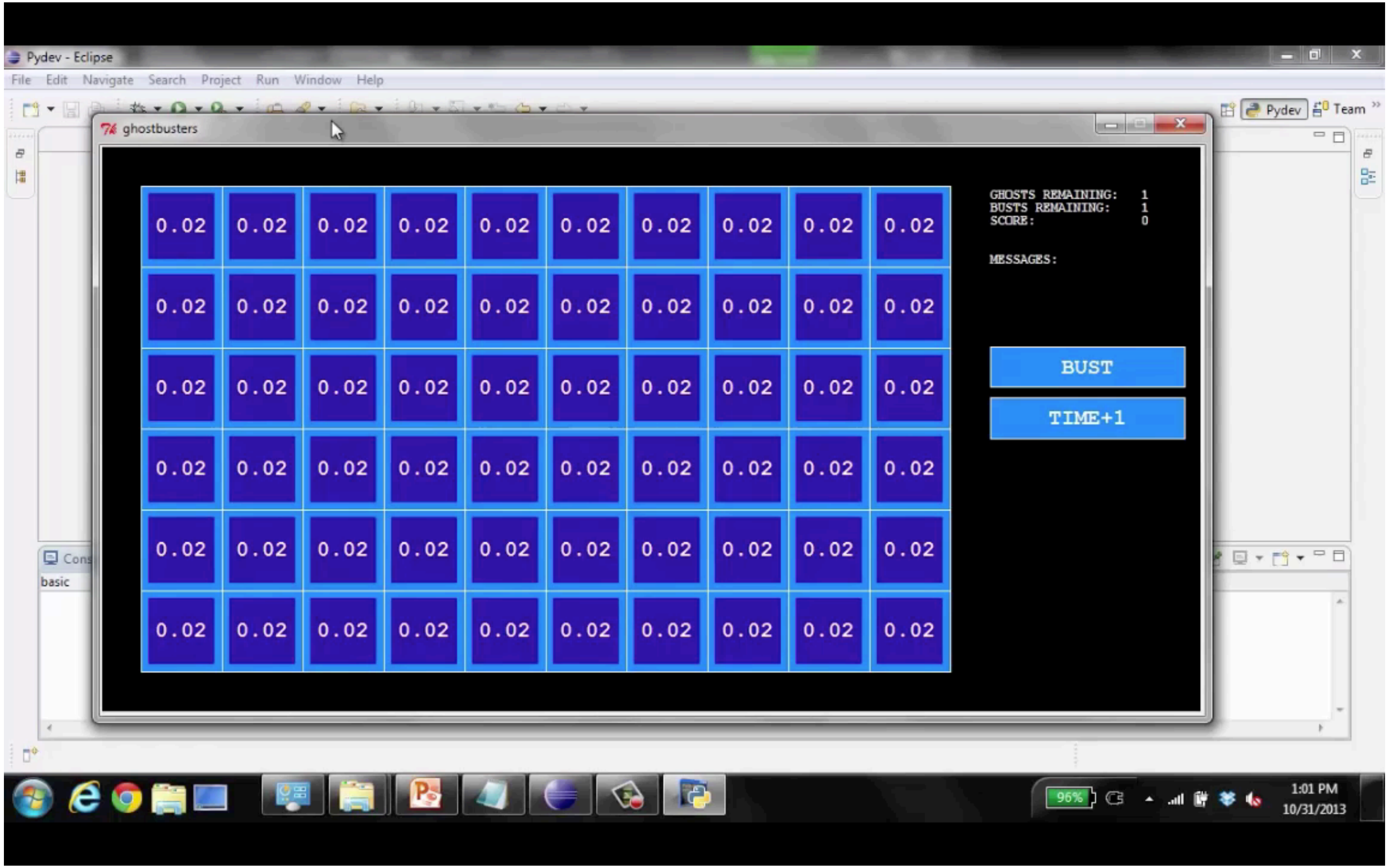
- From yet another initial distribution $P(X_1)$:

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} \quad \cdots \quad \rightarrow \begin{pmatrix} 0.75 \\ 0.25 \end{pmatrix}$$

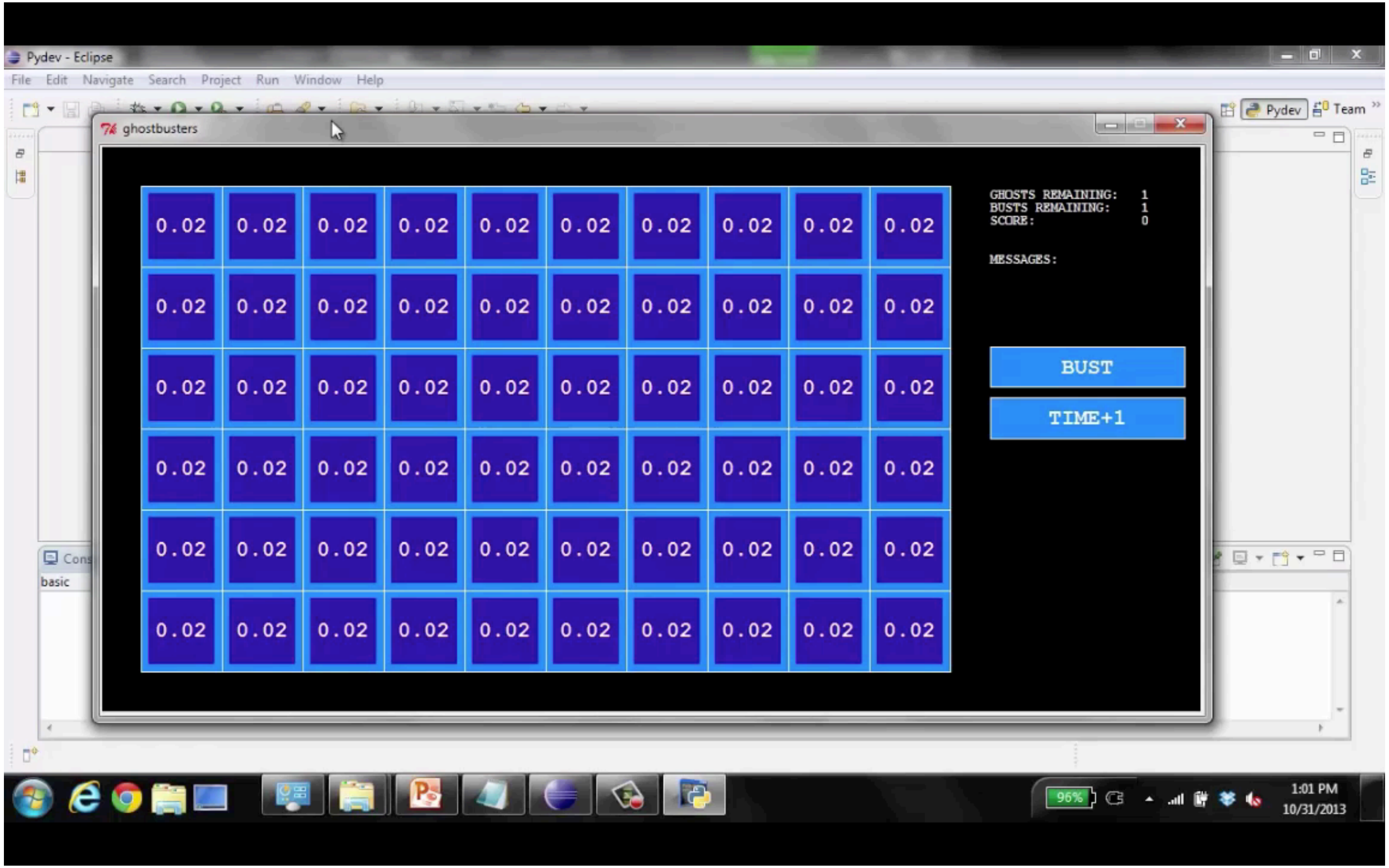
$$P(X_1) \qquad\qquad P(X_\infty)$$

[Demo: L13D1,2,3

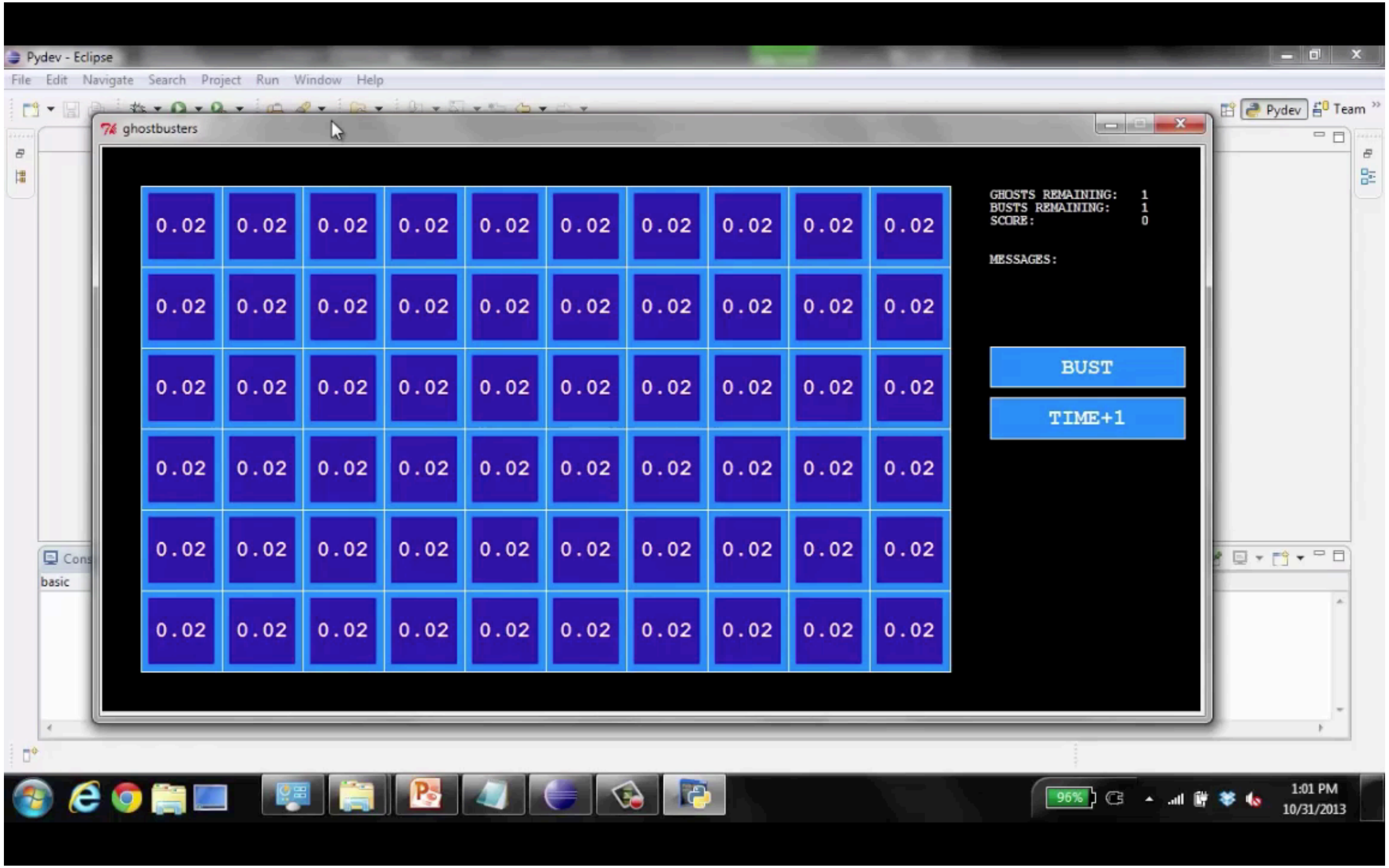
Video of Demo Ghostbusters Basic Dynamics



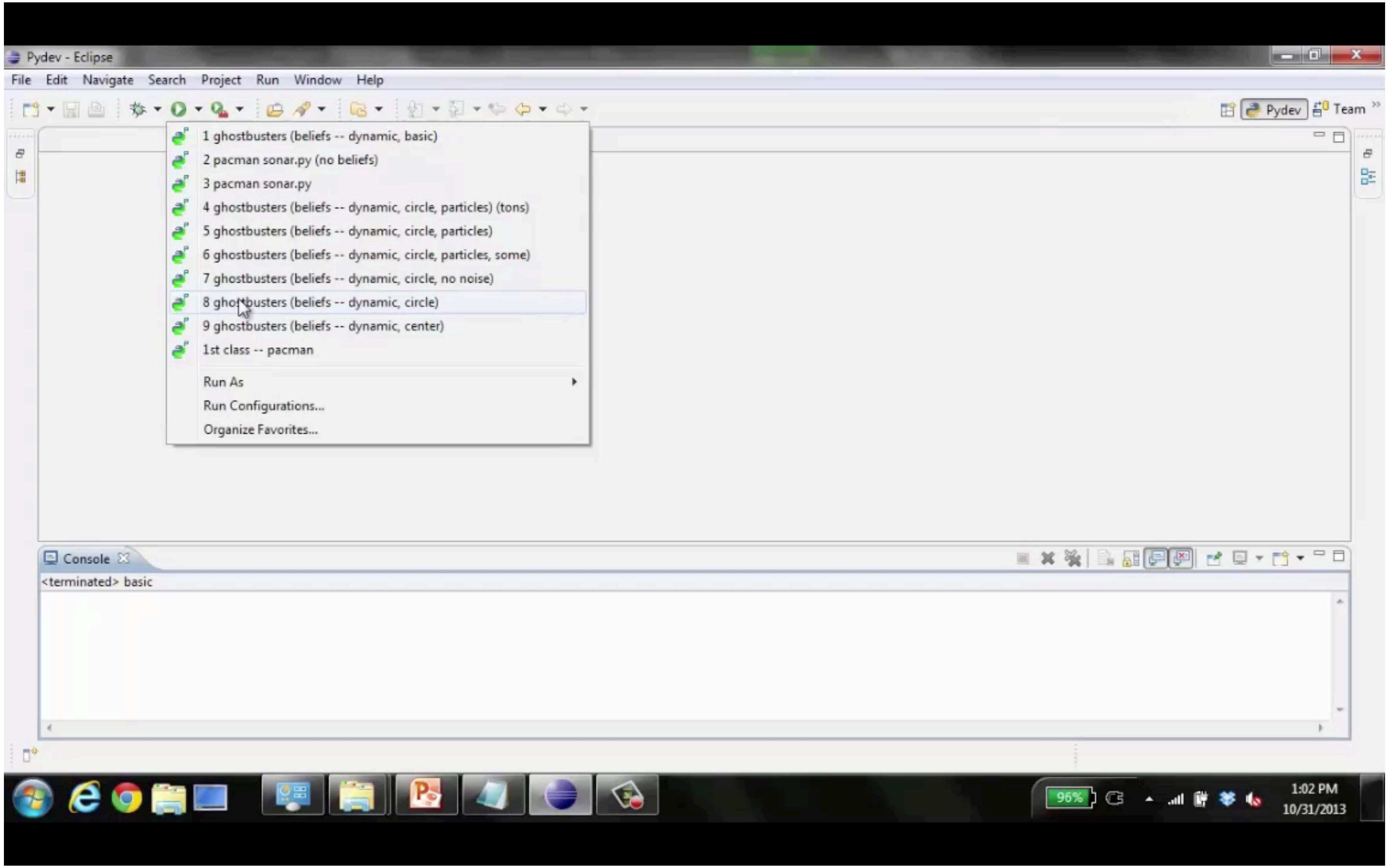
Video of Demo Ghostbusters Basic Dynamics



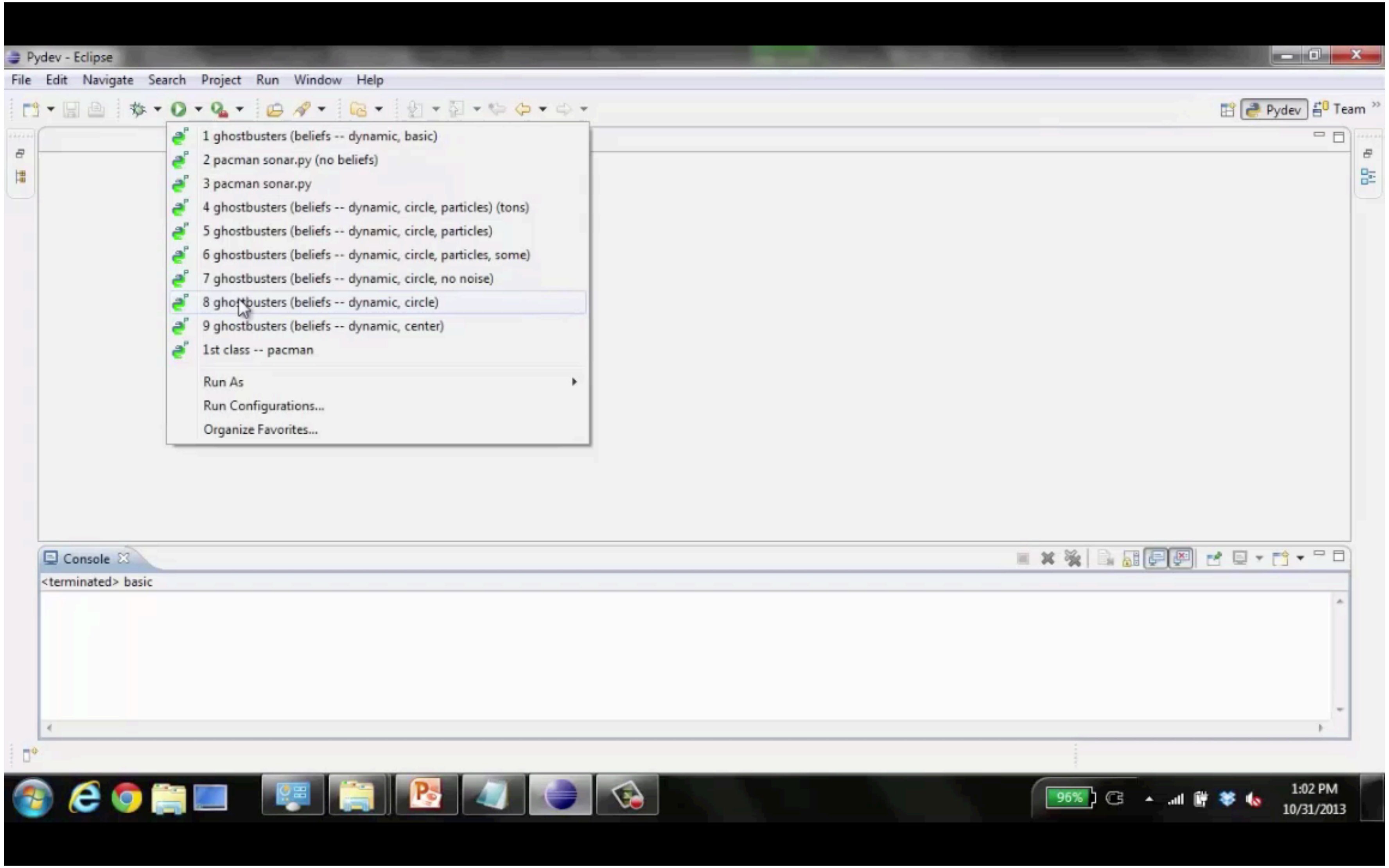
Video of Demo Ghostbusters Basic Dynamics



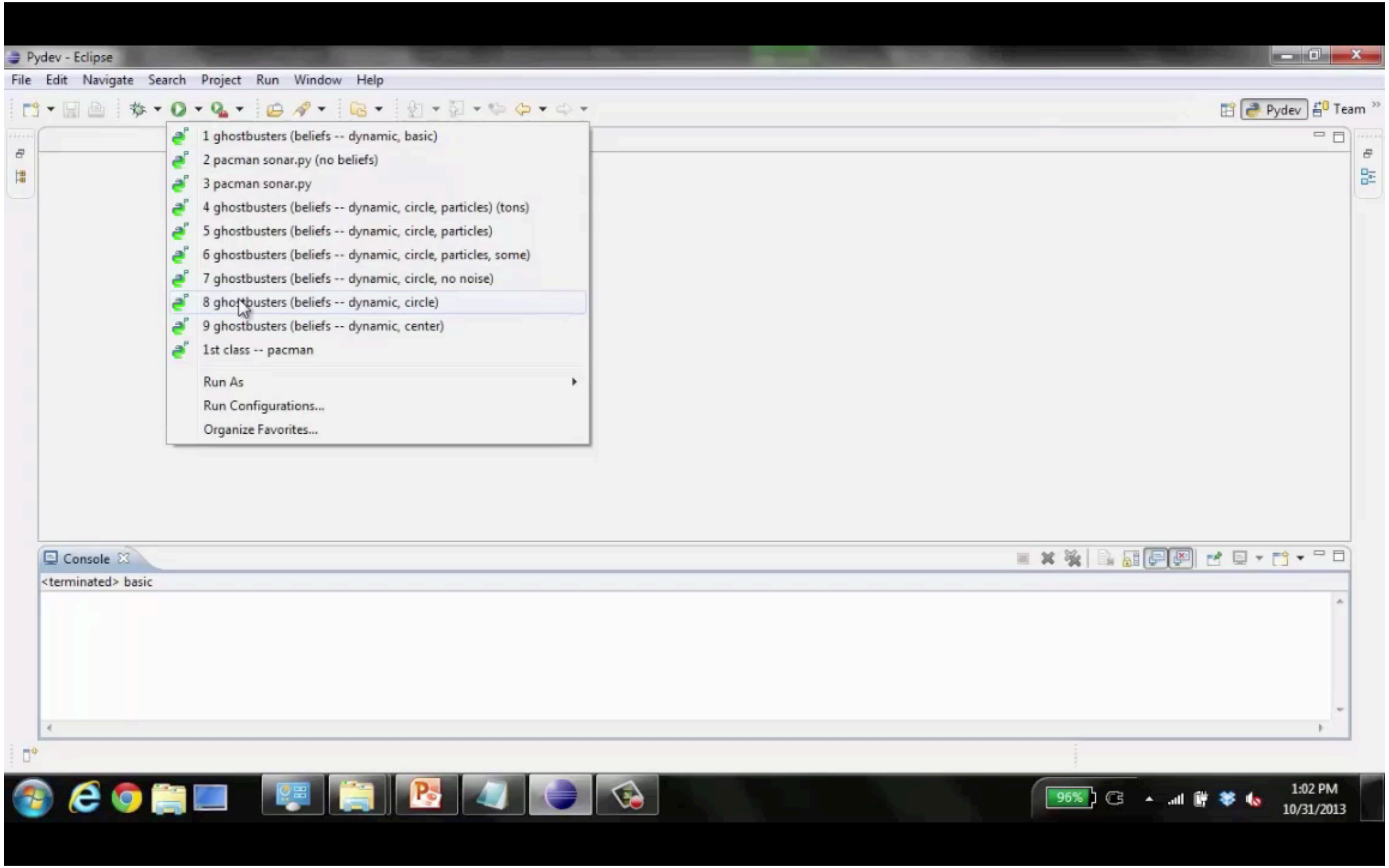
Video of Demo Ghostbusters Circular Dynamics



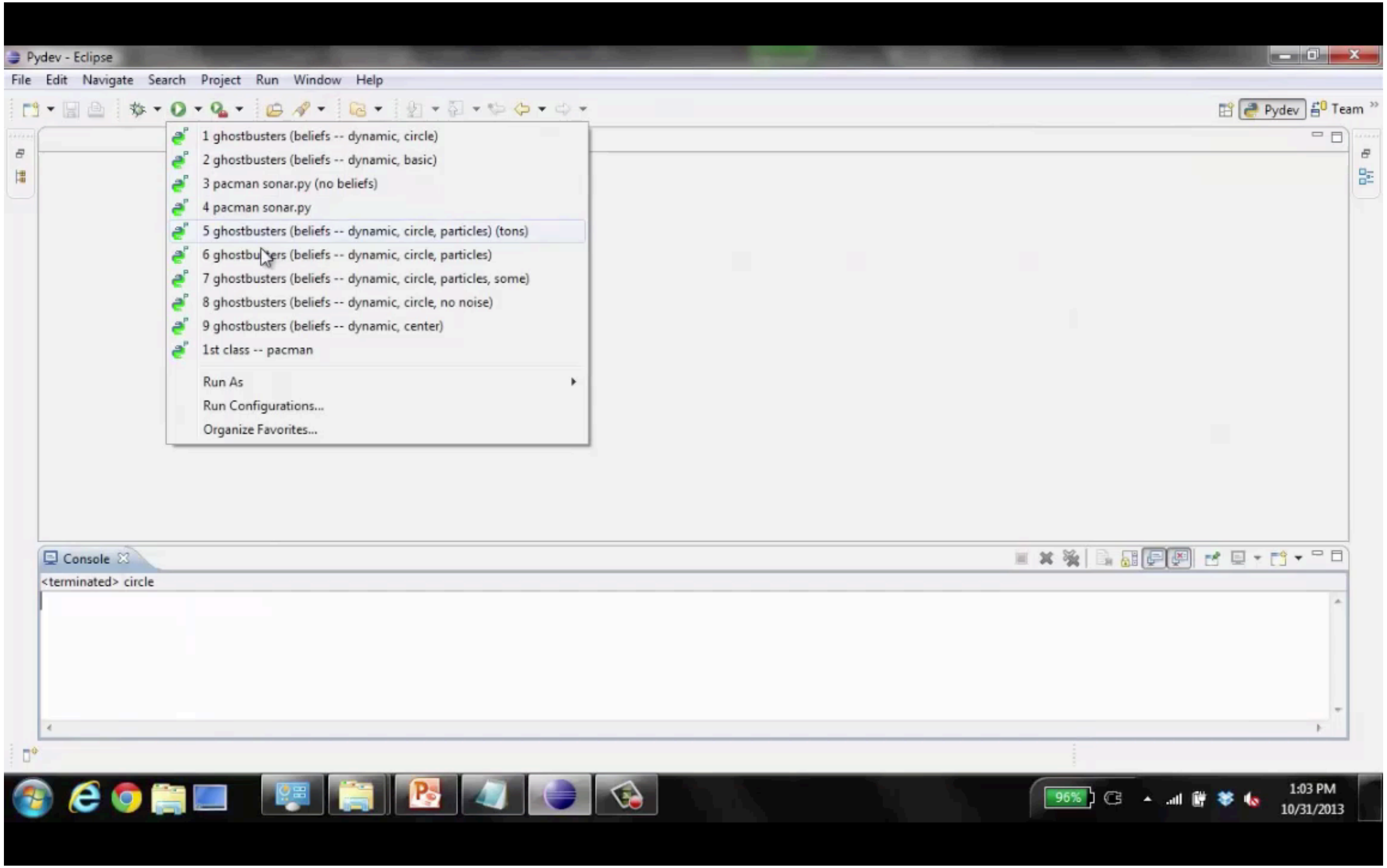
Video of Demo Ghostbusters Circular Dynamics



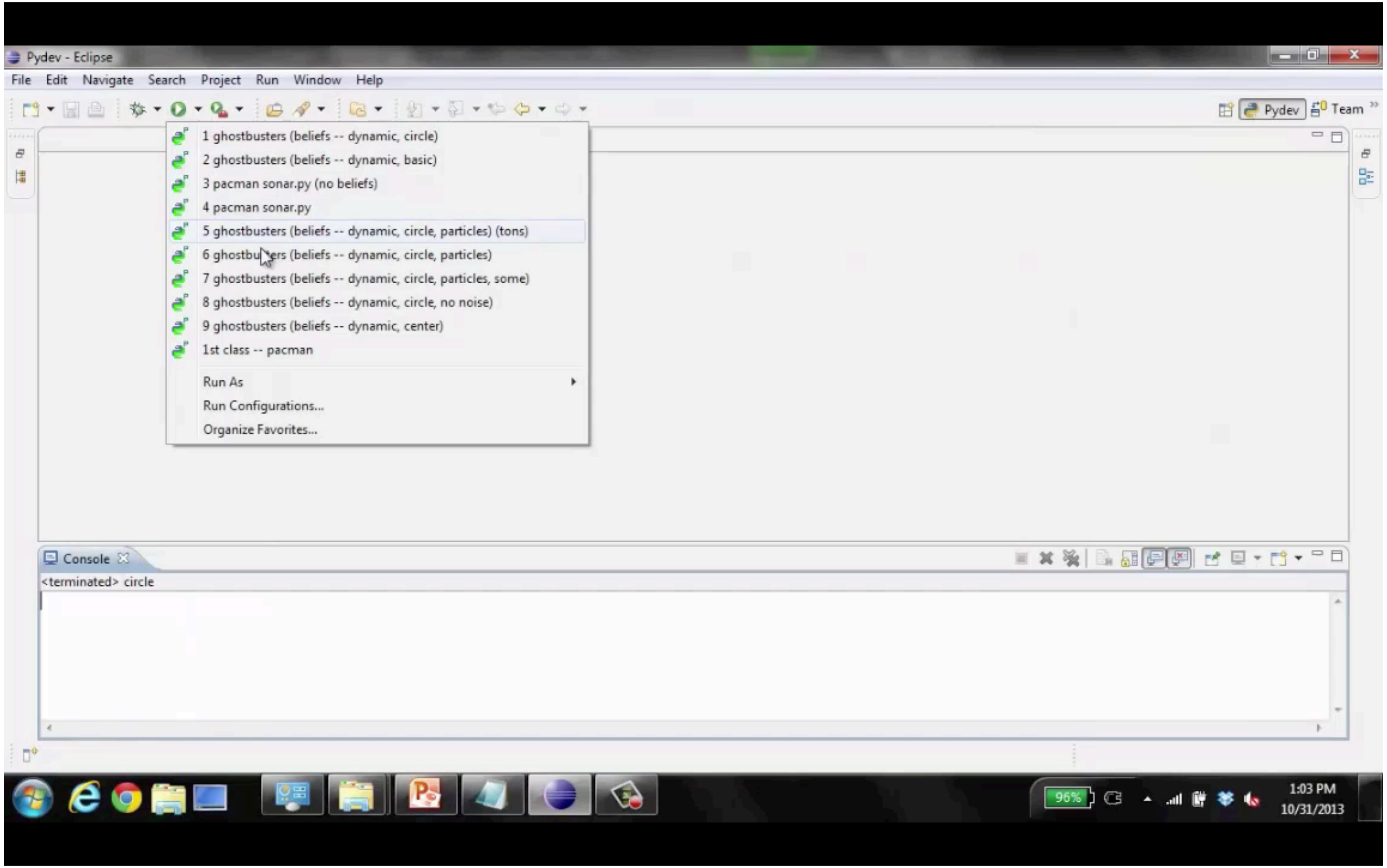
Video of Demo Ghostbusters Circular Dynamics



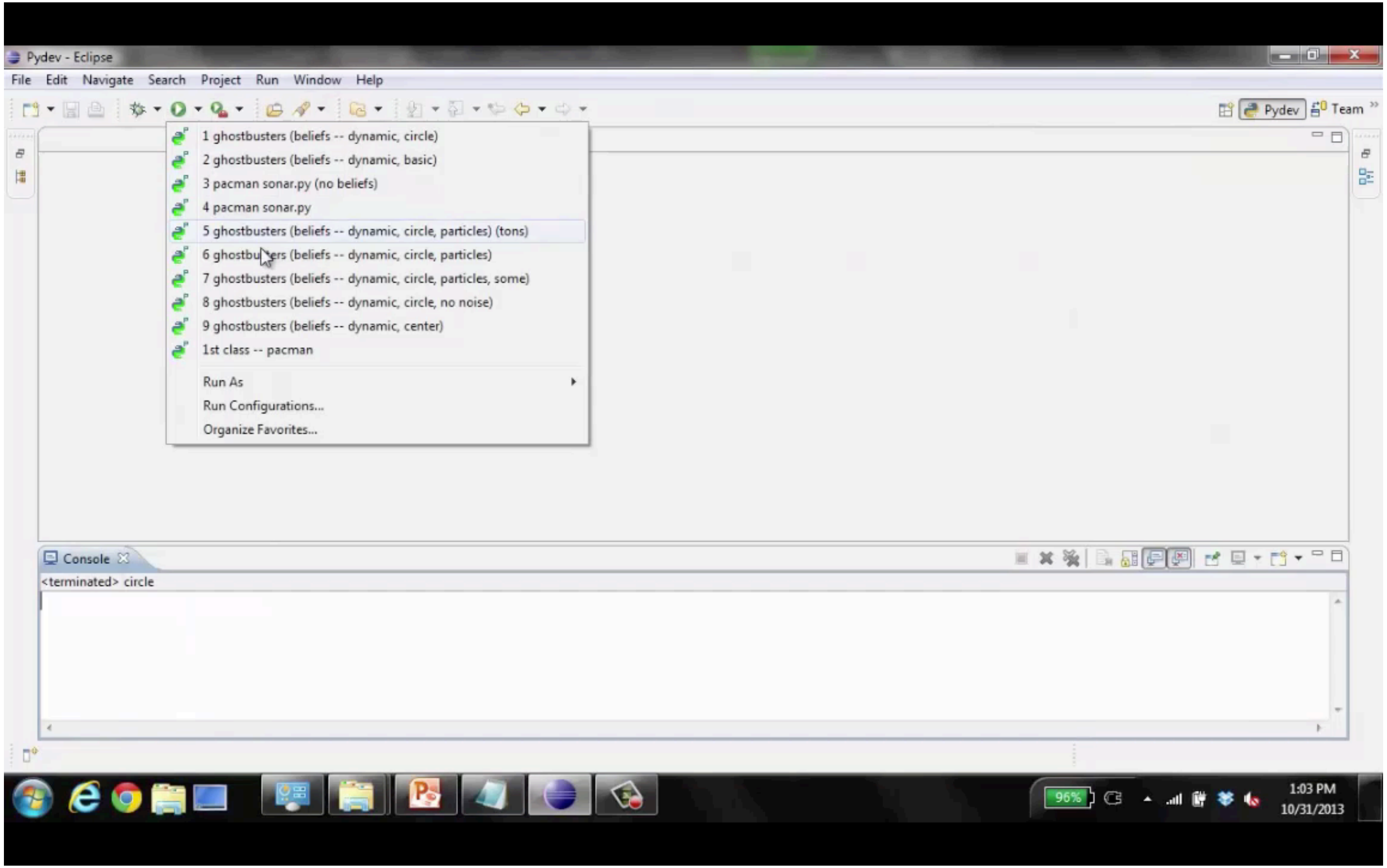
Video of Demo Ghostbusters Whirlpool Dynamics



Video of Demo Ghostbusters Whirlpool Dynamics



Video of Demo Ghostbusters Whirlpool Dynamics



Stationary Distributions

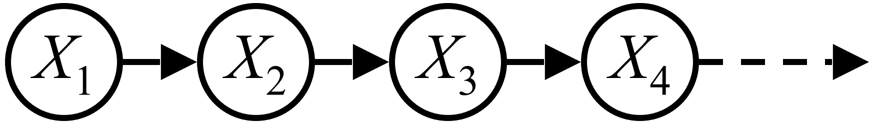
- For most chains:
  - Influence of the initial distribution gets less and less over time.
  - The distribution we end up in is independent of the initial distribution

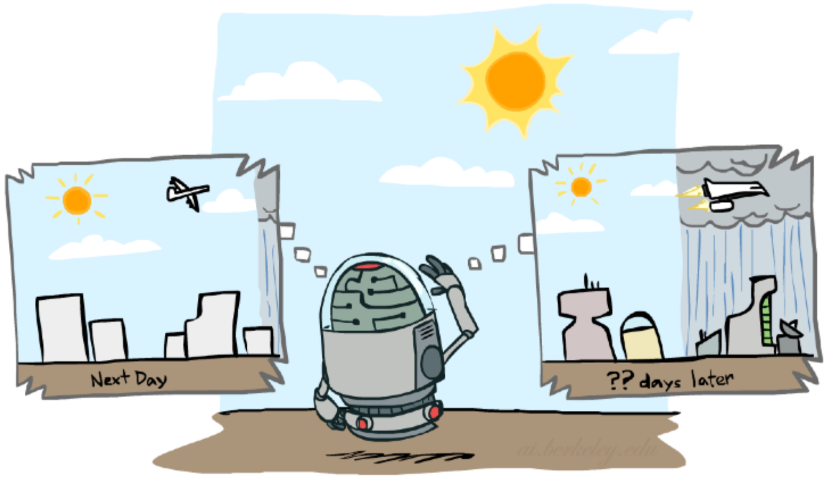
- Stationary distribution:
  - The distribution we end up with is called the stationary distribution P_∞ of the chain
  - It satisfies

$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$

Example: Stationary Distributions

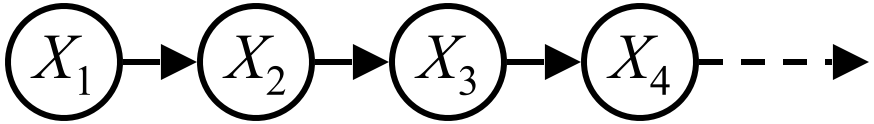
- Question: What's P(X) at time t = infinity?

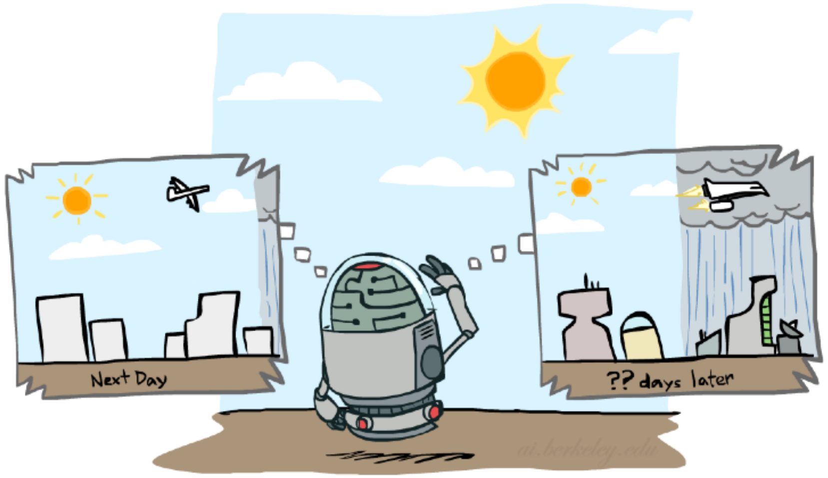




Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?

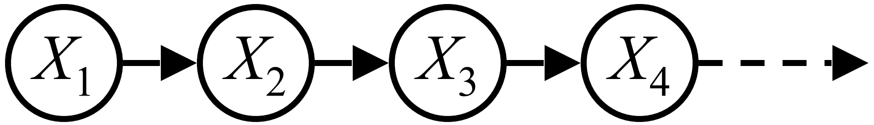




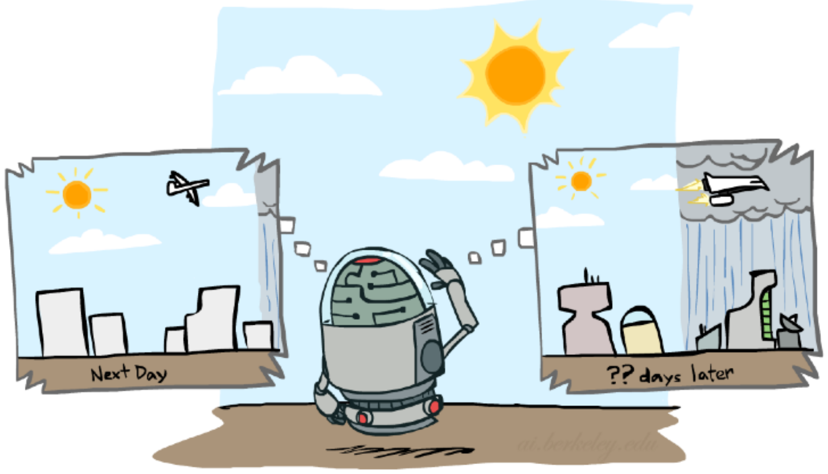
X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?



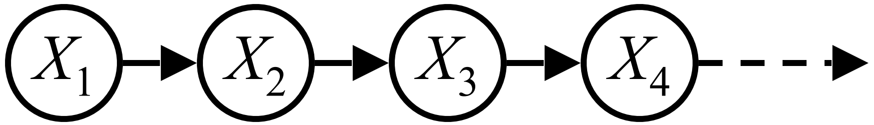
$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$



X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

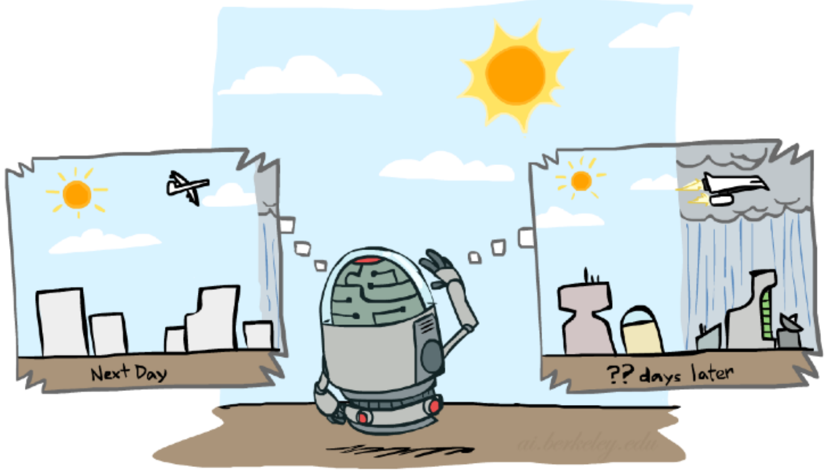
Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?



$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

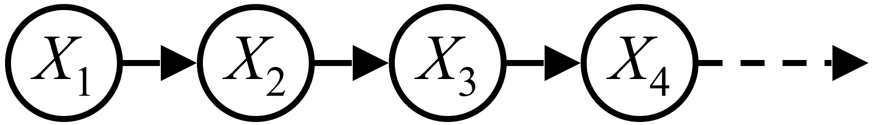
$$P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain)$$
$$P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain)$$

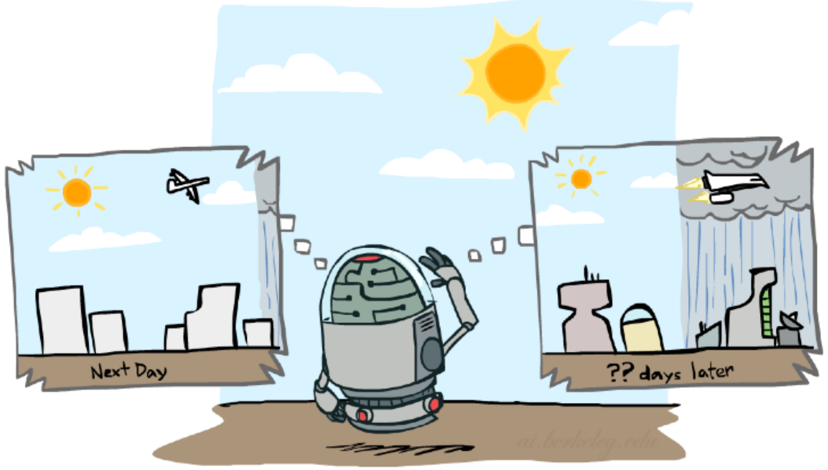


X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?





$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

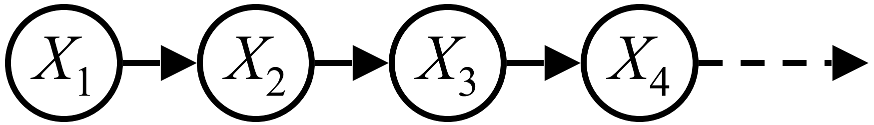
$$P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain)$$
$$P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain)$$

$$P_\infty(sun) = 3P_\infty(rain)$$
$$P_\infty(rain) = 1/3P_\infty(sun)$$

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?

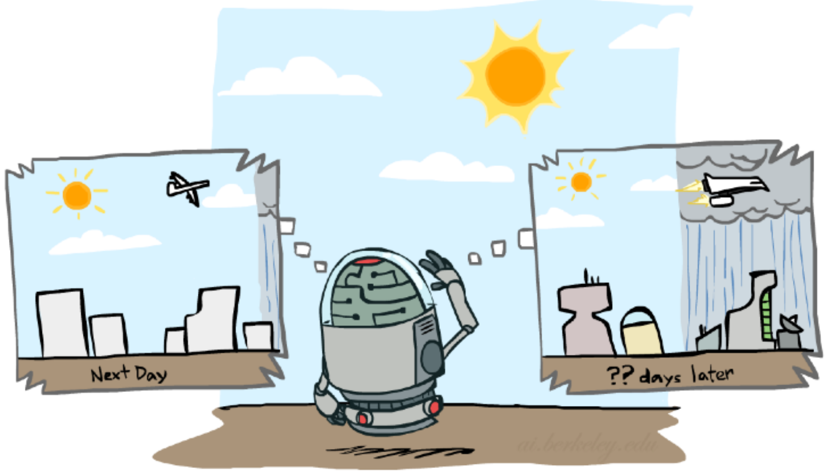


$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

$$P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain)$$
$$P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain)$$

$$P_\infty(sun) = 3P_\infty(rain)$$
$$P_\infty(rain) = 1/3P_\infty(sun)$$

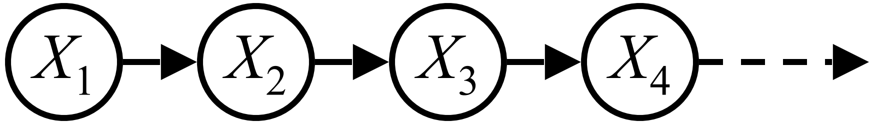
Also: $P_\infty(sun) + P_\infty(rain) = 1$

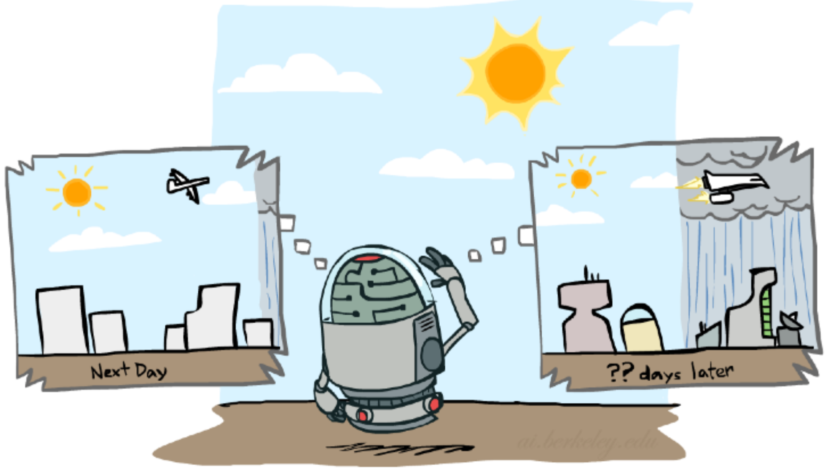


X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?





$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

$$P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain)$$
$$P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain)$$

$$P_\infty(sun) = 3P_\infty(rain)$$
$$P_\infty(rain) = 1/3P_\infty(sun)$$

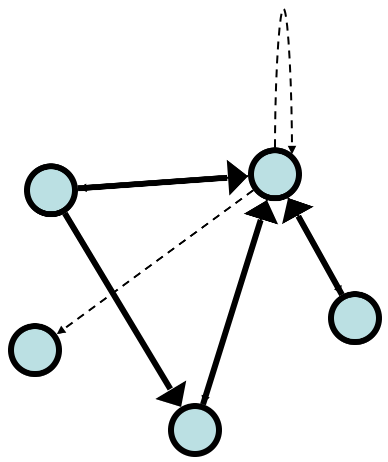
Also: $P_\infty(sun) + P_\infty(rain) = 1$

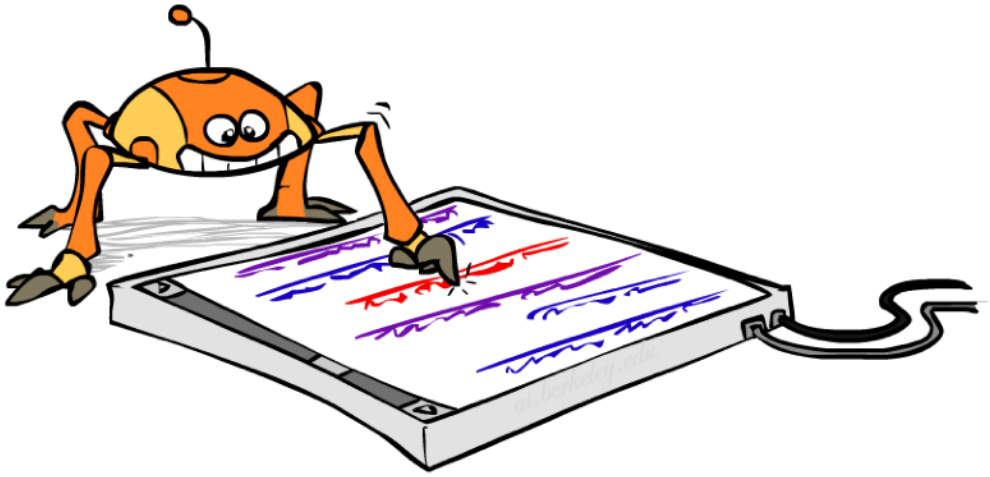
$$P_\infty(sun) = 3/4$$
$$P_\infty(rain) = 1/4$$

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Application of Stationary Distribution: Web Link Analysis

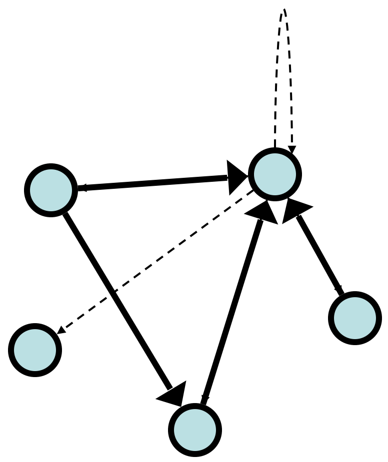
- PageRank over a web graph
  - Each web page is a state
  - Initial distribution: uniform over pages
  - Transitions:
    - With prob. c, uniform jump to a random page (dotted lines, not all shown)
    - With prob. 1-c, follow a random outlink (solid lines)



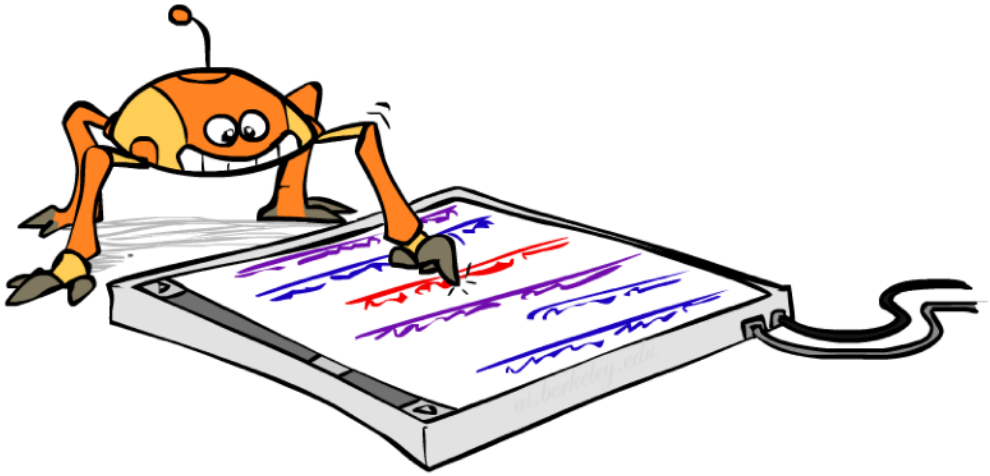


Application of Stationary Distribution: Web Link Analysis

- PageRank over a web graph
  - Each web page is a state
  - Initial distribution: uniform over pages
  - Transitions:
    - With prob. c, uniform jump to a random page (dotted lines, not all shown)
    - With prob. 1-c, follow a random outlink (solid lines)



- Stationary distribution
  - Will spend more time on highly reachable pages
  - E.g. many ways to get to the Acrobat Reader download page
  - Somewhat robust to link spam
  - Google 1.0 returned the set of pages containing all your keywords in decreasing rank, now all search engines use link analysis along with many other factors (rank actually getting less important over time)



Next Time: Hidden Markov Models!