



C4

Markov Models

COMP3270
Introduction to Artificial Intelligence

Outline

Probabilistic Reasoning

Probability refresher · Bayes' rule · Independence

Outline

Probabilistic Reasoning

Probability refresher · Bayes' rule · Independence

Reasoning over Time

Markov models · Forward prediction · Stationary distributions

Outline

Probabilistic Reasoning

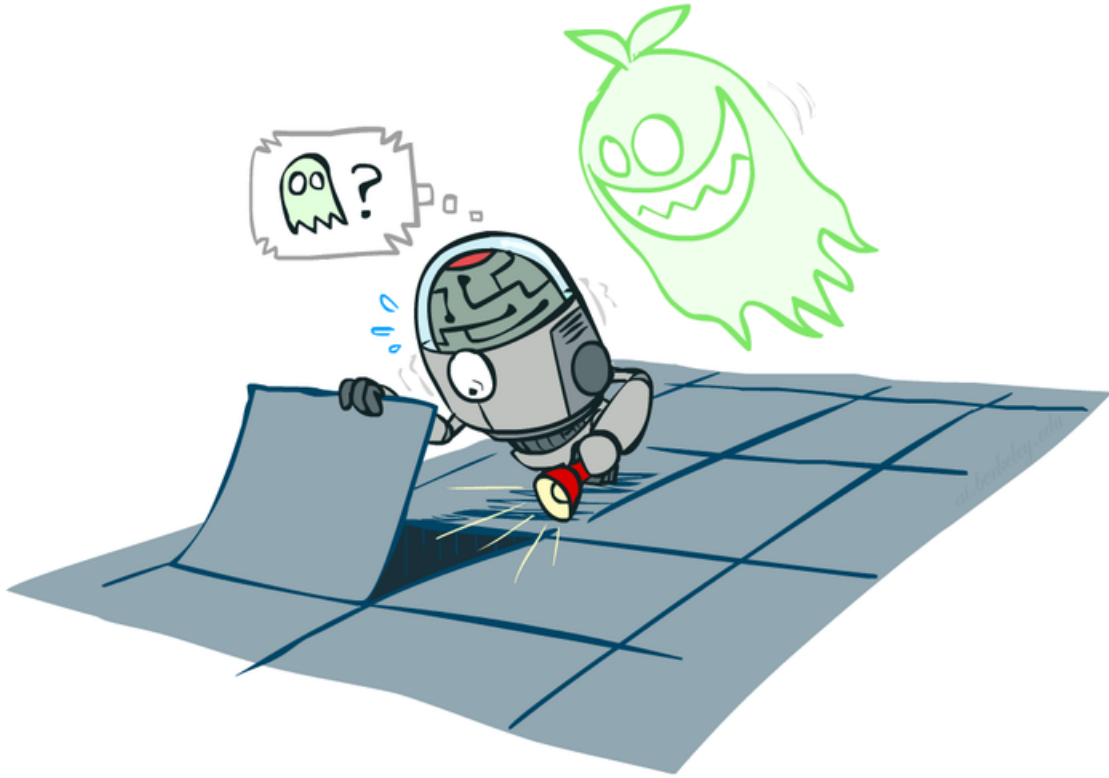
Probability refresher · Bayes' rule · Independence

Reasoning over Time

Markov models · Forward prediction · Stationary distributions

Next: Hidden Markov Models

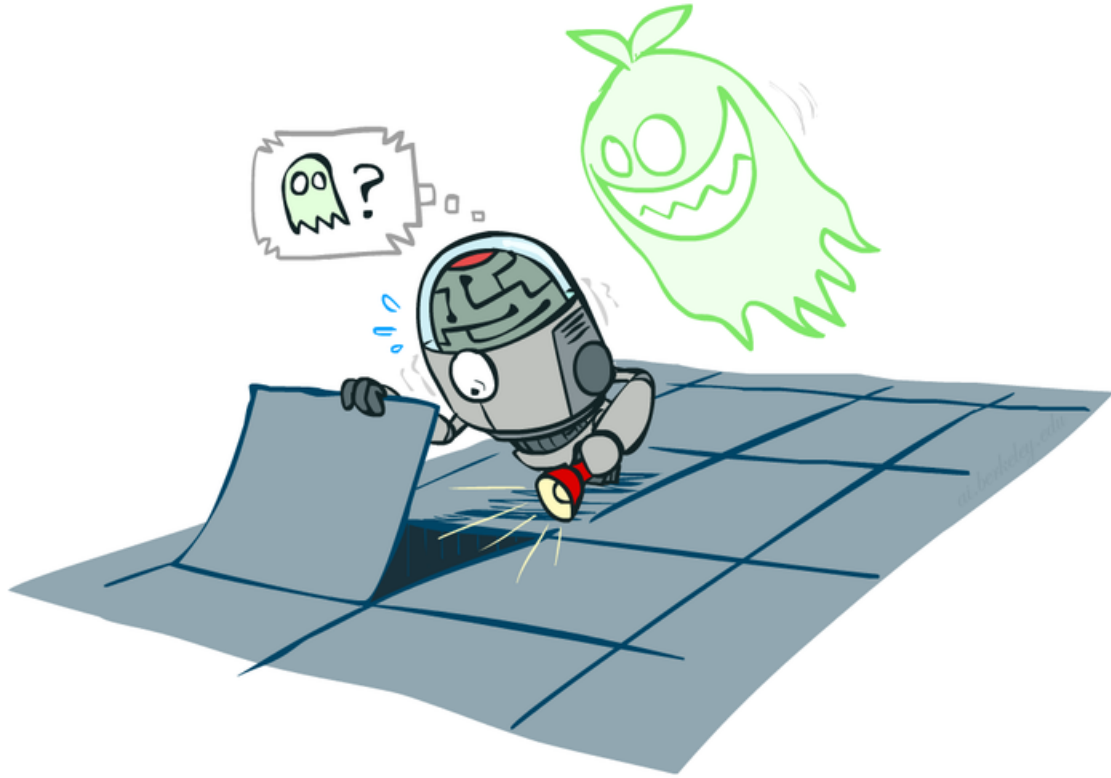
Ghostbusters



A ghost is hiding somewhere in the grid

A noisy sensor gives clues
about its distance

Ghostbusters

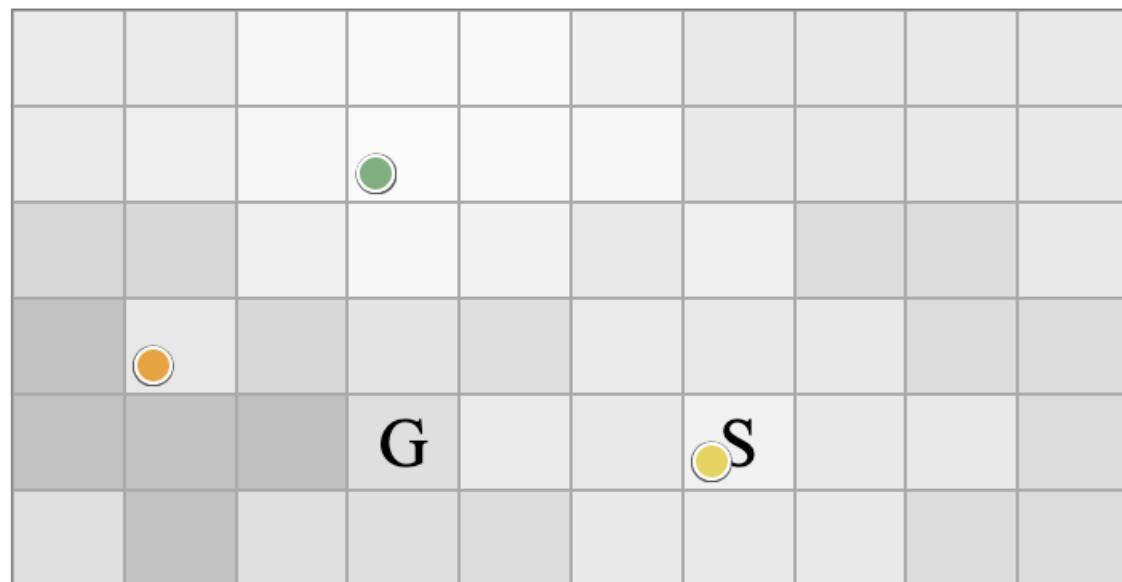


A ghost is hiding somewhere in the grid

A noisy sensor gives clues
about its distance

Where should we believe the ghost is?

Noisy Sensor Readings



S: sensor · G: ghost · Dots: latest readings

Latest reading from each of three cells · Darker = more likely

Combine noisy clues to narrow down likely locations

Typical reading	Distance
Red	0
Orange	1–2
Yellow	3–4
Green	5+

At distance 3:

Red	Orange	Yellow	Green
0.05	0.15	0.50	0.30

From Evidence to Belief

Ghostbusters

Observed	Sensor colour
Unknown	Ghost location
Model	Reading probabilities given distance

From Evidence to Belief

	Ghostbusters	Detective
Observed	Sensor colour	Puddle
Unknown	Ghost location	
Model	Reading probabilities given distance	



Does the puddle prove
it rained?

From Evidence to Belief

	Ghostbusters	Detective
Observed	Sensor colour	Puddle
Unknown	Ghost location	Whether it rained
Model	Reading probabilities given distance	How likely a puddle is with or without rain



Does the puddle prove
it rained?

From Evidence to Belief

	Ghostbusters	Detective
Observed	Sensor colour	Puddle
Unknown	Ghost location	Whether it rained
Model	Reading probabilities given distance	How likely a puddle is with or without rain



Does the puddle prove
it rained?

Use observed evidence to update beliefs
about what we cannot observe

Definition: Random Variables

A random variable names an uncertain quantity

Variable	Possible Values (Domain)
G : ghost location	Possible grid locations

Definition: Random Variables

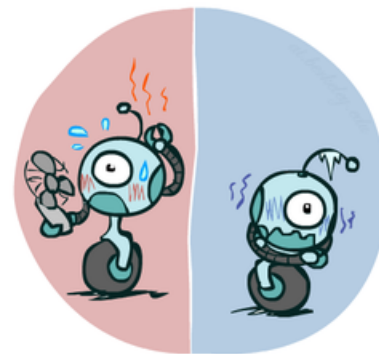
A random variable names an uncertain quantity

Variable	Possible Values (Domain)
G : ghost location	Possible grid locations
R : rain	{true, false}

Definition: Random Variables

A random variable names an uncertain quantity

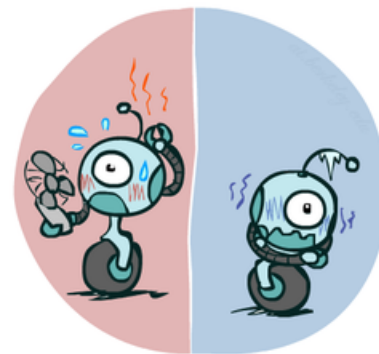
Variable	Possible Values (Domain)
G : ghost location	Possible grid locations
R : rain	{true, false}
T : temperature	{hot, cold}



Definition: Random Variables

A random variable names an uncertain quantity

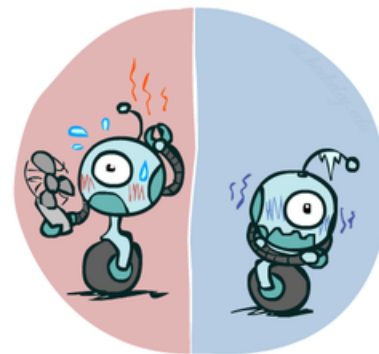
Variable	Possible Values (Domain)
G : ghost location	Possible grid locations
R : rain	{true, false}
T : temperature	{hot, cold}
W : weather	{sun, rain, fog, meteor}



Definition: Random Variables

A random variable names an uncertain quantity

Variable	Possible Values (Domain)
G : ghost location	Possible grid locations
R : rain	{true, false}
T : temperature	{hot, cold}
W : weather	{sun, rain, fog, meteor}



Capital letters name variables; lower-case values name outcomes

Boolean shorthand: $R = \text{true}$: $+r$ $R = \text{false}$: $-r$

Definition: Probability Distributions

A distribution assigns probabilities to possible values

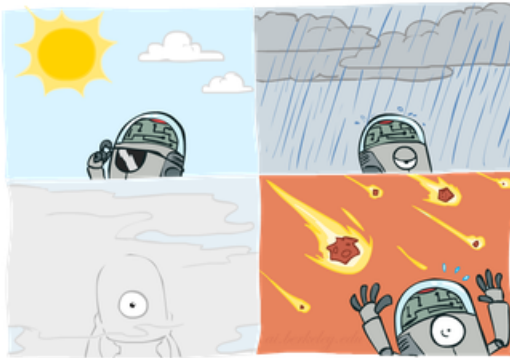
W

sun

rain

fog

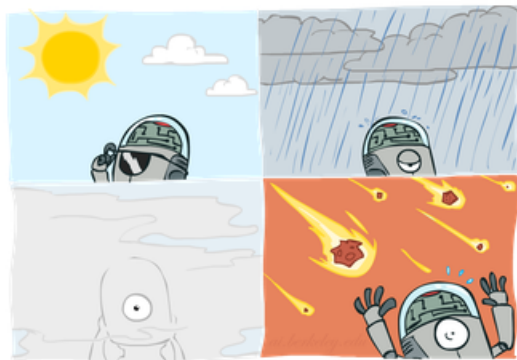
meteor



Definition: Probability Distributions

A distribution assigns probabilities to possible values

W	$P(W)$
sun	0.60
rain	0.10
fog	0.30
meteor	0.00



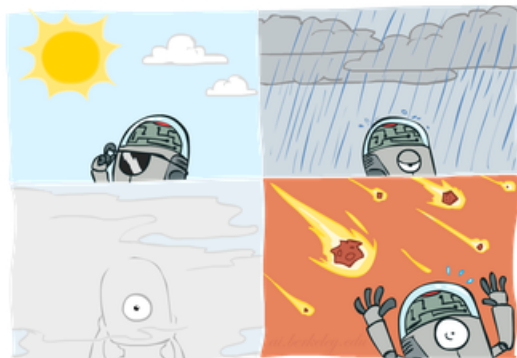
Definition: Probability Distributions

A distribution assigns probabilities to possible values

W	$P(W)$
sun	0.60
rain	0.10
fog	0.30
meteor	0.00

$$P(W = \text{rain}) = 0.10$$

One entry in the distribution



Definition: Probability Distributions

A distribution assigns probabilities to possible values

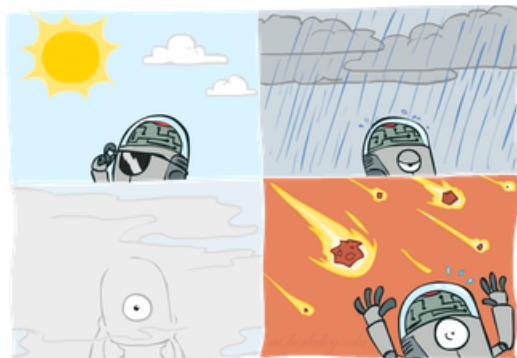
W	$P(W)$
sun	0.60
rain	0.10
fog	0.30
meteor	0.00

$$P(W = \text{rain}) = 0.10$$

One entry in the distribution

Nonnegative probabilities; total = 1

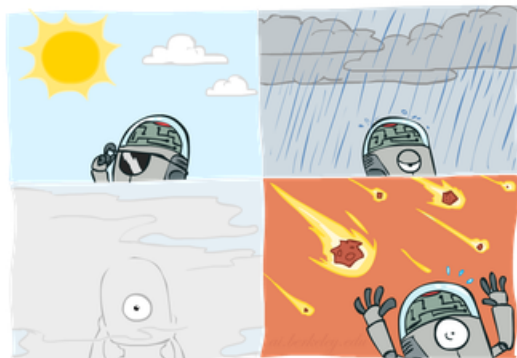
$$0.60 + 0.10 + 0.30 + 0.00 = 1$$



Definition: Probability Distributions

A distribution assigns probabilities to possible values

W	$P(W)$
sun	0.60
rain	0.10
fog	0.30
meteor	0.00



$$P(W = \text{rain}) = 0.10$$

One entry in the distribution

Nonnegative probabilities; total = 1

$$0.60 + 0.10 + 0.30 + 0.00 = 1$$

$$P(w) \geq 0 \quad \sum_w P(w) = 1$$

Definition: Joint Distributions

A joint distribution assigns probabilities to combinations of values

Definition: Joint Distributions

A joint distribution assigns probabilities to combinations of values

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Definition: Joint Distributions

A joint distribution assigns probabilities to combinations of values

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Probability: how likely is each assignment?



Definition: Joint Distributions

A joint distribution assigns probabilities to combinations of values

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Probability: how likely is each assignment?



$$P(T = \text{hot}, W = \text{sun}) = 0.40$$

Definition: Joint Distributions

A joint distribution assigns probabilities to combinations of values

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Probability: how likely is each assignment?



$$P(T = \text{hot}, W = \text{sun}) = 0.40$$

Entries ≥ 0 ; total = 1

Definition: Events

An event is a set of outcomes satisfying a condition

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Each row is one outcome

Definition: Events

An event is a set of outcomes satisfying a condition

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

$P(\text{hot, sun})$

Each row is one outcome

Definition: Events

An event is a set of outcomes satisfying a condition

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

$$P(\text{hot, sun}) = 0.40$$

Each row is one outcome

Definition: Events

An event is a set of outcomes satisfying a condition

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

$$P(\text{hot, sun}) = 0.40$$

$$P(\text{hot})$$

Each row is one outcome

Definition: Events

An event is a set of outcomes satisfying a condition

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

$$P(\text{hot, sun}) = 0.40$$

$$P(\text{hot}) = 0.40 + 0.10 = 0.50$$

Each row is one outcome

Definition: Events

An event is a set of outcomes satisfying a condition

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Each row is one outcome

$$P(\text{hot, sun}) = 0.40$$

$$P(\text{hot}) = 0.40 + 0.10 = 0.50$$

$$P(\text{hot or sun})$$

Definition: Events

An event is a set of outcomes satisfying a condition

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Each row is one outcome

$$P(\text{hot, sun}) = 0.40$$

$$P(\text{hot}) = 0.40 + 0.10 = 0.50$$

$$P(\text{hot or sun}) = 0.70$$

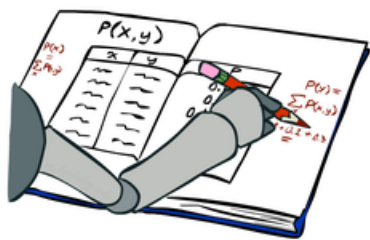
Definition: Marginal Distributions

A marginal distribution gives probabilities for one variable

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

We want the weather probabilities,
regardless of temperature

W	$P(W)$



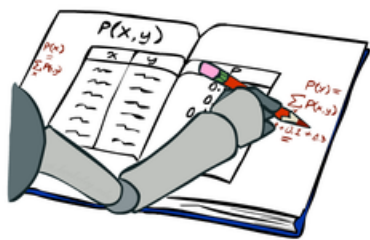
Definition: Marginal Distributions

A marginal distribution gives probabilities for one variable

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

We want the weather probabilities, regardless of temperature

W	$P(W)$
sun	$0.40 + 0.20 = 0.60$



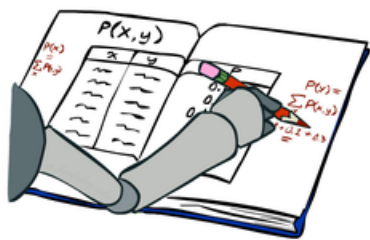
Definition: Marginal Distributions

A marginal distribution gives probabilities for one variable

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

We want the weather probabilities, regardless of temperature

W	$P(W)$
sun	$0.40 + 0.20 = 0.60$
rain	$0.10 + 0.30 = 0.40$



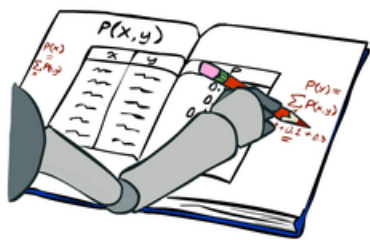
Definition: Marginal Distributions

A marginal distribution gives probabilities for one variable

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

We want the weather probabilities,
regardless of temperature

W	$P(W)$
sun	$0.40 + 0.20 = 0.60$
rain	$0.10 + 0.30 = 0.40$



Sum out temperature

$$P(w) = \sum_t P(t, w)$$

Activity: Events and Marginals

TASK

X	Y	$P(X, Y)$
+x	+y	0.20
+x	-y	0.30
-x	+y	0.40
-x	-y	0.10

$$P(+x, +y) = ?$$

$$P(+x) = ? \quad P(+y) = ?$$

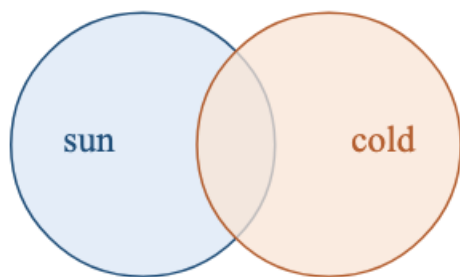
$$P(-y \text{ or } +x) = ?$$

Definition: Conditional Probability

Conditional probability describes likelihood given observed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Given cold, how likely is sun?



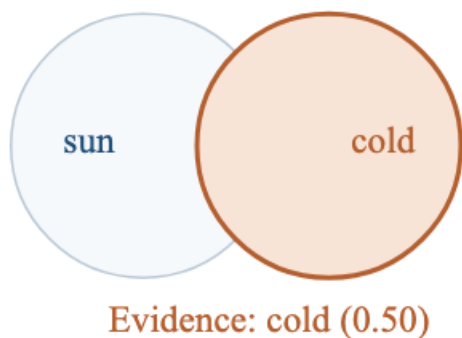
Definition: Conditional Probability

Conditional probability describes likelihood given observed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Given cold, how likely is sun?

$$P(\text{cold}) = 0.20 + 0.30 = 0.50$$



Definition: Conditional Probability

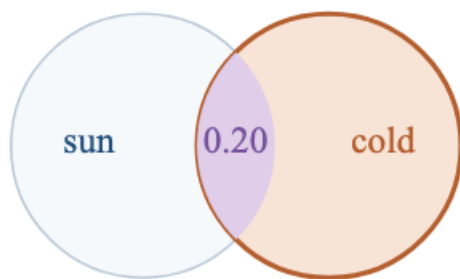
Conditional probability describes likelihood given observed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Given cold, how likely is sun?

$$P(\text{cold}) = 0.20 + 0.30 = 0.50$$

$$P(\text{sun, cold}) = 0.20$$



Evidence: cold (0.50)

Definition: Conditional Probability

Conditional probability describes likelihood given observed evidence

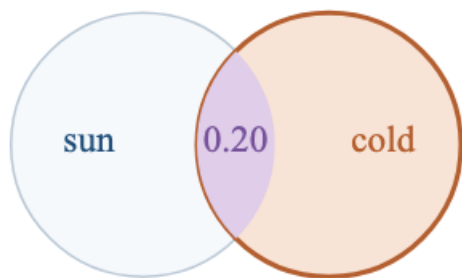
T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

Given cold, how likely is sun?

$$P(\text{cold}) = 0.20 + 0.30 = 0.50$$

$$P(\text{sun, cold}) = 0.20$$

$$P(\text{sun} \mid \text{cold}) = \frac{0.20}{0.50} = 0.40$$



Evidence: cold (0.50)

Definition: Conditional Probability

Conditional probability describes likelihood given observed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

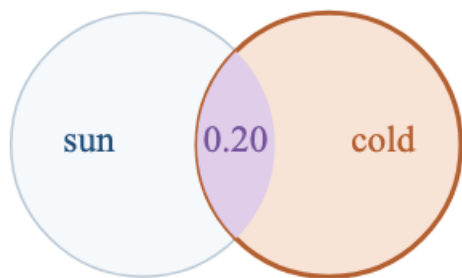
Given cold, how likely is sun?

$$P(\text{cold}) = 0.20 + 0.30 = 0.50$$

$$P(\text{sun, cold}) = 0.20$$

$$P(\text{sun} \mid \text{cold}) = \frac{0.20}{0.50} = 0.40$$

$$P(w \mid t) = \frac{P(w, t)}{P(t)} \quad P(t) > 0$$

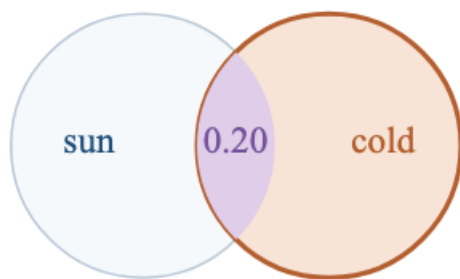


Evidence: cold (0.50)

Definition: Conditional Probability

Conditional probability describes likelihood given observed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30



Evidence: cold (0.50)

Given cold, how likely is sun?

$$P(\text{cold}) = 0.20 + 0.30 = 0.50$$

$$P(\text{sun, cold}) = 0.20$$

$$P(\text{sun} \mid \text{cold}) = \frac{0.20}{0.50} = 0.40$$

$$P(w \mid t) = \frac{P(w, t)}{P(t)} \quad P(t) > 0$$

Before observing cold: $P(\text{sun}) = 0.60$

Activity: Conditional Probabilities

TASK

X	Y	$P(X, Y)$
+x	+y	0.20
+x	-y	0.30
-x	+y	0.40
-x	-y	0.10

$$P(a \mid b) = \frac{P(a, b)}{P(b)} \quad P(b) > 0$$

$$P(+x \mid +y) = ?$$

$$P(-x \mid +y) = ?$$

$$P(-y \mid +x) = ?$$

Definition: Conditional Distributions

A conditional distribution gives probabilities for fixed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

One column for each evidence value

W
sun
rain

Definition: Conditional Distributions

A conditional distribution gives probabilities for fixed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

One column for each evidence value

W	$P(W \mid T = \text{cold})$
sun	0.40
rain	0.60

Definition: Conditional Distributions

A conditional distribution gives probabilities for fixed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

One column for each evidence value

W	$P(W \mid T = \text{cold})$	$P(W \mid T = \text{hot})$
sun	0.40	0.80
rain	0.60	0.20

Definition: Conditional Distributions

A conditional distribution gives probabilities for fixed evidence

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

One column for each evidence value

W	$P(W \mid T = \text{cold})$	$P(W \mid T = \text{hot})$
sun	0.40	0.80
rain	0.60	0.20
Total	1.00	1.00

Each column sums to 1

Definition: Normalization

Normalization rescales nonnegative weights so they sum to 1

Selected weights

W	$P(W, T = \text{cold})$
sun	0.20
rain	0.30

Divide each weight by the total

W	$P(W \mid T = \text{cold})$
sun	
rain	

Definition: Normalization

Normalization rescales nonnegative weights so they sum to 1

Selected weights

W	$P(W, T = \text{cold})$
sun	0.20
rain	0.30

$$Z = 0.20 + 0.30 = 0.50$$

Divide each weight by the total

W	$P(W \mid T = \text{cold})$
sun	
rain	

Definition: Normalization

Normalization rescales nonnegative weights so they sum to 1

Selected weights

W	$P(W, T = \text{cold})$
sun	0.20
rain	0.30

$$Z = 0.20 + 0.30 = 0.50$$

Divide each weight by the total

W	$P(W \mid T = \text{cold})$
sun	$\frac{0.20}{0.50} = 0.40$
rain	

Definition: Normalization

Normalization rescales nonnegative weights so they sum to 1

Selected weights

W	$P(W, T = \text{cold})$
sun	0.20
rain	0.30

$$Z = 0.20 + 0.30 = 0.50$$

Divide each weight by the total

W	$P(W \mid T = \text{cold})$
sun	$\frac{0.20}{0.50} = 0.40$
rain	$\frac{0.30}{0.50} = 0.60$

Definition: Normalization

Normalization rescales nonnegative weights so they sum to 1

Selected weights

W	$P(W, T = \text{cold})$
sun	0.20
rain	0.30

$$Z = 0.20 + 0.30 = 0.50$$

Divide each weight by the total

W	$P(W \mid T = \text{cold})$
sun	$\frac{0.20}{0.50} = 0.40$
rain	$\frac{0.30}{0.50} = 0.60$

Total	1.00
-------	------

$$p_i = \frac{w_i}{Z} \quad Z = \sum_j w_j > 0$$

Activity: Normalization

TASK

X	Y	$P(X, Y)$
+x	+y	0.20
+x	-y	0.30
-x	+y	0.40
-x	-y	0.10

$$P(X \mid -y) = ?$$

Select the entries, then normalize

$$p_i = \frac{w_i}{Z} \quad Z = \sum_j w_j > 0$$

Algorithm: Inference by Enumeration

Compute a query distribution from a joint distribution

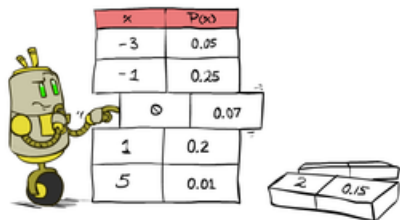
Role	Meaning
Q	Query: what we want to know
$E = e$	Evidence: what we observed
H	Hidden: other unobserved variables

Algorithm: Inference by Enumeration

Compute a query distribution from a joint distribution

Role	Meaning
Q	Query: what we want to know
$E = e$	Evidence: what we observed
H	Hidden: other unobserved variables

1. Select Matching Entries

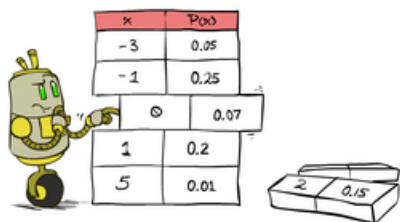


Algorithm: Inference by Enumeration

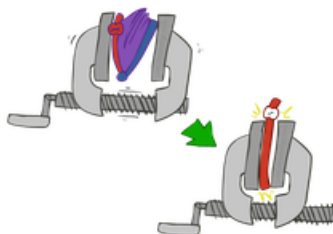
Compute a query distribution from a joint distribution

Role	Meaning
Q	Query: what we want to know
$E = e$	Evidence: what we observed
H	Hidden: other unobserved variables

1. Select Matching Entries



2. Sum Out Hidden Variables

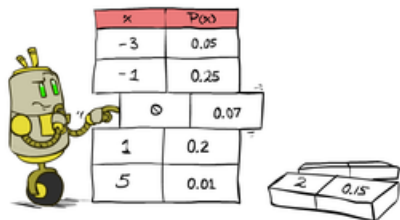


Algorithm: Inference by Enumeration

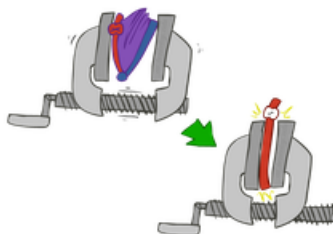
Compute a query distribution from a joint distribution

Role	Meaning
Q	Query: what we want to know
$E = e$	Evidence: what we observed
H	Hidden: other unobserved variables

1. Select Matching Entries



2. Sum Out Hidden Variables



3. Normalize

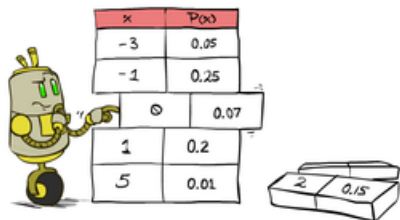
$$\times \frac{1}{Z}$$

Algorithm: Inference by Enumeration

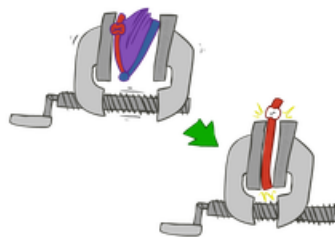
Compute a query distribution from a joint distribution

Role	Meaning
Q	Query: what we want to know
$E = e$	Evidence: what we observed
H	Hidden: other unobserved variables

1. Select Matching Entries



2. Sum Out Hidden Variables



3. Normalize

$$\times \frac{1}{Z}$$

$$P(Q \mid e) = \frac{1}{Z} \sum_h P(Q, e, h)$$

Example: Inference by Enumeration

Given: joint distribution

S	T	W	$P(S, T, W)$
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

S: season · T: temperature · W: weather

We know it is winter. How likely are sun and rain?

$$P(W \mid \text{winter}) = ?$$

Query: W · Evidence: S = winter · Hidden: T

Example: Inference by Enumeration

Given: joint distribution

S	T	W	$P(S, T, W)$
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

S: season · T: temperature · W: weather

We know it is winter. How likely are sun and rain?

$$P(W \mid \text{winter}) = ?$$

Query: W · Evidence: S = winter · Hidden: T

1. Select Matching Entries

Example: Inference by Enumeration

Given: joint distribution

S	T	W	$P(S, T, W)$
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

S: season · T: temperature · W: weather

We know it is winter. How likely are sun and rain?

$$P(W \mid \text{winter}) = ?$$

Query: W · Evidence: S = winter · Hidden: T

1. Select Matching Entries

2. Sum Out Hidden Variables

$$\text{sun: } 0.10 + 0.15 = 0.25$$

Example: Inference by Enumeration

Given: joint distribution

S	T	W	$P(S, T, W)$
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

S: season · T: temperature · W: weather

We know it is winter. How likely are sun and rain?

$$P(W \mid \text{winter}) = ?$$

Query: W · Evidence: S = winter · Hidden: T

1. Select Matching Entries

2. Sum Out Hidden Variables

$$\text{sun: } 0.10 + 0.15 = 0.25$$

$$\text{rain: } 0.05 + 0.20 = 0.25$$

Example: Inference by Enumeration

Given: joint distribution

S	T	W	$P(S, T, W)$
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

S: season · T: temperature · W: weather

We know it is winter. How likely are sun and rain?

$$P(W \mid \text{winter}) = ?$$

Query: W · Evidence: S = winter · Hidden: T

1. Select Matching Entries

2. Sum Out Hidden Variables

$$\text{sun: } 0.10 + 0.15 = 0.25$$

$$\text{rain: } 0.05 + 0.20 = 0.25$$

3. Normalize: $Z = 0.25 + 0.25 = 0.50$

Result: weather probabilities given winter

sun	rain
$\frac{0.25}{0.50} = 0.50$	$\frac{0.25}{0.50} = 0.50$

Sun and rain are equally likely, even though temperature is unknown

Problem: Enumeration Does Not Scale

Exact answers, but an exponentially large table

Binary variables	Entries in the joint table
3	8

Problem: Enumeration Does Not Scale

Exact answers, but an exponentially large table

Binary variables	Entries in the joint table
3	8
10	1,024

Problem: Enumeration Does Not Scale

Exact answers, but an exponentially large table

Binary variables	Entries in the joint table
3	8
10	1,024
30	1,073,741,824

Problem: Enumeration Does Not Scale

Exact answers, but an exponentially large table

Binary variables	Entries in the joint table
3	8
10	1,024
30	1,073,741,824

n variables, d values each $\Rightarrow d^n$ entries

Problem: Enumeration Does Not Scale

Exact answers, but an exponentially large table

Binary variables	Entries in the joint table
3	8
10	1,024
30	1,073,741,824

n variables, d values each $\Rightarrow d^n$ entries

Large models need a more compact representation

The Product Rule

Build a joint probability from a conditional and a marginal

$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

The Product Rule

Build a joint probability from a conditional and a marginal

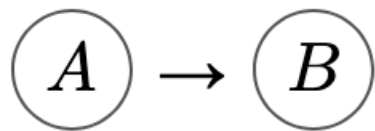
$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

$$P(a, b) = P(a)P(b \mid a)$$



The Product Rule

Build a joint probability from a conditional and a marginal

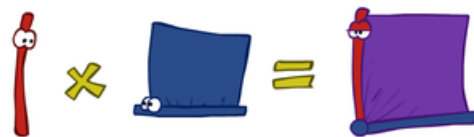


Circles: random variables

Arrow: B given A

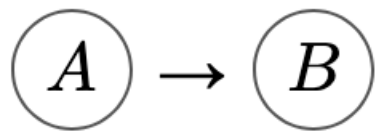
$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

$$P(a, b) = P(a)P(b \mid a)$$



The Product Rule

Build a joint probability from a conditional and a marginal

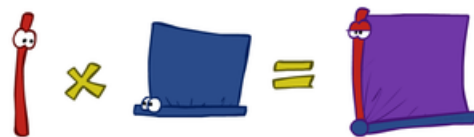


Circles: random variables

Arrow: B given A

$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

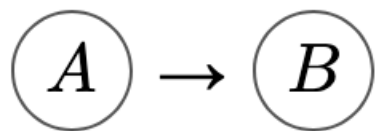
$$P(a, b) = P(a)P(b \mid a)$$



T		$P(T)$
hot		0.50
cold		0.50
W	Given hot	Given cold
sun	0.80	0.40
rain	0.20	0.60

The Product Rule

Build a joint probability from a conditional and a marginal

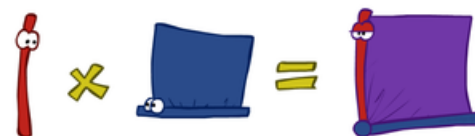


Circles: random variables

Arrow: B given A

$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

$$P(a, b) = P(a)P(b \mid a)$$

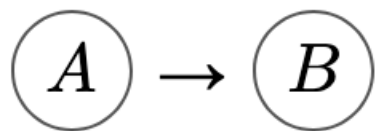


T		$P(T)$
hot		0.50
cold		0.50
W	Given hot	Given cold
sun	0.80	0.40
rain	0.20	0.60

T	W	$P(T, W)$
hot	sun	$0.50 \times 0.80 = 0.40$
hot	rain	
cold	sun	
cold	rain	

The Product Rule

Build a joint probability from a conditional and a marginal

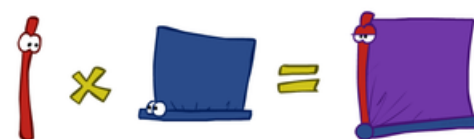


Circles: random variables

Arrow: B given A

$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

$$P(a, b) = P(a)P(b \mid a)$$

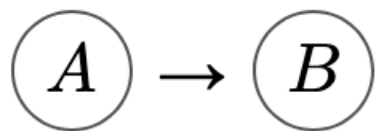


T		$P(T)$
hot		0.50
cold		0.50
W	Given hot	Given cold
sun	0.80	0.40
rain	0.20	0.60

T	W	$P(T, W)$
hot	sun	$0.50 \times 0.80 = 0.40$
hot	rain	$0.50 \times 0.20 = 0.10$
cold	sun	
cold	rain	

The Product Rule

Build a joint probability from a conditional and a marginal

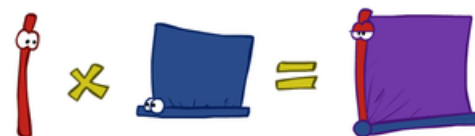


Circles: random variables

Arrow: B given A

$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

$$P(a, b) = P(a)P(b \mid a)$$

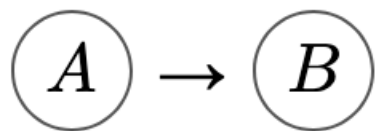


T		$P(T)$
hot		0.50
cold		0.50
W	Given hot	Given cold
sun	0.80	0.40
rain	0.20	0.60

T	W	$P(T, W)$
hot	sun	$0.50 \times 0.80 = 0.40$
hot	rain	$0.50 \times 0.20 = 0.10$
cold	sun	$0.50 \times 0.40 = 0.20$
cold	rain	

The Product Rule

Build a joint probability from a conditional and a marginal

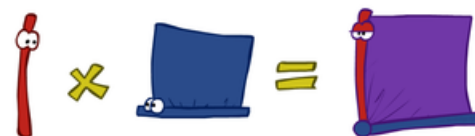


Circles: random variables

Arrow: B given A

$$P(b \mid a) = \frac{P(a, b)}{P(a)} \quad P(a) > 0$$

$$P(a, b) = P(a)P(b \mid a)$$



T		$P(T)$
hot		0.50
cold		0.50
W	Given hot	Given cold
sun	0.80	0.40
rain	0.20	0.60

T	W	$P(T, W)$
hot	sun	$0.50 \times 0.80 = 0.40$
hot	rain	$0.50 \times 0.20 = 0.10$
cold	sun	$0.50 \times 0.40 = 0.20$
cold	rain	$0.50 \times 0.60 = 0.30$

Activity: The Product Rule

TASK

W	D	$P(D W)$
sun	wet	0.10
sun	dry	0.90
rain	wet	0.70
rain	dry	0.30

W: weather · D: wet/dry condition

Find all four joint probabilities

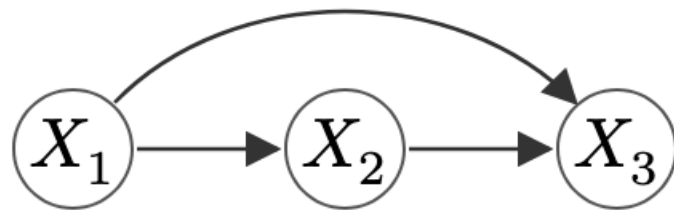
$$P(\text{sun}) = 0.80 \quad P(\text{rain}) = 0.20$$

$$P(D, W) = ?$$

$$P(d, w) = P(w)P(d | w)$$

The Chain Rule

Build a joint distribution by applying the product rule repeatedly

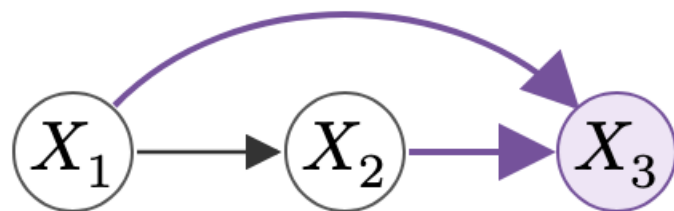


Three random variables

$$P(x_1, x_2, x_3) = ?$$

The Chain Rule

Build a joint distribution by applying the product rule repeatedly



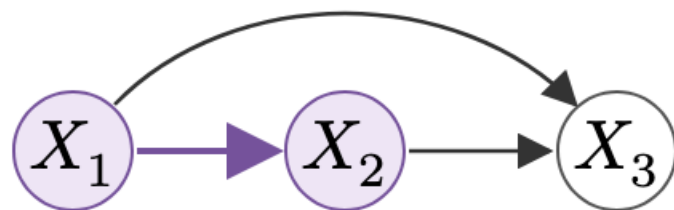
Three random variables

$$P(x_1, x_2, x_3) = ?$$

$$P(x_1, x_2, x_3) = P(x_1, x_2)P(x_3 \mid x_1, x_2)$$

The Chain Rule

Build a joint distribution by applying the product rule repeatedly



Three random variables

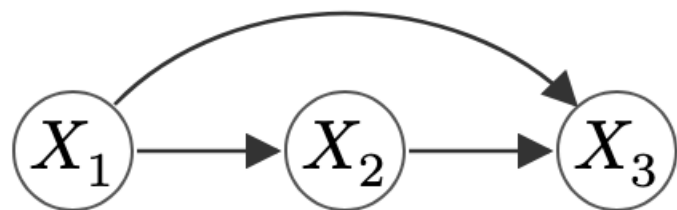
$$P(x_1, x_2, x_3) = ?$$

$$P(x_1, x_2, x_3) = P(x_1, x_2)P(x_3 \mid x_1, x_2)$$

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$$

The Chain Rule

Build a joint distribution by applying the product rule repeatedly



Three random variables

$$P(x_1, x_2, x_3) = ?$$

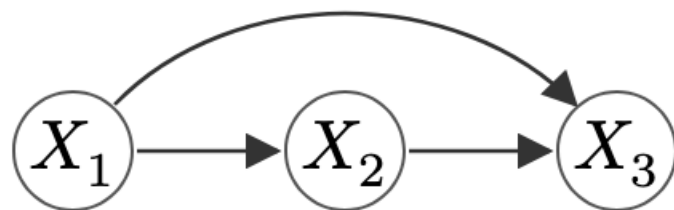
$$P(x_1, x_2, x_3) = P(x_1, x_2)P(x_3 \mid x_1, x_2)$$

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$$

$$P(x_1, \dots, x_n) = P(x_1) \prod_{i=2}^n P(x_i \mid x_1, \dots, x_{i-1})$$

The Chain Rule

Build a joint distribution by applying the product rule repeatedly



Three random variables

$$P(x_1, x_2, x_3) = ?$$

$$P(x_1, x_2, x_3) = P(x_1, x_2)P(x_3 \mid x_1, x_2)$$

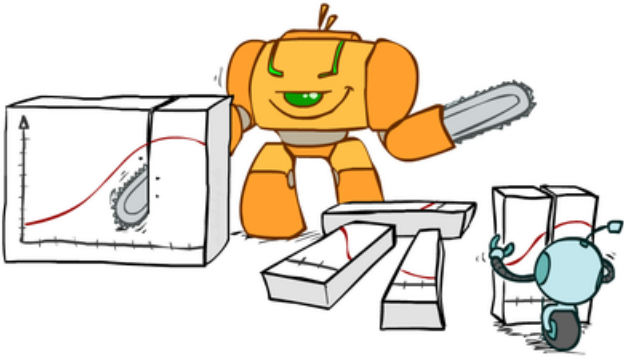
$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$$

$$P(x_1, \dots, x_n) = P(x_1) \prod_{i=2}^n P(x_i \mid x_1, \dots, x_{i-1})$$

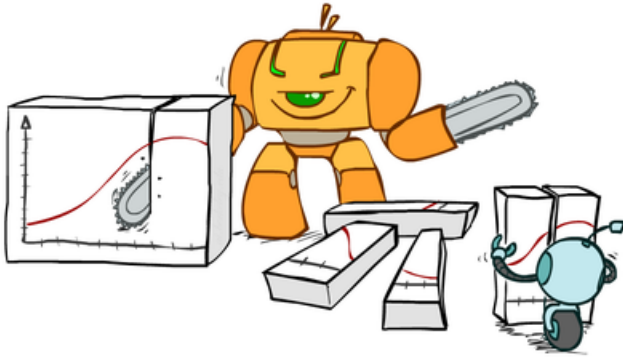
Each variable can depend on all earlier variables

Bayes' Rule

Known: how likely is wet ground given rain?



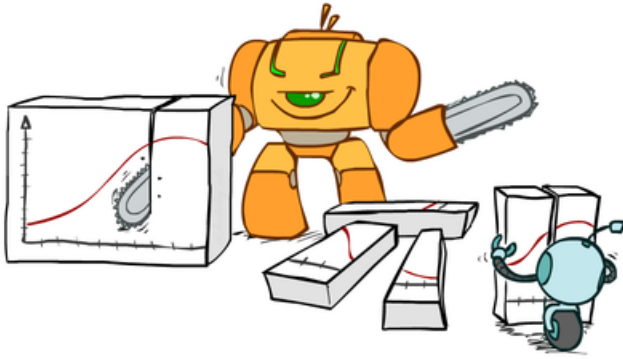
Bayes' Rule



Known: how likely is wet ground given rain?

Wet ground: how likely is rain?

Bayes' Rule



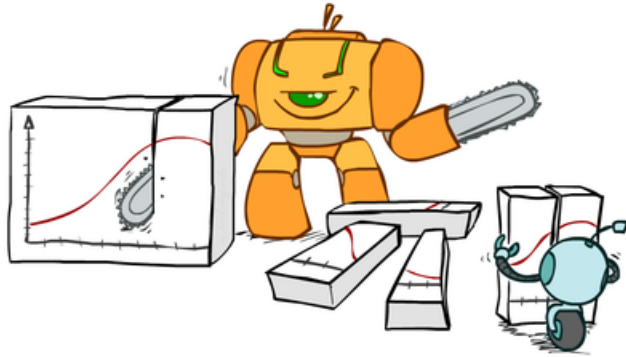
a : rain · b : wet ground

Known: how likely is wet ground given rain?

Wet ground: how likely is rain?

$$P(a, b) = P(a)P(b \mid a)$$

Bayes' Rule



a : rain · b : wet ground

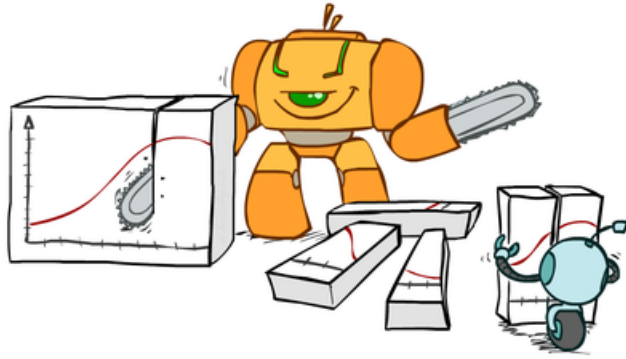
Known: how likely is wet ground given rain?

Wet ground: how likely is rain?

$$P(a, b) = P(a)P(b \mid a)$$

$$P(a, b) = P(b)P(a \mid b)$$

Bayes' Rule



a : rain · b : wet ground

Known: how likely is wet ground given rain?

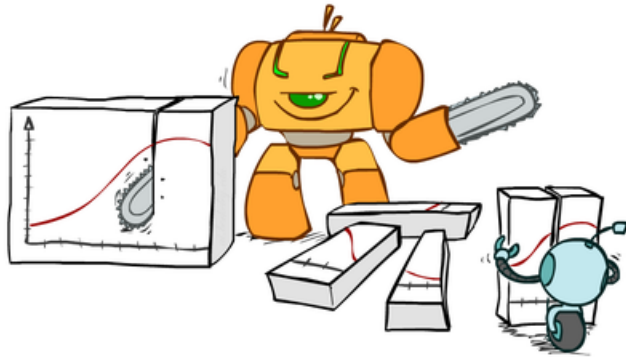
Wet ground: how likely is rain?

$$P(a, b) = P(a)P(b | a)$$

$$P(a, b) = P(b)P(a | b)$$

$$P(a | b) = \frac{P(b | a)P(a)}{P(b)}$$

Bayes' Rule



a : rain · b : wet ground

Known: how likely is wet ground given rain?

Wet ground: how likely is rain?

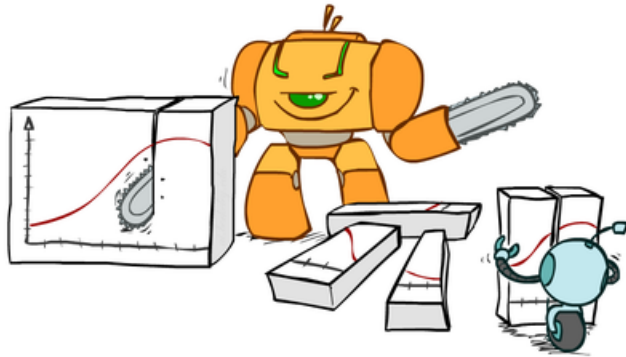
$$P(a, b) = P(a)P(b | a)$$

$$P(a, b) = P(b)P(a | b)$$

$$P(a | b) = \frac{P(b | a)P(a)}{P(b)}$$

Prior: $P(a)$ · Likelihood: $P(b | a)$ · Evidence probability: $P(b) > 0$

Bayes' Rule



a : rain · b : wet ground

Known: how likely is wet ground given rain?

Wet ground: how likely is rain?

$$P(a, b) = P(a)P(b | a)$$

$$P(a, b) = P(b)P(a | b)$$

$$P(a | b) = \frac{P(b | a)P(a)}{P(b)}$$

Prior: $P(a)$ · Likelihood: $P(b | a)$ · Evidence probability: $P(b) > 0$

Use the observation to update our belief

Example: Bayes' Rule

Given a stiff neck, how likely is meningitis?

Illustrative model · M: meningitis · S: stiff neck

Given probability	Value
Prior: $P(+m)$	0.0001
Likelihood: $P(+s \mid +m)$	0.80
Evidence: $P(+s)$	0.01

Example: Bayes' Rule

Given a stiff neck, how likely is meningitis?

Illustrative model · M: meningitis · S: stiff neck

Given probability	Value
Prior: $P(+m)$	0.0001
Likelihood: $P(+s \mid +m)$	0.80
Evidence: $P(+s)$	0.01

$$P(+m \mid +s) = \frac{P(+s \mid +m)P(+m)}{P(+s)}$$

Example: Bayes' Rule

Given a stiff neck, how likely is meningitis?

Illustrative model · M: meningitis · S: stiff neck

Given probability	Value
Prior: $P(+m)$	0.0001
Likelihood: $P(+s \mid +m)$	0.80
Evidence: $P(+s)$	0.01

$$P(+m \mid +s) = \frac{P(+s \mid +m)P(+m)}{P(+s)}$$

$$P(+m \mid +s) = \frac{0.80 \times 0.0001}{0.01}$$

Example: Bayes' Rule

Given a stiff neck, how likely is meningitis?

Illustrative model · M: meningitis · S: stiff neck

Given probability	Value
Prior: $P(+m)$	0.0001
Likelihood: $P(+s \mid +m)$	0.80
Evidence: $P(+s)$	0.01

$$P(+m \mid +s) = \frac{P(+s \mid +m)P(+m)}{P(+s)}$$

$$P(+m \mid +s) = \frac{0.80 \times 0.0001}{0.01}$$

$$P(+m \mid +s) = 0.008 = 0.8\%$$

Example: Bayes' Rule

Given a stiff neck, how likely is meningitis?

Illustrative model · M: meningitis · S: stiff neck

Given probability	Value
Prior: $P(+m)$	0.0001
Likelihood: $P(+s \mid +m)$	0.80
Evidence: $P(+s)$	0.01

$$P(+m \mid +s) = \frac{P(+s \mid +m)P(+m)}{P(+s)}$$

$$P(+m \mid +s) = \frac{0.80 \times 0.0001}{0.01}$$

$$P(+m \mid +s) = 0.008 = 0.8\%$$

A likely symptom does not imply a likely disease

Activity: Bayes' Rule

TASK

W	D	$P(D W)$
sun	wet	0.10
sun	dry	0.90
rain	wet	0.70
rain	dry	0.30

W: weather · D: wet/dry condition

It is dry. How likely are sun and rain?

$$P(\text{sun}) = 0.80 \quad P(\text{rain}) = 0.20$$

$$P(W | \text{dry}) = ?$$

$$P(w | \text{dry}) = \frac{P(\text{dry} | w)P(w)}{P(\text{dry})}$$

Example: Bayes' Rule in Ghostbusters

The sensor reports yellow: a noisy clue about distance



How does yellow change the probability of each possible ghost location?

G: location · E: sensor colour

Example: Bayes' Rule in Ghostbusters

The sensor reports yellow: a noisy clue about distance



How does yellow change the probability of each possible ghost location?

G: location · E: sensor colour

Apply the product rule: joint probability

$$P(G = g, E = \text{yellow}) = P(G = g)P(E = \text{yellow} \mid G = g)$$

Example: Bayes' Rule in Ghostbusters

The sensor reports yellow: a noisy clue about distance



How does yellow change the probability of each possible ghost location?

G: location · E: sensor colour

Apply the product rule: joint probability

$$P(G = g, E = \text{yellow}) = P(G = g)P(E = \text{yellow} \mid G = g)$$

Condition on yellow

$$P(G = g \mid E = \text{yellow}) = \frac{P(G = g, E = \text{yellow})}{P(E = \text{yellow})}$$

Example: Bayes' Rule in Ghostbusters

The sensor reports yellow: a noisy clue about distance



How does yellow change the probability of each possible ghost location?

G: location · E: sensor colour

Apply the product rule: joint probability

$$P(G = g, E = \text{yellow}) = P(G = g)P(E = \text{yellow} \mid G = g)$$

Condition on yellow

$$P(G = g \mid E = \text{yellow}) = \frac{P(G = g, E = \text{yellow})}{P(E = \text{yellow})}$$

Use the same two steps for every location

Ghostbusters with Probability

Example: a yellow reading at column 3, row 4

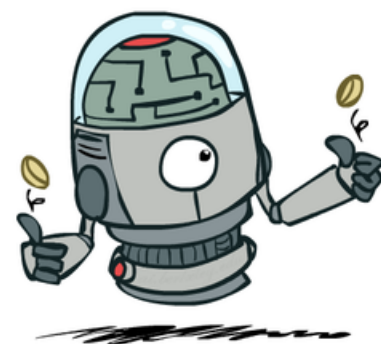
0.01	0.03	0.03	0.03	0.01	0.01	0.01	0.01	0.01	0.01
0.03	0.03	0.01	0.03	0.03	0.01	0.01	0.01	0.01	0.01
0.03	0.01	0.01	0.01	0.03	0.03	0.01	0.01	0.01	0.01
0.01	0.01	S < 0.01	0.01	0.01	0.03	0.03	0.01	0.01	0.01
0.03	0.01	0.01	0.01	0.03	0.03	0.01	0.01	0.01	0.01
0.03	0.03	0.01	0.03	0.03	0.01	0.01	0.01	0.01	0.01

Simplified model · Uniform prior · Rounded posterior probabilities

$$P(G = (3, 4) \mid \text{yellow}) = \frac{(1/60)(0.08)}{0.2930} \approx 0.0046$$

Definition: Independence

We toss two fair coins, X and Y

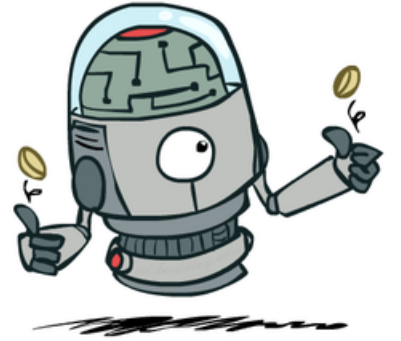


Definition: Independence

We toss two fair coins, X and Y

Y shows heads

Does this change the probability that X shows heads?



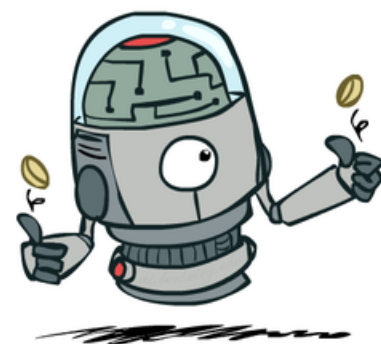
Definition: Independence

We toss two fair coins, X and Y

Y shows heads

Does this change the probability that X shows heads?

$$P(X = H) = \frac{1}{2} \quad \rightarrow \quad P(X = H \mid Y = H) = \frac{1}{2}$$



Definition: Independence

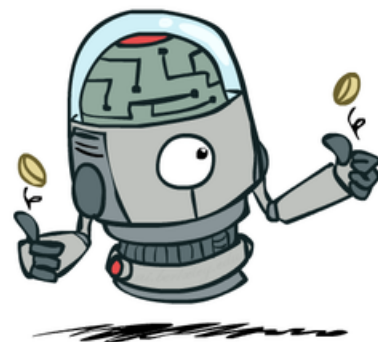
We toss two fair coins, X and Y

Y shows heads

Does this change the probability that X shows heads?

$$P(X = H) = \frac{1}{2} \quad \rightarrow \quad P(X = H \mid Y = H) = \frac{1}{2}$$

Independent: knowing Y does not change $P(X)$



Definition: Independence

We toss two fair coins, X and Y

Y shows heads

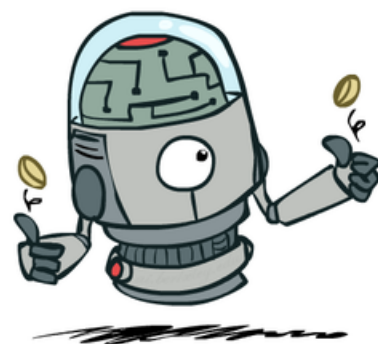
Does this change the probability that X shows heads?

$$P(X = H) = \frac{1}{2} \quad \rightarrow \quad P(X = H \mid Y = H) = \frac{1}{2}$$

Independent: knowing Y does not change $P(X)$

$$P(x \mid y) = P(x)$$

For every x and y with $P(y) > 0$



Definition: Independence

We toss two fair coins, X and Y

Y shows heads

Does this change the probability that X shows heads?

$$P(X = H) = \frac{1}{2} \quad \rightarrow \quad P(X = H \mid Y = H) = \frac{1}{2}$$

Independent: knowing Y does not change $P(X)$

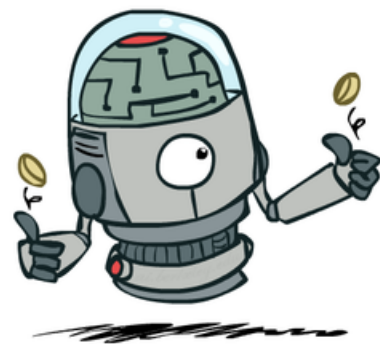
$$P(x \mid y) = P(x)$$

For every x and y with $P(y) > 0$

By the product rule:

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

For every x and y



Checking Independence

TASK

Are temperature and weather independent?

$$P(t, w) \stackrel{?}{=} P(t)P(w)$$

T	W	$P(T, W)$
hot	sun	0.40
hot	rain	0.10
cold	sun	0.20
cold	rain	0.30

T: temperature · W: weather

Why Independence Helps

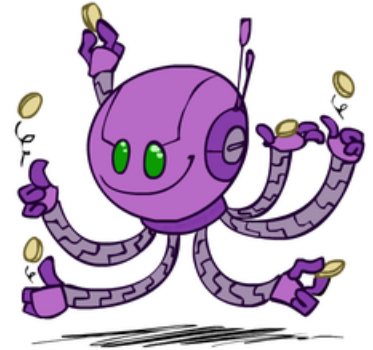
30 coin flips $\rightarrow$ over a billion joint outcomes

X_1	$P(X_1)$
H	0.50
T	0.50

X_2	$P(X_2)$
H	0.50
T	0.50

...

X_{30}	$P(X_{30})$
H	0.50
T	0.50

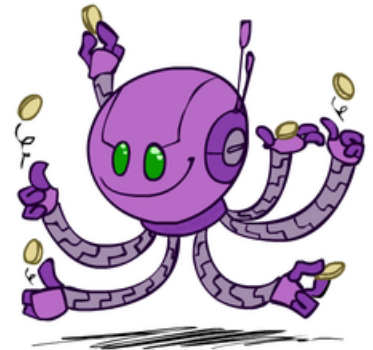


Why Independence Helps

30 coin flips $\rightarrow$ over a billion joint outcomes

X_1	$P(X_1)$		X_2	$P(X_2)$		X_{30}	$P(X_{30})$
H	0.50		H	0.50	...	H	0.50
T	0.50		T	0.50		T	0.50

With independence, multiply
30 individual probabilities



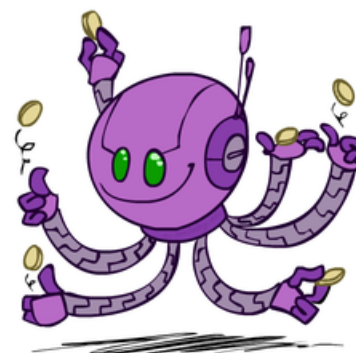
Why Independence Helps

30 coin flips $\rightarrow$ over a billion joint outcomes

X_1	$P(X_1)$		X_2	$P(X_2)$		X_{30}	$P(X_{30})$
H	0.50		H	0.50	...	H	0.50
T	0.50		T	0.50		T	0.50

With independence, multiply
30 individual probabilities

$$P(H, T, H) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$



Why Independence Helps

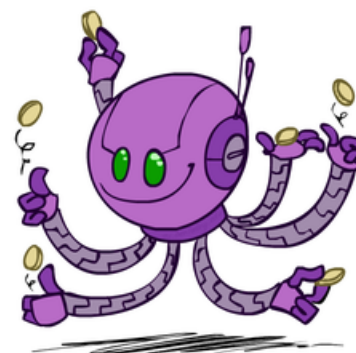
30 coin flips $\rightarrow$ over a billion joint outcomes

X_1	$P(X_1)$		X_2	$P(X_2)$		X_{30}	$P(X_{30})$
H	0.50		H	0.50	...	H	0.50
T	0.50		T	0.50		T	0.50

With independence, multiply
30 individual probabilities

$$P(H, T, H) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i) = \left(\frac{1}{2}\right)^n$$



Why Independence Helps

30 coin flips $\rightarrow$ over a billion joint outcomes

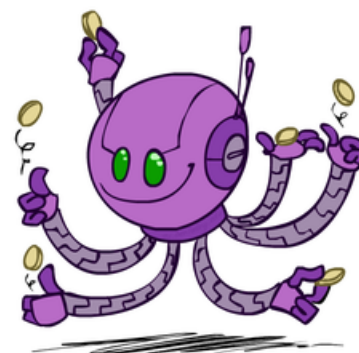
X_1	$P(X_1)$		X_2	$P(X_2)$		X_{30}	$P(X_{30})$
H	0.50		H	0.50	...	H	0.50
T	0.50		T	0.50		T	0.50

With independence, multiply
30 individual probabilities

$$P(H, T, H) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

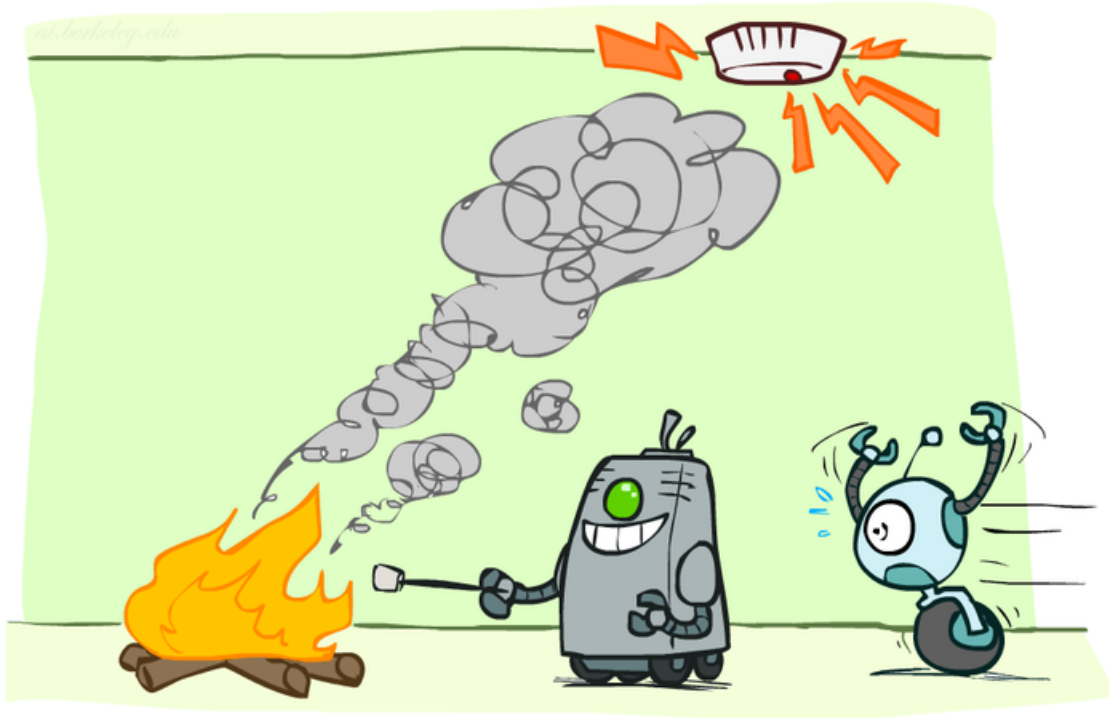
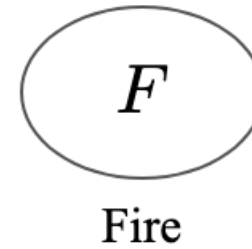
$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i) = \left(\frac{1}{2}\right)^n$$

Compute what you need. Don't store the whole joint table



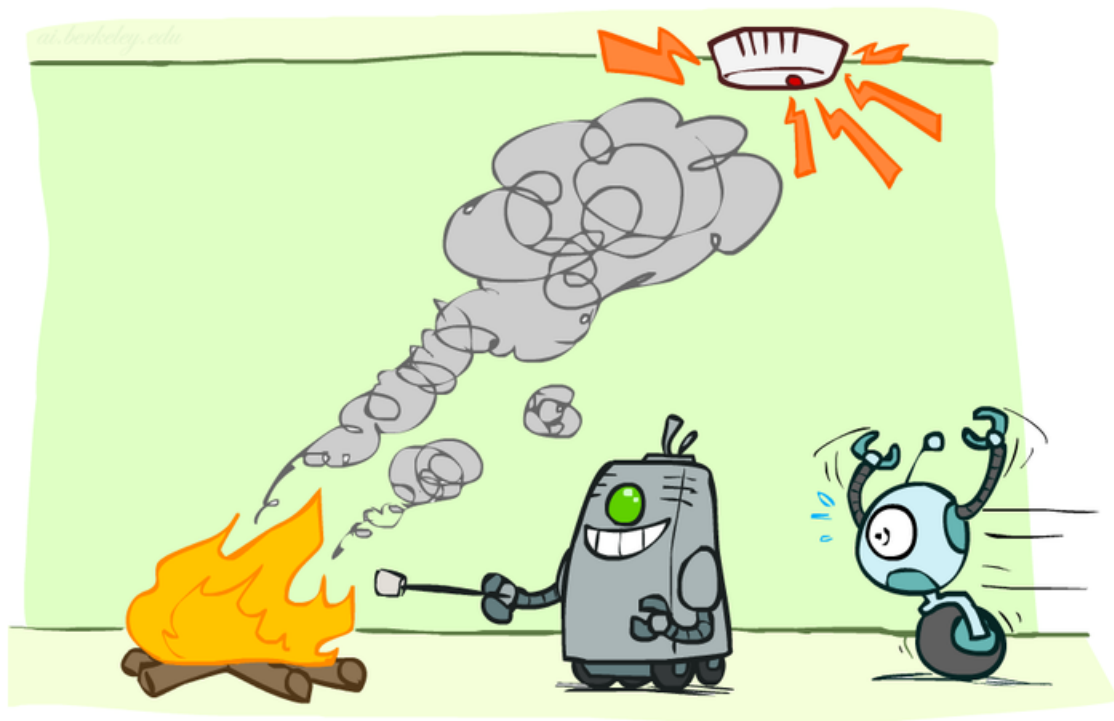
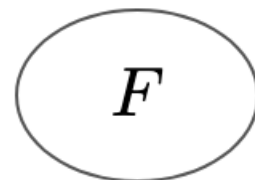
Conditional Independence

Example model: a smoke-triggered alarm

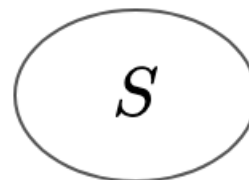


Conditional Independence

Example model: a smoke-triggered alarm



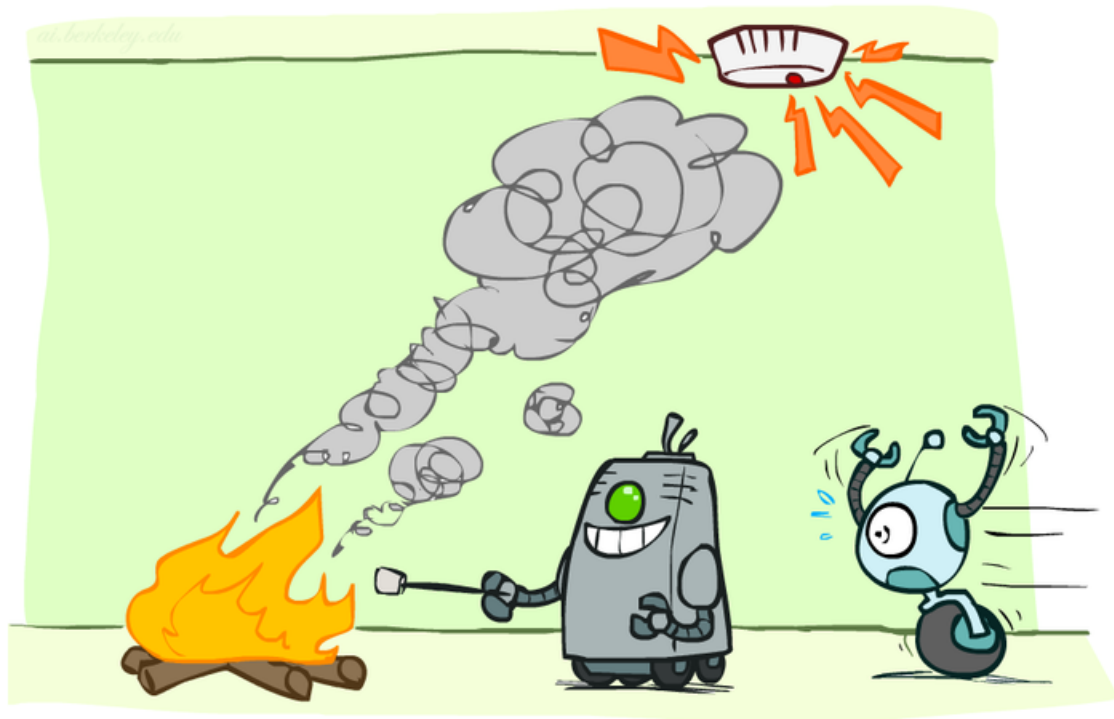
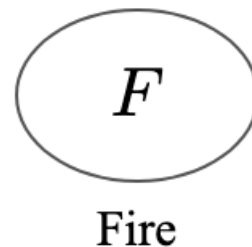
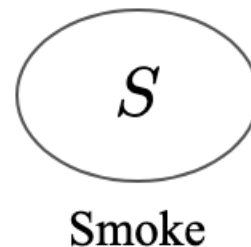
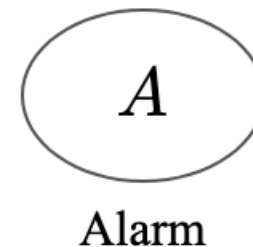
Fire



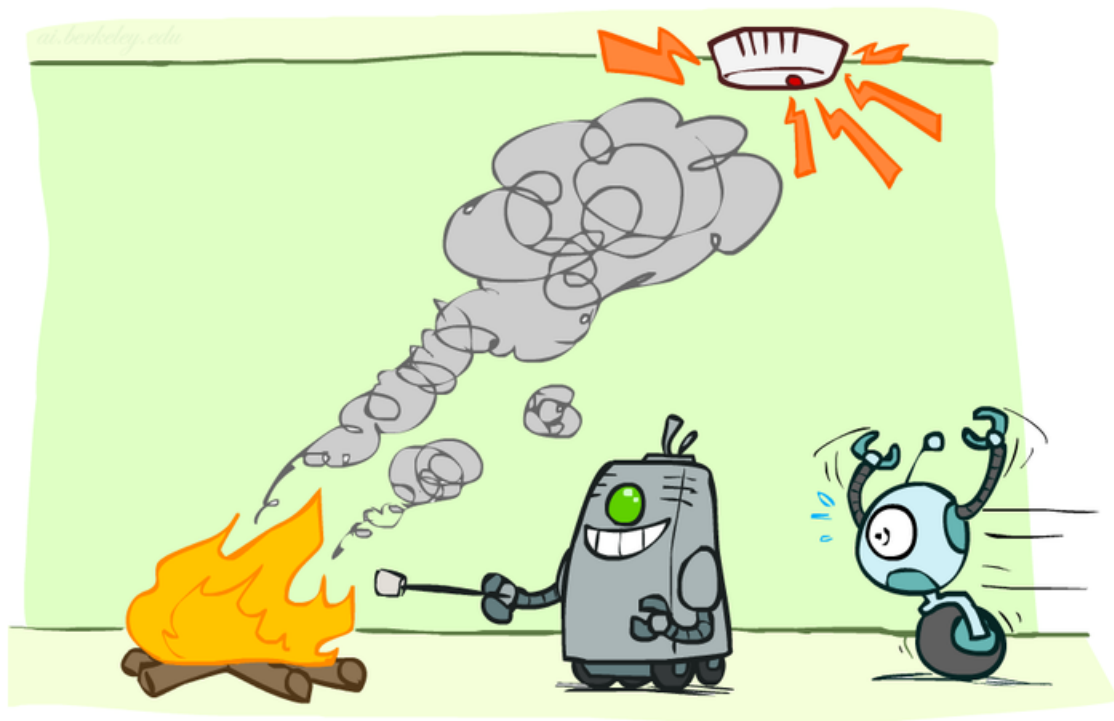
Smoke

Conditional Independence

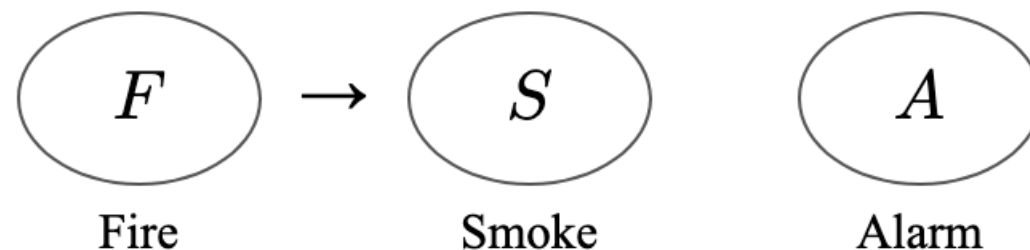
Example model: a smoke-triggered alarm



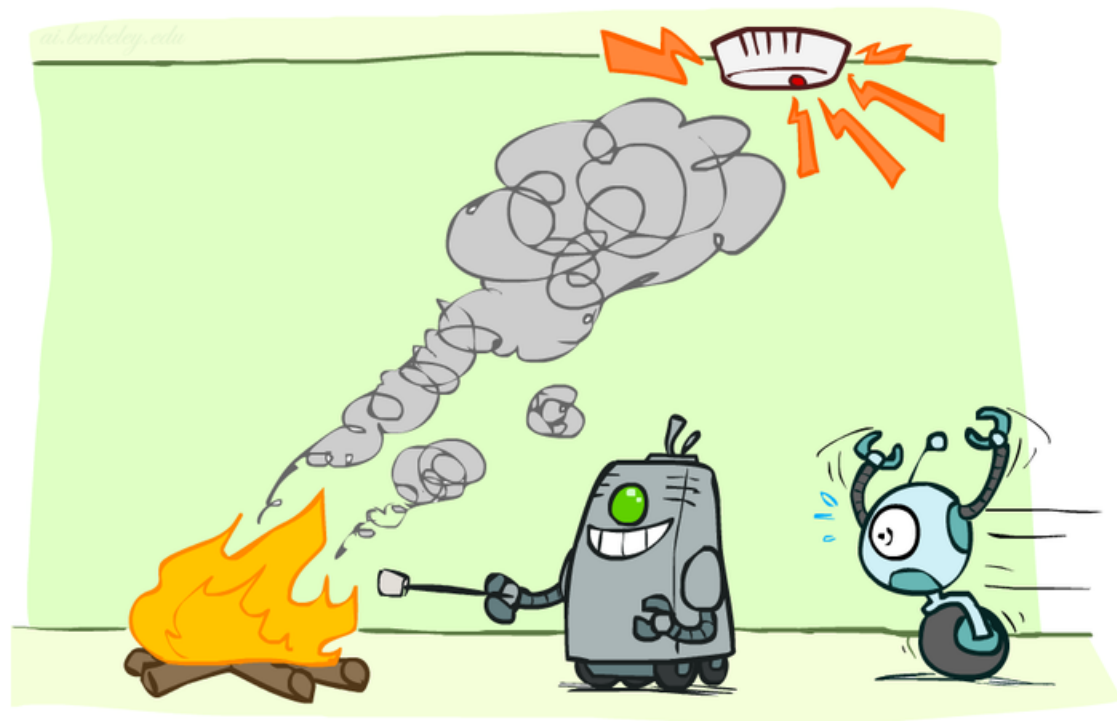
Conditional Independence



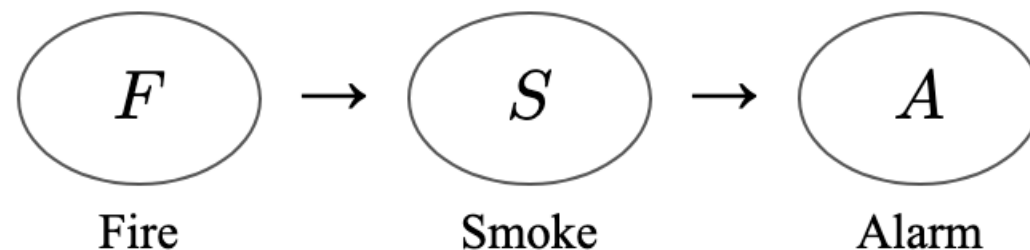
Example model: a smoke-triggered alarm



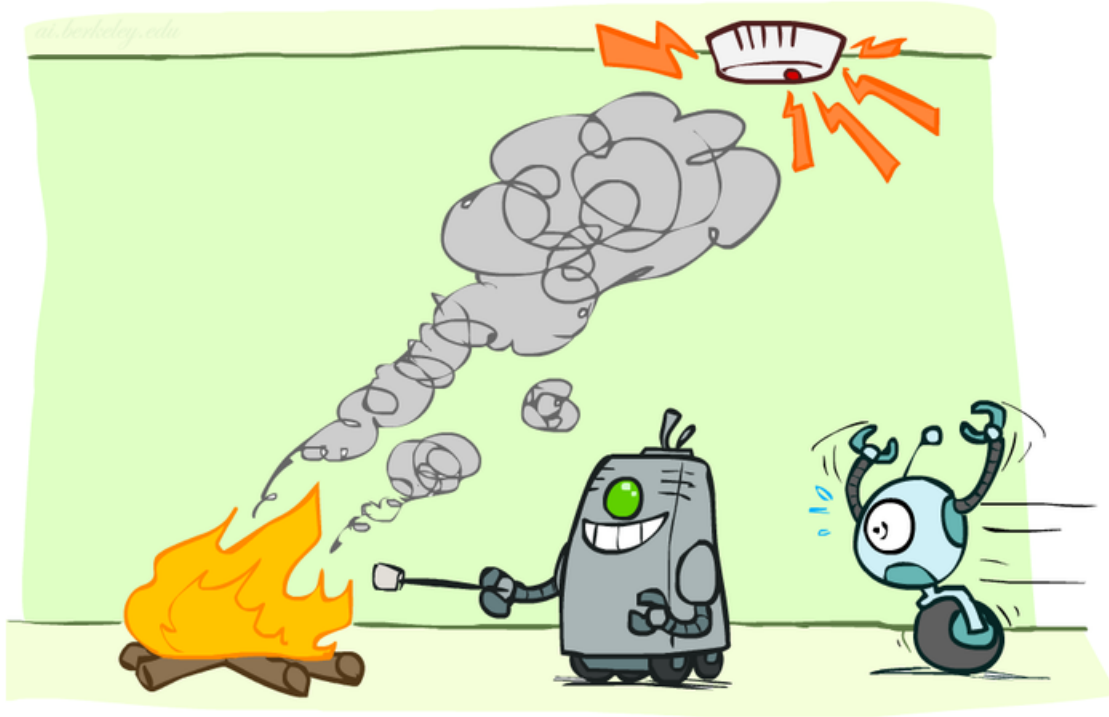
Conditional Independence



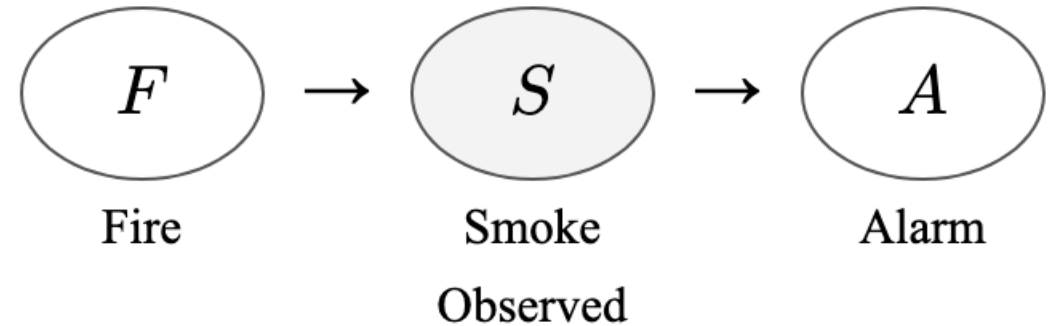
Example model: a smoke-triggered alarm



Conditional Independence

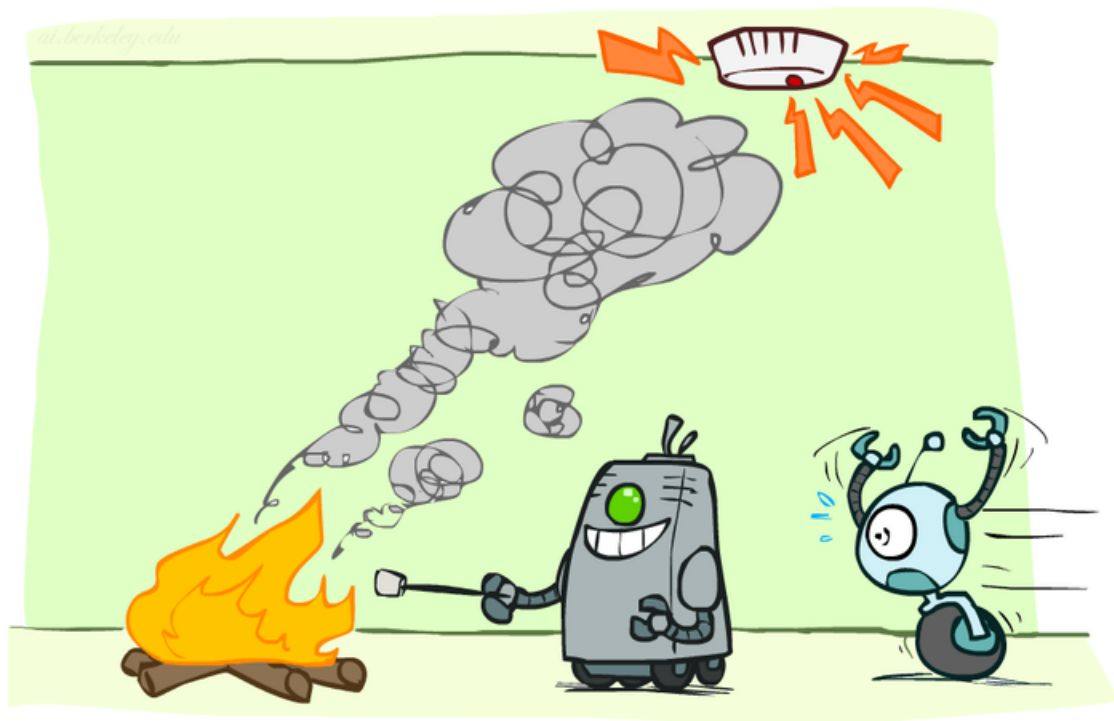


Example model: a smoke-triggered alarm

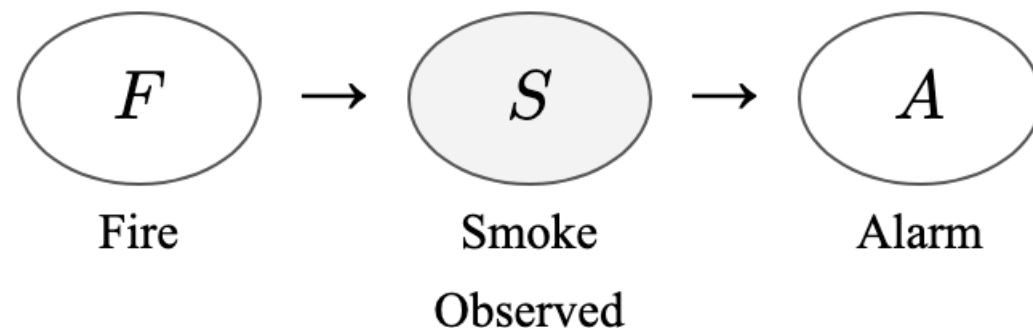


Once smoke is known, fire gives no extra information about the alarm

Conditional Independence



Example model: a smoke-triggered alarm



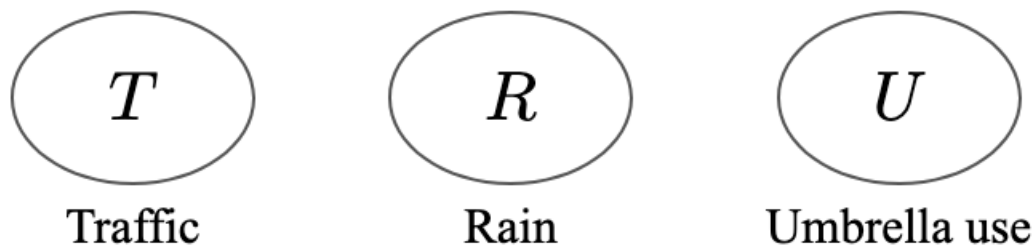
Once smoke is known, fire gives no extra information about the alarm

$$A \perp F \mid S \quad P(A \mid S, F) = P(A \mid S)$$

Activity: Conditional Independence

TASK

In this model, rain affects both traffic and umbrella use



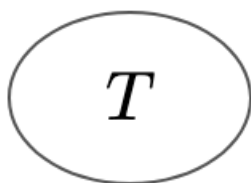
Does umbrella use tell us about traffic?



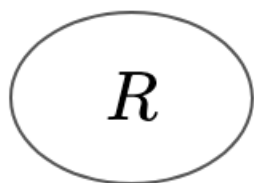
Activity: Conditional Independence

TASK

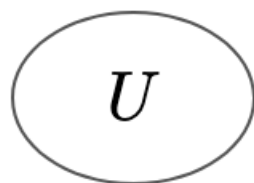
In this model, rain affects both traffic and umbrella use



Traffic



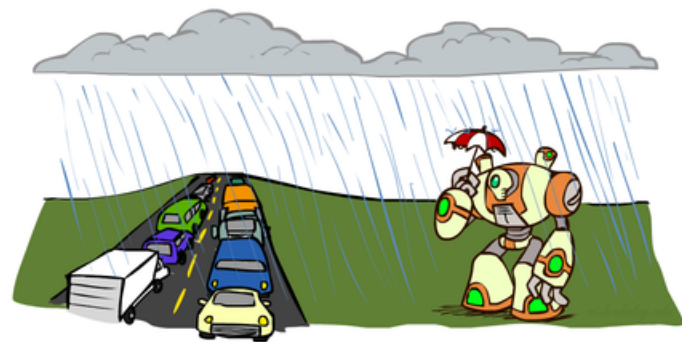
Rain



Umbrella use

Does umbrella use tell us about traffic?

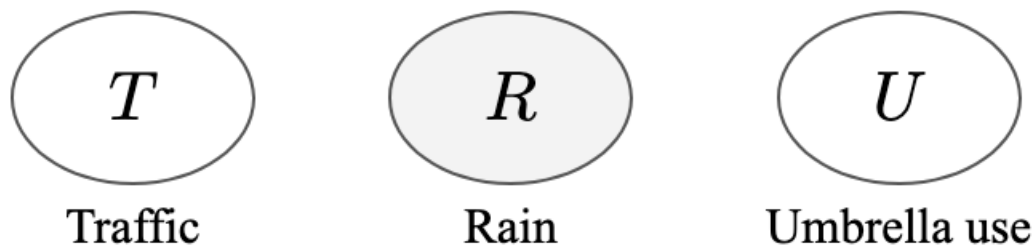
What if we already know whether it is raining?



Activity: Conditional Independence

TASK

In this model, rain affects both traffic and umbrella use



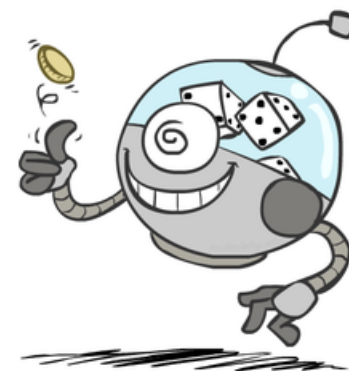
Does umbrella use tell us about traffic?

What if we already know whether it is raining?



Probability Recap

Operation	Rule
Marginalize	$P(x) = \sum_y P(x, y)$
Condition	$P(x \mid y) = \frac{P(x, y)}{P(y)}$



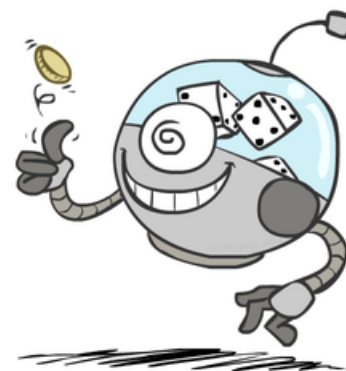
Probability Recap

Operation	Rule
Marginalize	$P(x) = \sum_y P(x, y)$
Condition	$P(x \mid y) = \frac{P(x, y)}{P(y)}$
Product rule	$P(x, y) = P(x \mid y)P(y)$
Chain rule	$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i \mid x_1, \dots, x_{i-1})$
Bayes' rule	$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)}$



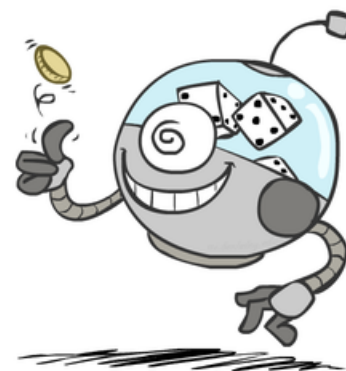
Probability Recap

Operation	Rule
Marginalize	$P(x) = \sum_y P(x, y)$
Condition	$P(x \mid y) = \frac{P(x, y)}{P(y)}$
Product rule	$P(x, y) = P(x \mid y)P(y)$
Chain rule	$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i \mid x_1, \dots, x_{i-1})$
Bayes' rule	$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)}$
Independence	$P(x, y) = P(x)P(y)$
Conditional independence	$P(x \mid y, z) = P(x \mid z)$



Probability Recap

Operation	Rule
Marginalize	$P(x) = \sum_y P(x, y)$
Condition	$P(x \mid y) = \frac{P(x, y)}{P(y)}$
Product rule	$P(x, y) = P(x \mid y)P(y)$
Chain rule	$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i \mid x_1, \dots, x_{i-1})$
Bayes' rule	$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)}$
Independence	$P(x, y) = P(x)P(y)$
Conditional independence	$P(x \mid y, z) = P(x \mid z)$



Next: use conditional independence to model a changing state

Reasoning over Time

The state can change over time



Reasoning over Time

The state can change over time

Robot localization



Reasoning over Time



The state can change over time

Robot localization

Speech recognition

Reasoning over Time



The state can change over time

Robot localization

Speech recognition

Monitoring a changing environment

Reasoning over Time



The state can change over time

Robot localization

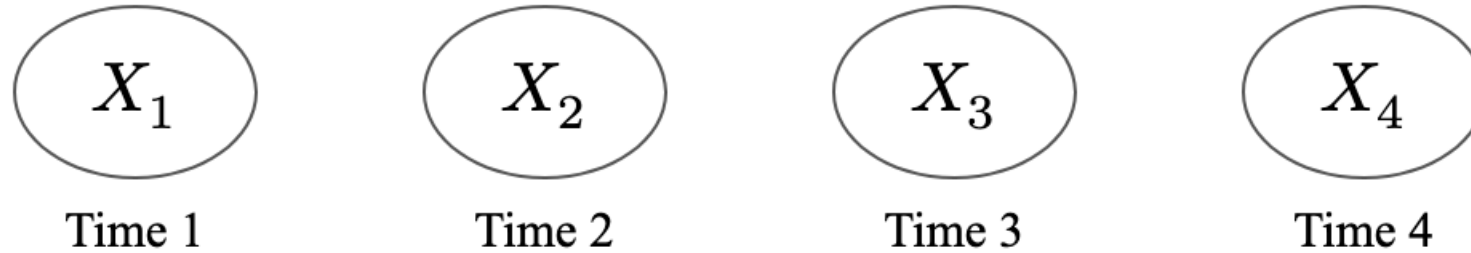
Speech recognition

Monitoring a changing environment

Represent a sequence of states

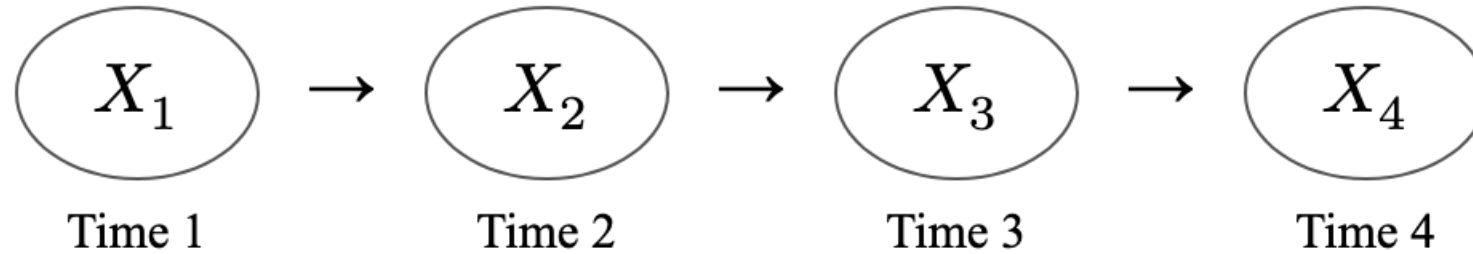
Definition: Markov Models

X_t : state at time t



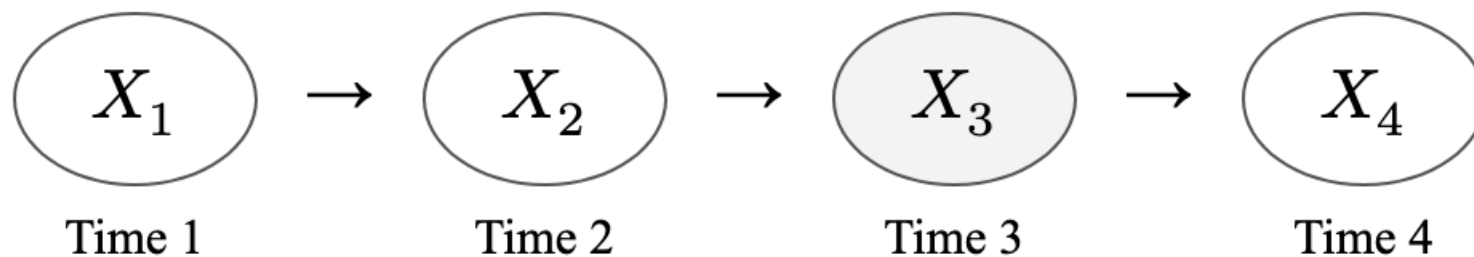
Definition: Markov Models

X_t : state at time t



Definition: Markov Models

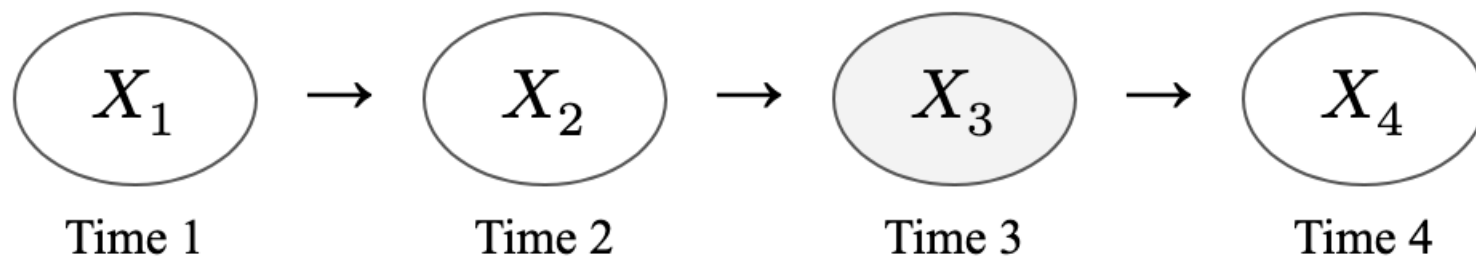
X_t : state at time t



Given the present, earlier history adds no information about the next state

Definition: Markov Models

X_t : state at time t

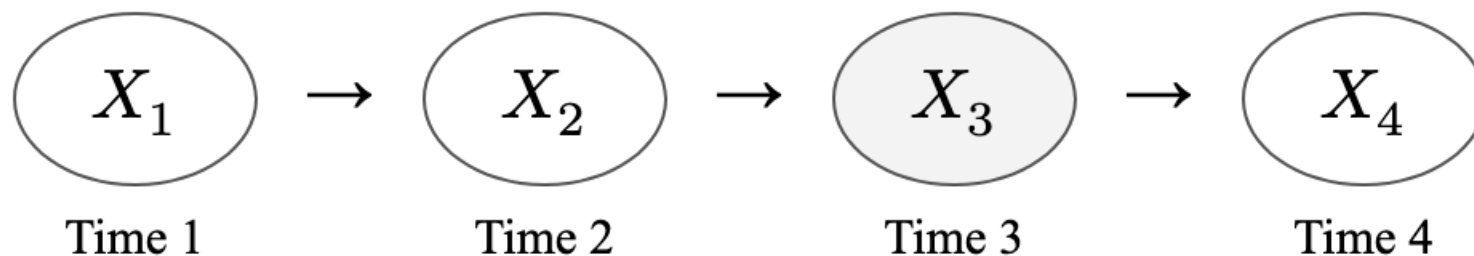


Given the present, earlier history adds no information about the next state

$$P(X_t \mid X_1, \dots, X_{t-1}) = P(X_t \mid X_{t-1})$$

Definition: Markov Models

X_t : state at time t



Given the present, earlier history adds no information about the next state

$$P(X_t \mid X_1, \dots, X_{t-1}) = P(X_t \mid X_{t-1})$$

Specify an **initial distribution** and **transition probabilities**

Joint Distribution of a Markov Model

What is the probability of a whole sequence?

Joint Distribution of a Markov Model

What is the probability of a whole sequence?

Start with the chain rule

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$$

Joint Distribution of a Markov Model

What is the probability of a whole sequence?

Start with the chain rule

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$$

For this three-state example

$$X_3 \perp X_1 \mid X_2$$

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_2)$$

Joint Distribution of a Markov Model

What is the probability of a whole sequence?

Start with the chain rule

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1, x_2)$$

For this three-state example

$$X_3 \perp X_1 \mid X_2$$

$$P(x_1, x_2, x_3) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_2)$$

Apply the same assumption at every step

$$P(x_1, \dots, x_n) = P(x_1) \prod_{t=2}^n P(x_t \mid x_{t-1})$$

Two Modelling Assumptions

1. Markov property

Given the present, earlier history adds no information about the next state

Two Modelling Assumptions

1. Markov property

Given the present, earlier history adds no information about the next state

2. Time-homogeneous transitions (additional assumption)

Use the same transition probabilities at every time step

Two Modelling Assumptions

1. Markov property

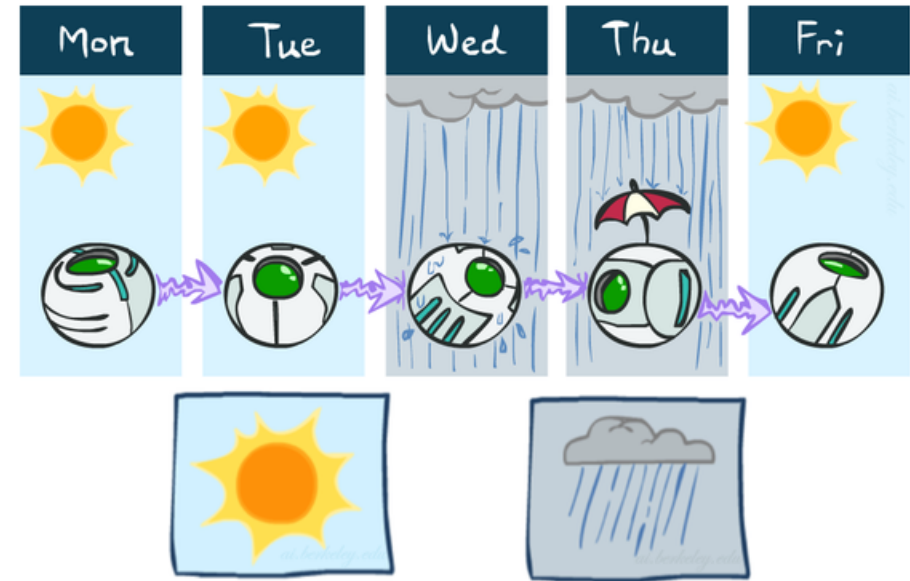
Given the present, earlier history adds no information about the next state

2. Time-homogeneous transitions (additional assumption)

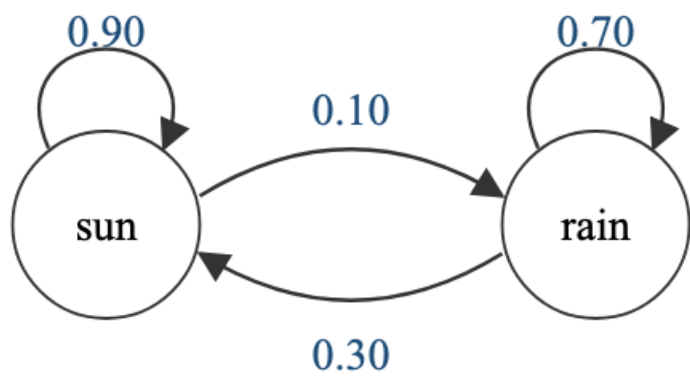
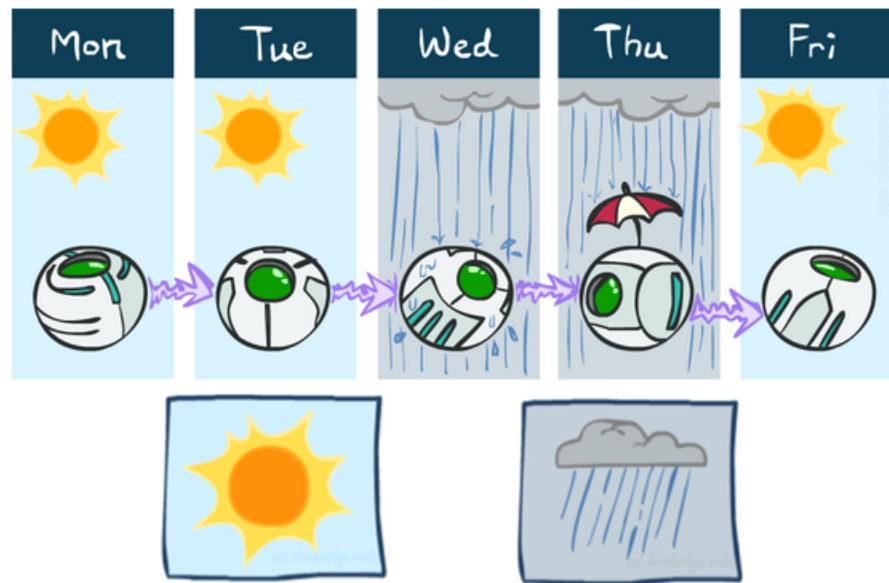
Use the same transition probabilities at every time step

Same transition model $\neq$ same state distribution

Example Markov Chain: Weather

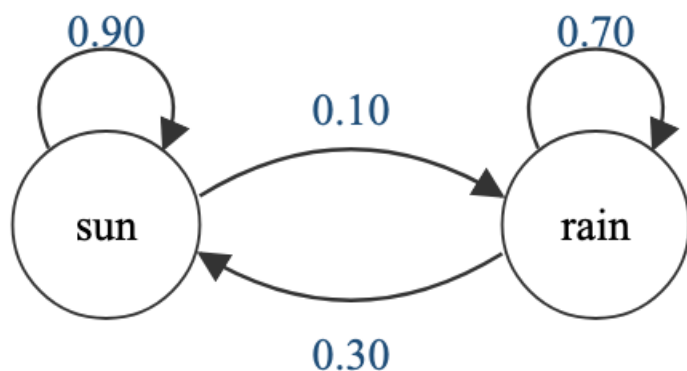
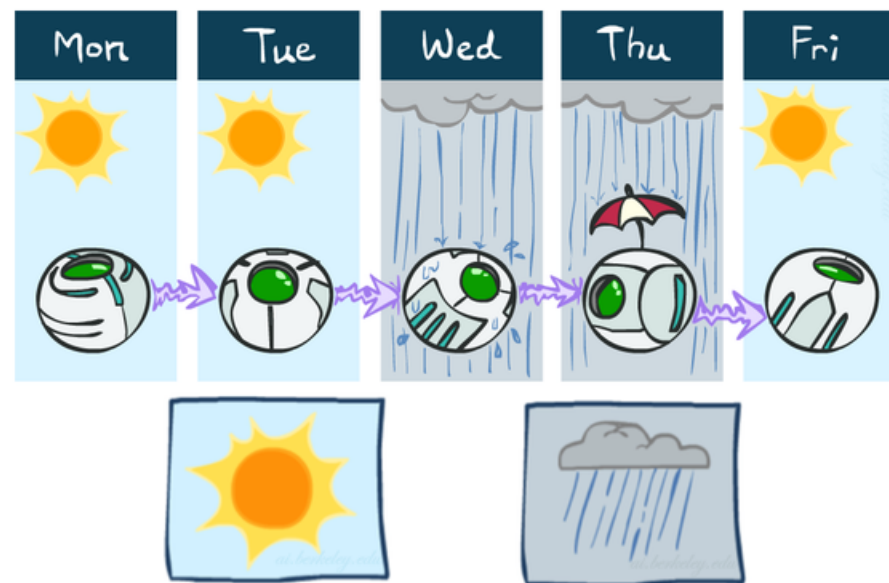


Example Markov Chain: Weather



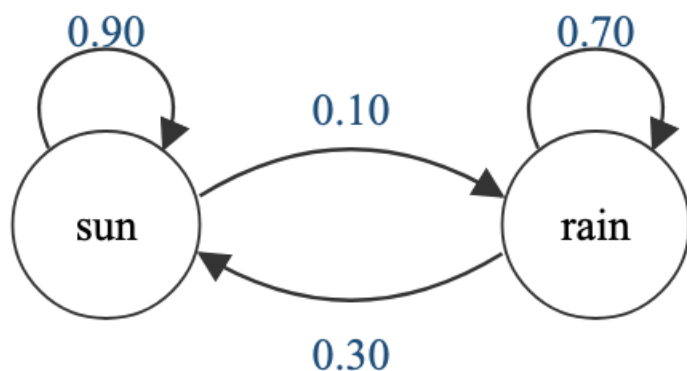
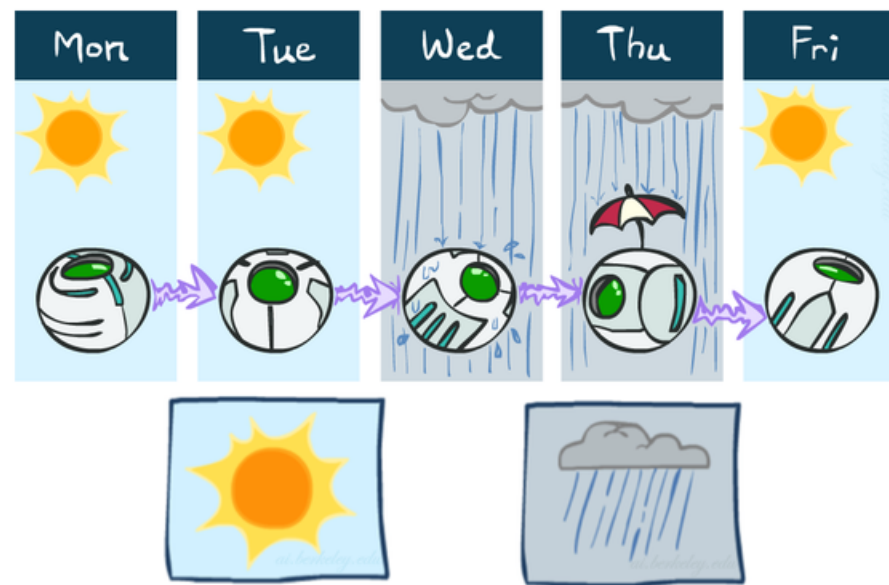
Example Markov Chain: Weather

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70



Example Markov Chain: Weather

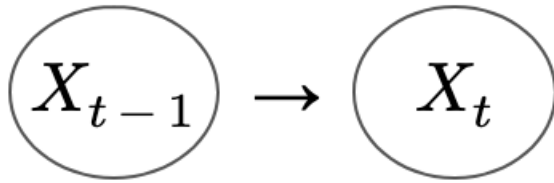
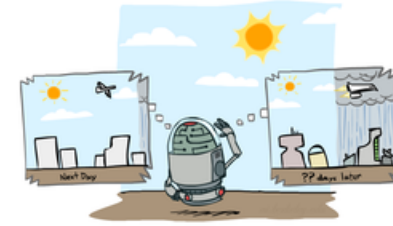
X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70



Initial distribution $P(X_1 = \text{sun}) = 1$ $P(X_1 = \text{rain}) = 0$

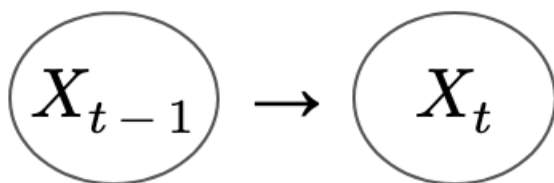
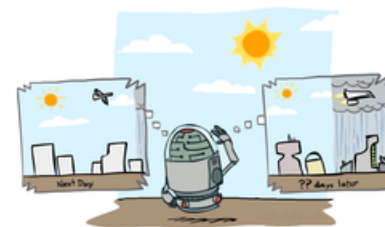
Algorithm: Mini-Forward

Predict the next state distribution



Algorithm: Mini-Forward

Predict the next state distribution

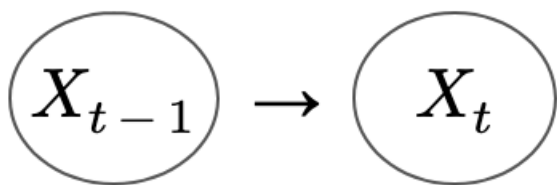
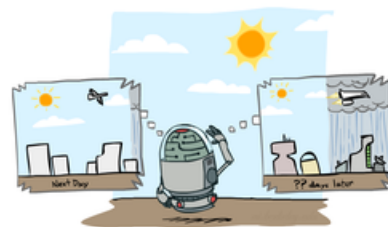


Marginalize over the previous state

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

Algorithm: Mini-Forward

Predict the next state distribution



Marginalize over the previous state

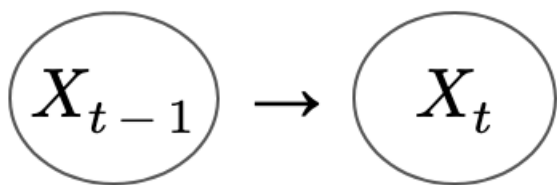
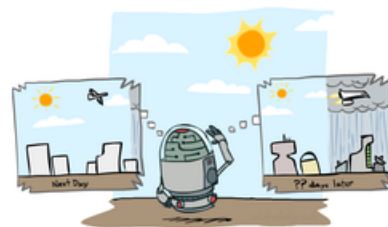
$$P(x_t) = \sum_{x_{t-1}} P(x_{t-1}, x_t)$$

Use the product rule

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

Algorithm: Mini-Forward

Predict the next state distribution



Marginalize over the previous state

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

Use the product rule

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

Repeat the update to predict further ahead

Weather: One Step Forward

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

Start: sun with probability 1

What is the distribution one day later?

$$P(X_t = x_t) = \sum_{x_{t-1}} P(X_t = x_t | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$

Weather: One Step Forward

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

Start: sun with probability 1

What is the distribution one day later?

$$P(X_t = x_t) = \sum_{x_{t-1}} P(X_t = x_t | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$

$$P(X_2 = \text{sun}): \underbrace{0.90 \times 1}_{\text{from sun}} + \underbrace{0.30 \times 0}_{\text{from rain}} = 0.90$$

Weather: One Step Forward

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

Start: sun with probability 1

What is the distribution one day later?

$$P(X_t = x_t) = \sum_{x_{t-1}} P(X_t = x_t | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$

$$P(X_2 = \text{sun}): \underbrace{0.90 \times 1}_{\text{from sun}} + \underbrace{0.30 \times 0}_{\text{from rain}} = 0.90$$

$$P(X_2 = \text{rain}): \underbrace{0.10 \times 1}_{\text{from sun}} + \underbrace{0.70 \times 0}_{\text{from rain}} = 0.10$$

Weather: Two Steps Forward

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

After one update:

$$P(X_2 = \text{sun}) = 0.90$$

$$P(X_2 = \text{rain}) = 0.10$$

Apply the same update once more

$$P(X_t = x_t) = \sum_{x_{t-1}} P(X_t = x_t | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$

Weather: Two Steps Forward

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

After one update:

$$P(X_2 = \text{sun}) = 0.90$$

$$P(X_2 = \text{rain}) = 0.10$$

Apply the same update once more

$$P(X_t = x_t) = \sum_{x_{t-1}} P(X_t = x_t | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$

$$P(X_3 = \text{sun}) = 0.90 \times 0.90 + 0.30 \times 0.10 = 0.84$$

Weather: Two Steps Forward

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

After one update:

$$P(X_2 = \text{sun}) = 0.90$$

$$P(X_2 = \text{rain}) = 0.10$$

Apply the same update once more

$$P(X_t = x_t) = \sum_{x_{t-1}} P(X_t = x_t | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$

$$P(X_3 = \text{sun}) = 0.90 \times 0.90 + 0.30 \times 0.10 = 0.84$$

$$P(X_3 = \text{rain}) = 0.10 \times 0.90 + 0.70 \times 0.10 = 0.16$$

Weather over Time

Compare Two Starting Distributions

Time	sunny start: sun	sunny start: rain	rainy start: sun	rainy start: rain
1	1.00	0.00	0.00	1.00
2	0.90	0.10	0.30	0.70
3	0.84	0.16	0.48	0.52
4	0.80	0.20	0.59	0.41
Limit	0.75	0.25	0.75	0.25

Both starts approach the same stationary distribution for this model

Definition: Stationary Distributions

A stationary distribution is unchanged by one transition

p_* : stationary distribution



Definition: Stationary Distributions

A stationary distribution is unchanged by one transition

p_* : stationary distribution

$$p_*(x) = \sum_{x'} P(x \mid x') p_*(x')$$



Definition: Stationary Distributions

A stationary distribution is unchanged by one transition

p_* : stationary distribution

$$p_*(x) = \sum_{x'} P(x \mid x') p_*(x')$$

The distribution stays the same; individual states can still change



Definition: Stationary Distributions

A stationary distribution is unchanged by one transition

p_* : stationary distribution

$$p_*(x) = \sum_{x'} P(x \mid x') p_*(x')$$

The distribution stays the same; individual states can still change

In this weather model, different starting distributions approach it



Stationary Weather Distribution

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

Which distribution is unchanged by one update?

$$p_*(\text{sun}) + p_*(\text{rain}) = 1$$

Stationary Weather Distribution

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

Which distribution is unchanged by one update?

$$p_*(\text{sun}) + p_*(\text{rain}) = 1$$

$$p_*(\text{sun}) = 0.90p_*(\text{sun}) + 0.30p_*(\text{rain})$$

Stationary Weather Distribution

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

Which distribution is unchanged by one update?

$$p_*(\text{sun}) + p_*(\text{rain}) = 1$$

$$p_*(\text{sun}) = 0.90p_*(\text{sun}) + 0.30p_*(\text{rain})$$

$$p_*(\text{sun}) = 3p_*(\text{rain})$$

Stationary Weather Distribution

X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.90
sun	rain	0.10
rain	sun	0.30
rain	rain	0.70

Which distribution is unchanged by one update?

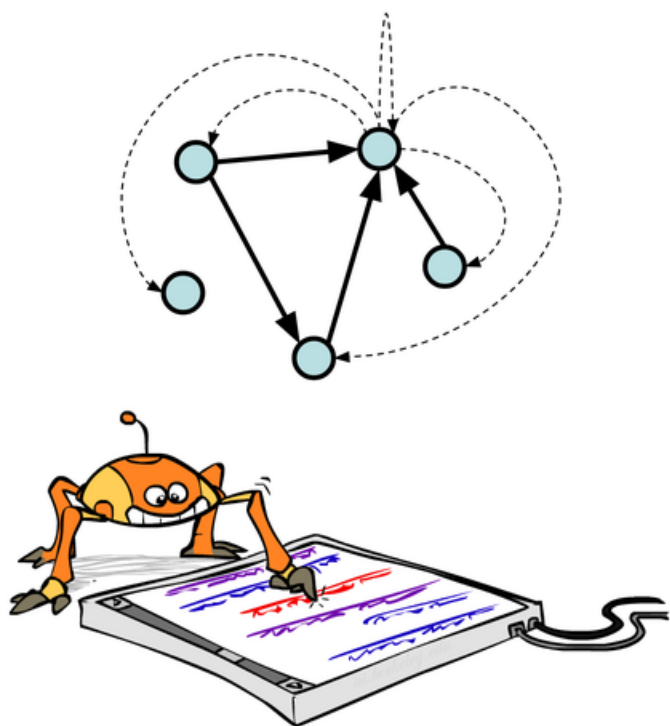
$$p_*(\text{sun}) + p_*(\text{rain}) = 1$$

$$p_*(\text{sun}) = 0.90p_*(\text{sun}) + 0.30p_*(\text{rain})$$

$$p_*(\text{sun}) = 3p_*(\text{rain})$$

$$p_*(\text{sun}) = 0.75 \quad p_*(\text{rain}) = 0.25$$

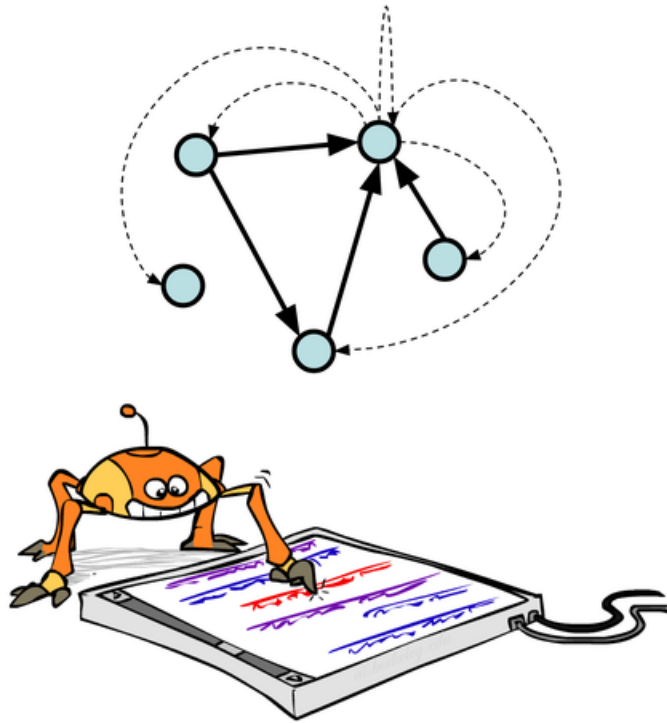
Application: Google PageRank



From weather prediction to web page ranking

State $\rightarrow$ Web page

Application: Google PageRank

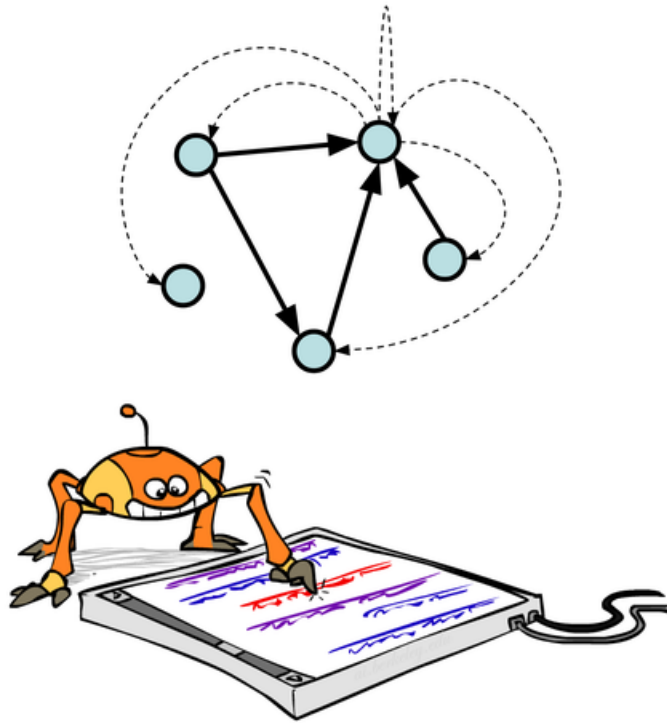


From weather prediction to web page ranking

State $\rightarrow$ Web page

Transition $\rightarrow$ Link or random jump

Application: Google PageRank



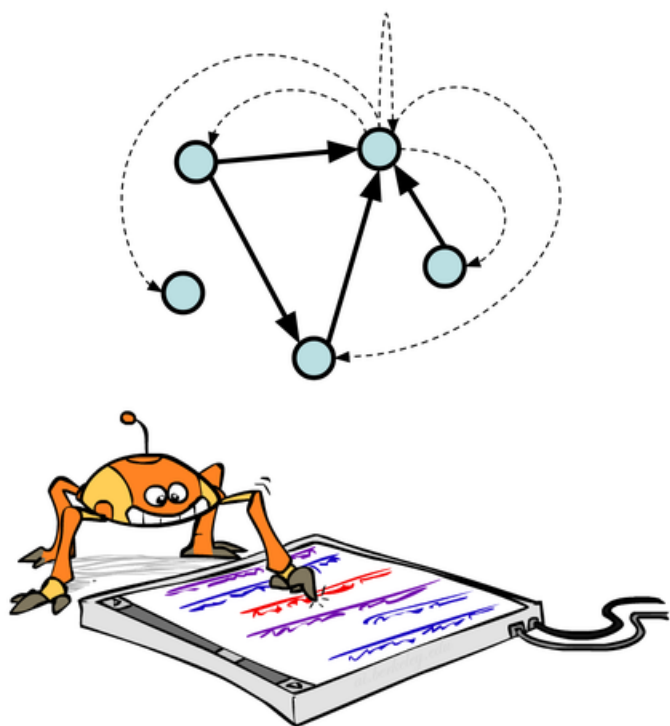
From weather prediction to web page ranking

State $\rightarrow$ Web page

Transition $\rightarrow$ Link or random jump

Stationary probability $\rightarrow$ PageRank

Application: Google PageRank



From weather prediction to web page ranking

State $\rightarrow$ Web page

Transition $\rightarrow$ Link or random jump

Stationary probability $\rightarrow$ PageRank

More long-run visits $\rightarrow$ Higher PageRank