Lecture 6

Things Markov

- Markov Chains
  - Gibbs Sampling
- Hidden Markov Models
  - State Estimation,
  - Prediction,
  - Smoothing,
  - Most Probable Path

Background

In dynamic systems:

State Estimation – Estimating the current state of the system given current knowledge.

Prediction – Estimating future state(s) of the system given current knowledge.

Smoothing – Estimating prior states of the system given current knowledge.

Background

Independence

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

Conditional Independence

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

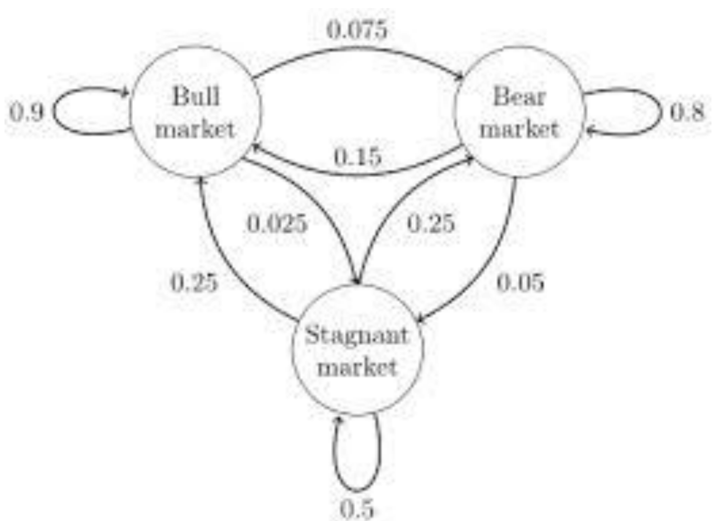
$$=P(A|B,C)=P(A|C)$$

Markov Blanket

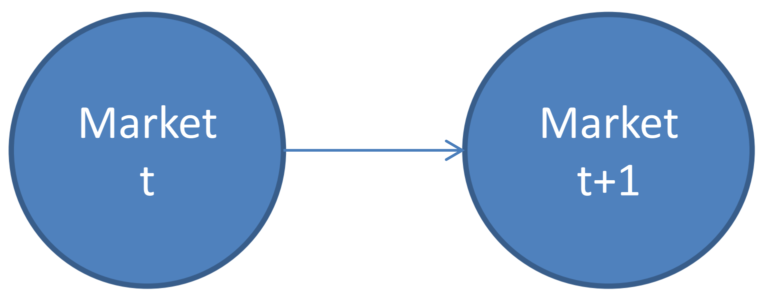
The variables given which a variable is conditionally independent of all others in the model.

Markov Chains

- Initial state
- Transition probabilities
  - Markov Condition: State at time t+1 depends only on state at time t. (Leads to higher order MCs)
  - Ie Current state conditionally independent of all prior states except preceeding.



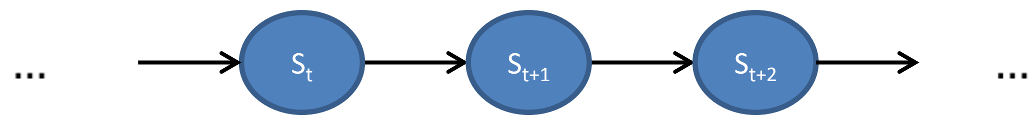
	Bull	Stagnant	Bear
Bull	.9	.025	.075
Stagnant	.25	.5	.25
Bear	.15	.05	.8



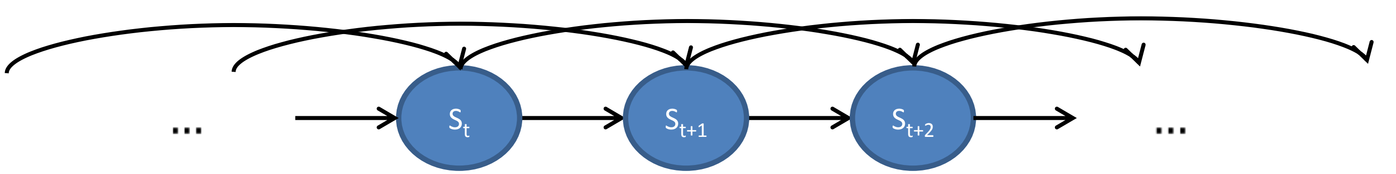
Markov Chains

- Using nodes to represent variables & conditional distributions
- Conditioned upon variables indicated by edges.

1st Order Markov Chain



2nd Order Markov Chain



Markov Chain

Transition probabilities for transition matrix T.

Simple state prediction (1st Order):

$$\boldsymbol{X}_{t+n} = \boldsymbol{X}_t{}^T \boldsymbol{T}^n$$

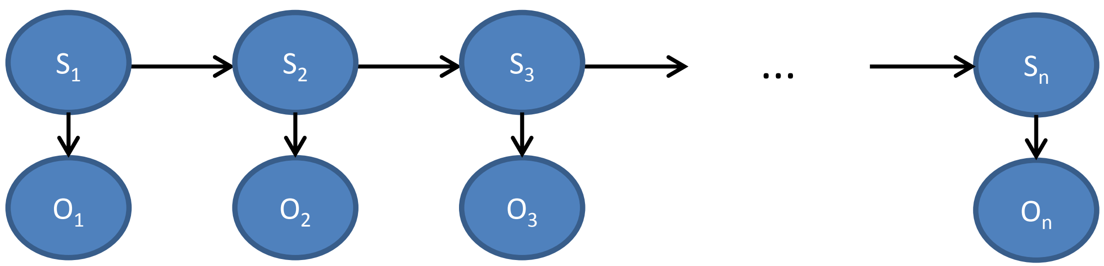
The eigenvector to the eigen value 1 gives the steady equilibrium distribution. (Ie ’long run’ distribution of the MC).

Markov Chain Monte Carlo

- Generate a (1st order) MC that (in its equilibrium state) represents the target distribution.
- Proceed to generate samples from it by evolving the MC.
    - Note: Samples not independent.
- As the number of samples approaches infinity, the sampled distribution approaches the actual equilibrium distribution.
    - Burn period
    - *n*th sample

Hidden Markov Models

- State of system is hidden from us.
- Some observation related to the state is available to us.
  - Require sensor/emission probabilities, E.
  - Assume observations depend only on current state. (Conditionally indepent of all other states and observations.)



Hidden Markov Models

- Prediction: Just as in Markov Chains...

$$P(S_{t+n}|S_t\ ) = P(S_t)\boldsymbol{T}^n$$

Hidden Markov Models

Note * notation:

$$P^*(S_t|S_{t-1}) = \sum_{i=1}^{m} P(S_t|S_{t-1} = m)\, P(S_{t-1} = m)$$

Hidden Markov Models

We will make use of Bayes Rule:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

When Y is observed, this becomes:

$$P(X|Y = y) = \frac{P(Y = y|X)P(X)}{P(Y = y)}$$

Hidden Markov Models

State Estimation, t>0:

$$P(S_t|O_{1:t},S_0) = P(S_t|S_{t-1},O_t)P(S_{t-1}|O_{1:t-1},S_0)$$

Note the recursion:

$$\boldsymbol{P}(\boldsymbol{S_t}|\boldsymbol{O_{1:t}},\boldsymbol{S_0}) = P(S_t|S_{t-1},O_t)\boldsymbol{P}(\boldsymbol{S_{t-1}}|\boldsymbol{O_{1:t-1}},\boldsymbol{S_0})$$

So we can proceed iteratively through, basising our estimation of S_t only on our estimation of S_{t-1} and observation O_t.

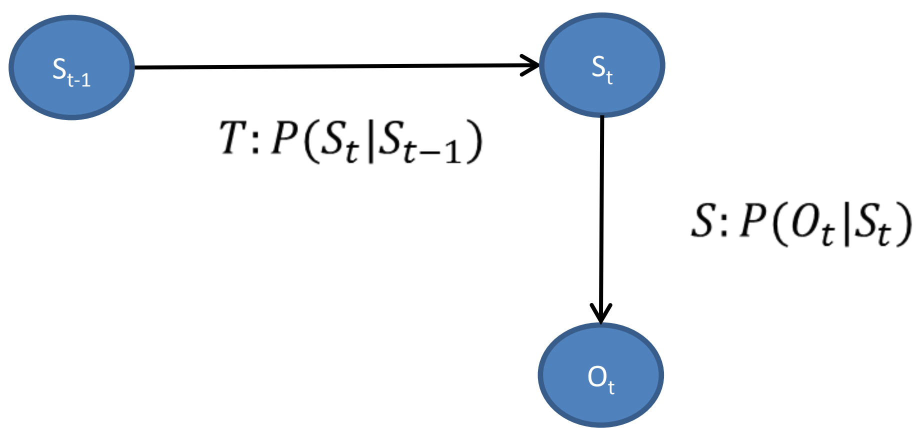
Hidden Markov Models

- State Estimation

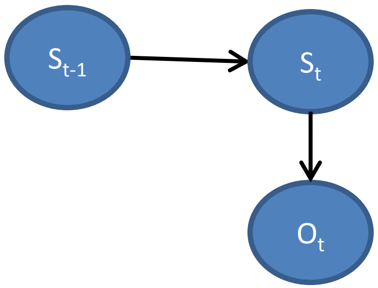
$$P(S_t|S_{t-1}, O_t) = \frac{P(O_t|S_t)P^*(S_t|S_{t-1})}{P(O_t)}$$

$$\propto P(O_t|S_t)P^*(S_t|S_{t-1})$$

Remember: The previous state estimation has <u>all</u> relevant information from the past!



Hidden Markov Models



S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

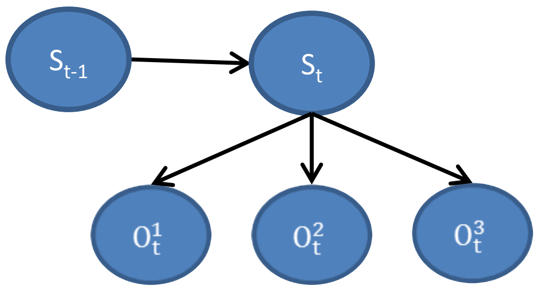
S_t	O_t=T	O_t=F
T	.3	.7
F	.1	.9

$$P(S_t|S_{t-1},O_t) \quad \propto P(O_t|S_t)P^*(S_t|S_{t-1})$$

- *Let our belief regarding* S_0 *be that it is 80% likely* S_0=T.
- *Let us observe* O_1=F.

$$P(S_1|S_0) =< (.8)(.9) + (.2)(.3), (.8)(.1) + (.2)(.7) > = < .78, .22 >$$
$$P(O_1 = F|S_1) = < .7, .9 >$$
$$P(S_1|S_0,O_1 = F) \propto <(.78)(.7),(.22)(.9)>=<.546,.198>$$
$$P(S_1|S_0,O_1 = F) = < \frac{.546}{.546+.198}, \frac{.198}{.546+.198} > \approx <.734, .266 >$$

Hidden Markov Models



S_{t-1}	S_t=T	S_t=F
T	.6	.4
F	.5	.5

S_t	O_t^1
T	$\mathcal{N}(3.5,10)$
F	$\mathcal{N}(5,5)$

S_t	O_t^2
T	$\mathcal{N}(45,100)$
F	$\mathcal{N}(55,225)$

S_t	O_t^3
T	$\mathcal{N}(0,.1)$
F	$\mathcal{N}(0,.5)$

$$P(S_t|S_{t-1},O_t) \propto \rho(O_t^1|S_t)\rho(O_t^2|S_t)\rho(O_t^3|S_t)P^*(S_t|S_{t-1})$$

- *Let our belief regarding* S_0 *be that it is 50% likely* S_0=T.
- *Let us observe* O_1^1=6.103, O_1^2=54.7 and O_1^3=.154

$$P(S_1|S_0) = < (.5)(.6) + (.5)(.5), (.5)(.4) + (.5)(.5) > = < .55, .45 >$$
$$P(O_1^1 = 6.103|S_1) \approx < .089, .158 >$$
$$P(O_1^2 = 54.7|S_1) \approx < .025, .027 >$$
$$P(O_1^3 = .154|S_1) \approx < 1.120, .551 >$$
$$P(S_1|S_0, O_1^1 = 6.103, O_1^2 = 54.7, O_1^3 = .154) \propto <(.55)(.089)(.025)(1.120),(.45)(.158)(.027)(.551)> \approx <.00137,.00106>$$
$$P(S_1|S_0, O_1^1 = 6.103, O_1^2 = 54.7, O_1^3 = .154) \approx < \frac{.00137}{.00137+.00106}, \frac{.00106}{.00137+.00106} > \approx <.564, .436 >$$

Hidden Markov Models: Lab B

- State Estimation
  - Given an initial state, transition and sensor probabilities, we can iteratively calculate the distribution at each subsequent state.
  - We can do this online.

Note that for Lab B

- Vector of 3 observations (as in last example)
- Sparse transition matrix (many impossible transitions).
- Presumably uniform initial state
- NOT real time.

Hidden Markov Models

Smoothing: The Forward-Backward Algorithm

$$P(S_{s\le t}|O_{0:t},S_0) = P(S_{s\le t}|O_{0:s},S_0)P(S_{s\le t}|O_{s+1:t})$$

The Forward Algorithm:

- We have seen how, given an initial state, transition and sensor probabilities, we can iteratively calculate $P(S_s|O_{0:s},S_0)$ for $1 \le s \le t$.

The Backward Algorithm:

- Starting at t, we can iteratively calculate:

$$P(S_s|O_{s+1:t}) \propto P(O_{s+1:t}|S_s)P(S_s)$$

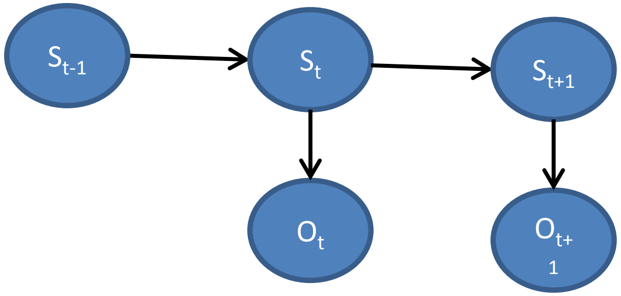
Since the forward algorithm gives us $P(S_s)$, we are interest in:

$$\boldsymbol{P(O_{s+1:t}|S_s)} = \qquad \boldsymbol{P(O_{s+2:t}|S_{s+1})}P(O_{s+1}|S_{s+1})P(S_{s+1}|S_s)$$

Notice the recursion, and we have a base case since:

$$P(O_{t+1:t}|S_t) = P(\emptyset|S_t) = 1$$

Hidden Markov Models



S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.3	.7
F	.1	.9

$$P(S_t|S_{t-1},O_t,O_{t+1}) \propto P(S_t|S_{t-1},O_t)P(O_{t+1}|S_t)$$

- *Let our belief regarding* S_0 *be that it is 80% likely* S_1=T.
- *Let us observe* O_1=F, O_2=F.
- From a previous example, we have $P(S_1|S_0,O_1=F) \approx <.734,.266>$

$$P(O_2=F|S_1) = P(O_2=F|S_2)P(S_2|S_1)$$

$P(S_2|S_1) \approx <(.734)(.9)+(.266)(.3),(.734)(.1)+(.266)(.7)> = <.74,.26>$

$P(O_2=F|S_2) = <.7,.9>$

$P(S_1|S_0,O_1=F,O_2=F) \propto <(.74)(.7),(.26)(.9)>=<.518,.234>$

$P(S_1|S_0,O_1=F,O_2=F) \approx <\frac{.518}{.518+.234},\frac{.234}{.518+.234}> \approx <.689,.311>$

So O_2 gives additional reason to think S_1 is false.

Hidden Markov Models

Most Probable Path: Viterbi Algorithm

- Similar to Dynamic Programming for Path Finding. Differences:
  – Multiplicative instead of additive accumulation function.
  – Normalize over observations at each step.

Hidden Markov Models

Let $\varphi(n, v) = P(S_{0:n} = s_{0:n}^{v} | O_{1:n} = o_{1:n})$
Let $s_{0:n}^{v}$ be the most probable sequence of states for $0 \leq t \leq n$ where $S_n = v$.

We have:
$s_{0:n+1}^{w} =$
$\text{argmax}_{v,w}\, \varphi(n, v)\, P(S_{n+1} = w | S_n = v) P(O_{n+1} = o_{n+1} | S_{n+1} = w)$
$\varphi(n+1, w) \propto \varphi(n, v)\, P(S_{n+1} = w | S_n = v) P(O_{n+1} = o_{n+1} | S_{n+1} = w)$

And our base cases are:

- $s_{0:n}^{v} =< v >$
- $\varphi(0, v) = P(S_{0:0} = s_{0:0}^{v} | \emptyset) = P(S_0 = v)$.

Hidden Markov Models

We proceed iteratively:

Given our base cases, for each time, $n > 0$, for each state value, v, find and store $s_{0:n}^{v}$ and $\varphi(n, v)$.

Our most probable sequece of states is given by $s_{0:t}^{v}$, where:

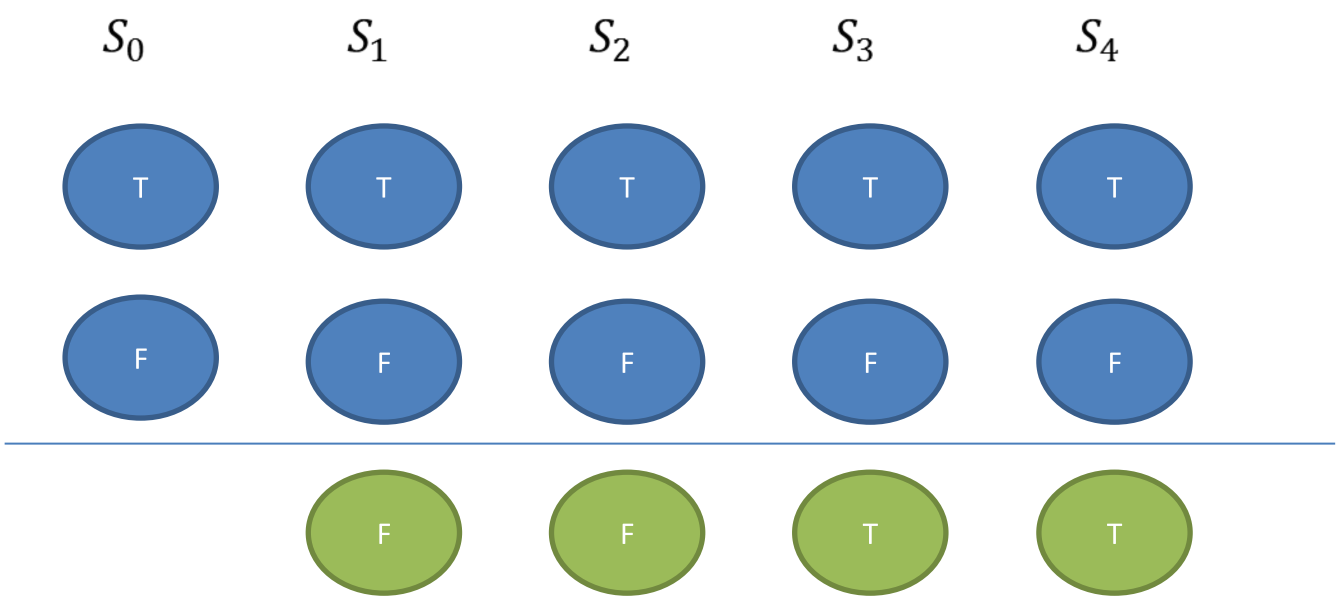
- v maximises $\varphi(t, v)$
- t is the last time slice.

Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

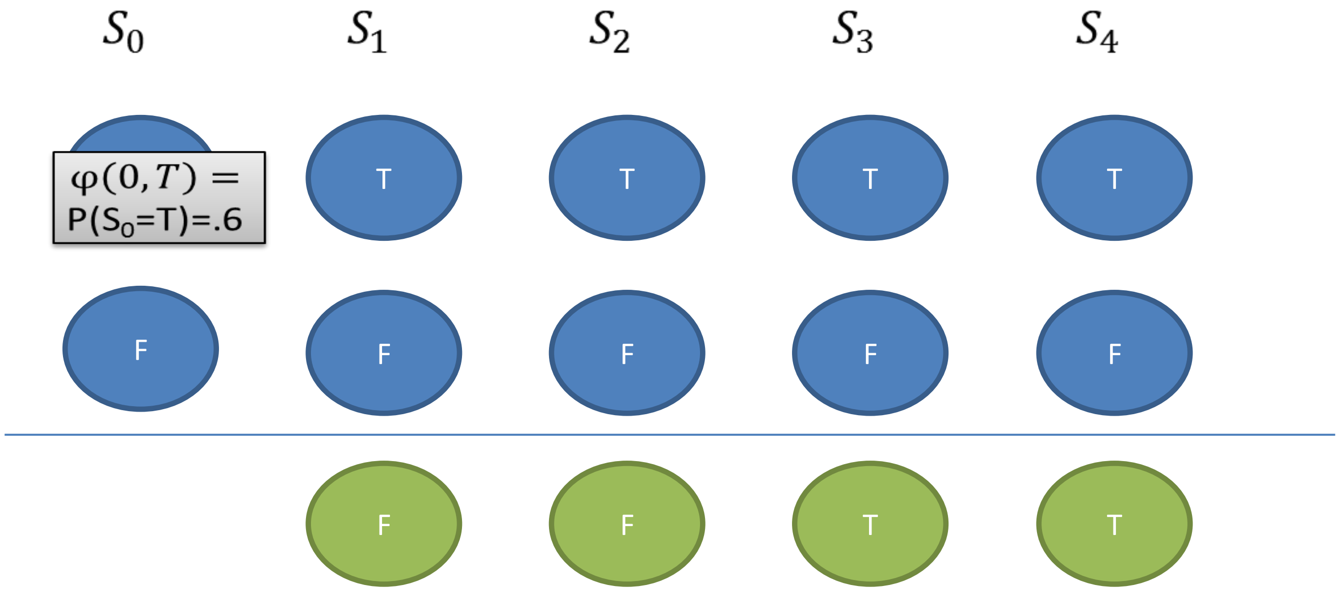


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

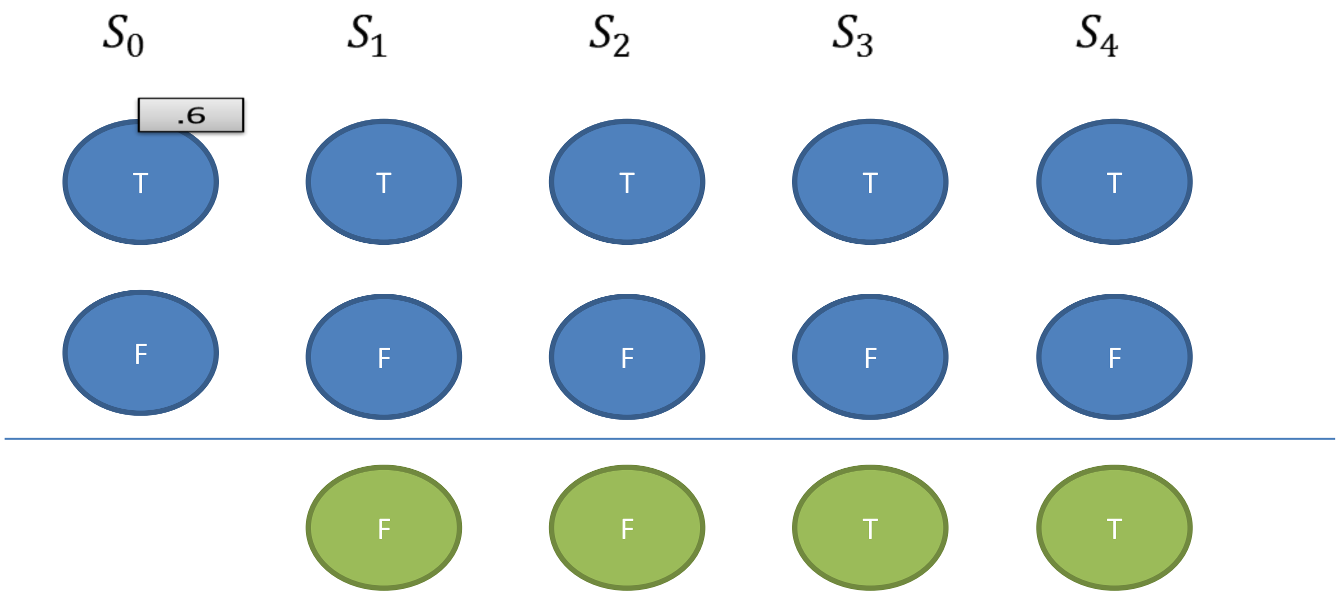


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

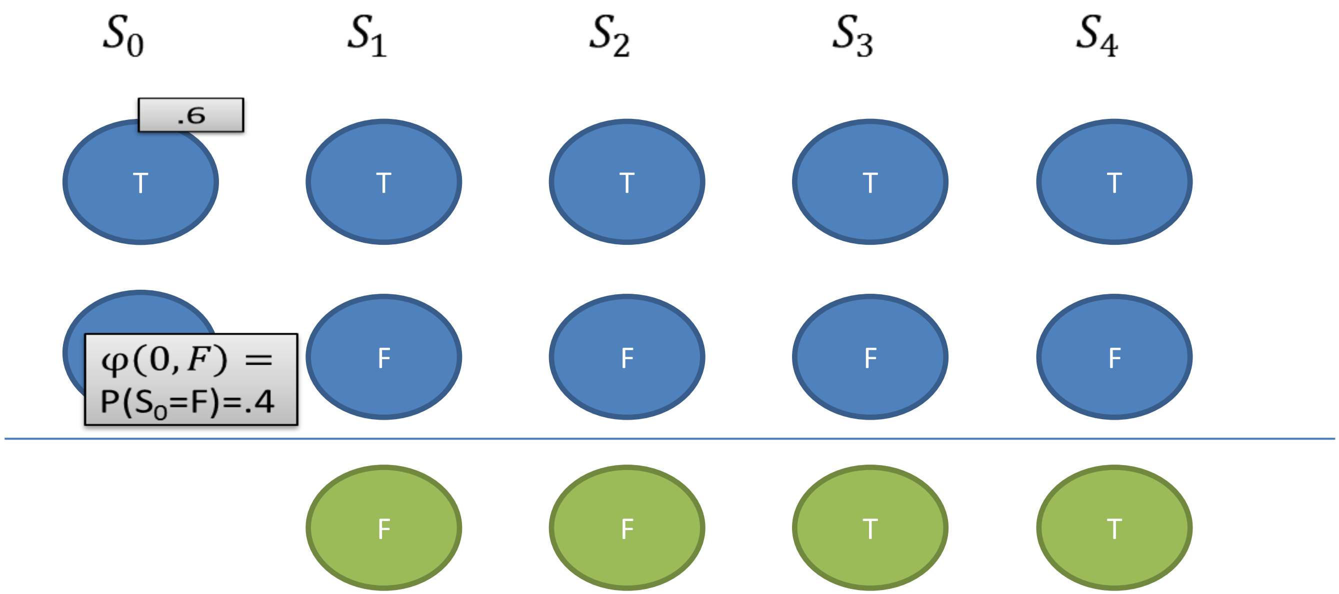


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

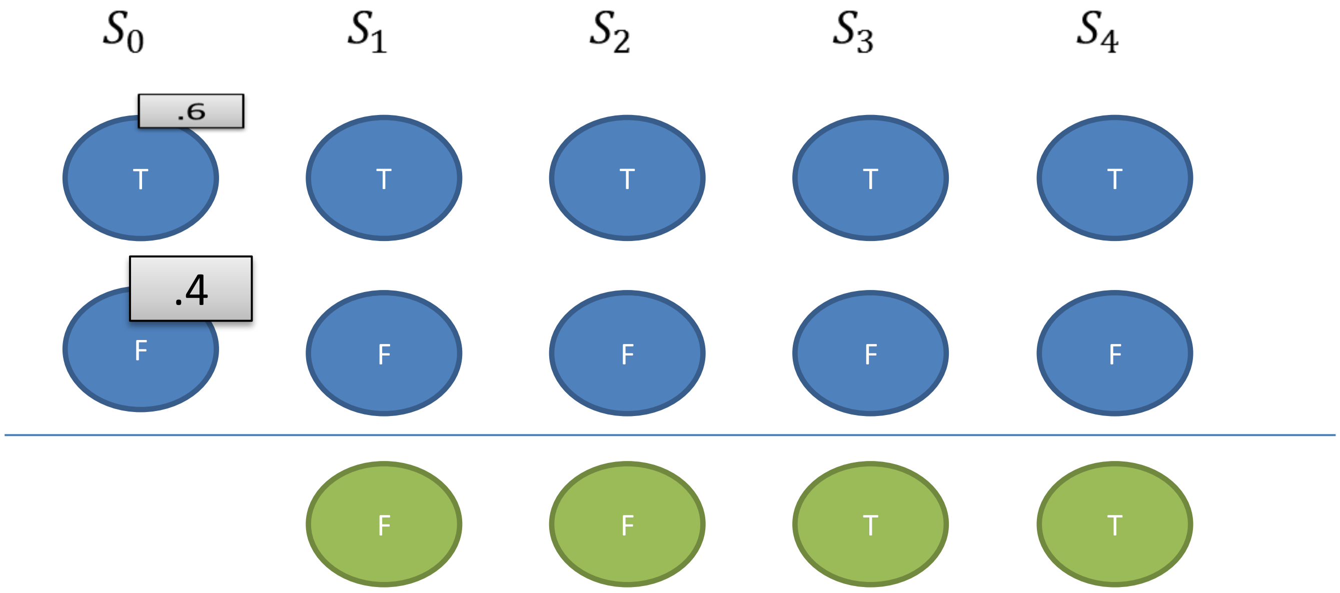


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

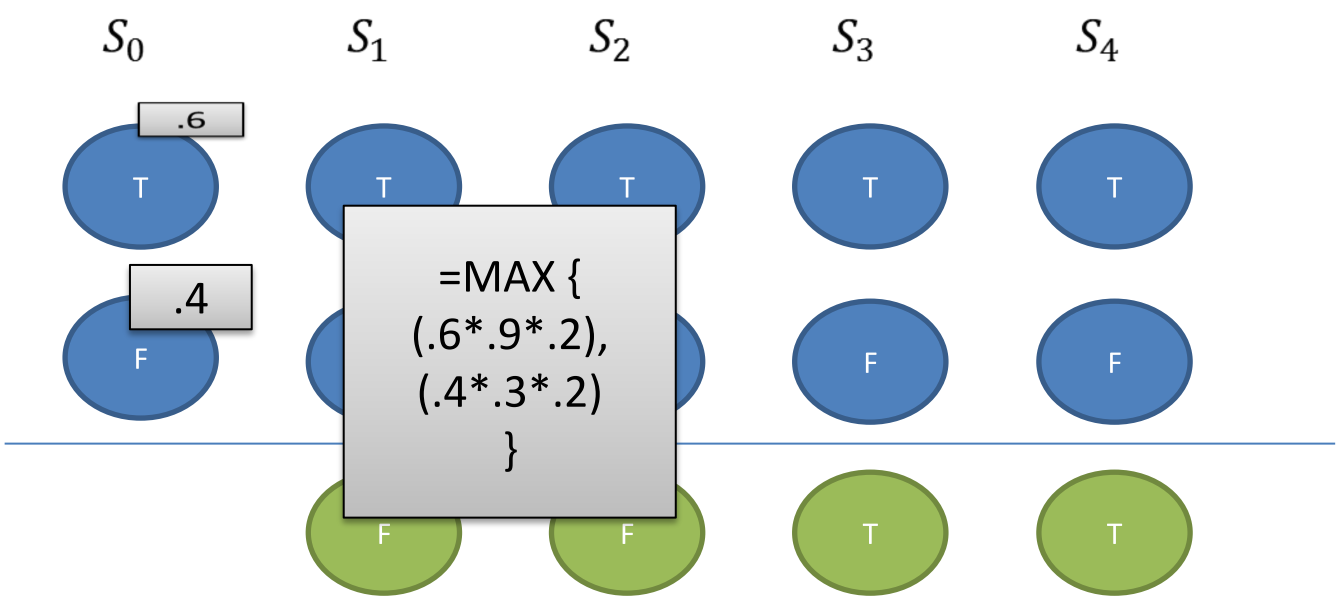


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

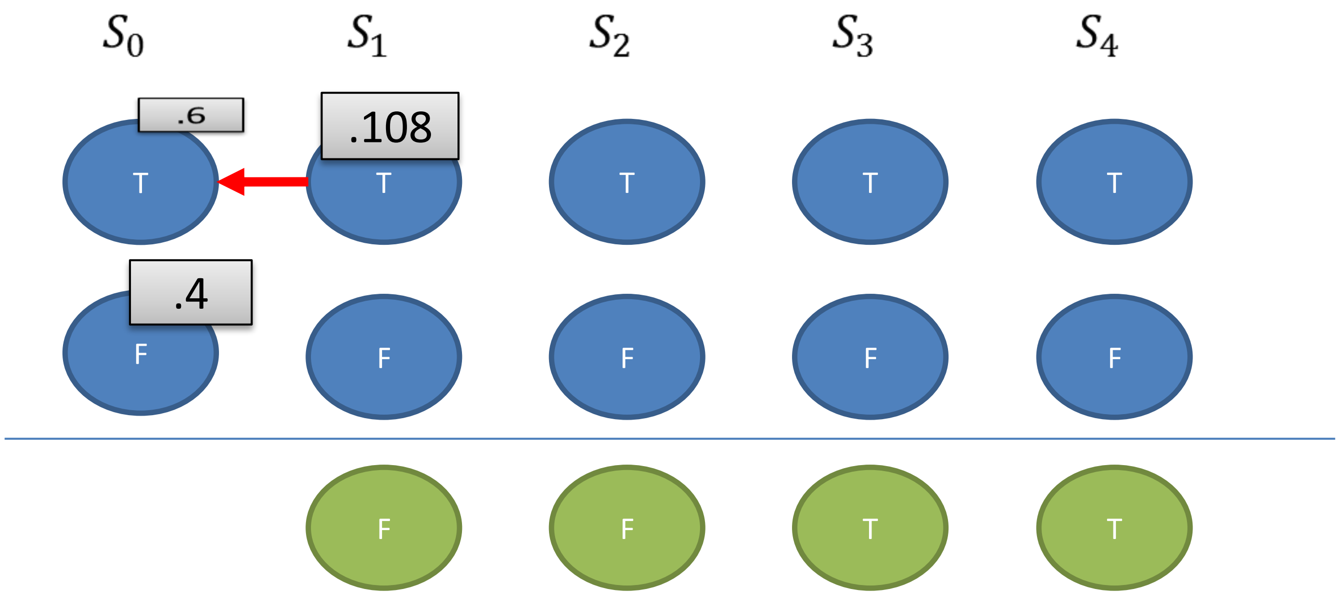


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

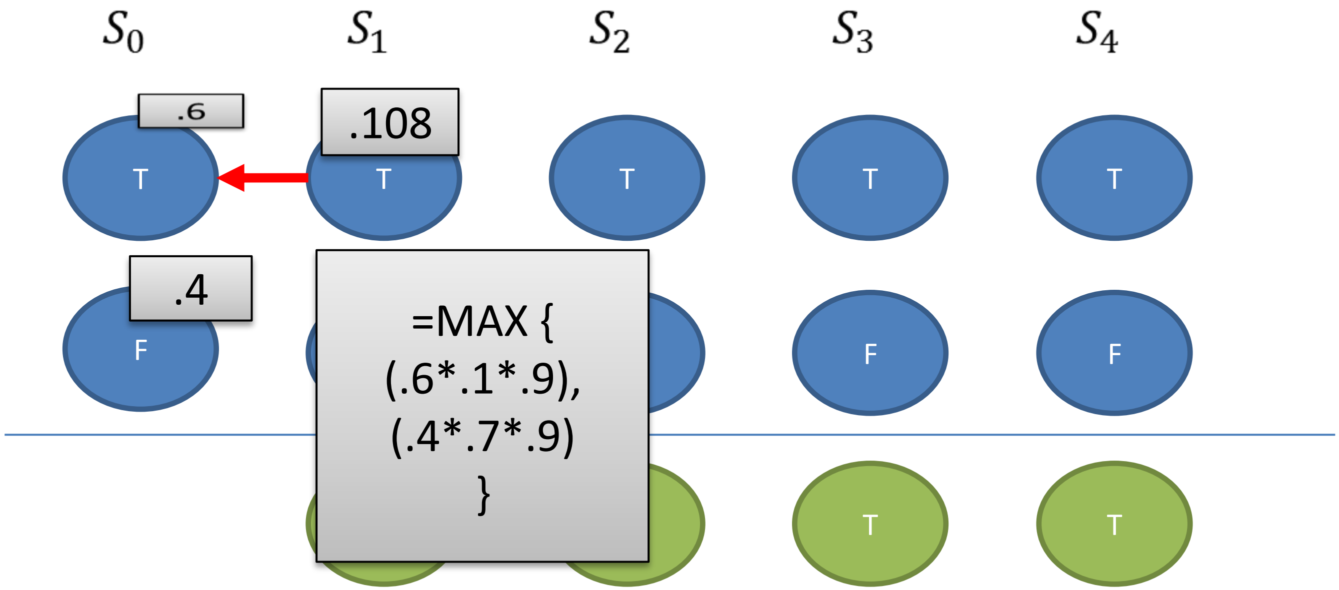


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

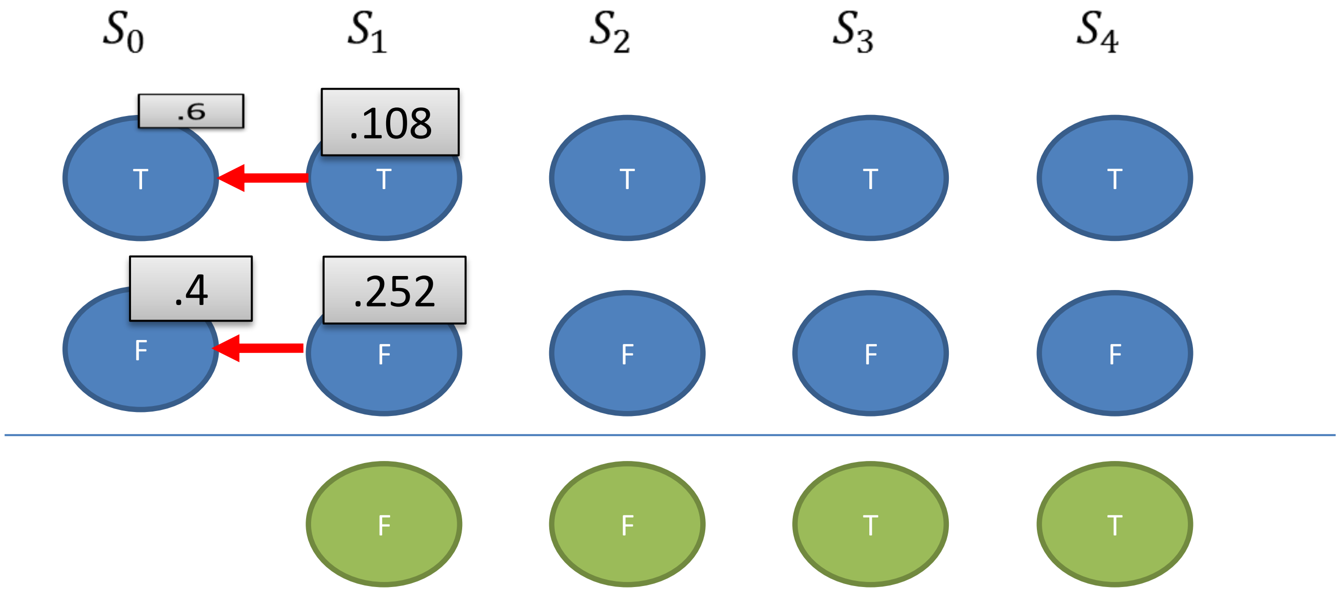


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

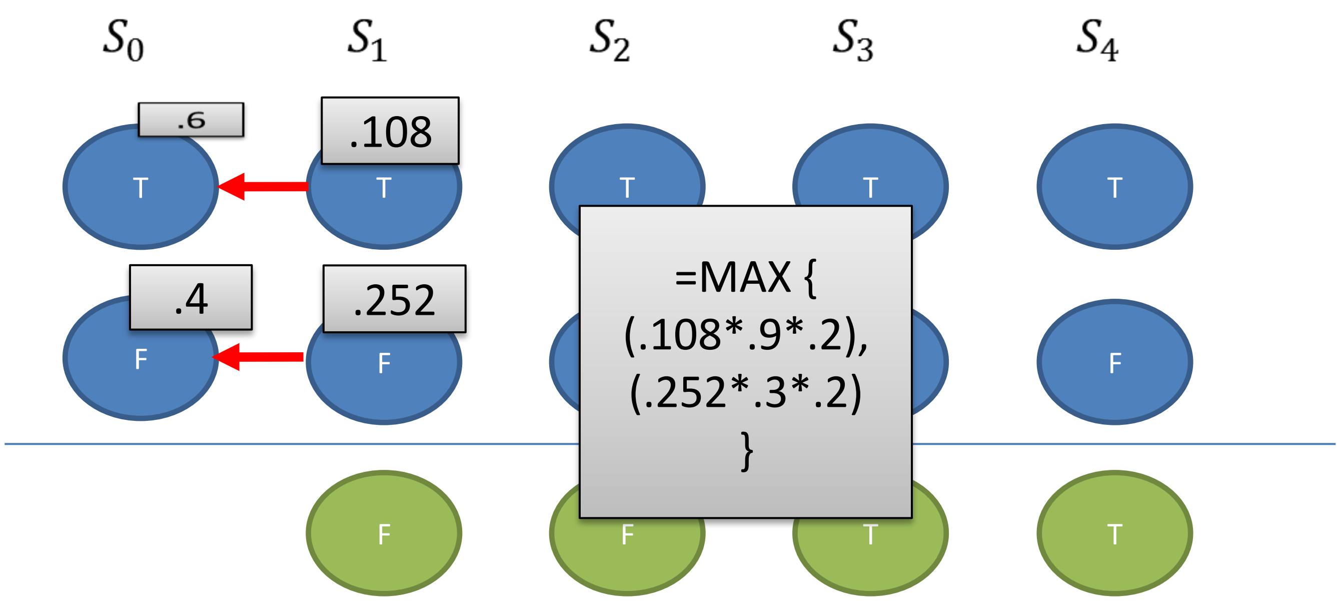


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

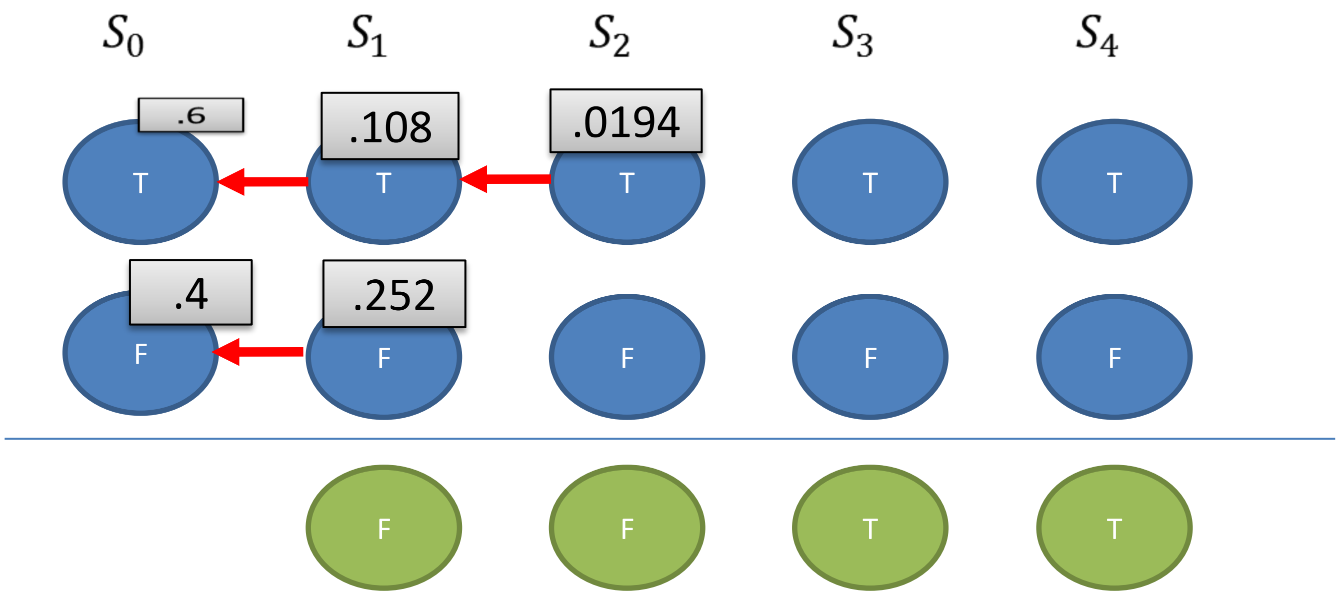


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

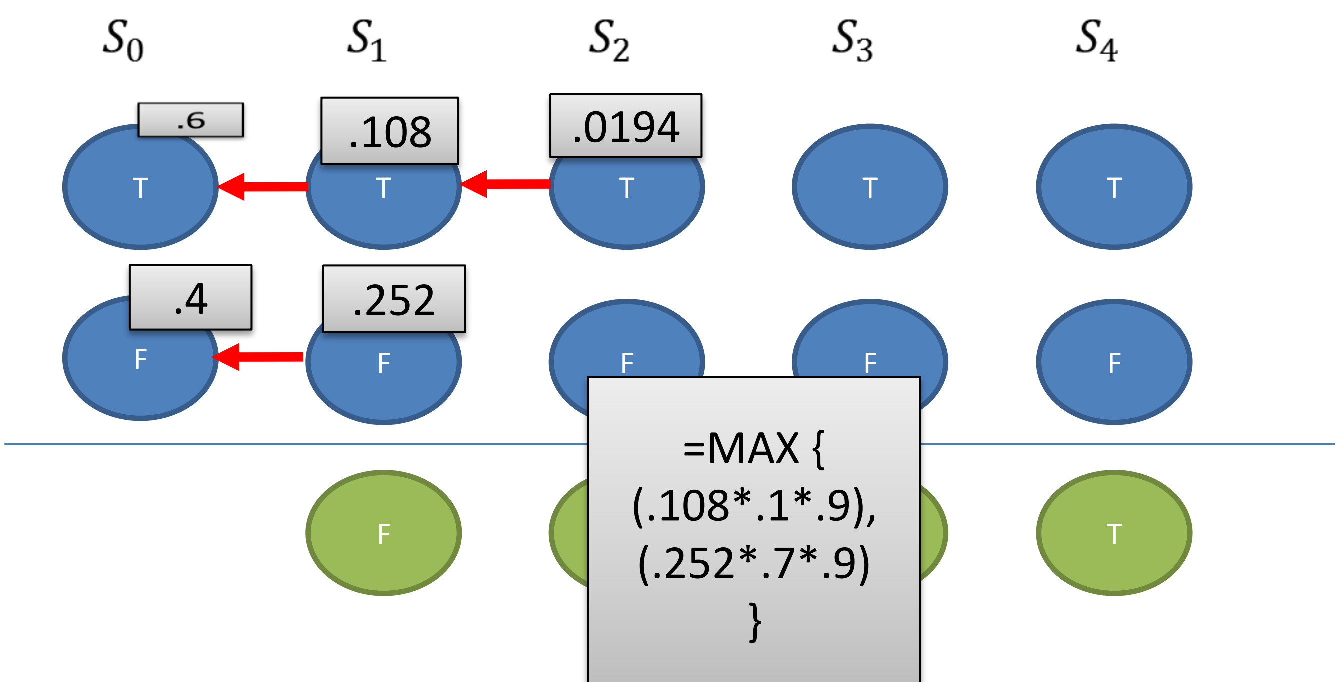


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

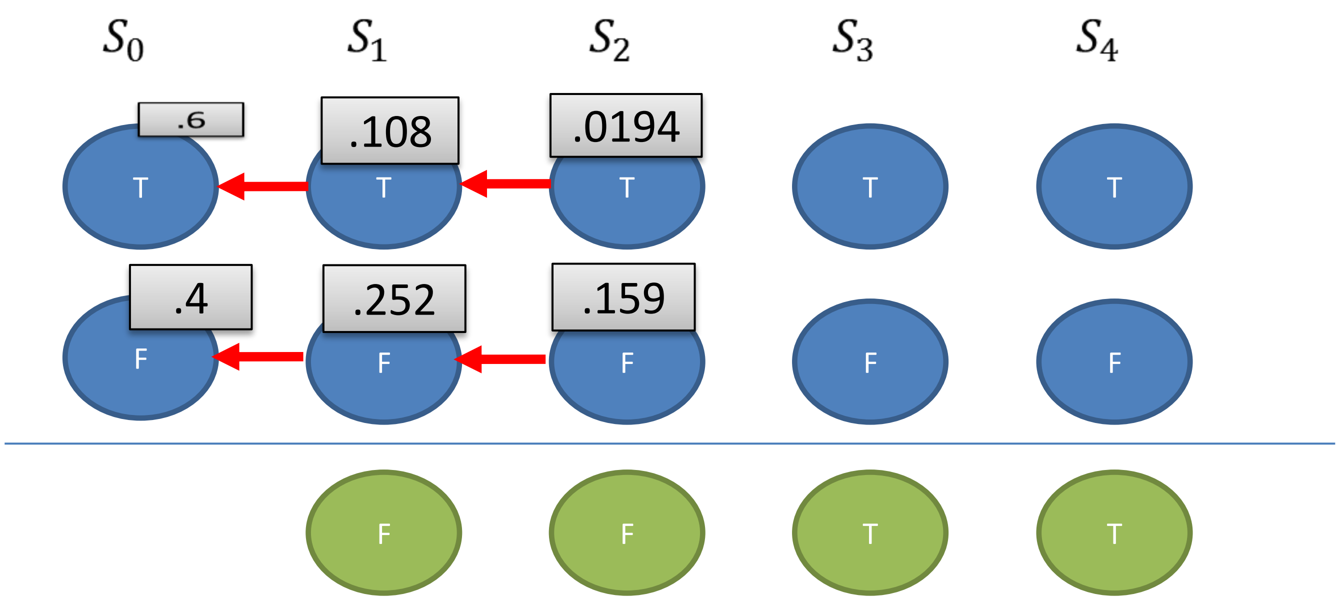


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

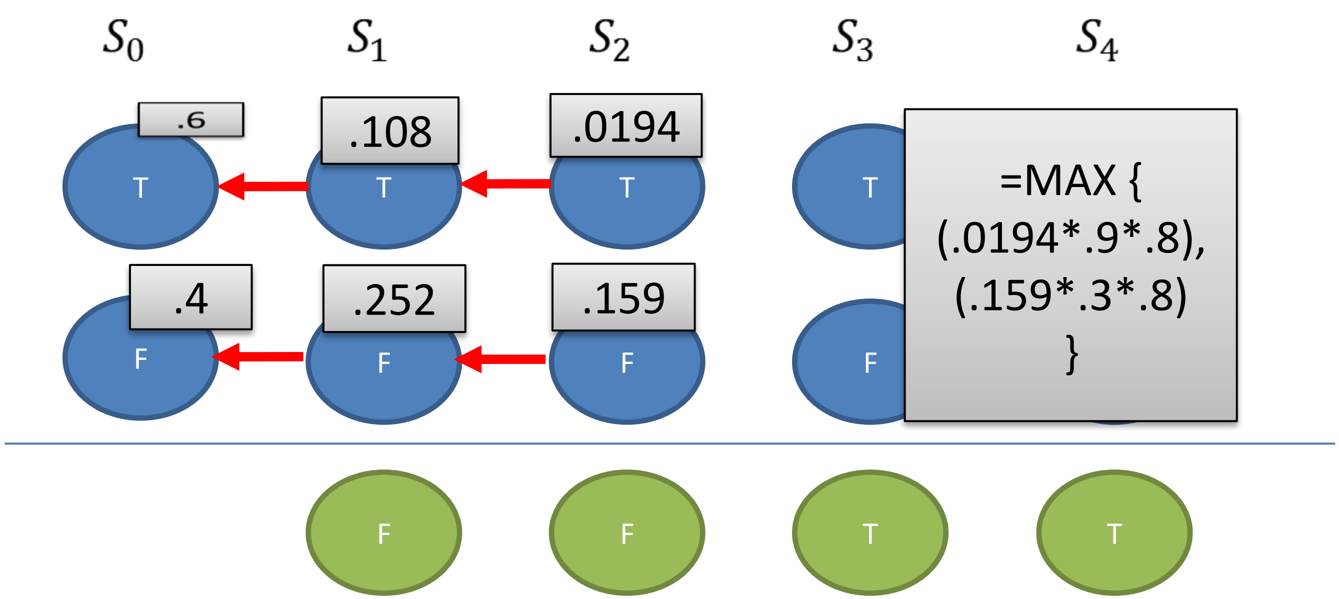


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

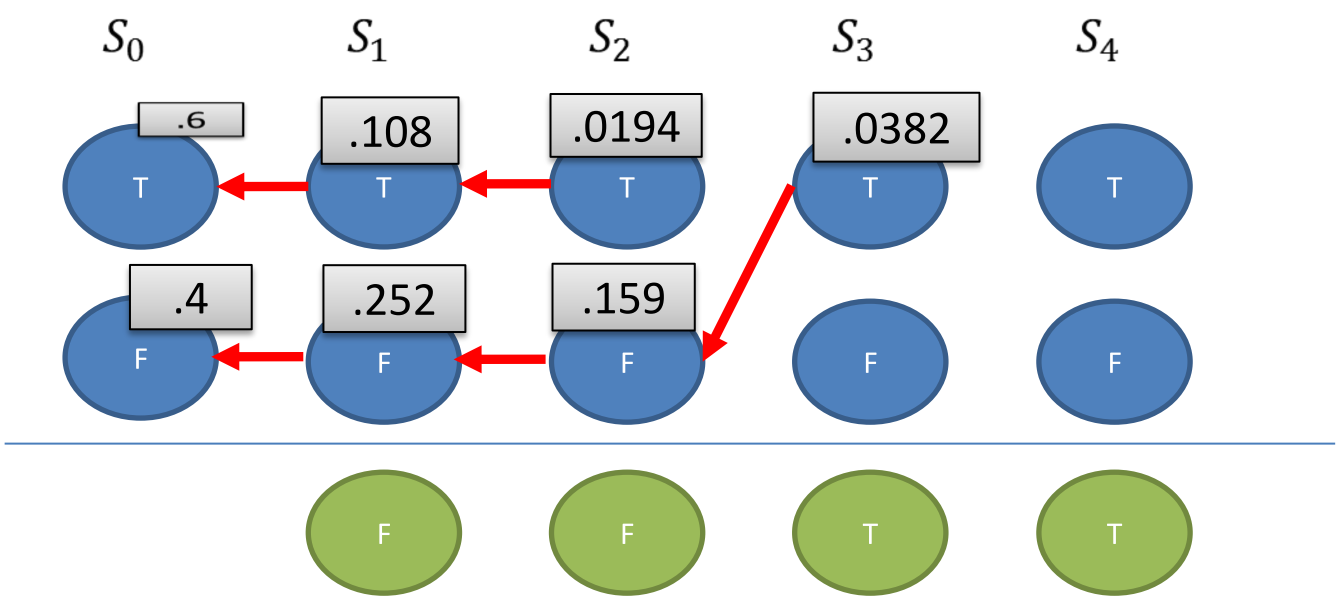


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

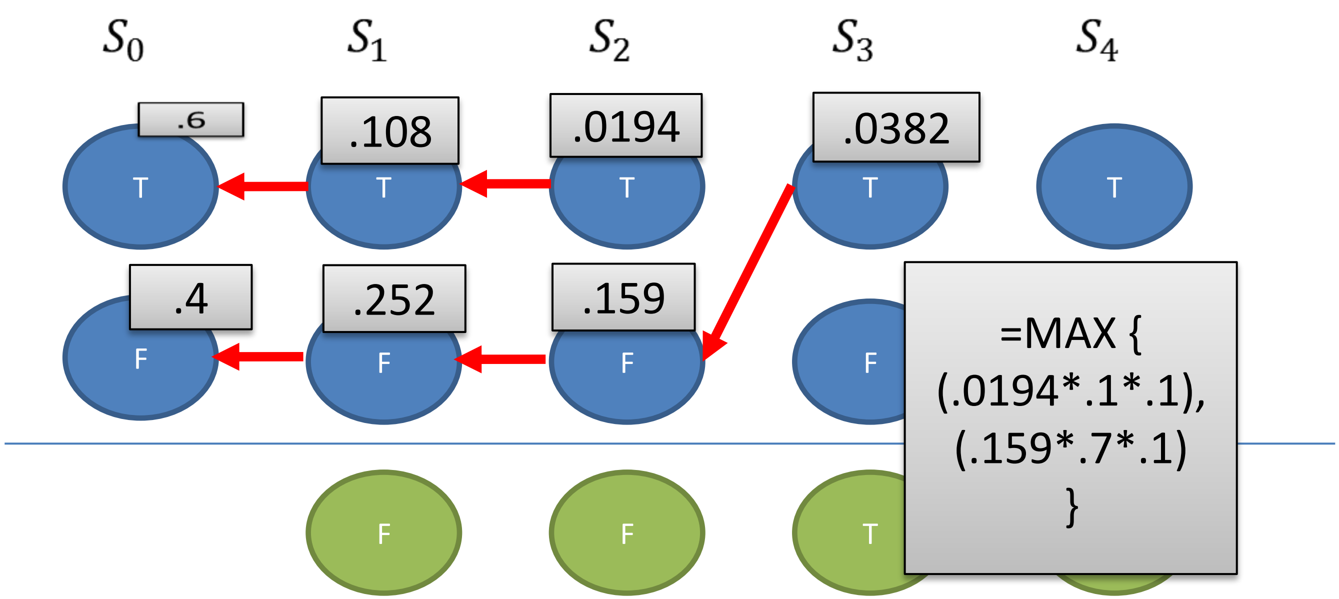


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

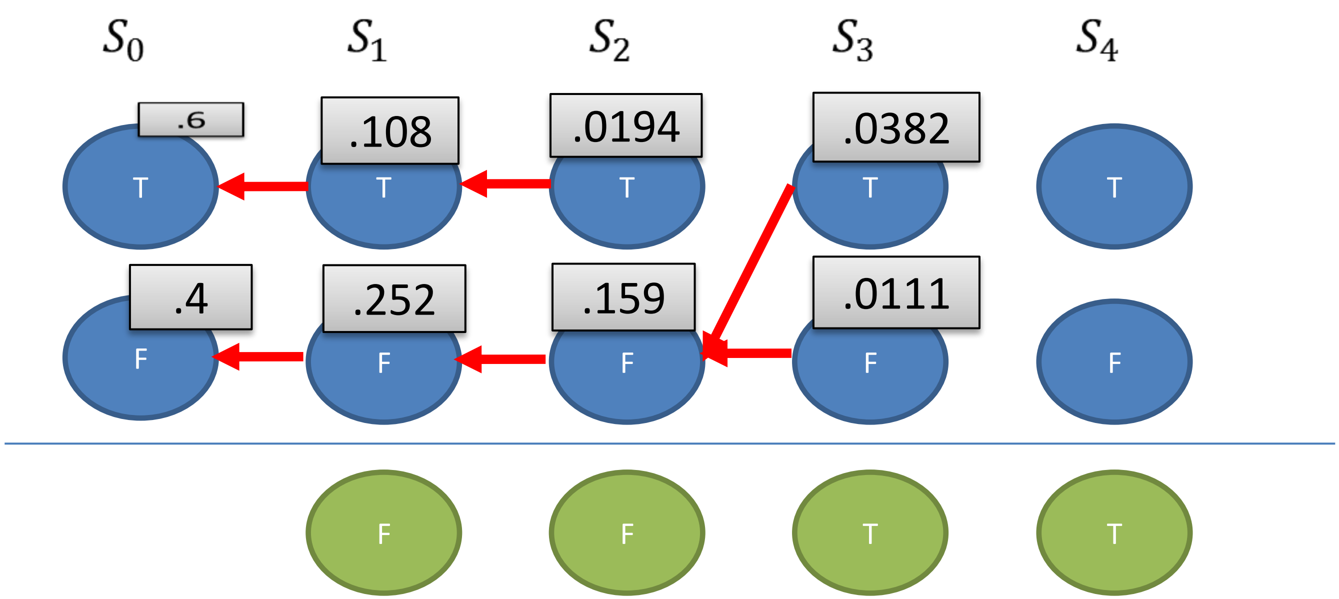


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

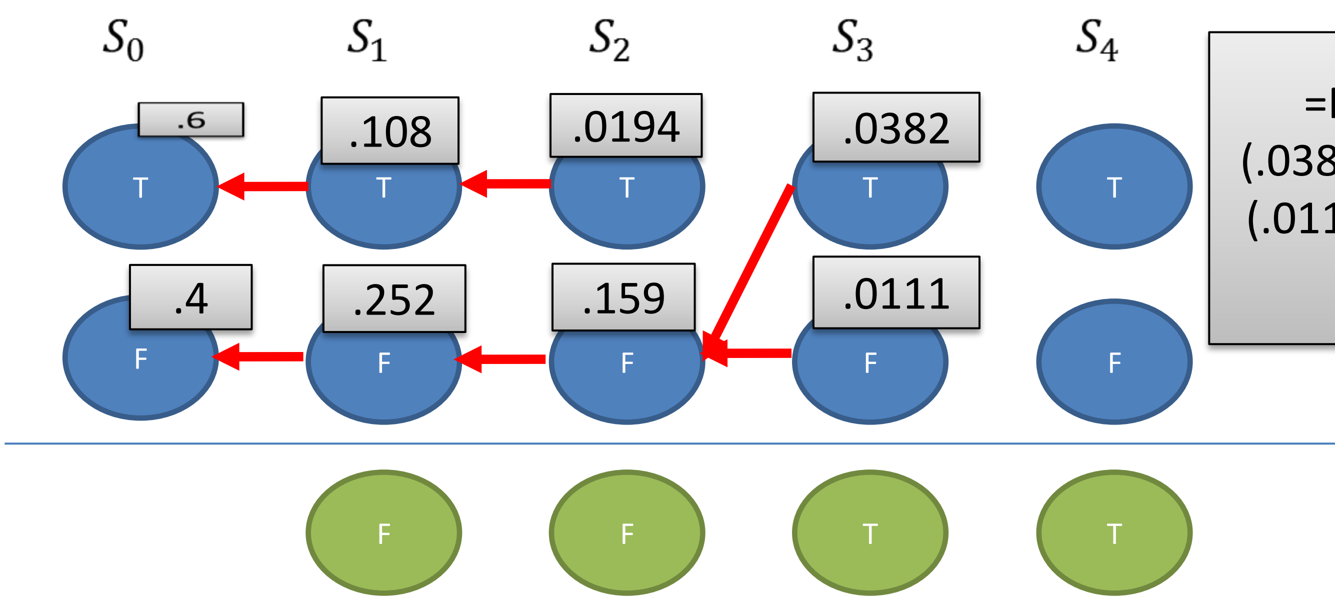


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9



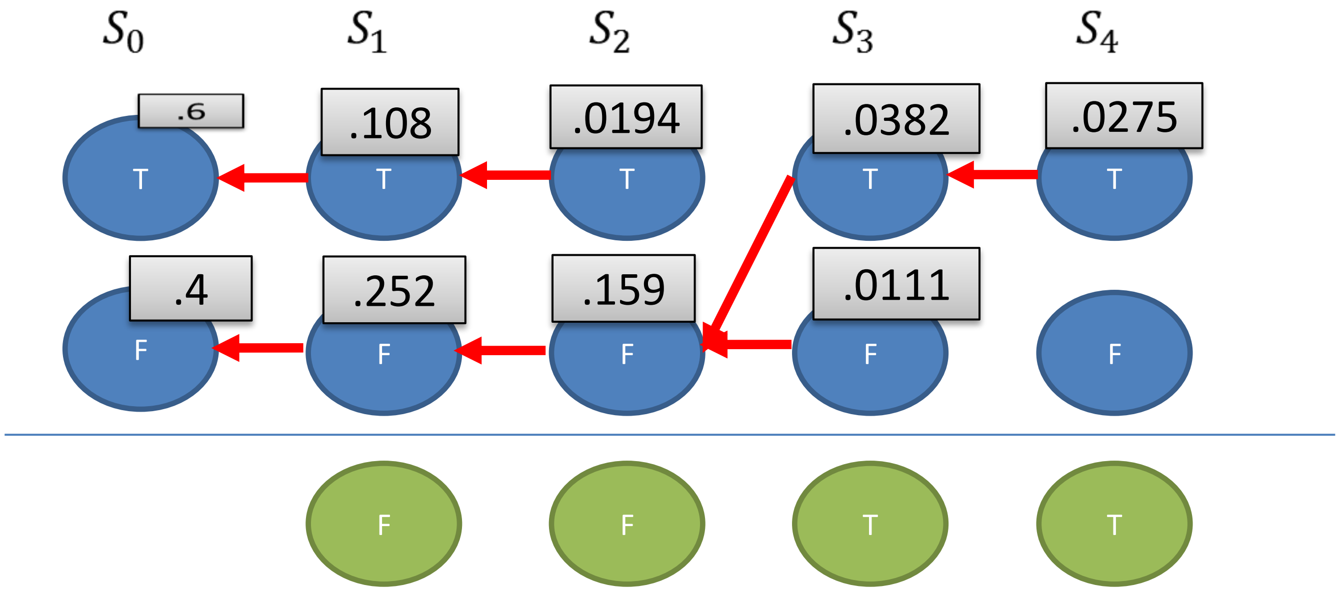
=MAX {
(.0382*.9*.8),
(.0111*.3*.8)
}

Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

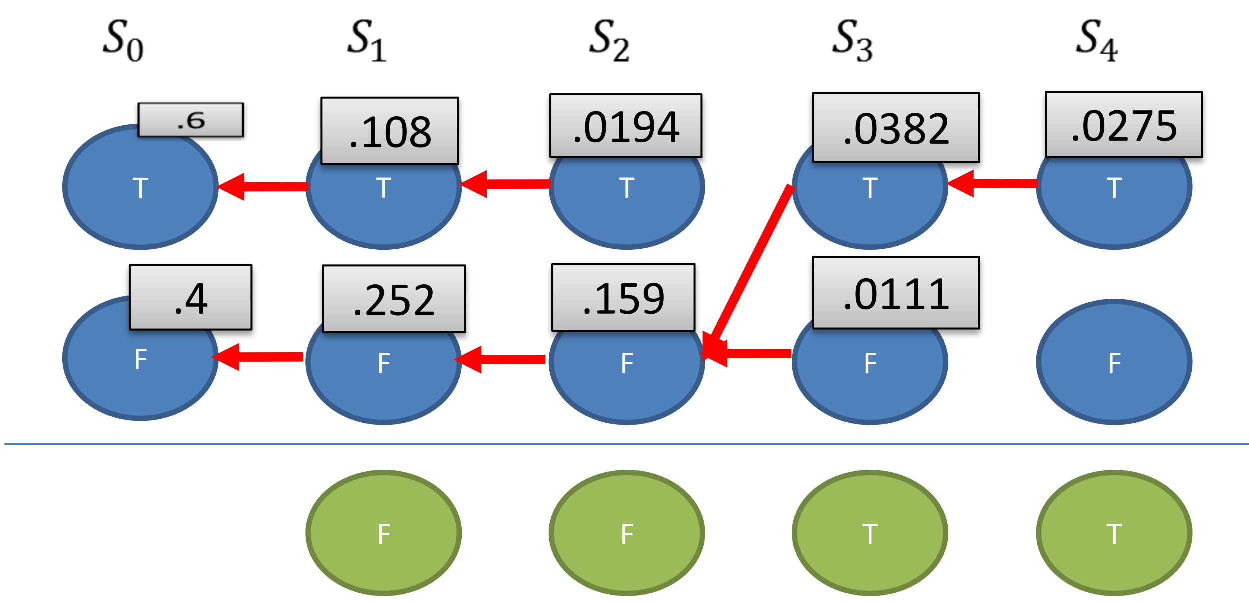


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9



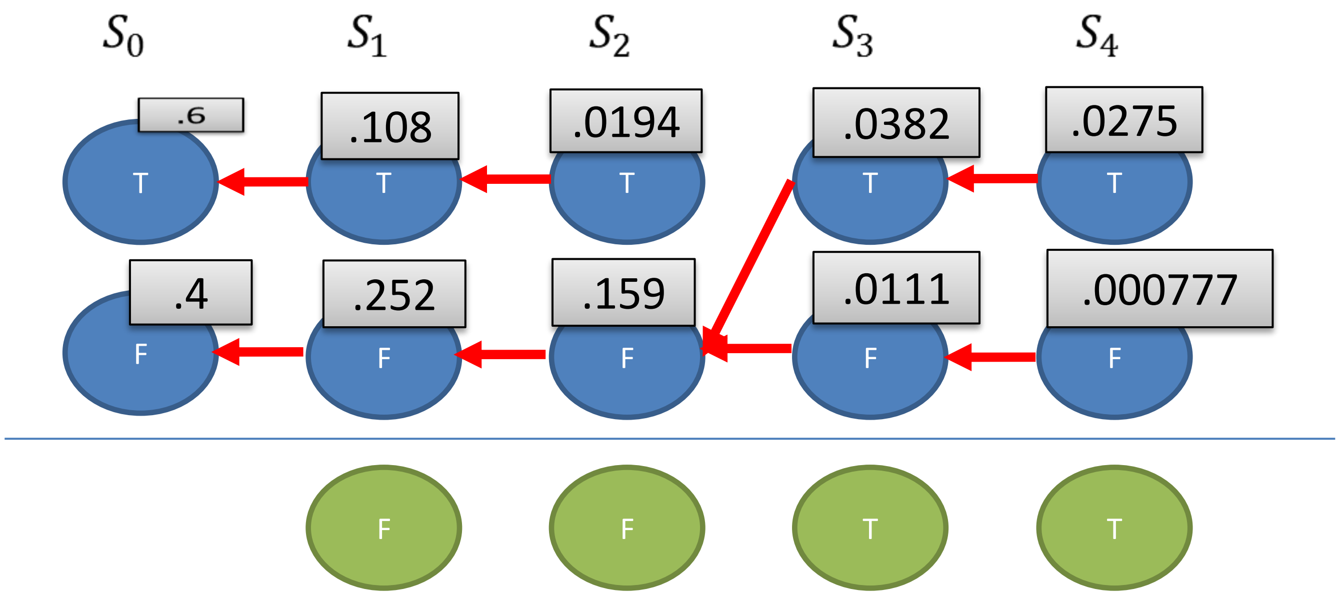
=MAX { (.0382*.1*.1), (.0111*.7*.1) }

Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9

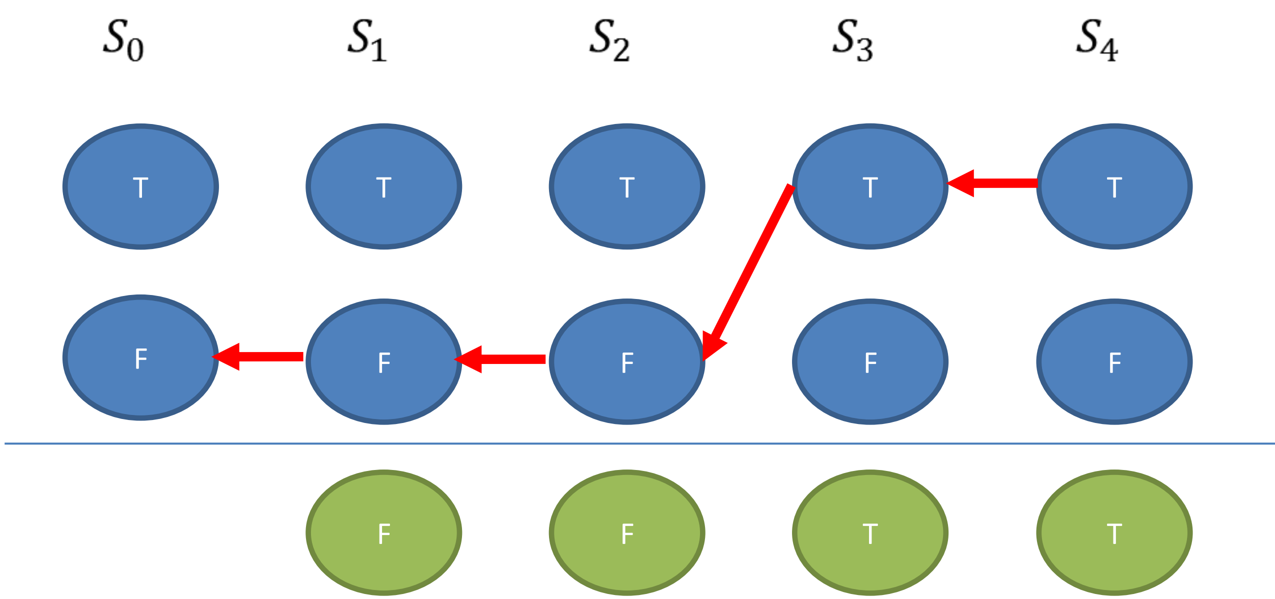


Hidden Markov Models

S_0=T	S_0=F
.6	.4

S_{t-1}	S_t=T	S_t=F
T	.9	.1
F	.3	.7

S_t	O_t=T	O_t=F
T	.8	.2
F	.1	.9



.0275 is **NOT** the probability of the path, since we didn't normalize observation likelihoods as we went…