CSC242: Introduction to Artificial Intelligence

Lecture 3.5

Please put away all electronic devices

Announcements

- Unit 3 Exam next class
- Project 3 due that night 1159PM
  - Don't wait to be finished

The Goal

- Query variable X
- Evidence variables $E_1, \dots, E_m$
  - Observed values: $\boldsymbol{e} = \langle e_1 , \dots, e_m \rangle$
- Non-evidence, non-query ("hidden") variables: $\boldsymbol{Y}$
- Approximate: $\mathbf{P}(X \mid \mathbf{e})$

Sampling

- Generate an assignment of values to the random variables
- That is consistent with the full joint distribution encoded in the network
  - In the sense that in the limit, the probability of any event is equal to the frequency of its occurrence

Rejection Sampling

- Generate sample from the prior distribution specified by the network
- Reject sample if inconsistent with the evidence
- Use remaining samples to estimate probability of event

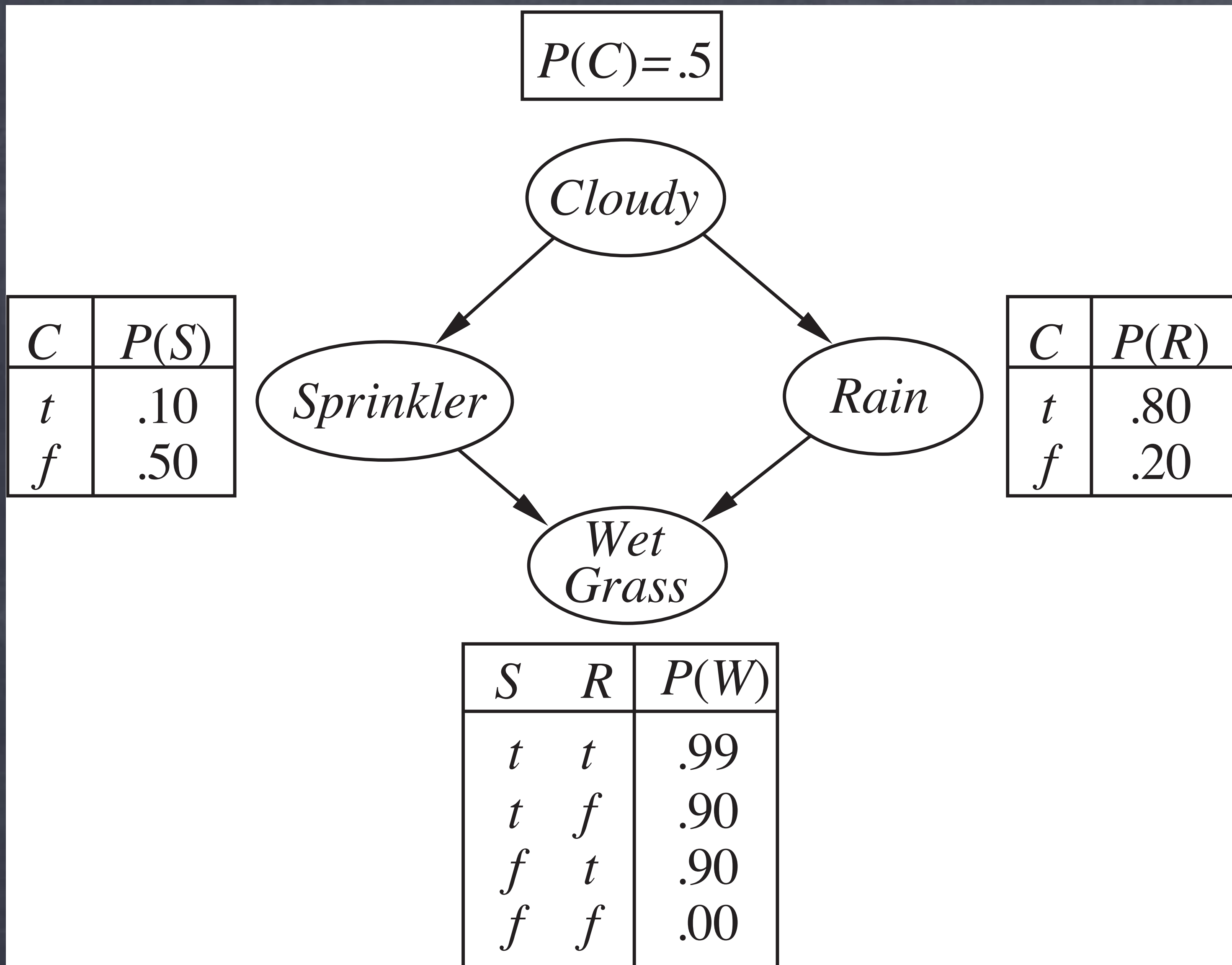
Rejection Sampling

- Generate sample from the prior distribution specified by the network
- Reject sample if inconsistent with the evidence
- Use remaining samples to estimate probability of event

Fraction of samples consistent with the evidence drops exponentially with number of evidence variables

Likelihood Weighting

- Generate only samples consistent with the evidence
  - i.e., fix values of evidence variables
- Instead of counting 1 for each non-rejected sample, weight the count by the likelihood (probability) of the sample given the evidence

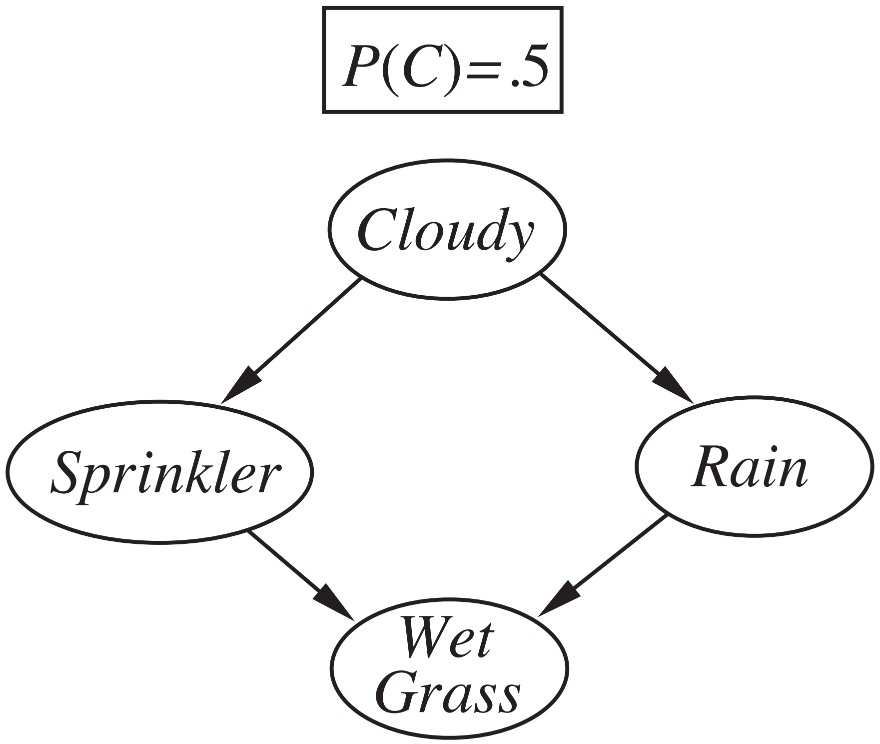


Cloudy	*true*
Sprinkler	*false*
Rain	*true*
WetGrass	*true*

$w = 0.45$

$$\mathbf{P}(Rain \mid Cloudy{=}true,\ WetGrass{=}true)$$

$N_{Rain=true}$ **+=** 0.45



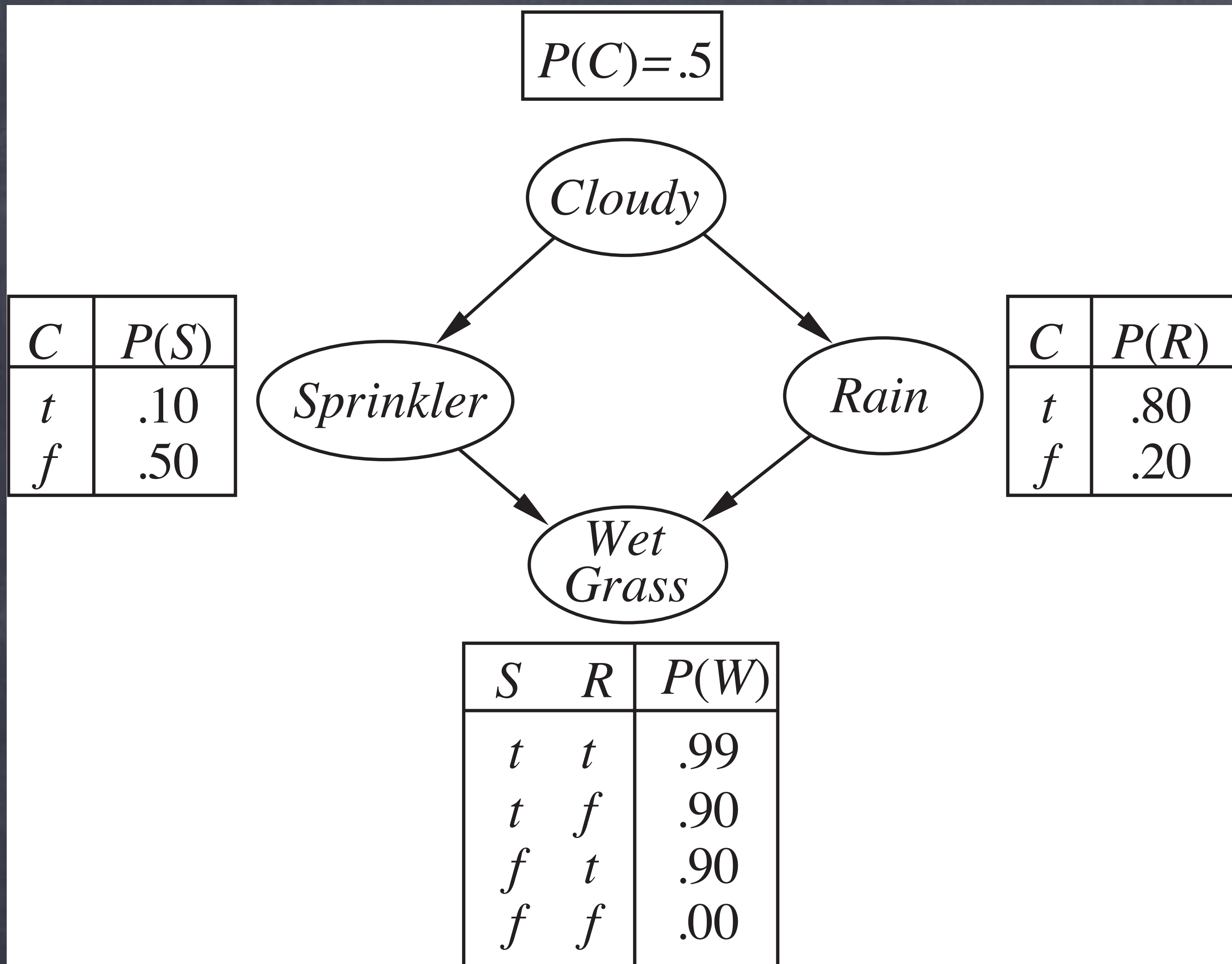
C	P(S)
t	.10
f	.50

C	P(R)
t	.80
f	.20

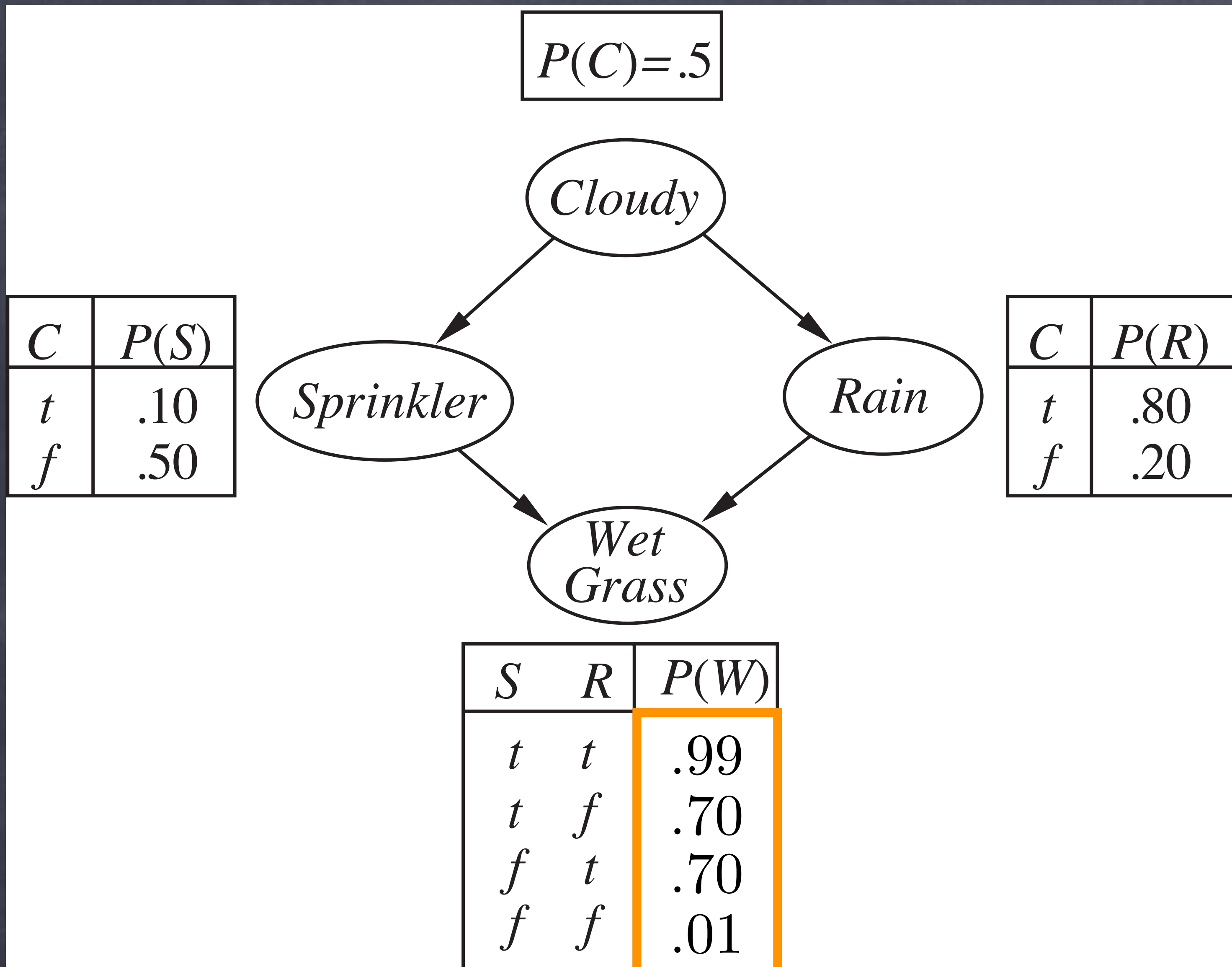
S	R	P(W)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

$Cloudy$	$true$
$Sprinkler$	$false$
$Rain$	$true$
$WetGrass$	

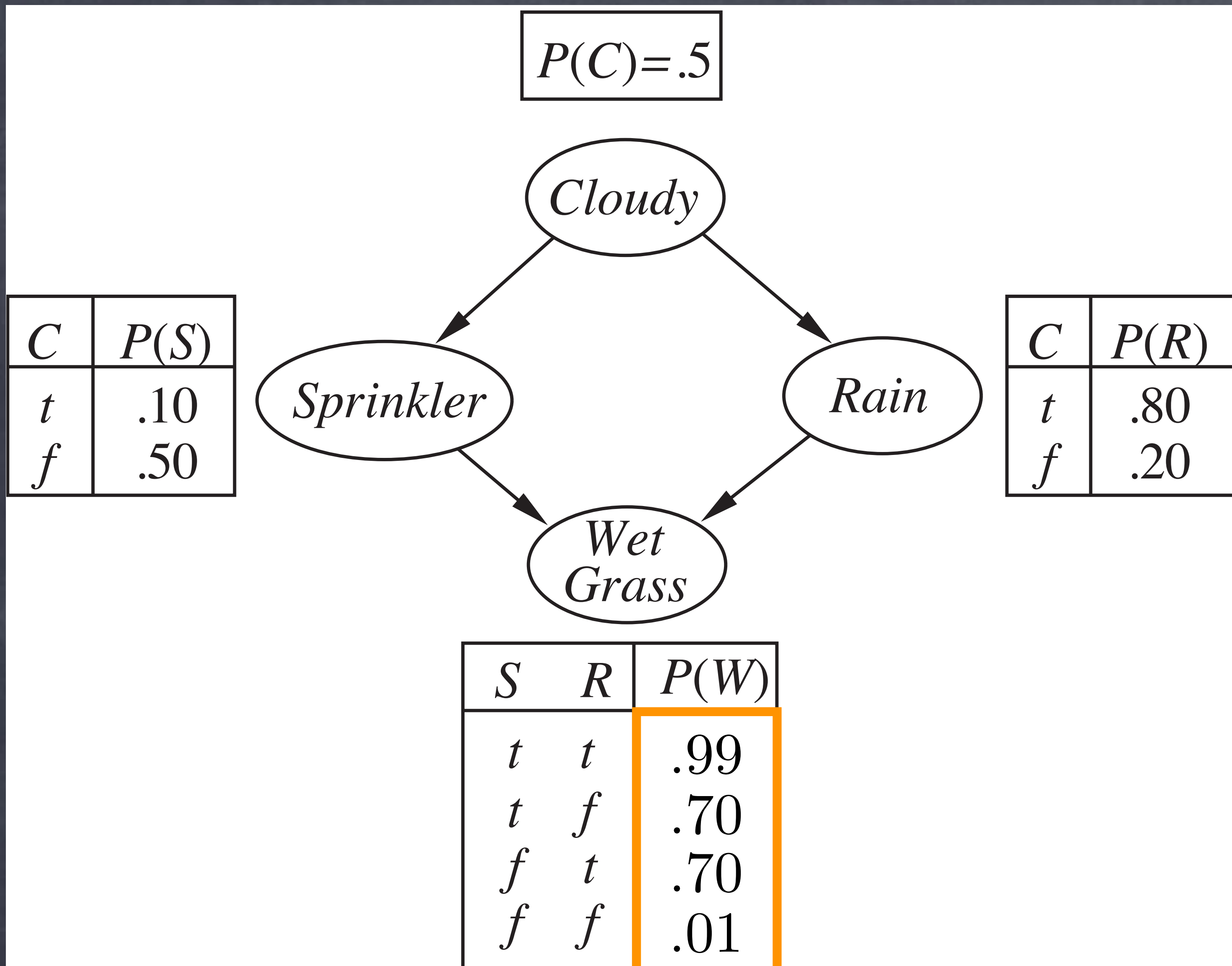
$$\mathbf{P}(Rain \mid Cloudy{=}true)$$



$$\mathbf{P}(Cloudy \mid WetGrass{=}true)$$

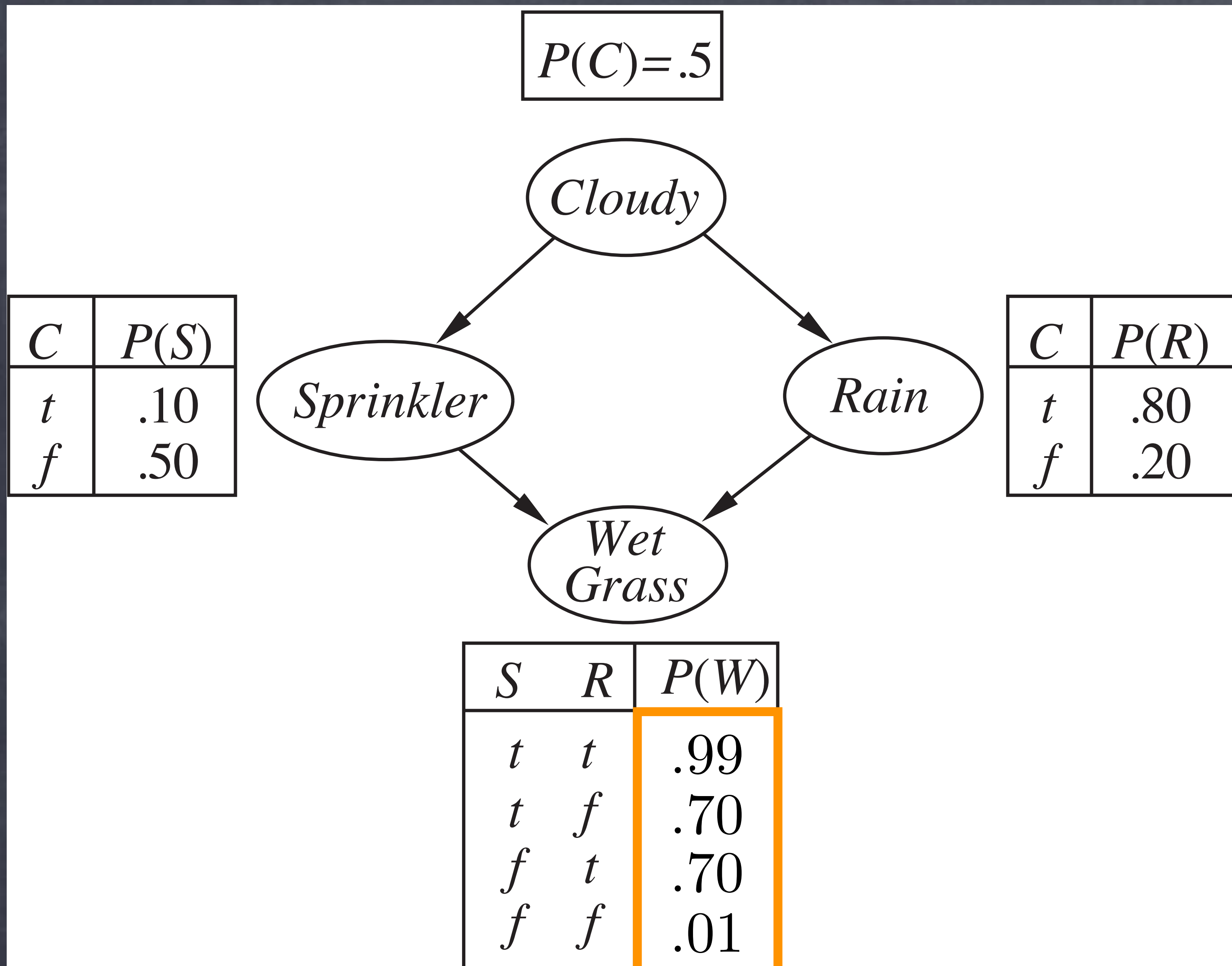


$$\mathbf{P}(Cloudy \mid WetGrass = true)$$



Cloudy	$\langle 0.5, 0.5 \rangle$	F
Sprinkler	$\langle 0.5, 0.5 \rangle$	F
Rain	$\langle 0.2, 0.8 \rangle$	F
WetGrass		

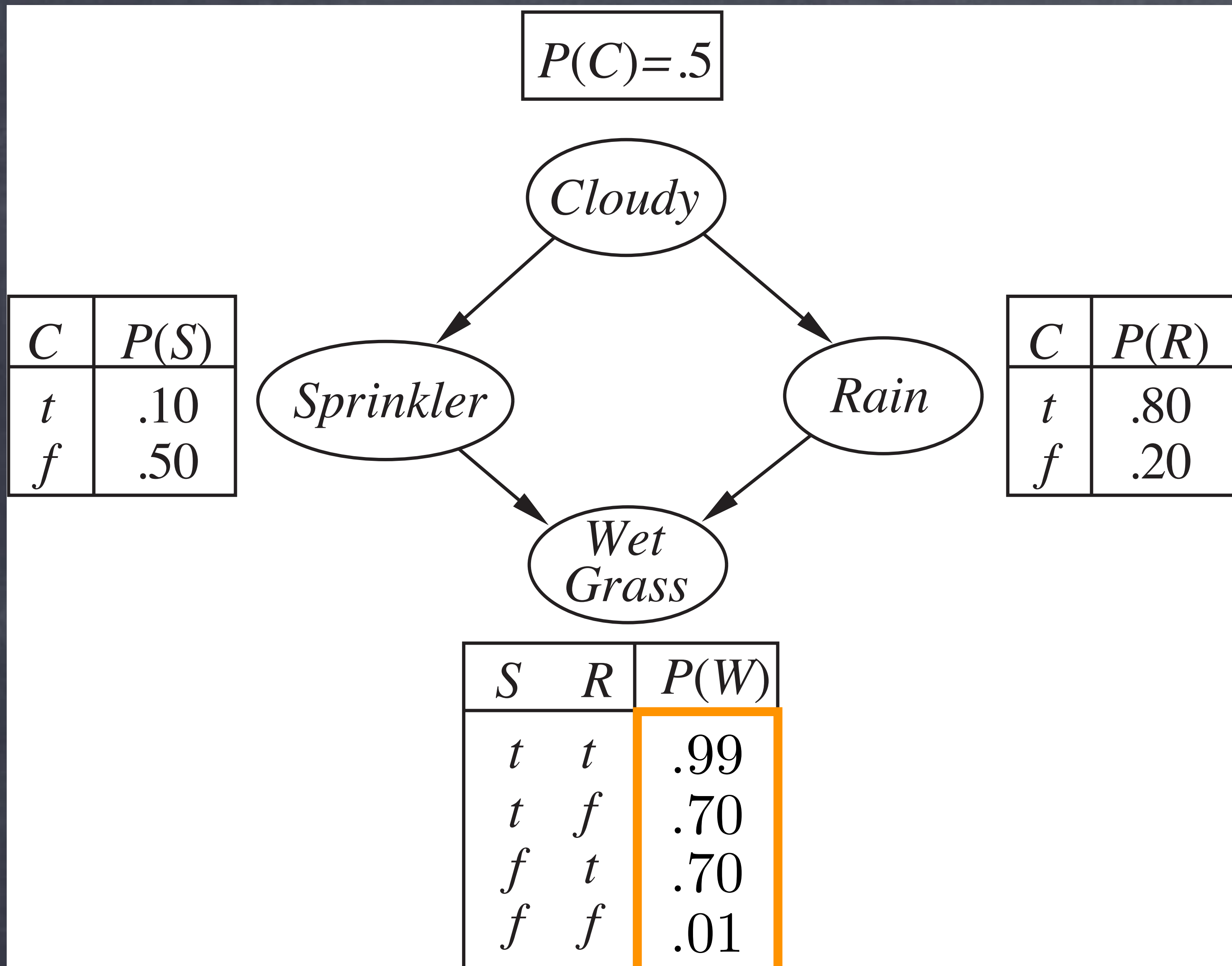
$$\mathbf{P}(Cloudy \mid WetGrass{=}true)$$



Cloudy	$\langle 0.5, 0.5 \rangle$	F
Sprinkler	$\langle 0.5, 0.5 \rangle$	F
Rain	$\langle 0.2, 0.8 \rangle$	F
WetGrass	Evidence	T

$w = 0.01$

$$\mathbf{P}(Cloudy \mid WetGrass{=}true)$$



Cloudy	$\langle 0.5, 0.5 \rangle$	F
Sprinkler	$\langle 0.5, 0.5 \rangle$	F
Rain	$\langle 0.2, 0.8 \rangle$	F
WetGrass	Evidence	T

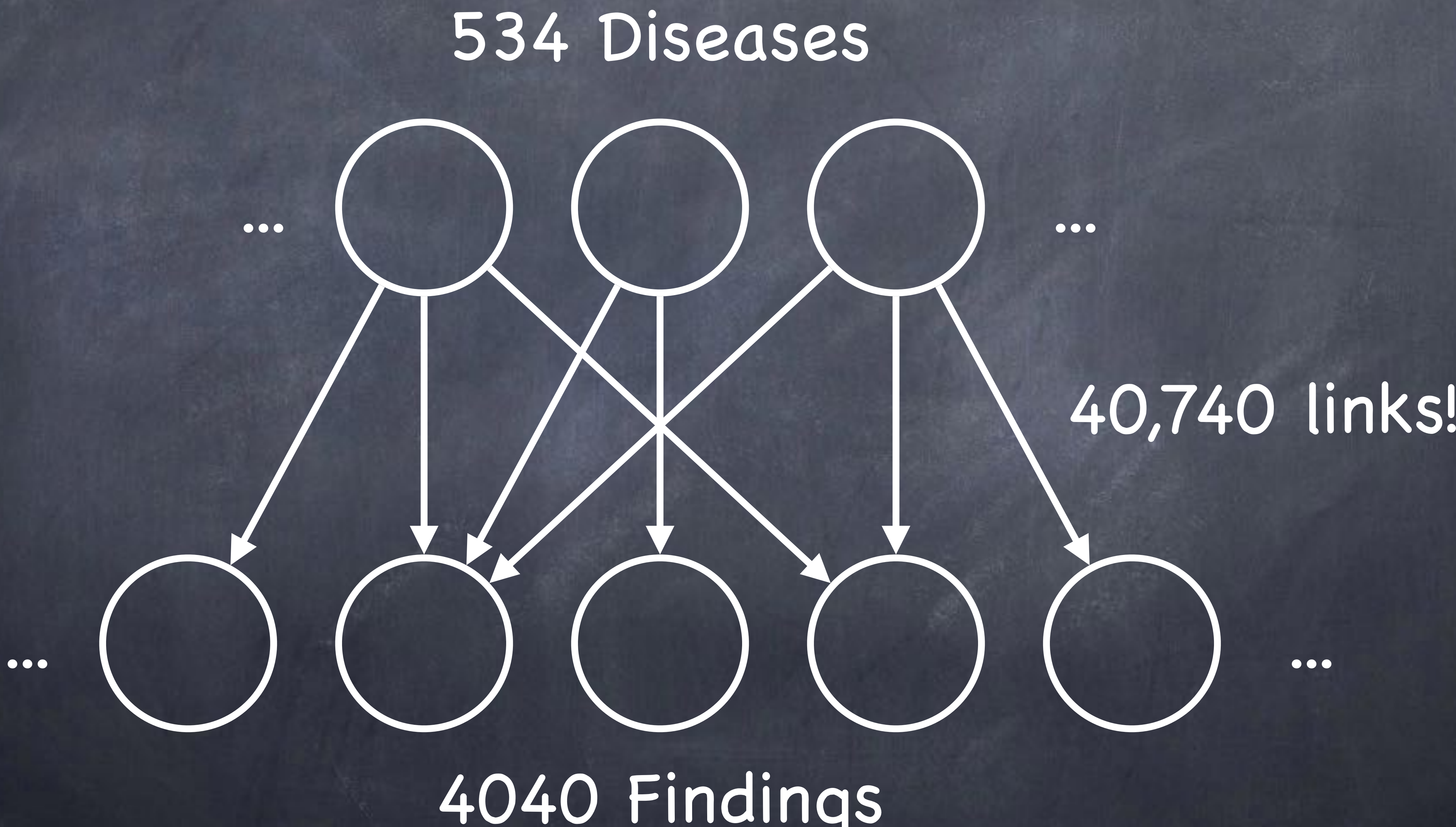
$$w = 0.01$$

Cloudy	$\langle 0.5, 0.5 \rangle$	T
Sprinkler	$\langle 0.1, 0.9 \rangle$	F
Rain	$\langle 0.8, 0.2 \rangle$	T
WetGrass	Evidence	T

$$w = 0.7$$

$$\mathbf{P}(Cloudy \mid WetGrass{=}true)$$

QRM-DT



*Citation at end of slides

Likelihood Weighting

- Pros:
  - Doesn't reject any samples
- Cons:
  - More evidence $\Rightarrow$ lower weight
  - Affected by order of evidence vars in topsort (later = worse)

Gibbs Sampling

- To approximate: $\mathbf{P}(X \mid \mathrm{e})$
- Start in a state with evidence variables set to evidence values (others arbitrary)
- On each step, sample non-evidence variables conditioned on the values of the variables in their Markov Blanket
  - Order irrelevant
- Similar to local search!

Approximate Inference in Bayesian Networks

- Sampling consistent with a distribution
- Rejection Sampling: rejects too much
- Likelihood Weighting: weights get too small
- Gibbs Sampling: MCMC algorithm
  - A form of local search
- All generate consistent estimates (equal to exact probability in the large-sample limit)

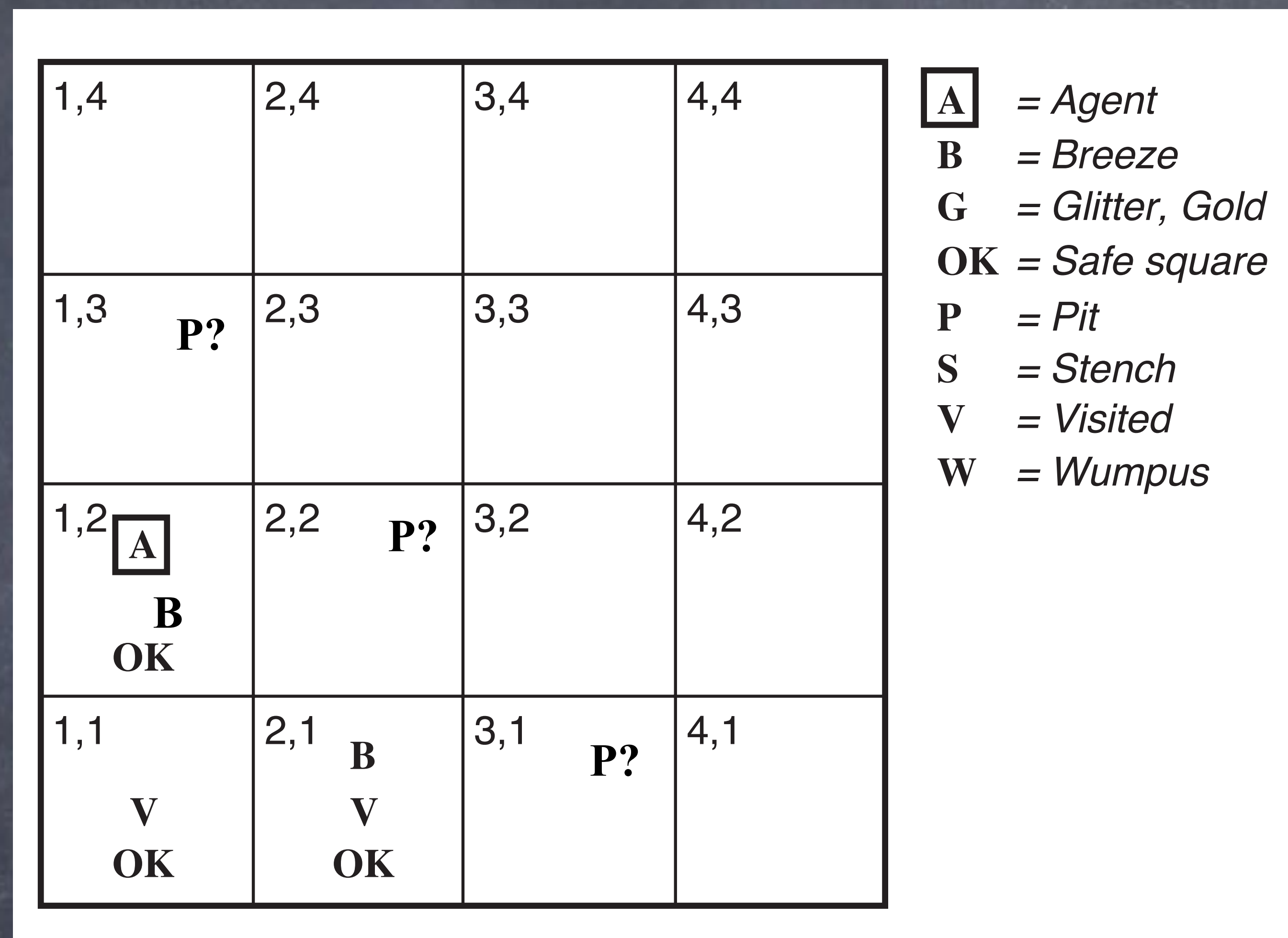
Probabilistic Reasoning (Uncertain Inference) Over Time

What An Agent Believes

- Logic: Defined in terms of possible (or impossible) worlds
- Probability: Defined in terms of more (or less) likely possible worlds

What An Agent Believes

- Logic: Defined in terms of possible (or impossible) worlds
- Probability: Defined in terms of more (or less) likely possible worlds
- STUFF HAPPENS!


1,4
2,4
3,4
4,4
1,3
P?
2,3
3,3
4,3
1,2
A
B
OK
2,2
P?
3,2
4,2
1,1
V
OK
2,1
B
V
OK
3,1
P?
4,1
A = Agent
B = Breeze
G = Glitter, Gold
OK = Safe square
P = Pit
S = Stench
V = Visited
W = Wumpus

- Query: $BloodSugar$, $InsulinLevel$
- Evidence: $MeasuredBloodSugar$, $InsulinTaken$, $FoodEaten$, ...
- Hidden: $MetabolicActivity$, ...

- Query: $Price$ (in the future)
- Evidence: $Price$ (in the past), $InflationRate$, $InterestRates$, ...
- Hidden: Almost everything...

Uncertain Inference Over Time

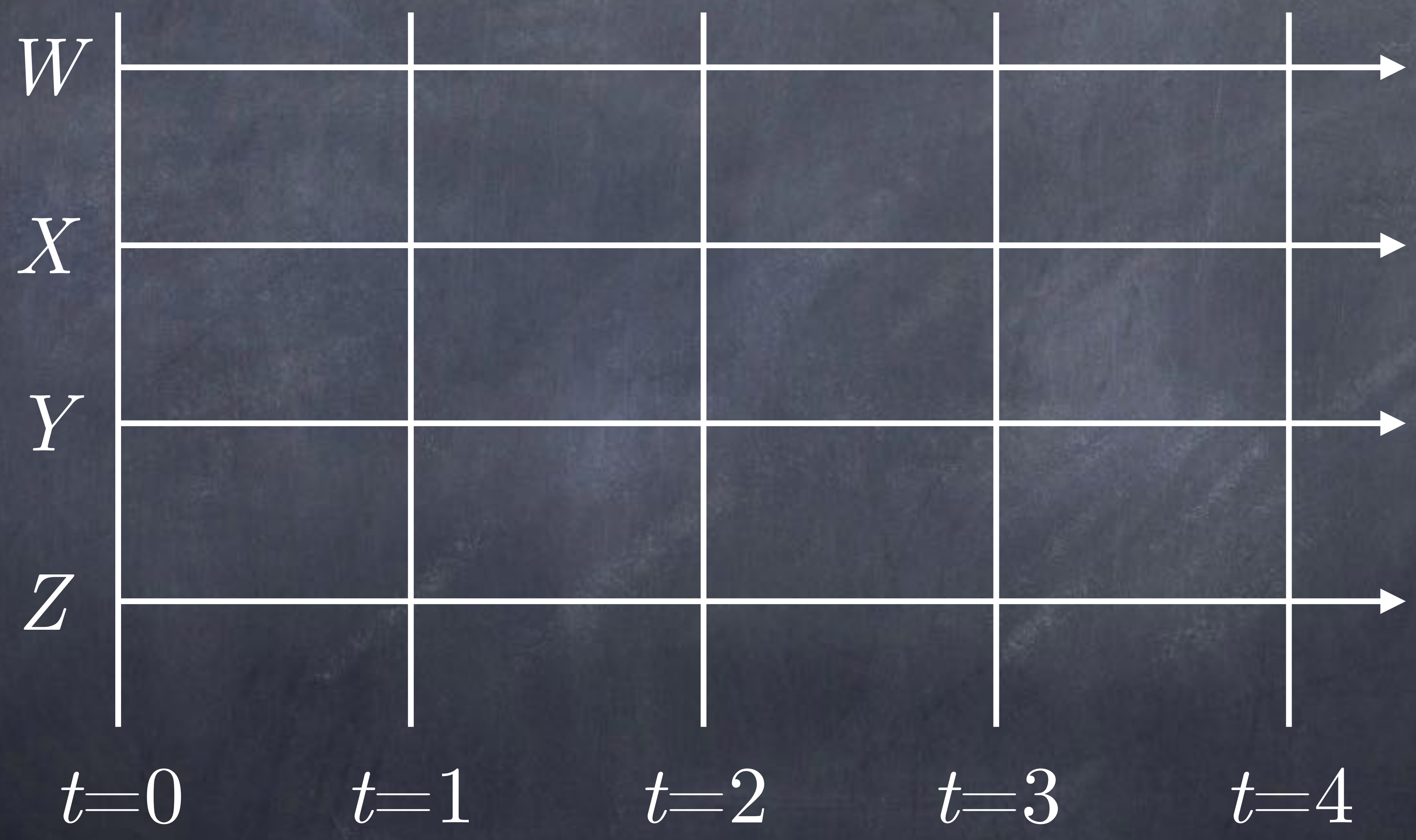
- Given history of observable values (evidence):
  - Assess current state
  - Predict future states
  - Revise earlier state estimates
  - Determine the sequence of states that caused our observations

Time

“The only reason for time is so that everything doesn't happen at once.”
– Albert Einstein

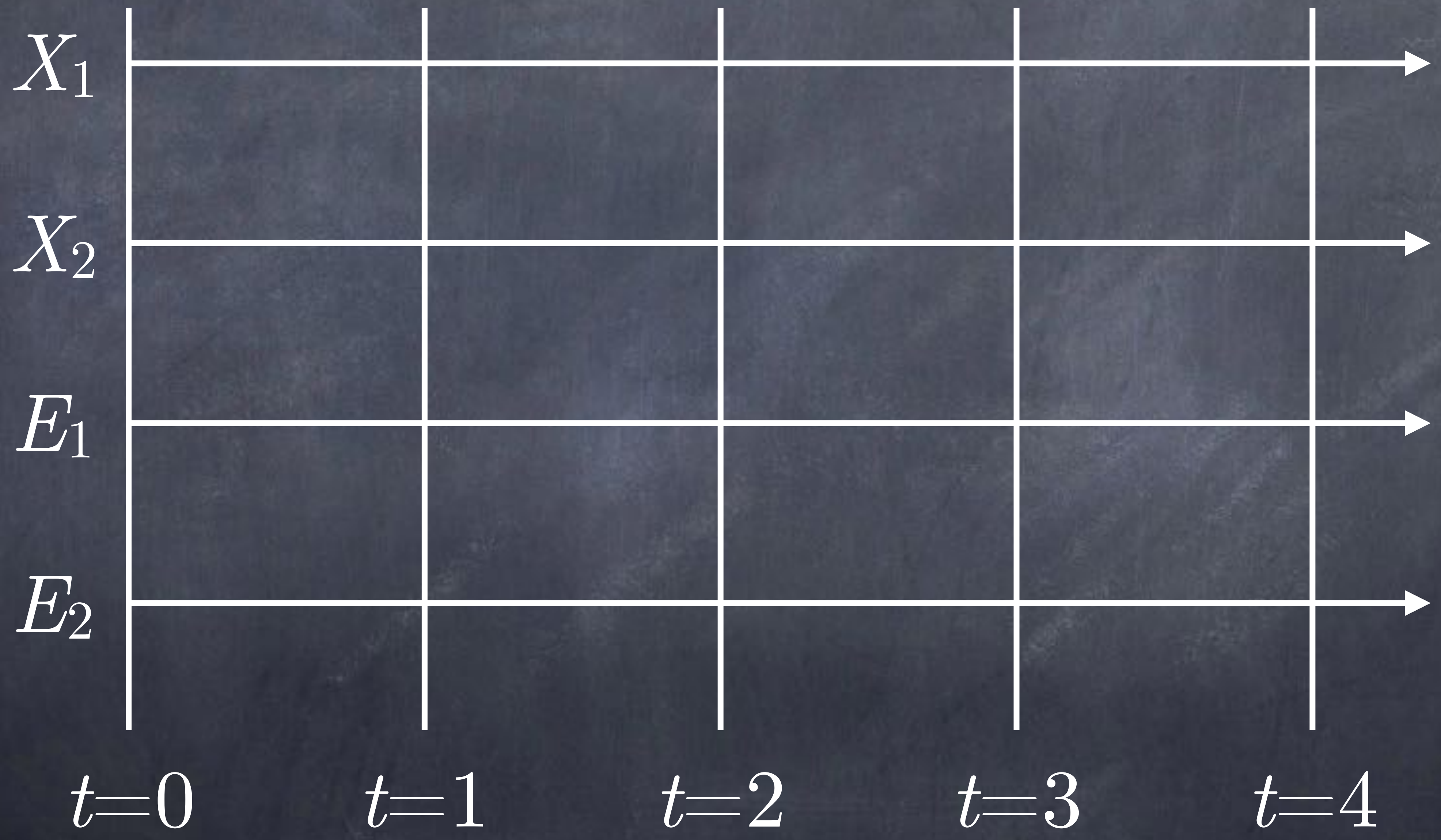
“Time is an illusion. Lunchtime doubly so.”
– Douglas Adams

Time Slices



Time Slices

State variables: $\mathbf{X}$



Evidence variables: $\mathbf{E}$

Time Slices

- Model the world as a series of <u>time slices</u>
- Unobservable state variables $\mathbf{X}$
- Observable evidence variables $\mathbf{E}$
- State at time t: $\mathbf{X}_t$
- Observations at time t: $\mathbf{E}_t = \mathbf{e}_t$


HELP IS ON THE WAY.
UNTIL THEN,
THERE'S PAUL BLART
MALL COP
IN THEATERS
JANUARY 16

R_t

R_t U_t

R_t U_t

$$\mathbf{X}_t : R_0, R_1, R_2, \ldots$$
$$\mathbf{E}_t : \quad U_1, U_2, \ldots$$

Modeling Change

- The world changes between time steps

$$\mathbf{X}_t \;, \mathbf{E}_t \quad \Longrightarrow \quad \mathbf{X}_{t+1} \;, \mathbf{E}_{t+1}$$

- Model how the world changes
- Model how the state of the world determines what we observe

Modeling Change

Given state from time 0 until time t, need to know distribution of state variables for time $t{+}1$

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{X}_{0:t}) = \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{X}_0, \mathbf{X}_1, \ldots \mathbf{X}_t)$$

Modeling Change

Given state from time 0 until time t, need to know distribution of state variables for time $t{+}1$

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{X}_{0:t}) = \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{X}_0, \mathbf{X}_1, \ldots \mathbf{X}_t)$$

$O(m^t)$ entries!


Reebok
SAMPLE
79
A


A
MARKOV
79

Andrey (Andrei) Andreyevich Markov

(1856 – 1922)

Markov Process (Assumption)

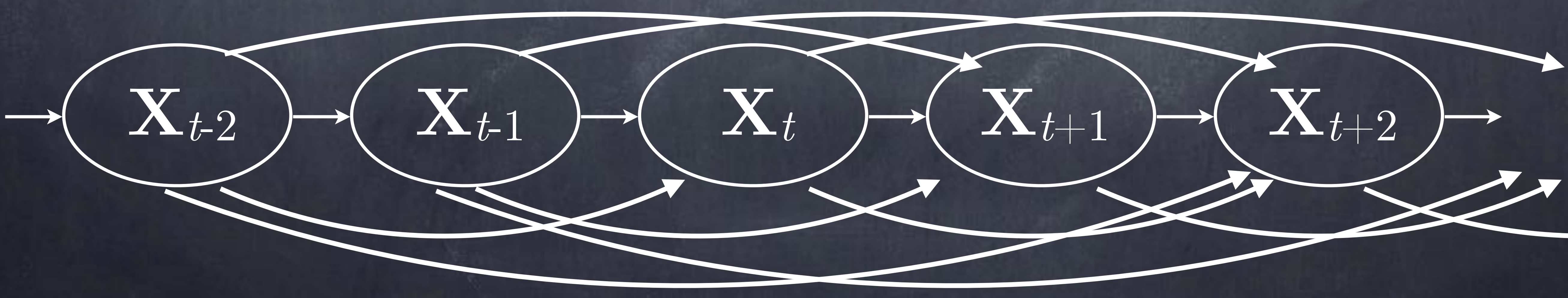
- Current state depends only on the previous state and not on earlier states
- "The future is conditionally independent of the past, given the present"

$$\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{0:t-1}) = \mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t-1})$$

Markov Process

$$\rightarrow \mathbf{X}_{t\text{-}2} \rightarrow \mathbf{X}_{t\text{-}1} \rightarrow \mathbf{X}_{t} \rightarrow \mathbf{X}_{t+1} \rightarrow \mathbf{X}_{t+2} \rightarrow$$

Markov Process



Markov Process

$$\mathbf{P}(R_t \mid R_{t-1})$$

Markov Process

$\mathbf{P}(R_{t-2} \mid R_{t-3})$ $\mathbf{P}(R_t \mid R_{t-1})$ $\mathbf{P}(R_{t+2} \mid R_{t-1})$

$$\rightarrow R_{t\text{-}2} \rightarrow R_{t\text{-}1} \rightarrow R_t \rightarrow R_{t+1} \rightarrow R_{t+2} \rightarrow$$

$\mathbf{P}(R_{t-1} \mid R_{t-2})$ $\mathbf{P}(R_{t+1} \mid R_t)$

Stationary Process

- Changes in the state are caused by a process that does not itself change
- Can use the same model to compute the changes for any pair of states $\mathbf{X}_{t\text{-}1}$, $\mathbf{X}_t$
- $\mathbf{P}(X_t \mid X_{t\text{-}1})$ is the same for any t

Transition Model

$$\mathbf{P}(R_t \mid R_{t\text{-}1})$$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$$\mathbf{P}(R_t \mid R_{t\text{-}1})$$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$$\rightarrow R_{t\text{-}1} \rightarrow R_t \rightarrow R_{t+1} \rightarrow$$

Is this realistic?

- Representation of state: $\mathbf{X}_t$, $\mathbf{E}_t$
- Transition model: $\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t\text{-}1})$
  - Markov assumption, stationary process

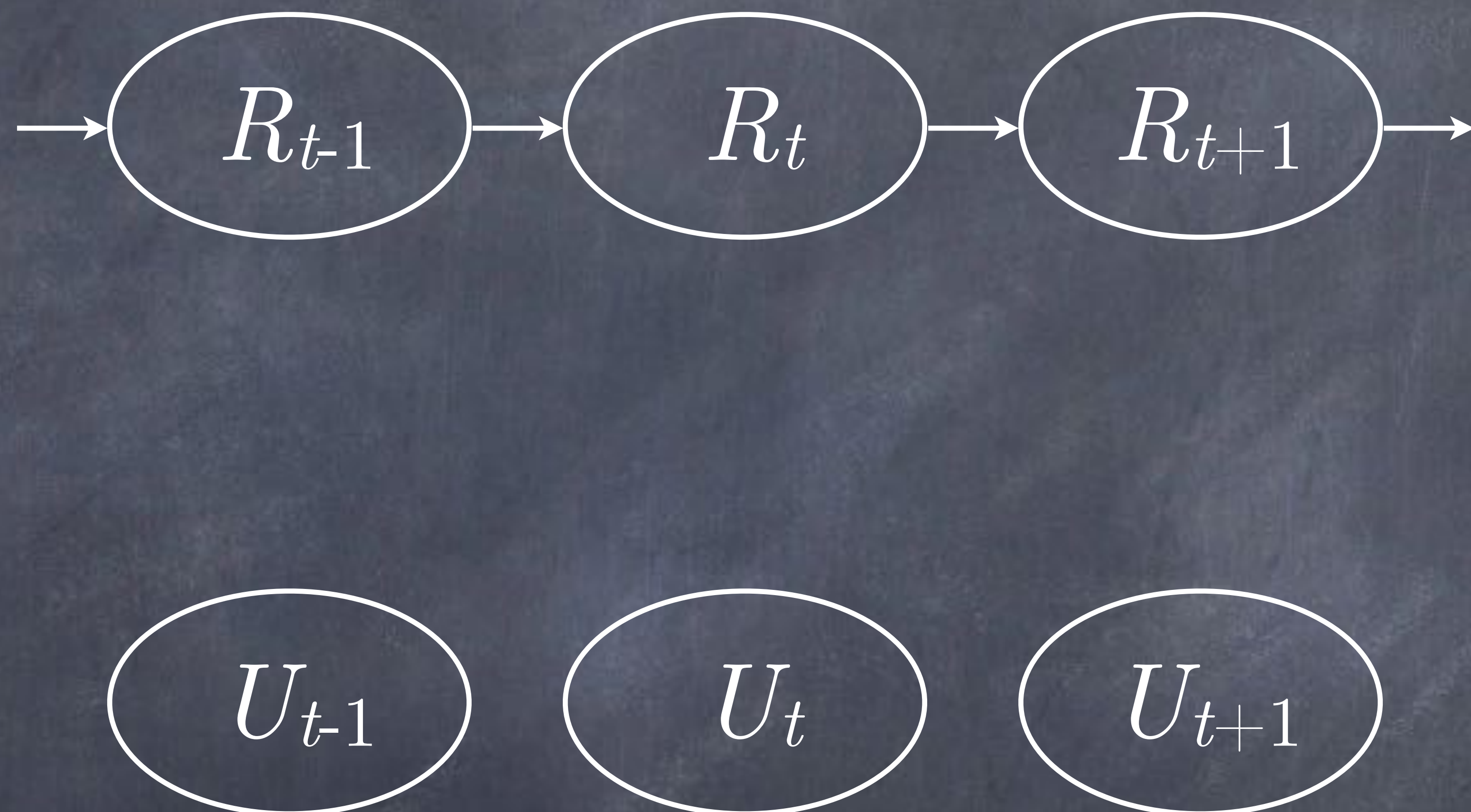
  

R_t U_t

$$\mathbf{X}_t : R_0, R_1, R_2, \ldots$$
$$\mathbf{E}_t : \qquad U_1, U_2, \ldots$$

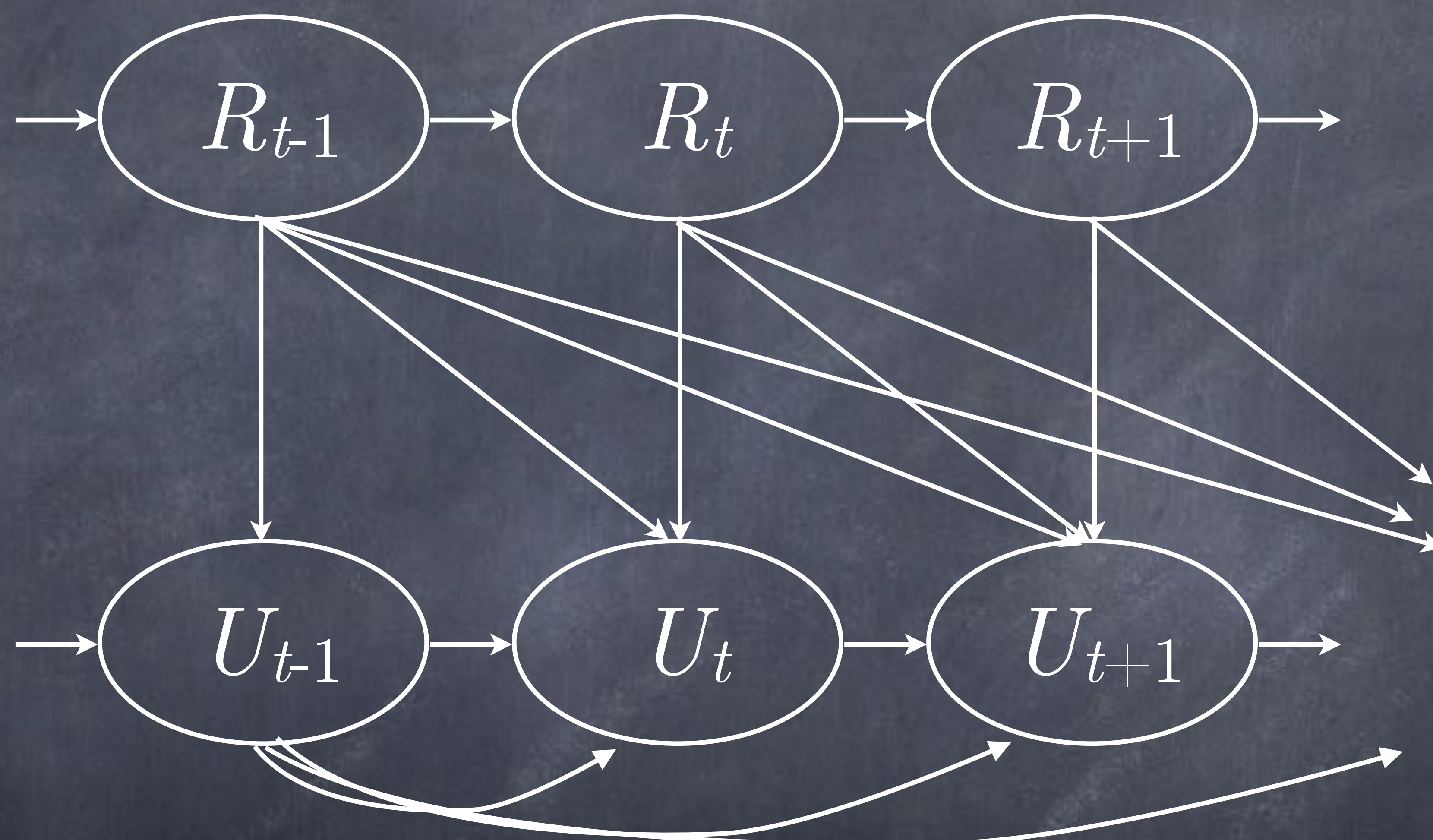
$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3



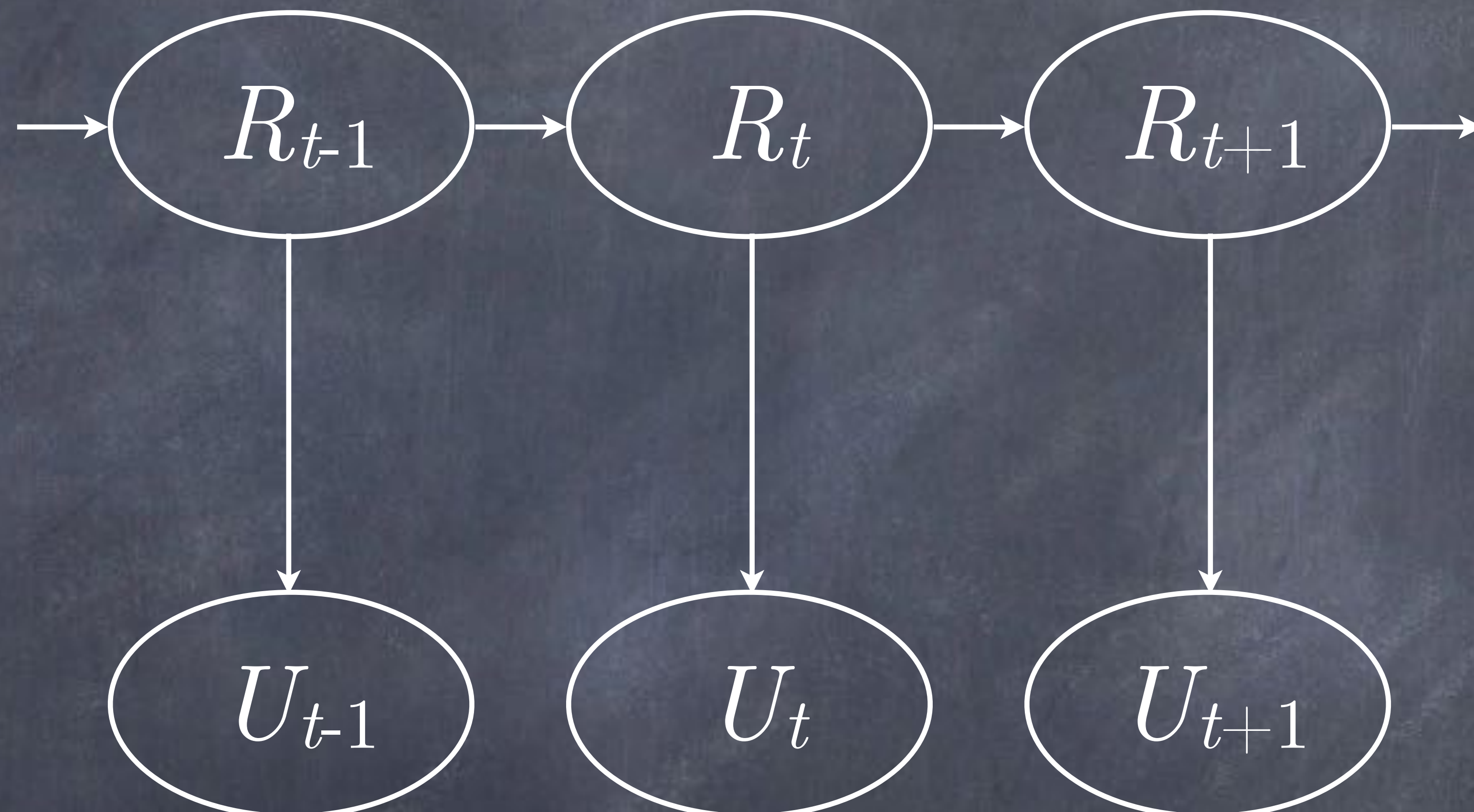
$$\mathbf{P}(R_t \mid R_{t\text{-}1})$$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3



$$\mathbf{P}(R_t \mid R_{t\text{-}1})$$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3



Sensor Markov Assumption

- Observed values of evidence variables depend only on the current state
- Evidence is conditionally independent of the past, given the present

$$\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}) = \mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$$

Sensor Markov Assumption

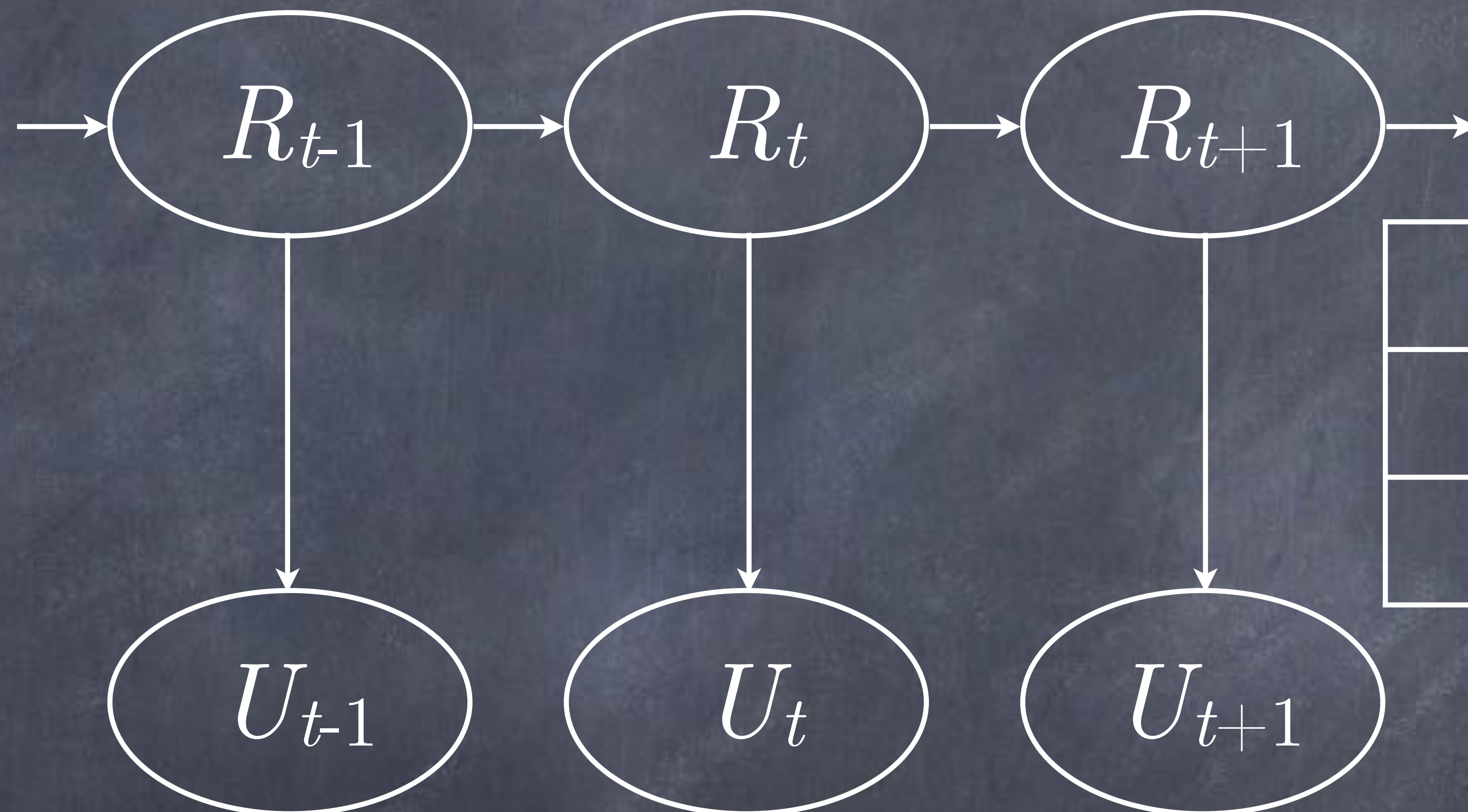
- Observed values of evidence variables depend only on the current state
- Evidence is conditionally independent of the past, given the present

$$\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_{0:t}, \mathbf{E}_{0:t-1}) = \mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$$

Sensor Model

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

- Representation of state: $\mathbf{X}_t$, $\mathbf{E}_t$
- Transition model: $\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t\text{-}1})$
  - Markov assumption, stationary process
- Sensor model: $\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$
  - Sensor Markov assumption, stationary process

- Representation of state: $\mathbf{X}_t$, $\mathbf{E}_t$
- Transition model: $\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t\text{-}1})$
  - Markov assumption, stationary process
- Sensor model: $\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$
  - Sensor Markov assumption, stationary process
- Prior distribution at time 0: $\mathbf{P}(\mathbf{X}_0)$

Temporal Model

- Representation of state: $\mathbf{X}_t$, $\mathbf{E}_t$
- Transition model: $\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t\text{-}1})$
  - Markov assumption, stationary process
- Sensor model: $\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$
  - Sensor Markov assumption, stationary process
- Prior distribution at time 0: $\mathbf{P}(\mathbf{X}_0)$

Temporal Model

$$\mathbf{P}(\mathbf{X}_{0:t}, \mathbf{E}_{1:t}) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i \mid \mathbf{X}_{i-1}) \mathbf{P}(\mathbf{E}_i \mid \mathbf{X}_i)$$

Initial State Model

Transition Model

Sensor Model

Inference

- Filtering (State Estimation)
- Prediction
- Smoothing
- Most Likely Explanation

Filtering (State Estimation)

- Compute current belief state given all evidence to date

Filtering (State Estimation)

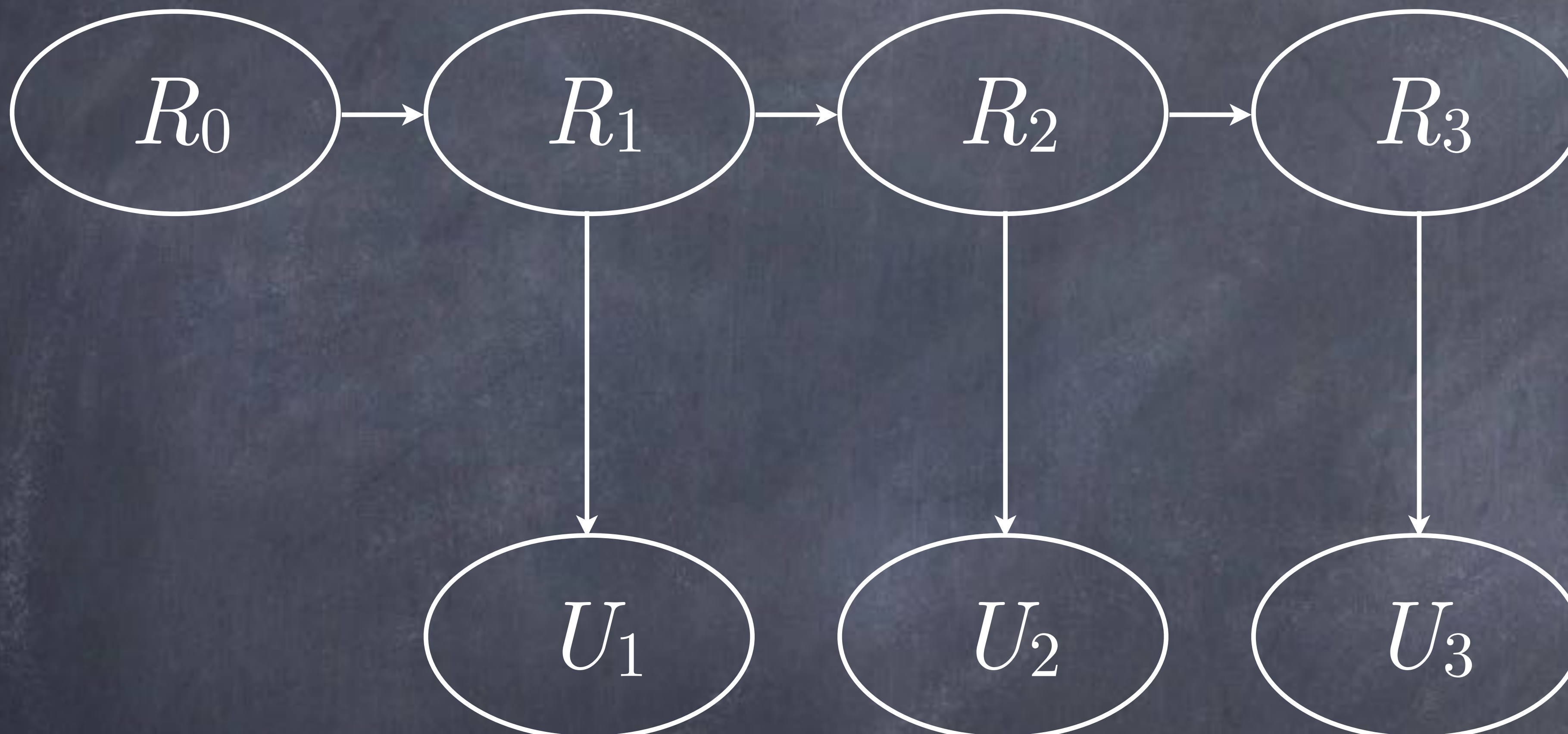
- Compute current belief state given all evidence to date
- Method: Build network incrementally, do inference on it

$\mathbf{P}(R_0)$

$P(R_t)$
0.5

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

Filtering (State Estimation)

- Compute current belief state given all evidence to date
- ~~Method: Build network incrementally, do inference on it~~
- Have to maintain current state estimate and update it rather than recomputing over history of observations every time

Filtering (State Estimation)

$$\begin{aligned}\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \\ &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \\ &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t})\end{aligned}$$

Bayes' Rule

Sensor Markov assumption

Filtering (State Estimation)

$$
\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t})
\end{aligned}
$$

Prediction of next state

Update with evidence
(using sensor model)

One-Step Prediction

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) = \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

$$= \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

Condition on $\mathbf{X}_t$

Markov assumption

Filtering (State Estimation)

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

Prediction of next state

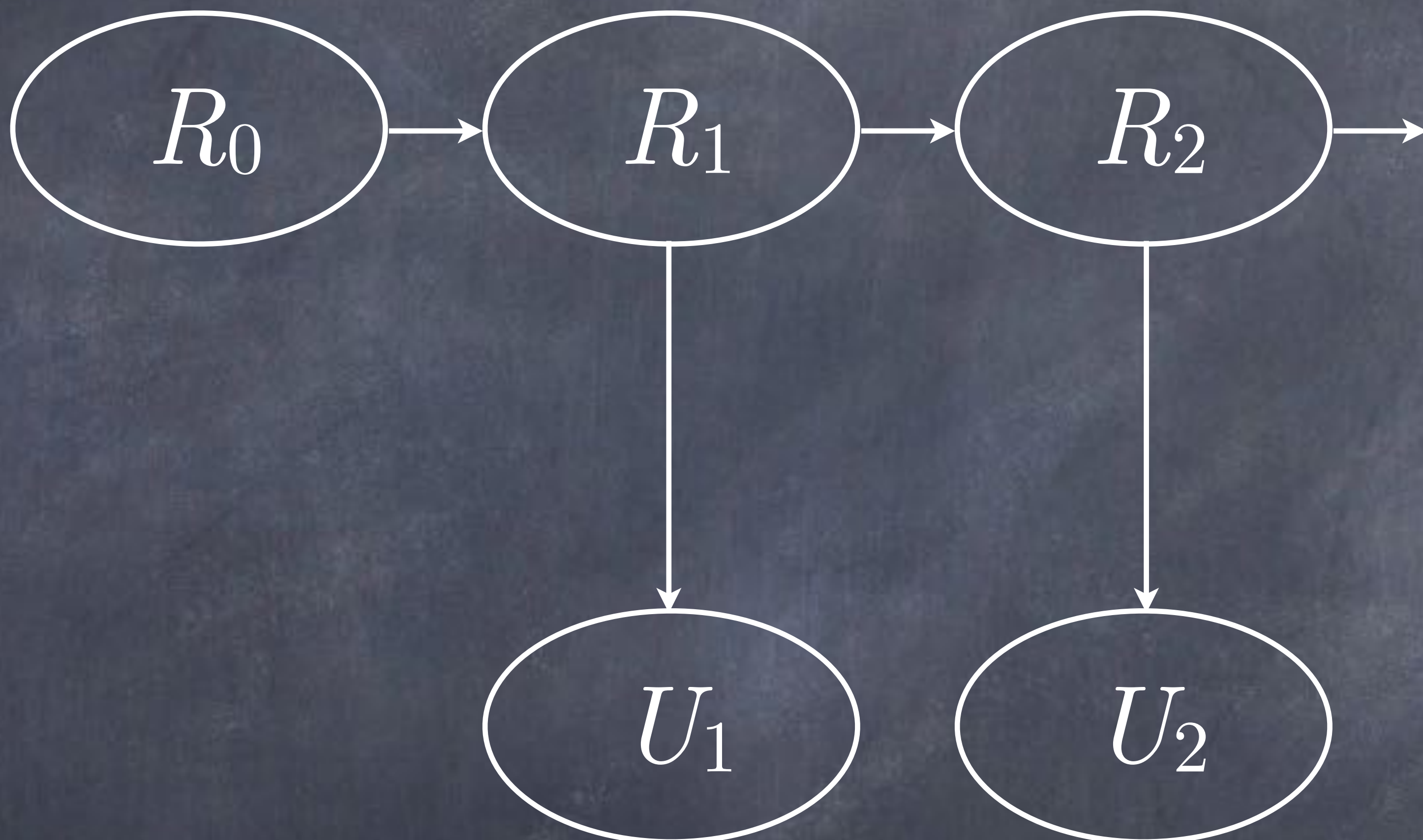
Update with evidence
(using sensor model)

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5

R_0

$\mathbf{P}(R_0)=\langle 0.5,0.5\rangle$

$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5

$$R_0 \longrightarrow R_1$$

$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$$\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$$

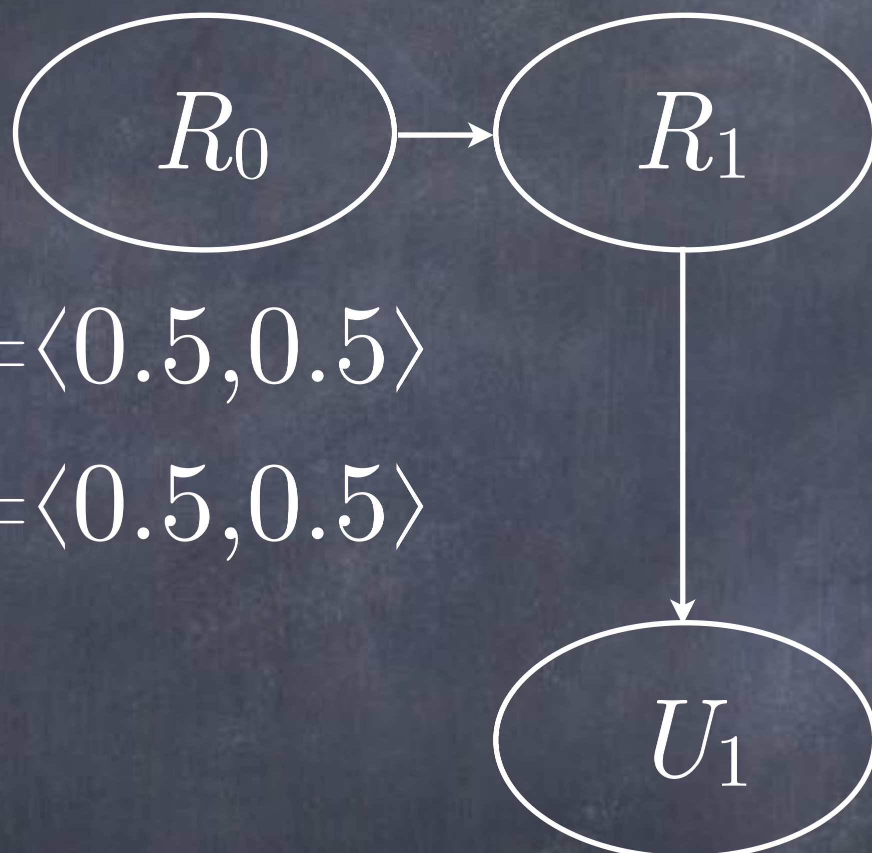
$$\begin{aligned}
\mathbf{P}(R_1) &= \sum_{r_0} \mathbf{P}(R_1 \mid r_0) P(r_0) \\
&= \langle 0.7, 0.3 \rangle \times 0.5 + \langle 0.3, 0.7 \rangle \times 0.5 \\
&= \langle 0.5, 0.5 \rangle
\end{aligned}$$

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$U_1 = true$

$\mathbf{P}(R_0)=\langle 0.5,0.5\rangle$

$\mathbf{P}(R_1)=\langle 0.5,0.5\rangle$

$\mathbf{P}(U_t \mid R_t)$

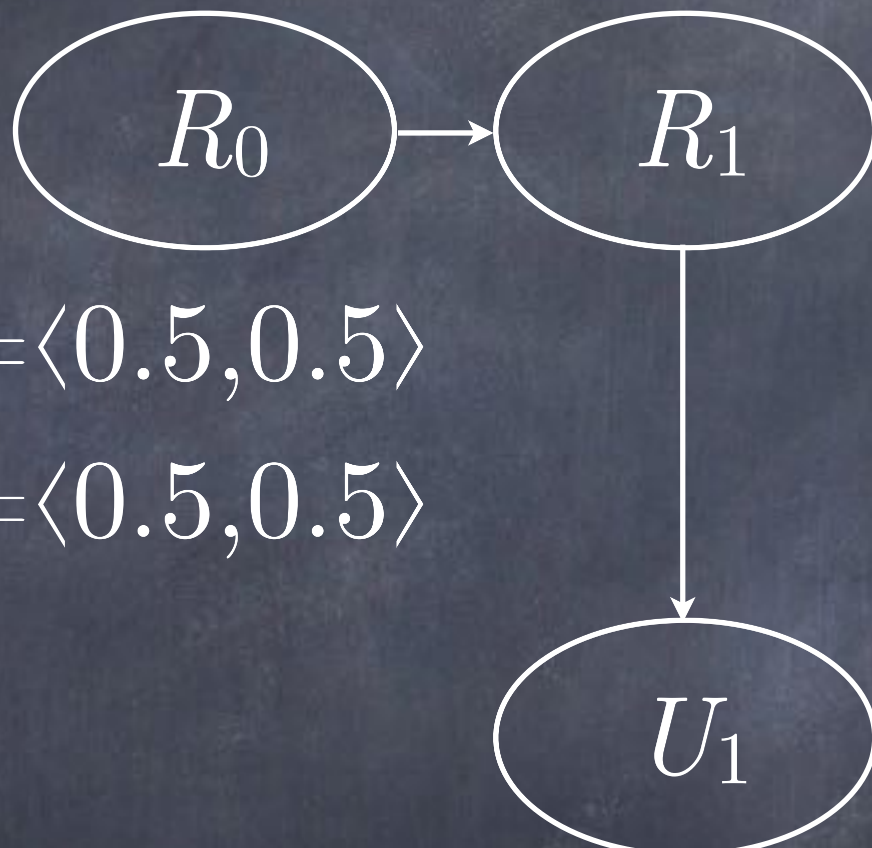
R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_0)=\langle 0.5,0.5\rangle$

$\mathbf{P}(R_1)=\langle 0.5,0.5\rangle$

$U_1 = true$

$$\mathbf{P}(R_1 \mid u_1) = \alpha\, \mathbf{P}(u_1 \mid R_1)\mathbf{P}(R_1)$$

$$= \alpha\, \langle 0.9, 0.2\rangle\langle 0.5, 0.5\rangle$$

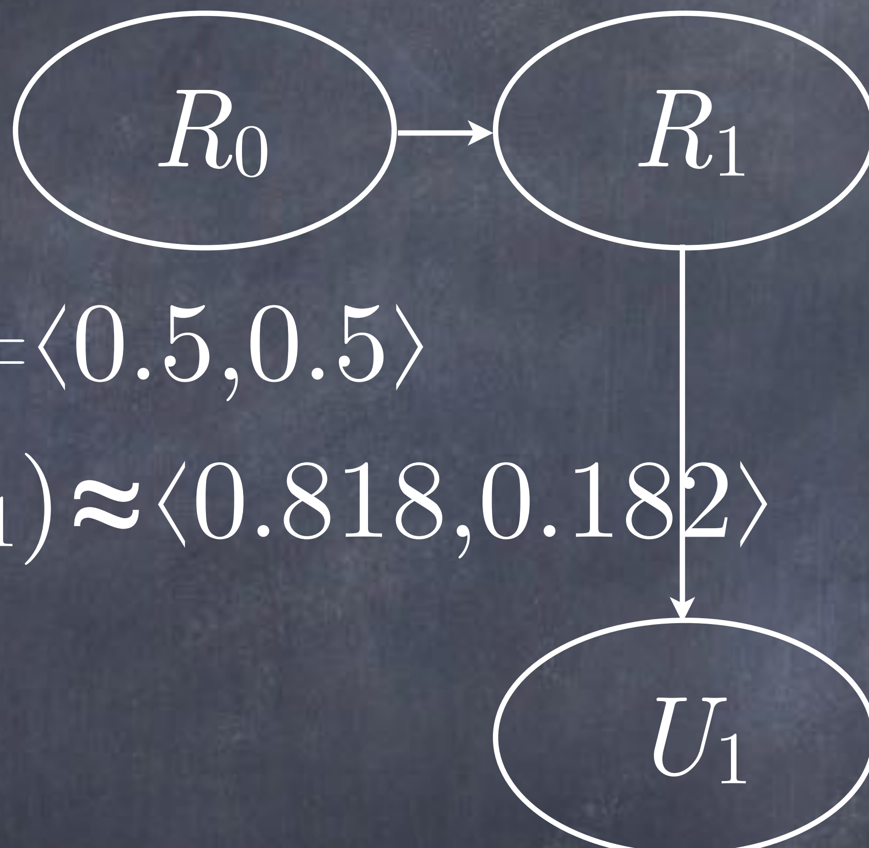
$$= \alpha\, \langle 0.45, 0.1\rangle \approx \langle 0.818, 0.182\rangle$$

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$\mathbf{P}(R_0)=\langle 0.5,0.5\rangle$

$\mathbf{P}(R_{1}|u_1)\approx\langle 0.818,0.182\rangle$

$U_1 = true$

$\mathbf{P}(U_t \mid R_t)$

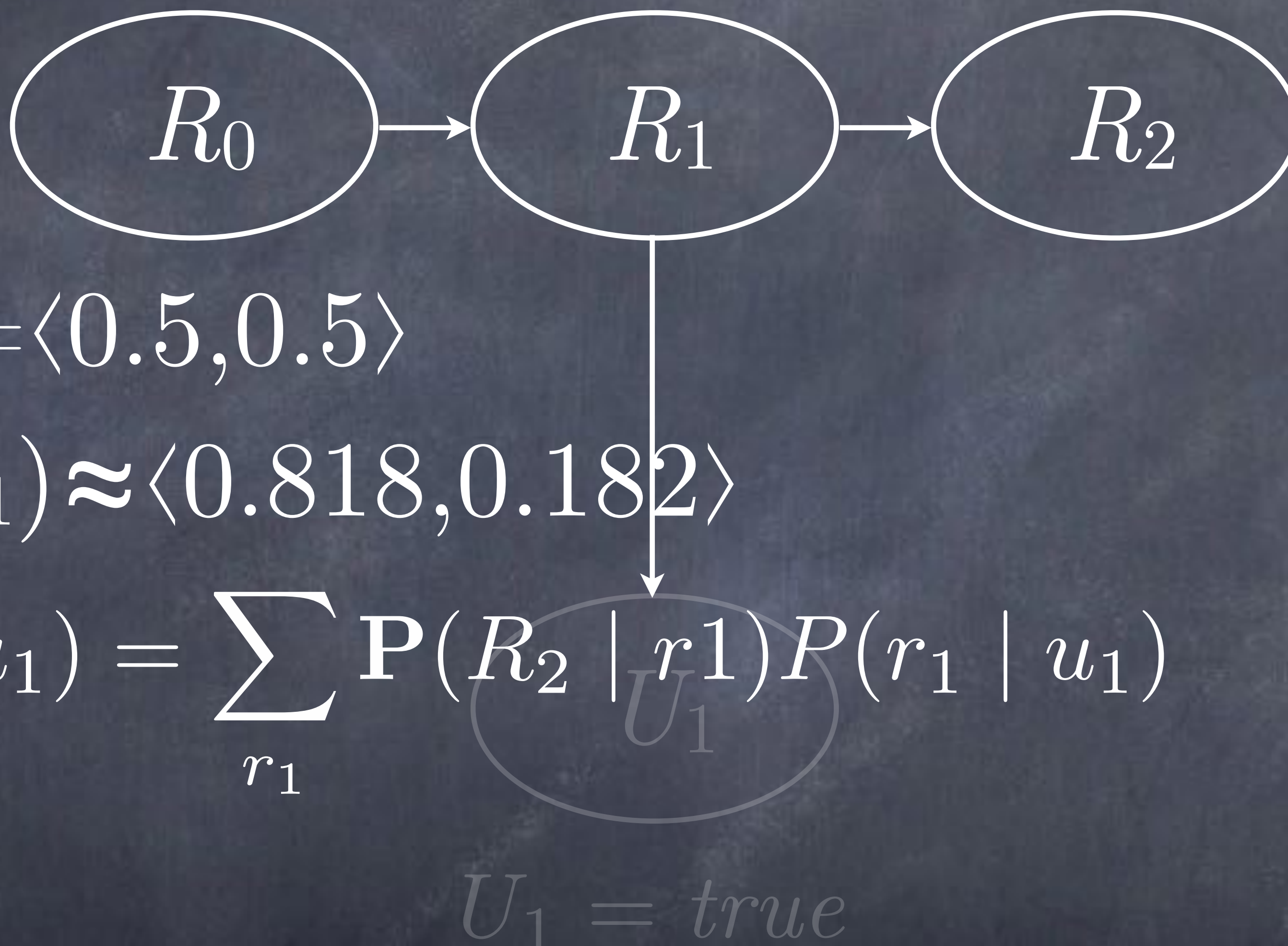
R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_0)=\langle 0.5,0.5\rangle$

$\mathbf{P}(R_{1|}u_1)\approx\langle 0.818,0.182\rangle$

$$\mathbf{P}(R_2 \mid u_1) = \sum_{r_1} \mathbf{P}(R_2 \mid r1)P(r_1 \mid u_1)$$

$$= \langle 0.7, 0.3\rangle \times 0.818 + \langle 0.3, 0.7\rangle \times 0.182$$

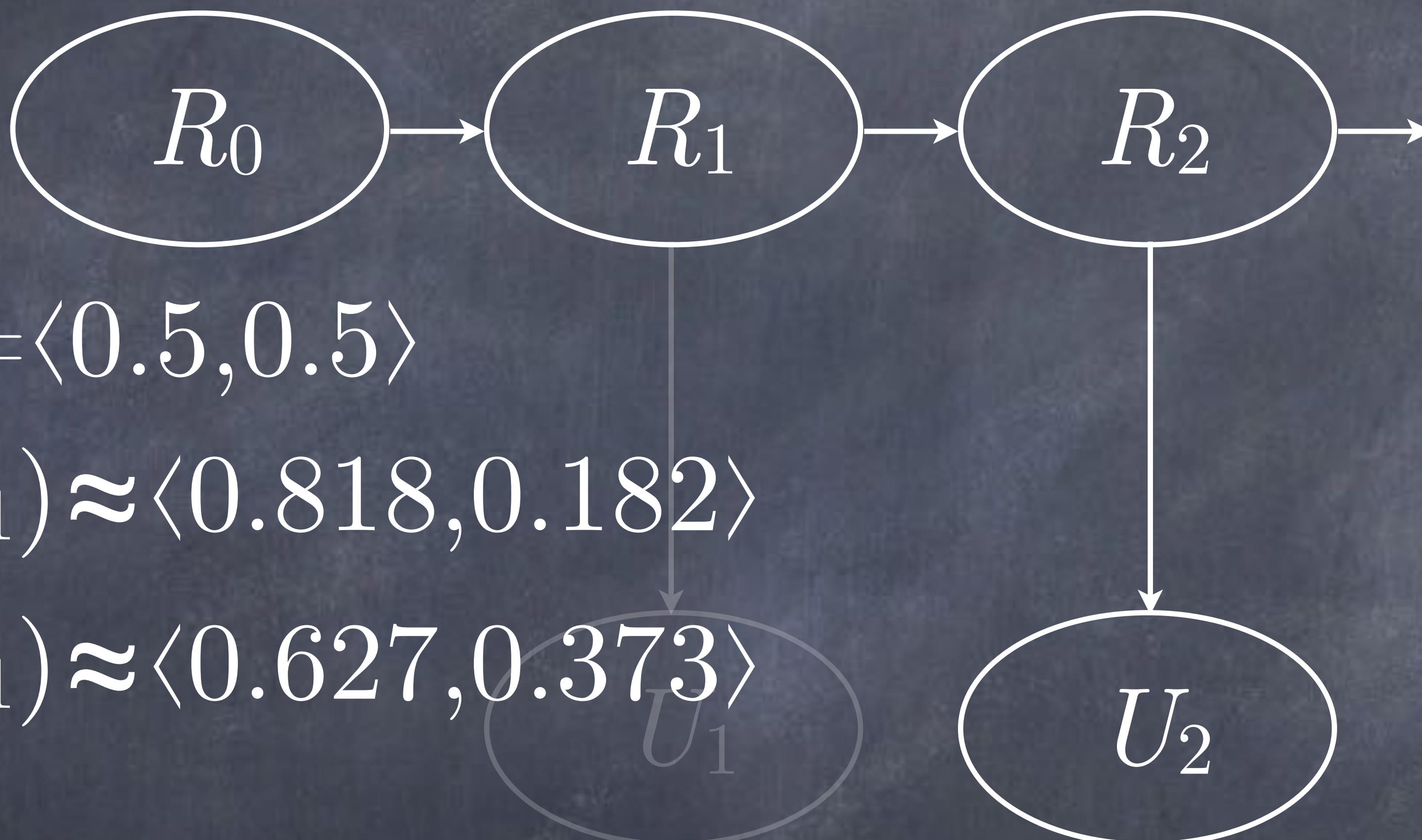
$$\approx \langle 0.627, 0.373\rangle$$

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$

$\mathbf{P}(R_{1|}u_1) \approx \langle 0.818, 0.182 \rangle$

$\mathbf{P}(R_{2|}u_1) \approx \langle 0.627, 0.373 \rangle$

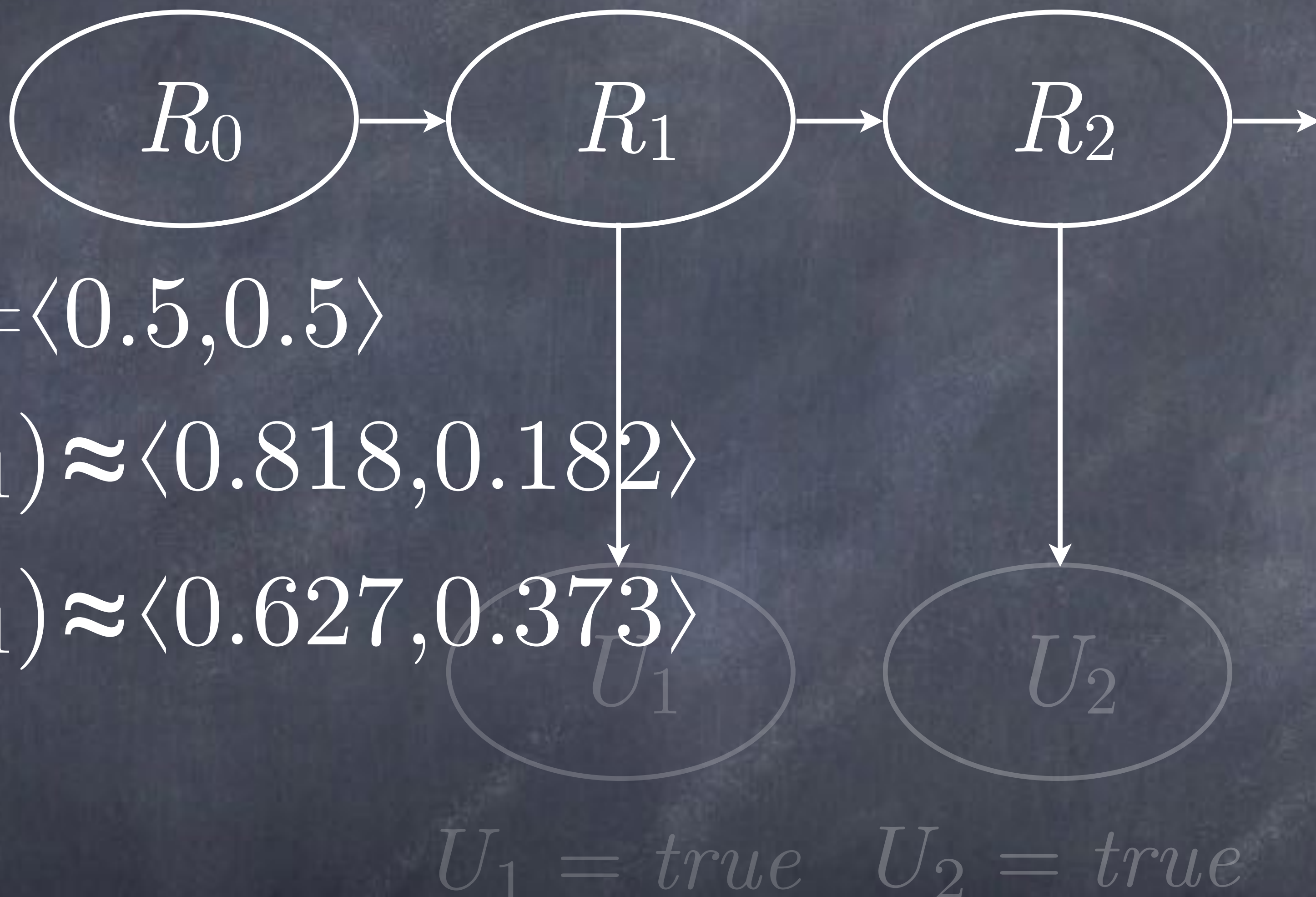
$U_1 = true$ $\quad U_2 = true$

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$

$\mathbf{P}(R_{1|}u_1) \approx \langle 0.818, 0.182 \rangle$

$\mathbf{P}(R_{2|}u_1) \approx \langle 0.627, 0.373 \rangle$

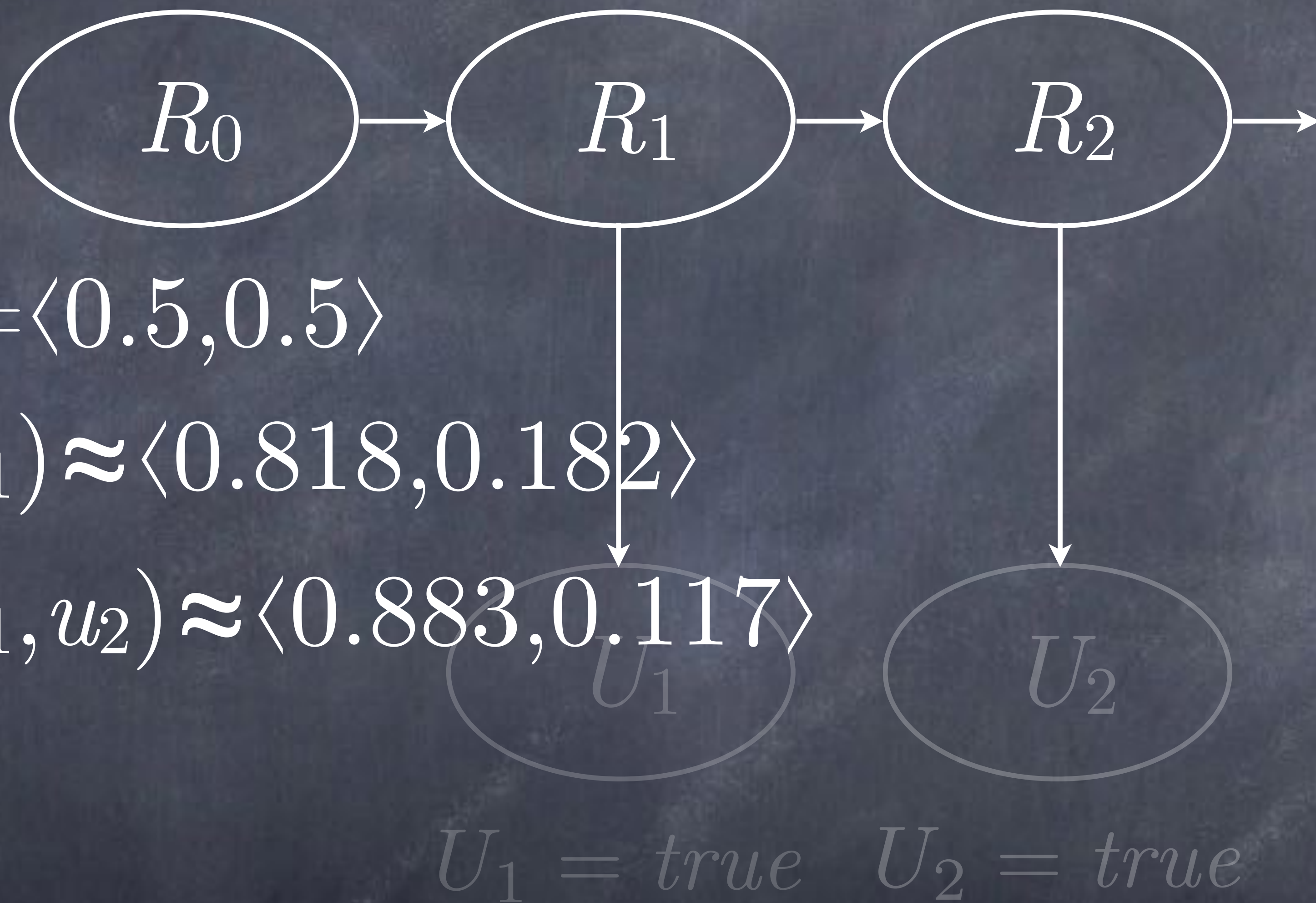
$$\mathbf{P}(R_2 \mid u_1, u_2) = \alpha\, \mathbf{P}(u_2 \mid R_2)\mathbf{P}(R_2 \mid u_1)$$
$$= \alpha\, \langle 0.9, 0.2 \rangle \langle 0.627, 0.373 \rangle$$
$$= \alpha\, \langle 0.565, 0.075 \rangle \approx \langle 0.883, 0.117 \rangle$$

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$

$\mathbf{P}(R_1 | u_1) \approx \langle 0.818, 0.182 \rangle$

$\mathbf{P}(R_2 | u_1, u_2) \approx \langle 0.883, 0.117 \rangle$

Filtering (State Estimation)

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

Implement as recursive procedure:

$$\mathbf{P}(\mathbf{X}_0 \mid \mathbf{e}_{1:0}) = \mathbf{P}(\mathbf{X}_0)$$

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha\, \textsc{Forward}(\mathbf{P}(\mathbf{X}_t \mid \mathbf{e}_{1:t}), \mathbf{e}_{t+1})$$

Updates in constant time and space (w.r.t. t)!

Inference

- Filtering (State Estimation)
- Prediction
- Smoothing
- Most Likely Explanation

Inference

- Filtering (State Estimation)
- Prediction
- Smoothing
- Most Likely Explanation

Prediction

- Compute posterior distribution for future state, given all evidence to date
- This is filtering (state estimation) without the addition of any new evidence

$$\mathbf{P}(X_{t+k+1} \mid e_{1:t+1})$$

Prediction

$$\mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{e}_{1:t+1}) = \sum_{\mathbf{x}_{t+k}} \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{x}_{t+k}) \mathbf{P}(\mathbf{x}_{t+k} \mid \mathbf{e}_{1:t})$$

Prediction

$$\mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{e}_{1:t+1}) = \sum_{\mathbf{x}_{t+k}} \mathbf{P}(\mathbf{X}_{t+k+1} \mid \mathbf{x}_{t+k}) \mathbf{P}(\mathbf{x}_{t+k} \mid \mathbf{e}_{1:t})$$

Converges to "stationary distribution" of the Markov process defined by the model

Inference

- Filtering (State Estimation)
- Prediction
- Smoothing
- Most Likely Explanation

Smoothing

- Compute posterior over past state(s) given evidence up to the present
- May allow you to improve the estimate you made at the time, since you know now what was then in the future

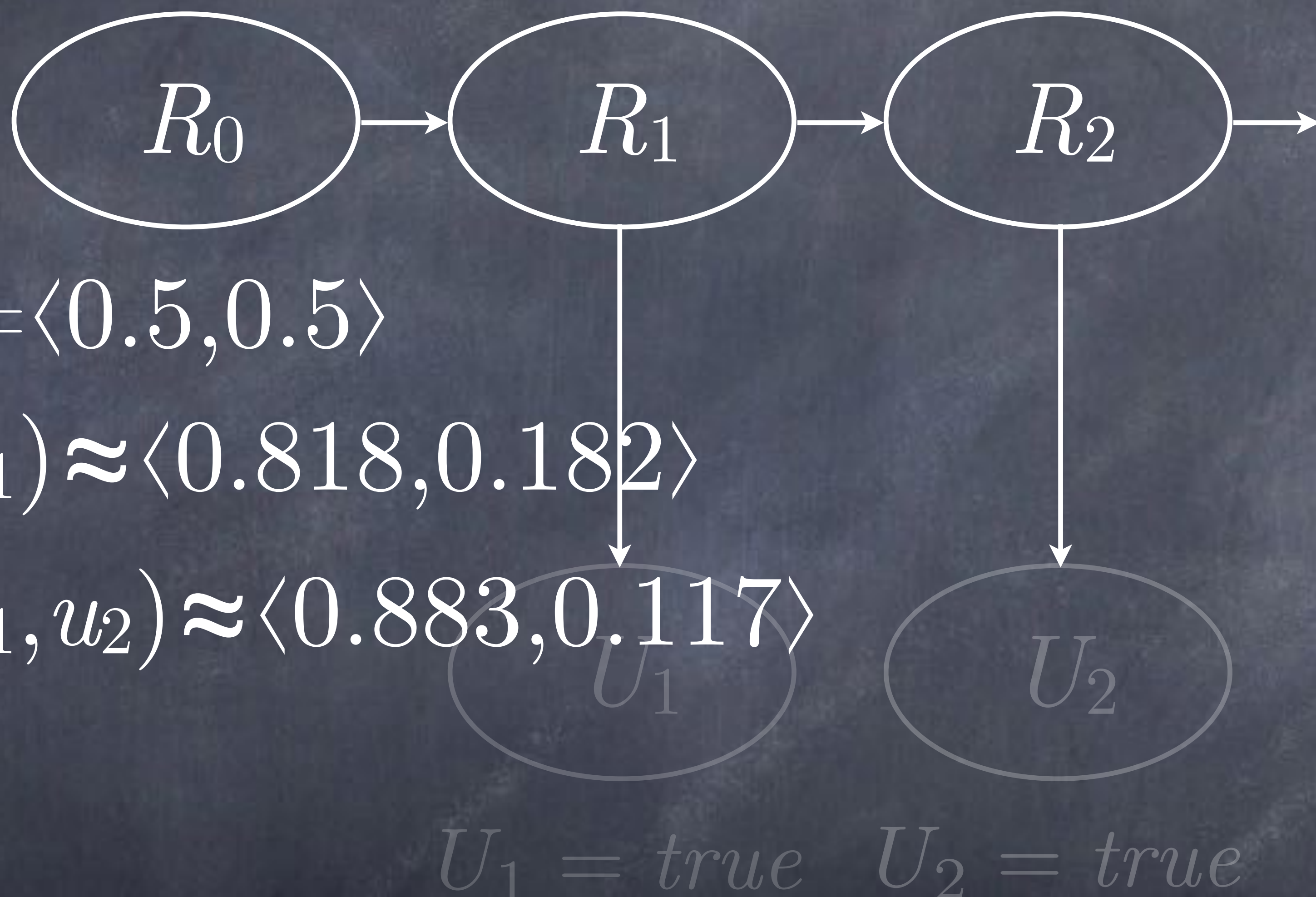
$$\mathbf{P}(X_{t-k} \mid e_{1:t+1})$$

$\mathbf{P}(R_0)$

$P(R_t)$
0.5

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3



$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$

$\mathbf{P}(R_{1|}u_1) \approx \langle 0.818, 0.182 \rangle$

$\mathbf{P}(R_{2|}u_1, u_2) \approx \langle 0.883, 0.117 \rangle$

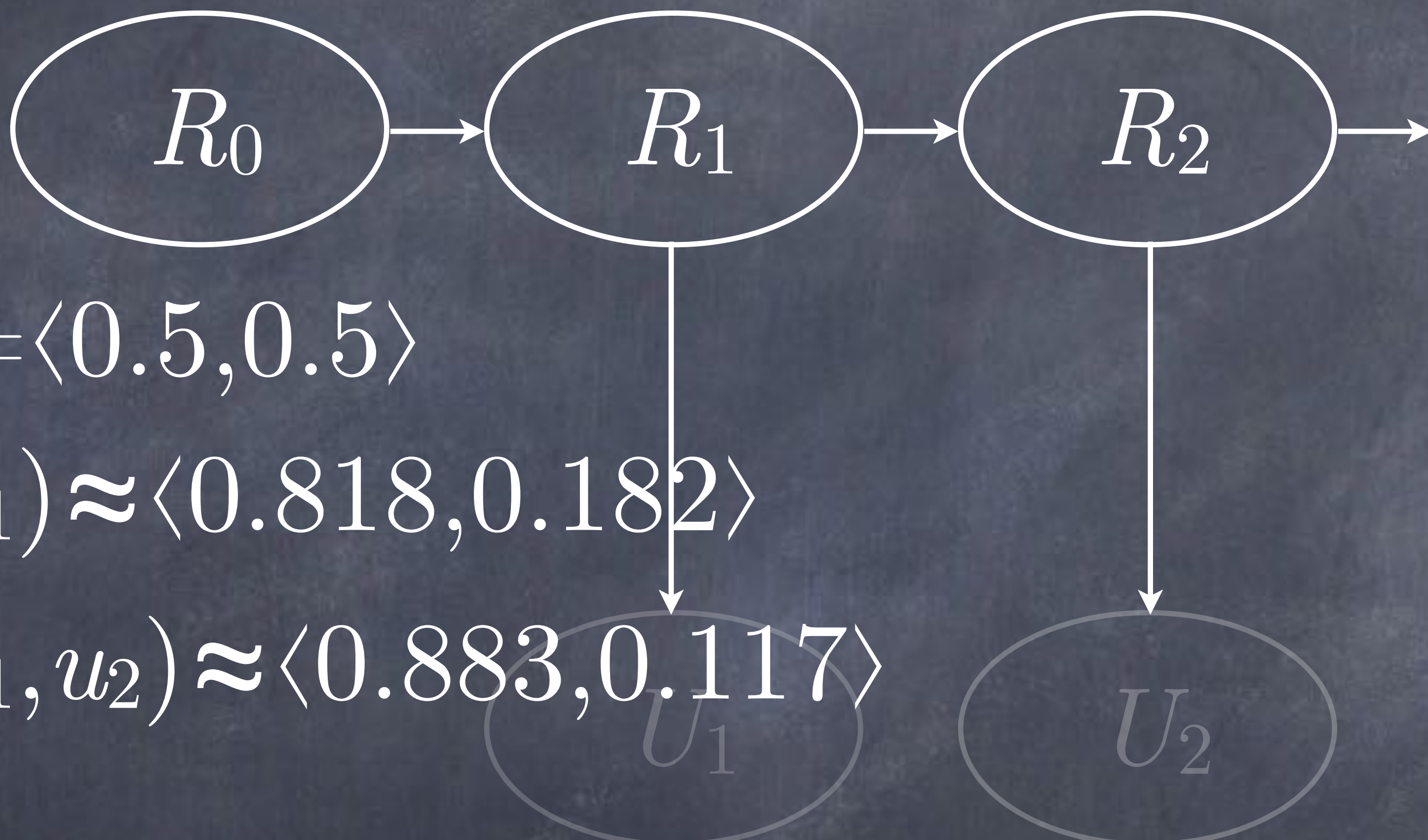
$\mathbf{P}(R_1 \mid u_1, u_2)$

$\mathbf{P}(R_t \mid R_{t\text{-}1})$

$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3

$\mathbf{P}(R_0)$

$P(R_t)$
0.5



$U_1 = true \quad U_2 = true$

$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2

$\mathbf{P}(R_0) = \langle 0.5, 0.5 \rangle$

$\mathbf{P}(R_{1|}u_1) \approx \langle 0.818, 0.182 \rangle$

$\mathbf{P}(R_{2|}u_1, u_2) \approx \langle 0.883, 0.117 \rangle$

$$\mathbf{P}(R_1 \mid u_1, u_2) = \alpha\, \mathbf{P}(R_1 \mid u_1)\mathbf{P}(u_2 \mid R_1)$$
$$\approx \langle 0.883, 0.117 \rangle$$

Inference

- Filtering (State Estimation)
- Prediction
- Smoothing
- Most Likely Explanation

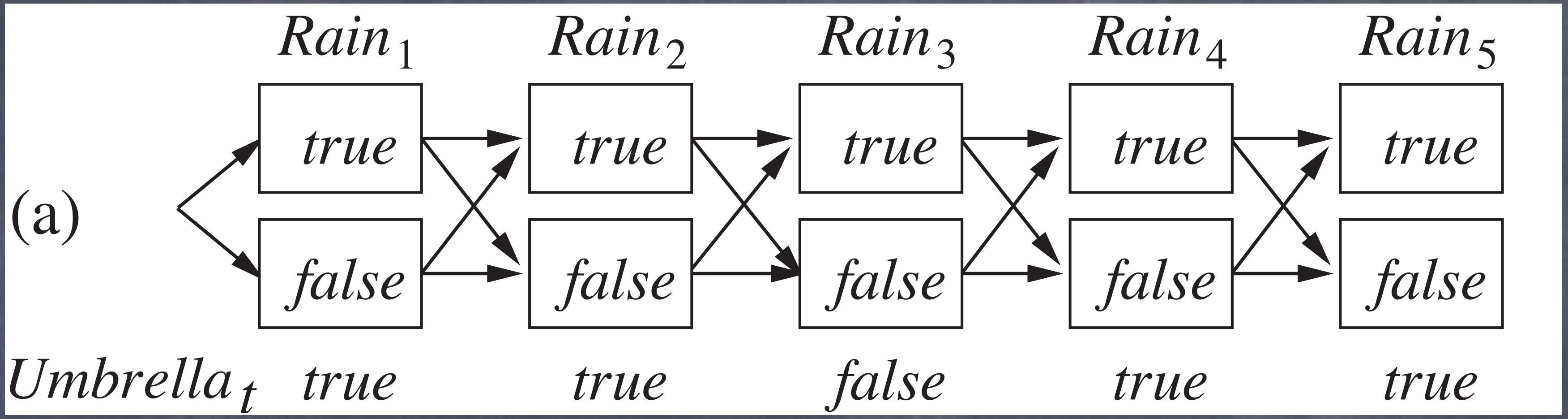
Inference

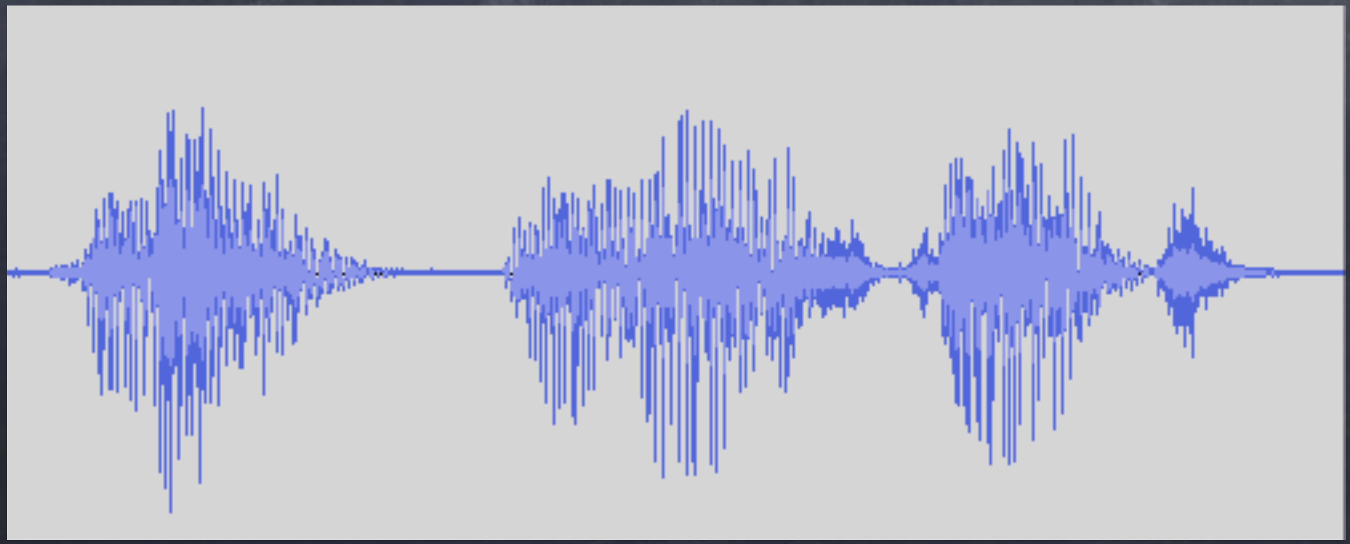
- Filtering (State Estimation)
- Prediction
- Smoothing
- Most Likely Explanation

Finding the Most Likely Sequence

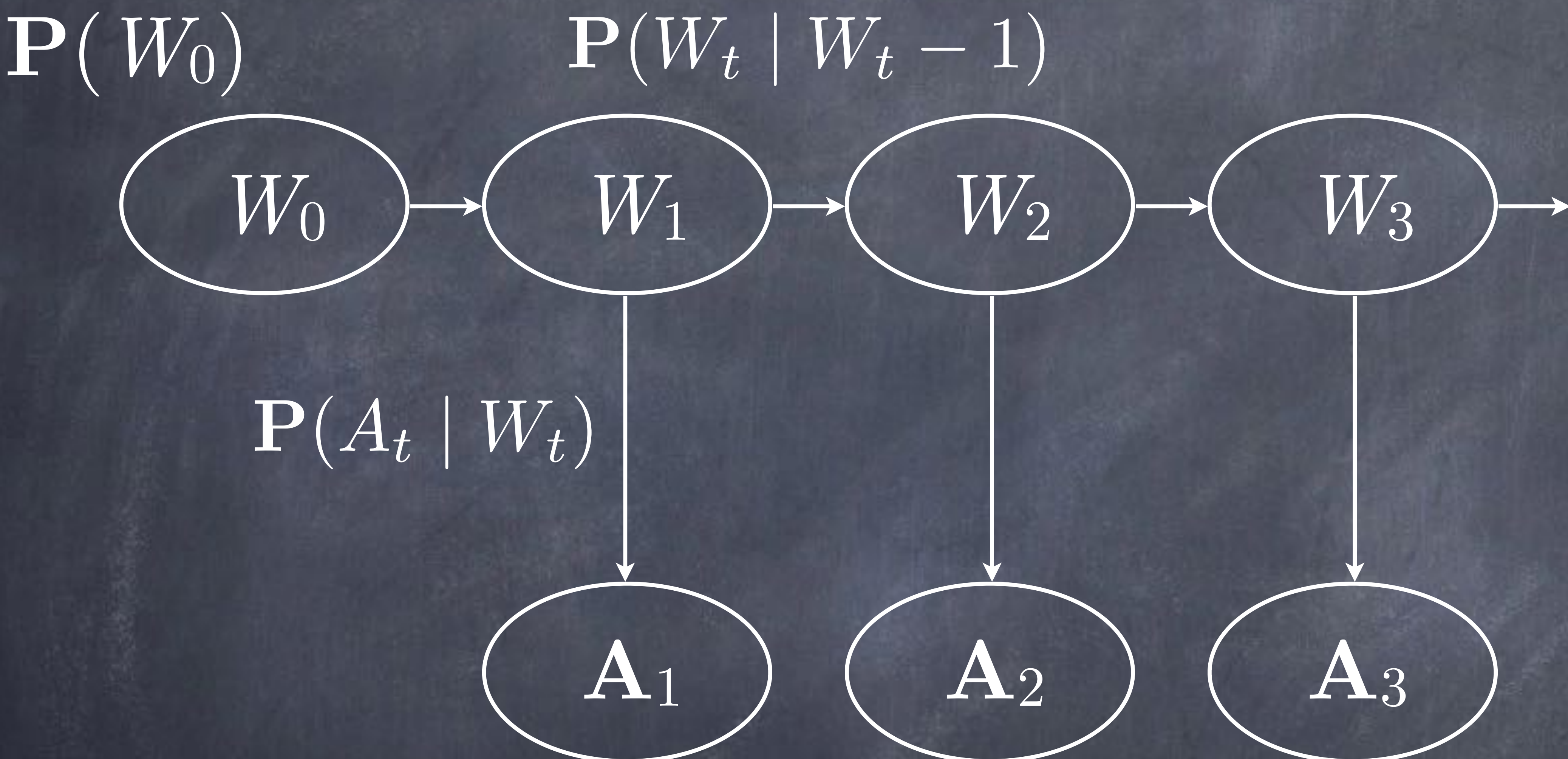
- Infer most likely sequence of states that could have generated observations

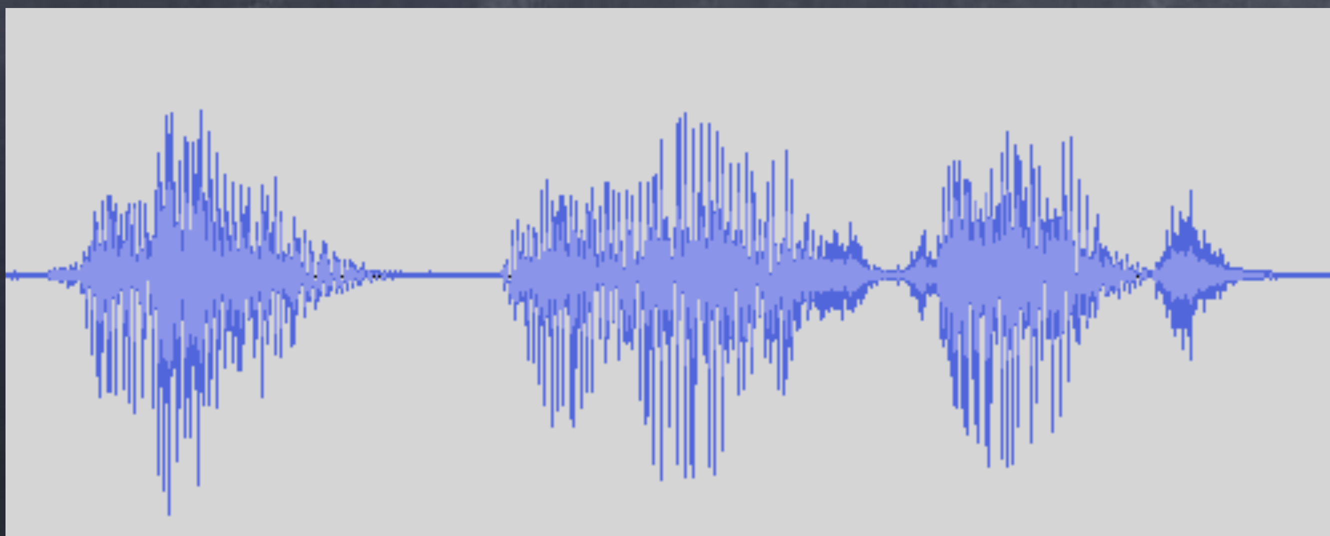
$$\underset{\mathbf{X}_{0:t}}{\operatorname{argmax}}\ \mathbf{P}(\mathbf{X}_{0:t} \mid \mathbf{e}_{1:t})$$

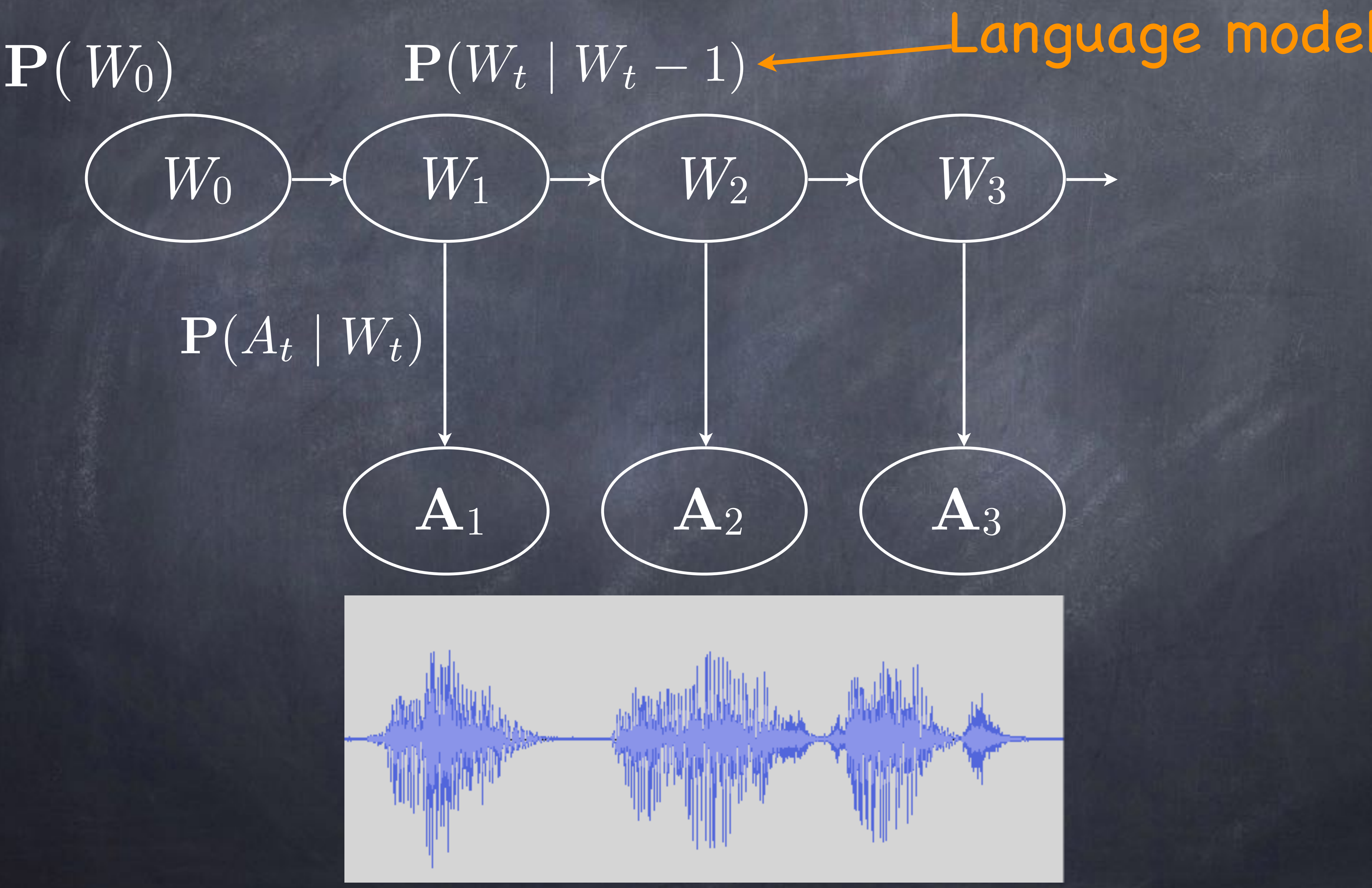

Rain1
Rain2
Rain3
Rain4
Rain5
(a)
true
false
Umbrellat
true
true
false
true
true

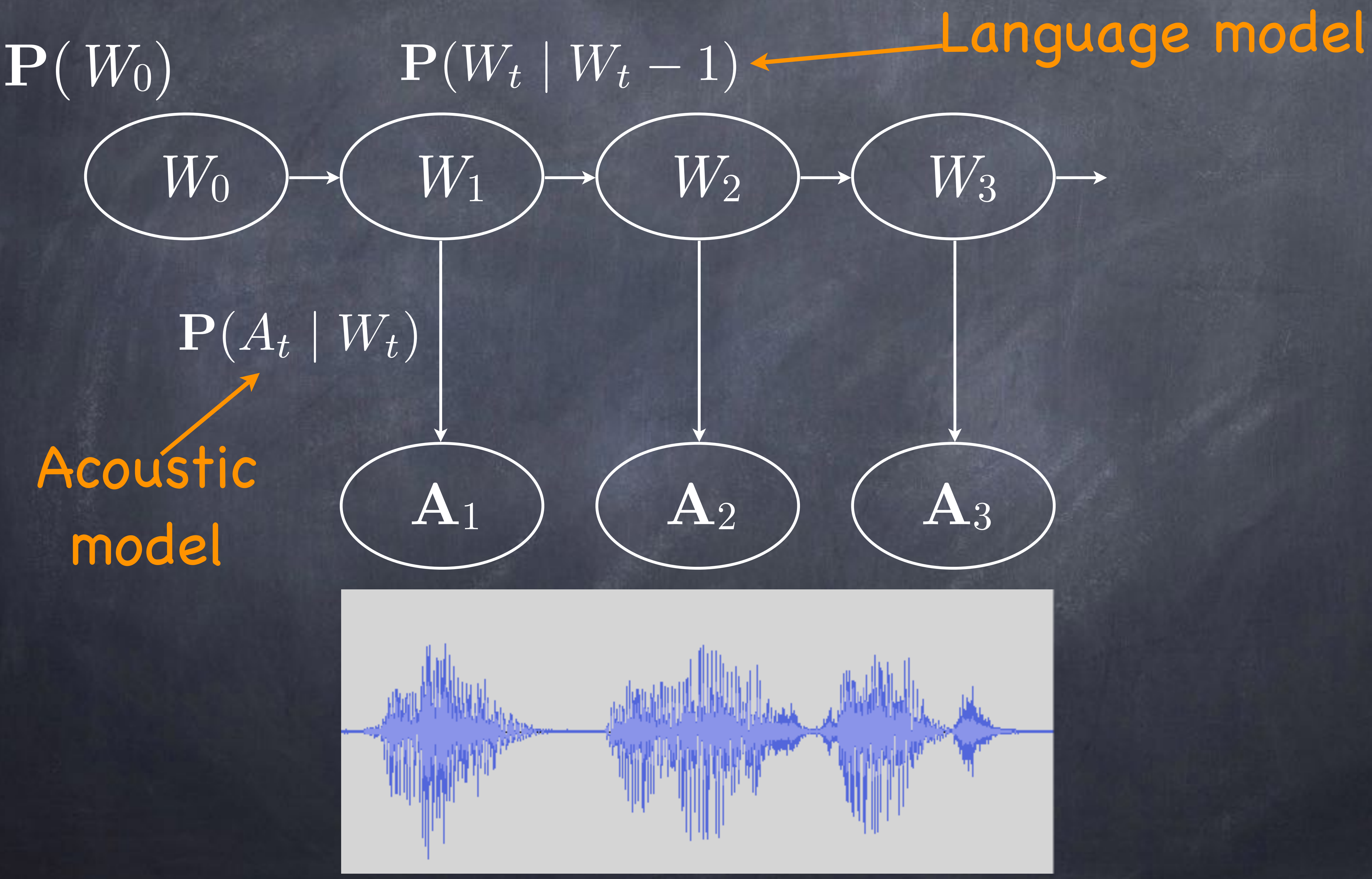



$\mathbf{A}_1$
$\mathbf{A}_2$
$\mathbf{A}_3$

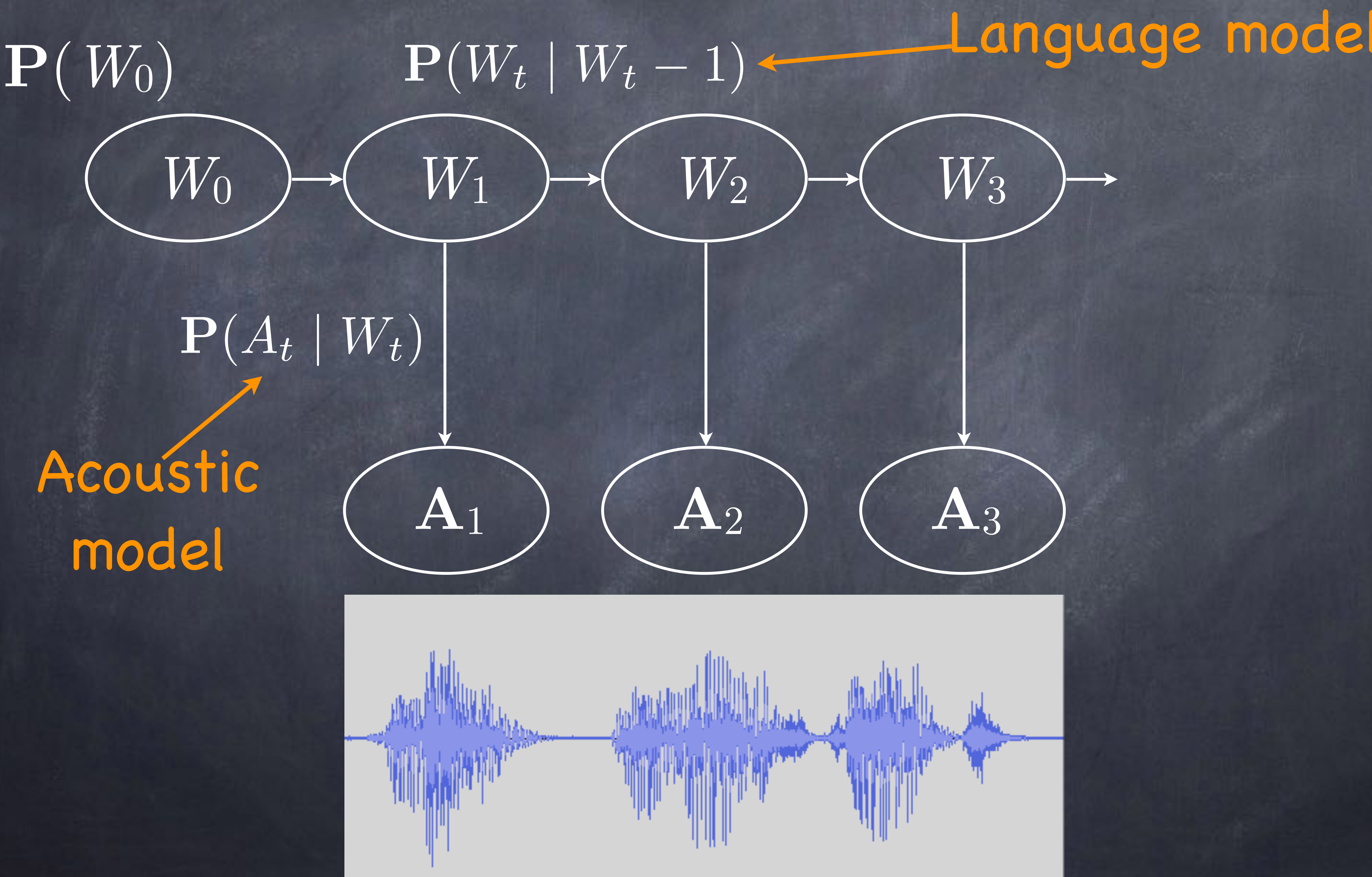

$\mathbf{P}(W_0)$
$\mathbf{P}(W_t \mid W_t - 1)$
W_0
W_1
W_2
W_3
$\mathbf{P}(A_t \mid W_t)$
$\mathbf{A}_1$
$\mathbf{A}_2$
$\mathbf{A}_3$

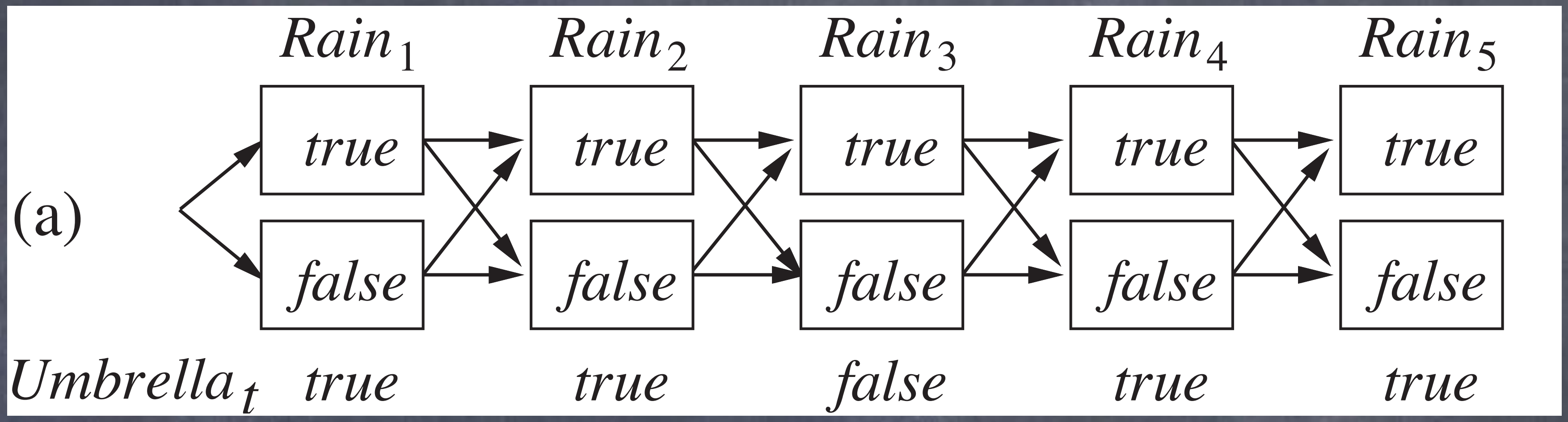



Language model
$\mathbf{P}(W_0)$
$\mathbf{P}(W_t \mid W_t - 1)$
W_0
W_1
W_2
W_3
$\mathbf{P}(A_t \mid W_t)$
$\mathbf{A}_1$
$\mathbf{A}_2$
$\mathbf{A}_3$


$\mathbf{P}(W_0)$
$\mathbf{P}(W_t \mid W_t - 1)$
Language model
W_0
W_1
W_2
W_3
$\mathbf{P}(A_t \mid W_t)$
Acoustic model
$\mathbf{A}_1$
$\mathbf{A}_2$
$\mathbf{A}_3$

HELLO ... MY NAME IS GEORGE





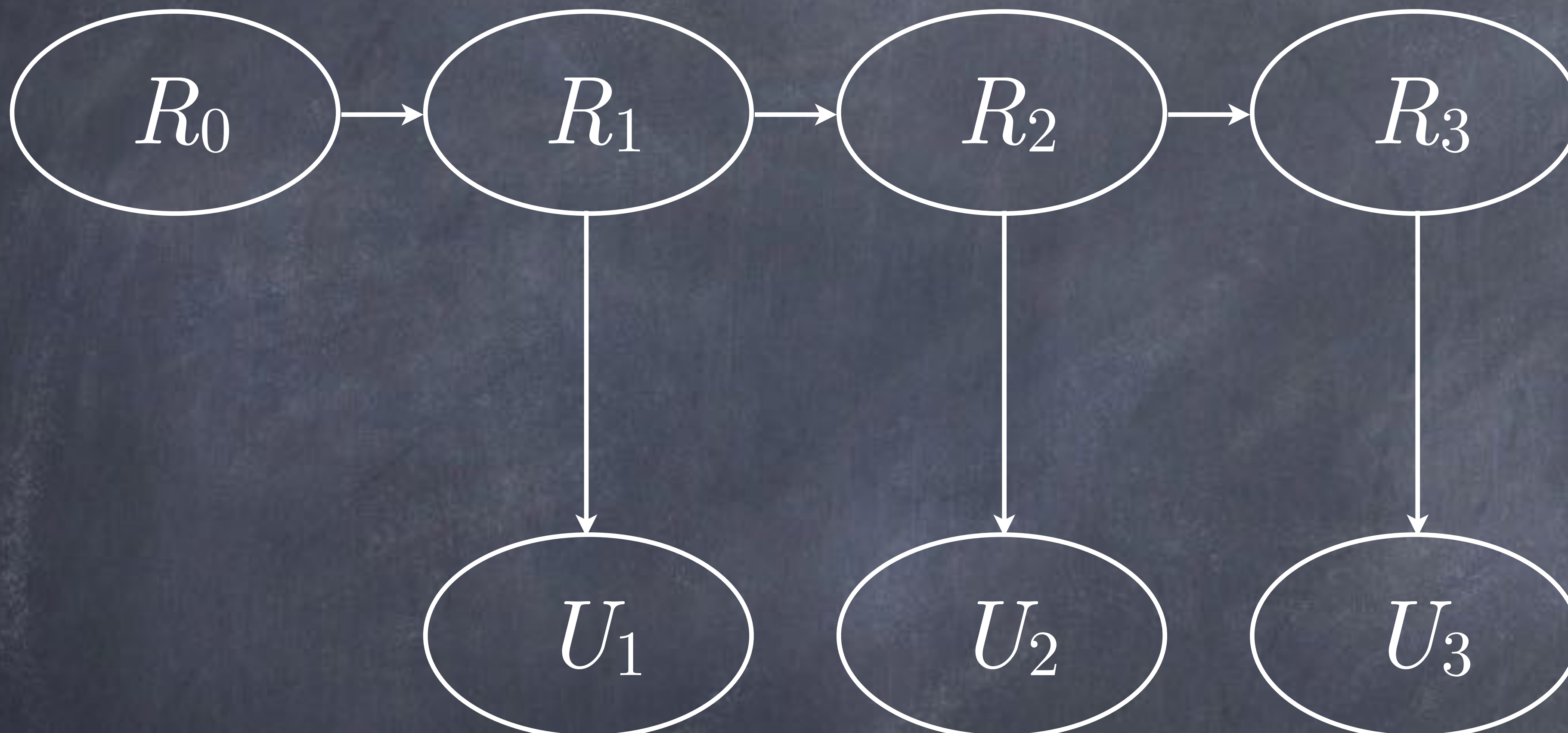
Rain 1
Rain 2
Rain 3
Rain 4
Rain 5
true
false
(a)
Umbrella t
true
true
false
true
true

$\mathbf{P}(R_0)$

$P(R_t)$
0.5

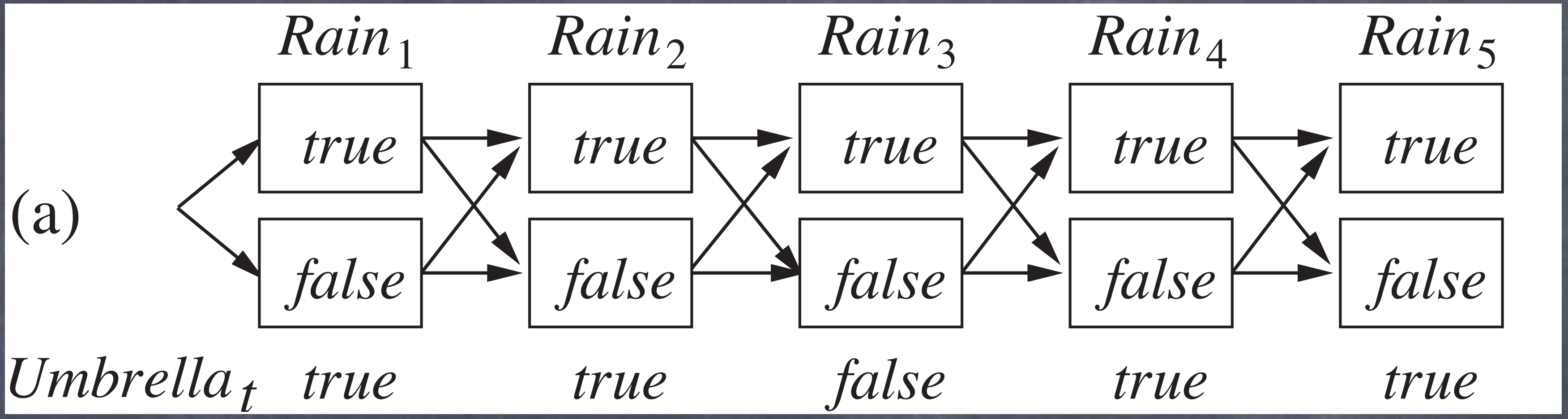
$\mathbf{P}(R_t \mid R_{t\text{-}1})$

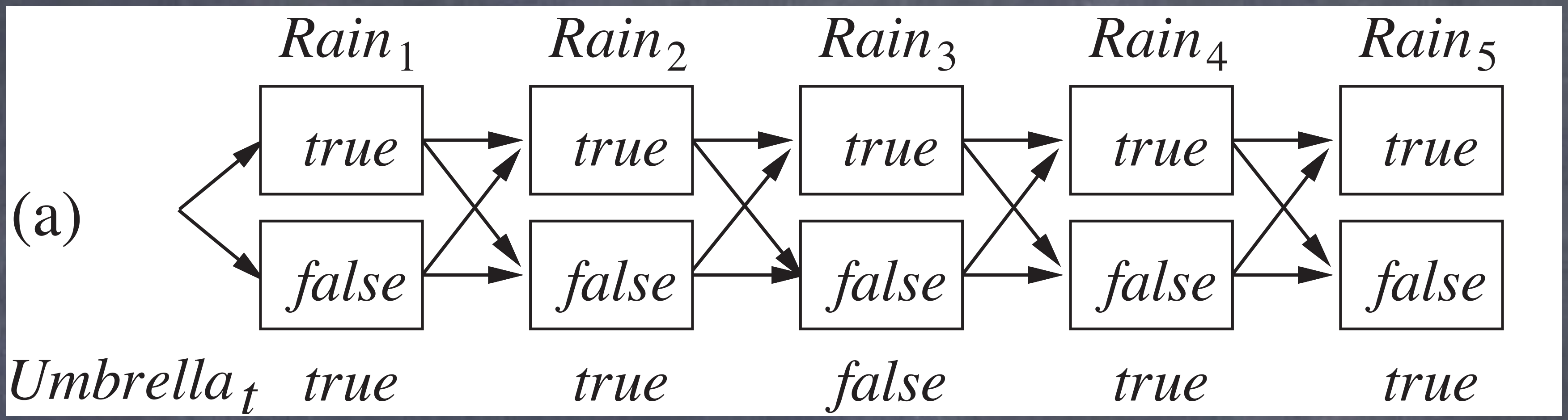
$R_{t\text{-}1}$	$P(R_t)$
t	0.7
f	0.3



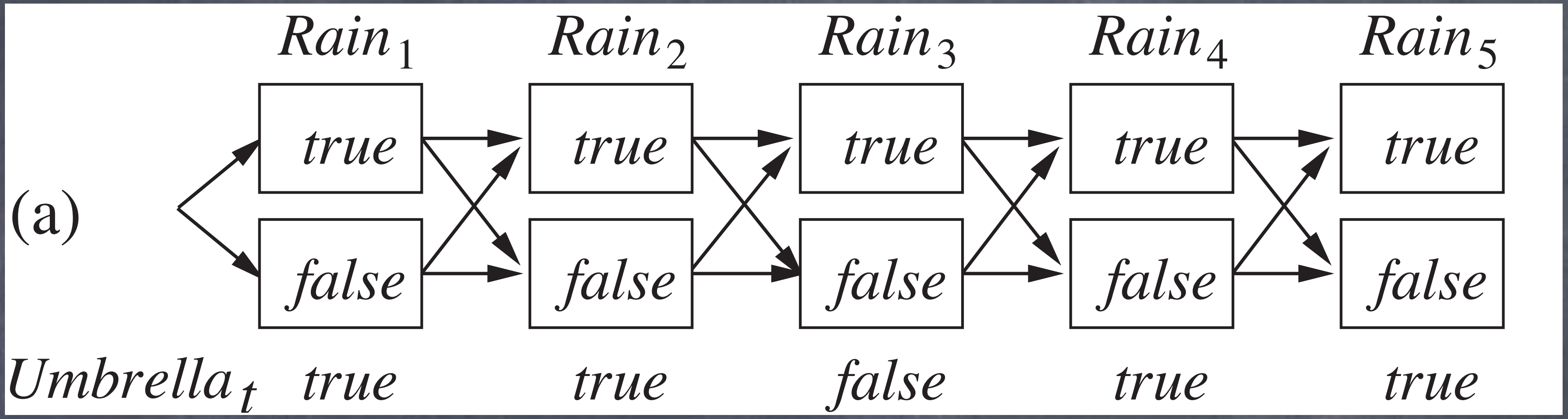
$\mathbf{P}(U_t \mid R_t)$

R_t	$P(U_t)$
t	0.9
f	0.2


Rain 1
Rain 2
Rain 3
Rain 4
Rain 5
true
true
true
true
true
(a)
false
false
false
false
false
Umbrella t
true
true
false
true
true



- Compute all possible sequences of values
- Use the model to evaluate each sequence



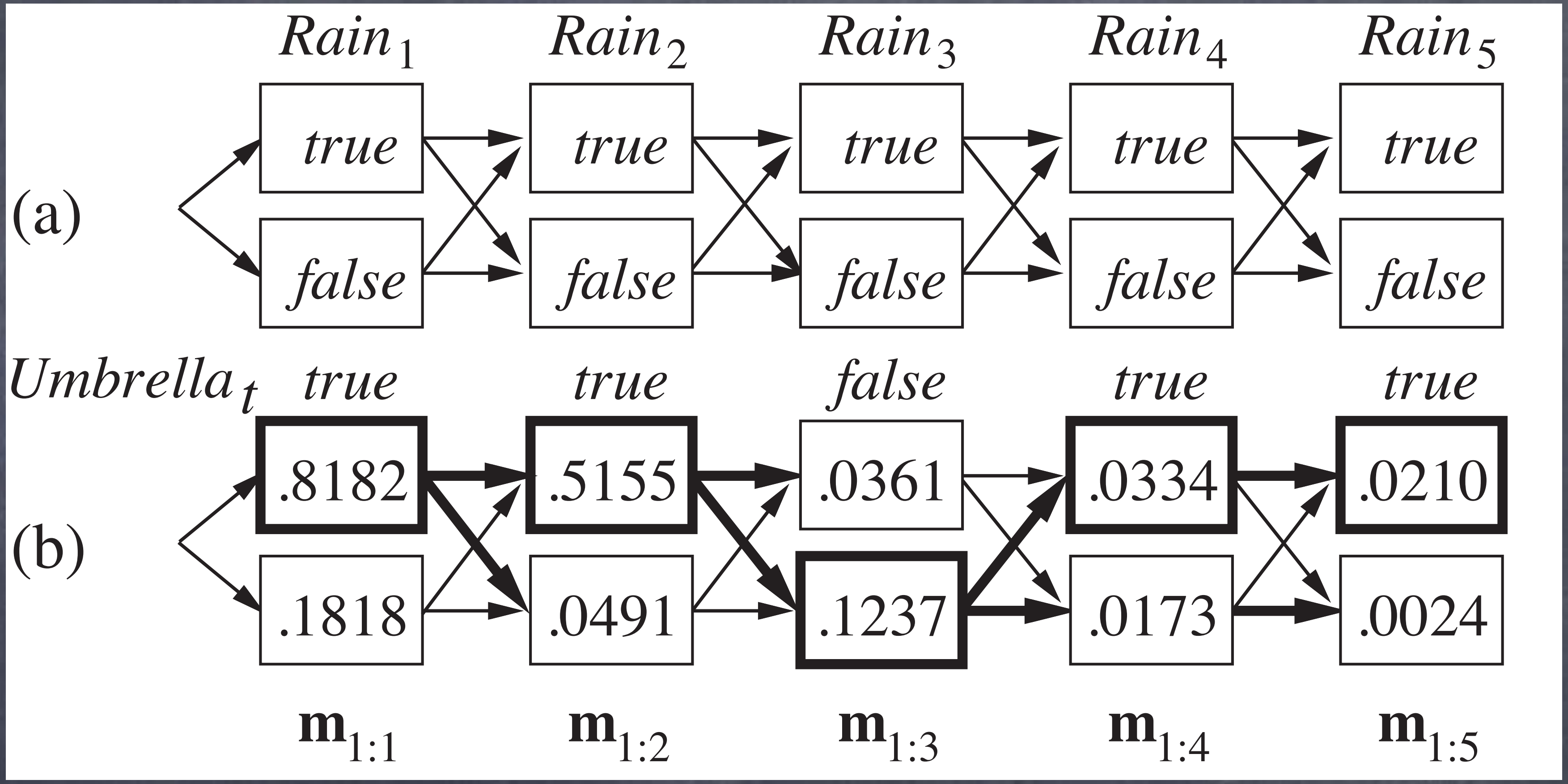
- Compute all possible sequences of values
- Use the model to evaluate each sequence

Exponential # of possible sequences!

Finding the Most Likely Sequence

- Infer most likely sequence of states that could have generated observations
- Without enumerating all possible sequences of states and evaluating their likelihood

Viterbi Algorithm



Time complexity: $O(t)$

Space complexity: $O(t)$

Temporal Model

- Representation of state: $\mathbf{X}_t$, $\mathbf{E}_t$
- Transition model: $\mathbf{P}(\mathbf{X}_t \mid \mathbf{X}_{t\text{-}1})$
  - Markov assumption, stationary process
- Sensor model: $\mathbf{P}(\mathbf{E}_t \mid \mathbf{X}_t)$
  - Sensor Markov assumption, stationary process
- Prior distribution at time 0: $\mathbf{P}(\mathbf{X}_0)$

Temporal Model

$$\mathbf{P}(\mathbf{X}_{0:t}, \mathbf{E}_{1:t}) = \mathbf{P}(\mathbf{X}_0) \prod_{i=1}^{t} \mathbf{P}(\mathbf{X}_i \mid \mathbf{X}_{i-1}) \mathbf{P}(\mathbf{E}_i \mid \mathbf{X}_i)$$

Initial State Model

Sensor Model

Transition Model

Inference

- Filtering (State Estimation)
- Prediction
- Smoothing
- Most Likely Explanation

Uncertainty

- Probability
- Probabilistic Inference with full joint prob. dist.
- Bayesian Networks
- Exact and Approximate inference in BNs
- Temporal Models

For Next Time:

Unit 3 Exam

QRM-DT

Shwe, Middletone, Heckerman, Henrion, Horvitz, Lehmann, Cooper (1991). Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base: I. The Probabilistic Model and Inference. Methods of Information in Medicine: 241-255.

https://www.researchgate.net/publication/237128541_Probabilistic_Diagnosis_Using_a_Reformulation_of_the_INTERNIST1QMR_Knowledge_Base