Guiding Question

- What does "Markov" mean in "Markov Decision Process"?

Stochastic Process

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$$x_0 = 0$$

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$$x_0 = 0 \qquad\qquad x_{t+1} = x_t + v_t$$

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$$x_0 = 0 \qquad x_{t+1} = x_t + v_t$$

$$v_t \sim \mathcal{U}(\{0, 1\}) \text{ (i.i.d.)}$$

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$x_0 = 0$ $\quad x_{t+1} = x_t + v_t$

$v_t \sim \mathcal{U}(\{0, 1\})$ (i.i.d.)

Shorthand:

$x' = x + v$

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

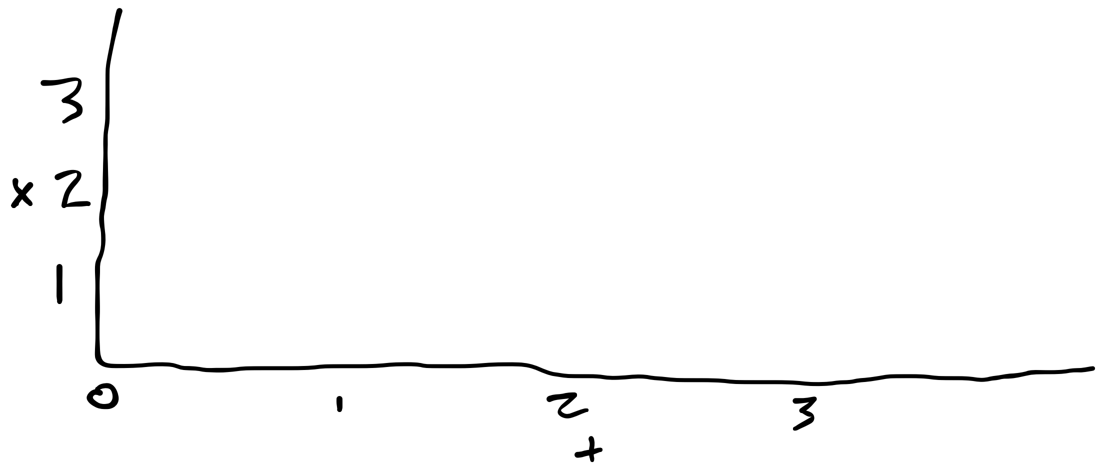
Example:

$x_0 = 0$ $x_{t+1} = x_t + v_t$

$v_t \sim \mathcal{U}(\{0, 1\})$ (i.i.d.)

Shorthand:

$x' = x + v$



Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

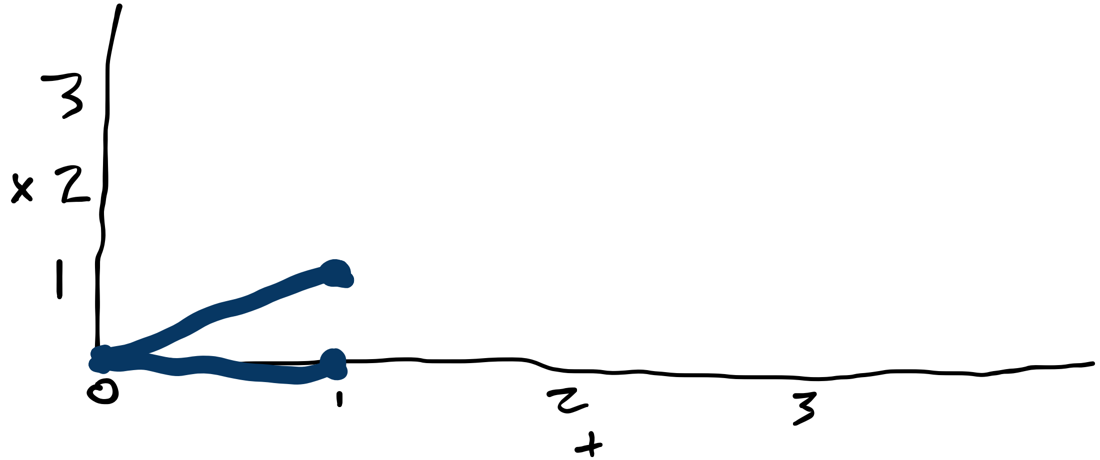
Example:

$$x_0 = 0 \qquad x_{t+1} = x_t + v_t$$

$$v_t \sim \mathcal{U}(\{0, 1\}) \text{ (i.i.d.)}$$

Shorthand:

$$x' = x + v$$



Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

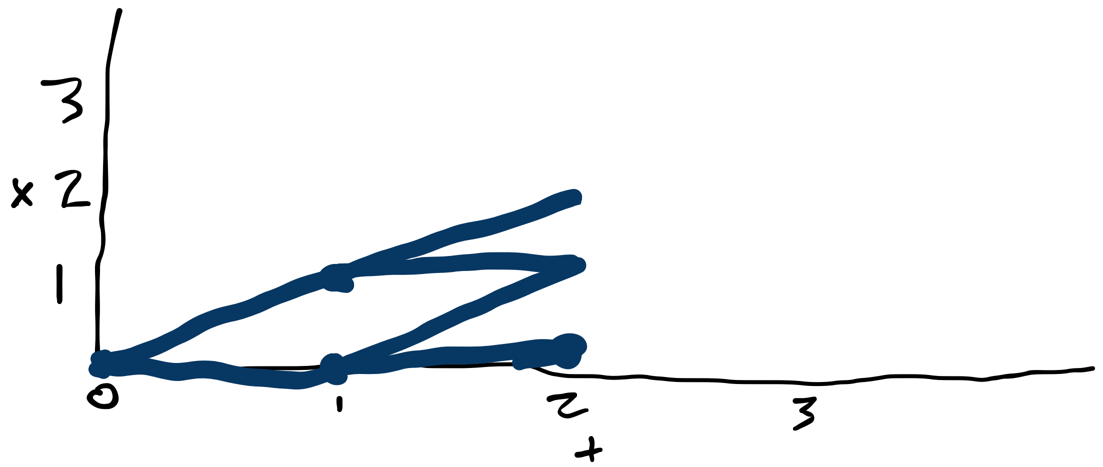
$x_0 = 0$

$$x_{t+1} = x_t + v_t$$

$$v_t \sim \mathcal{U}(\{0, 1\}) \text{ (i.i.d.)}$$

Shorthand:

$$x' = x + v$$



Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

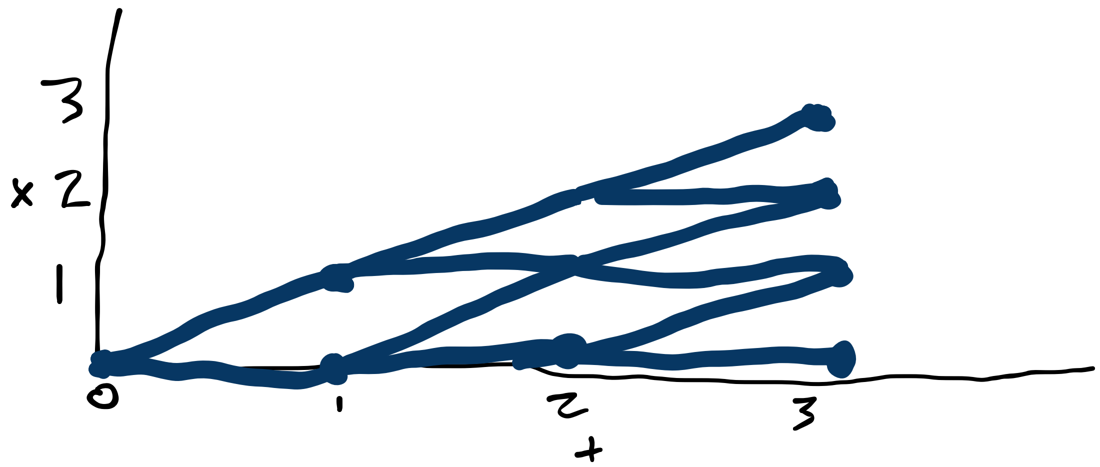
$x_0 = 0$

$x_{t+1} = x_t + v_t$

$v_t \sim \mathcal{U}(\{0,1\})$ (i.i.d.)

Shorthand:

$x' = x + v$



Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

Conditional

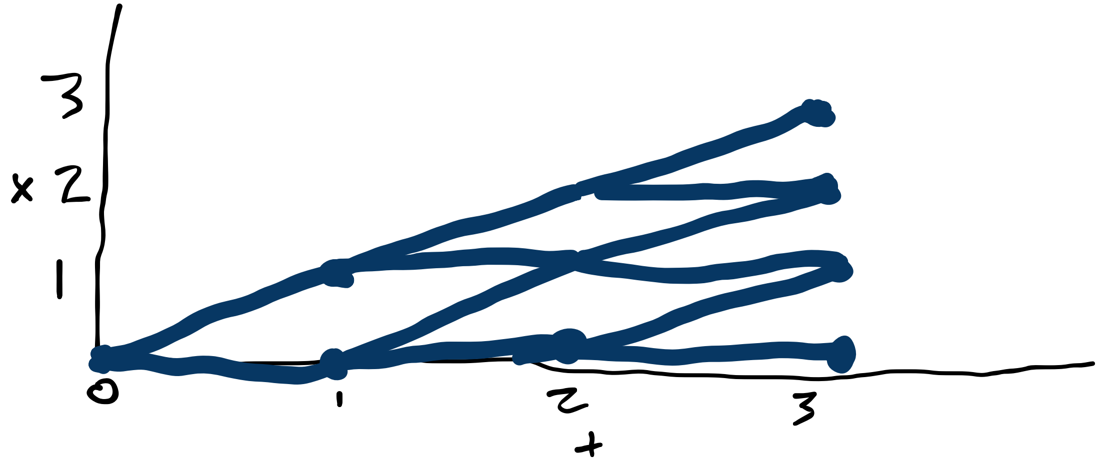
$x_0 = 0$

$x_{t+1} = x_t + v_t$

$v_t \sim \mathcal{U}(\{0, 1\})$ (i.i.d.)

Shorthand:

$x' = x + v$



Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$$x_0 = 0 \qquad x_{t+1} = x_t + v_t$$

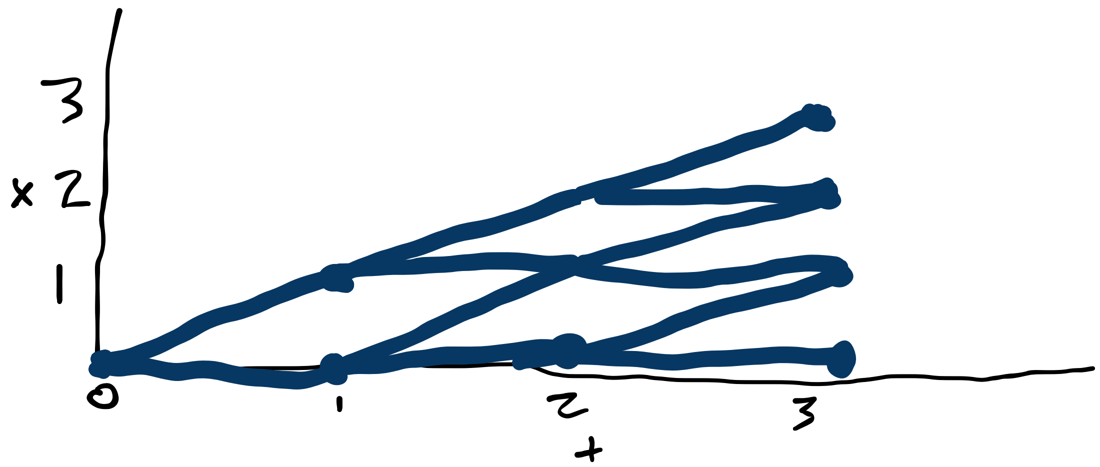
$$v_t \sim \mathcal{U}(\{0, 1\}) \text{ (i.i.d.)}$$

Shorthand:

$$x' = x + v$$

Conditional

x_{t+1}	$P(x_{t+1} \mid x_t)$
x_t	0.5
$x_t + 1$	0.5



Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$x_0 = 0$

$$x_{t+1} = x_t + v_t$$

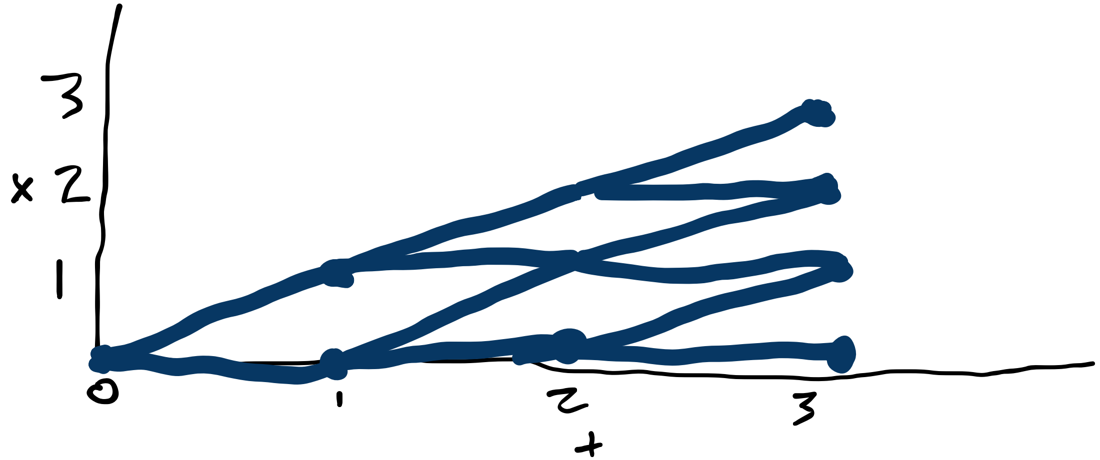
$v_t \sim \mathcal{U}(\{0, 1\})$ (i.i.d.)

Shorthand:

$$x' = x + v$$

Conditional

x_{t+1}	$P(x_{t+1} \mid x_t)$
x_t	0.5
$x_t + 1$	0.5



Joint

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$x_0 = 0$

$x_{t+1} = x_t + v_t$

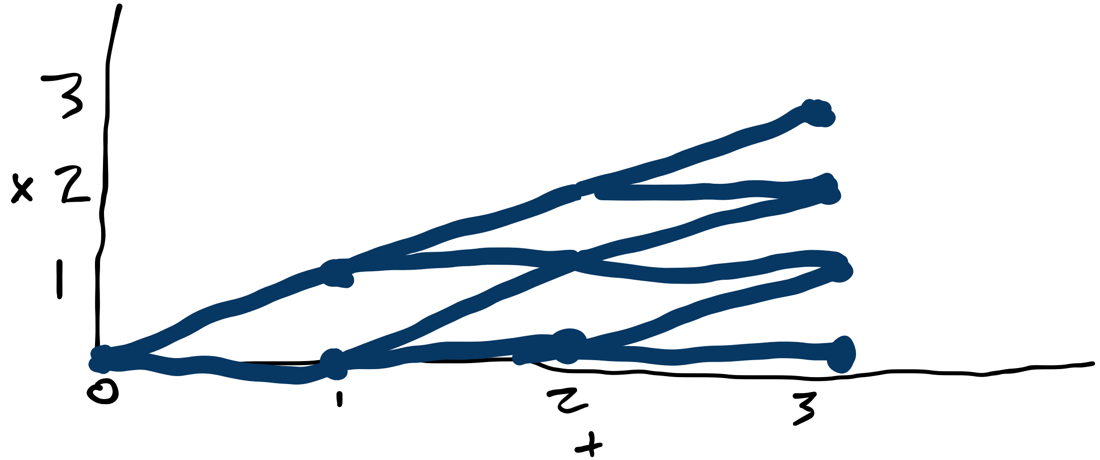
$v_t \sim \mathcal{U}(\{0, 1\})$ (i.i.d.)

Shorthand:

$x' = x + v$

Conditional

x_{t+1}	$P(x_{t+1} \mid x_t)$
x_t	0.5
$x_t + 1$	0.5



Joint

x0	x1	x2	P(x1, x2, x3)
0	0	0	0.25
0	0	1	0.25
0	1	1	0.25
0	1	2	0.25

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$$x_0 = 0 \qquad x_{t+1} = x_t + v_t$$

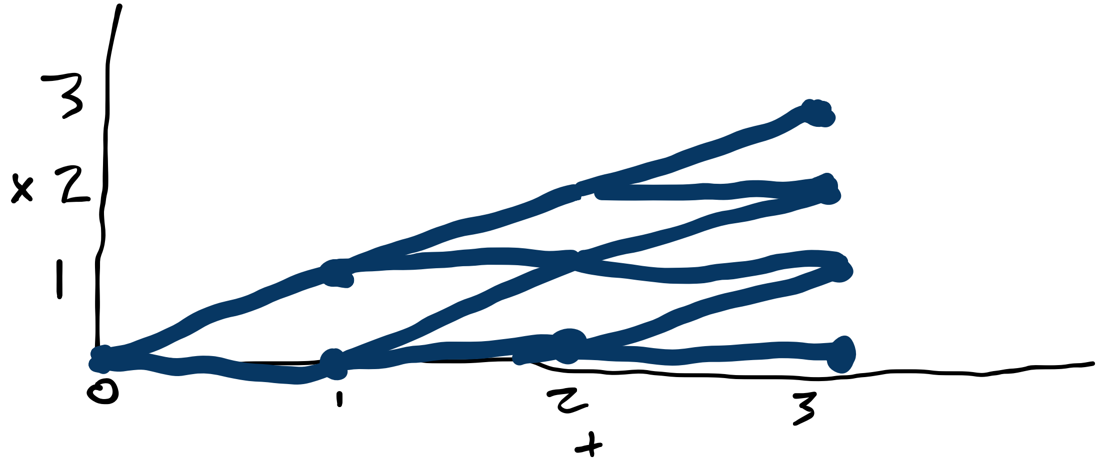
$$v_t \sim \mathcal{U}(\{0, 1\}) \text{ (i.i.d.)}$$

Shorthand:

$$x' = x + v$$

Conditional

x_{t+1}	$P(x_{t+1} \mid x_t)$
x_t	0.5
x_t+1	0.5



Joint

x0	x1	x2	P(x1, x2, x3)
0	0	0	0.25
0	0	1	0.25
0	1	1	0.25
0	1	2	0.25

Marginal

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$x_0 = 0$

$x_{t+1} = x_t + v_t$

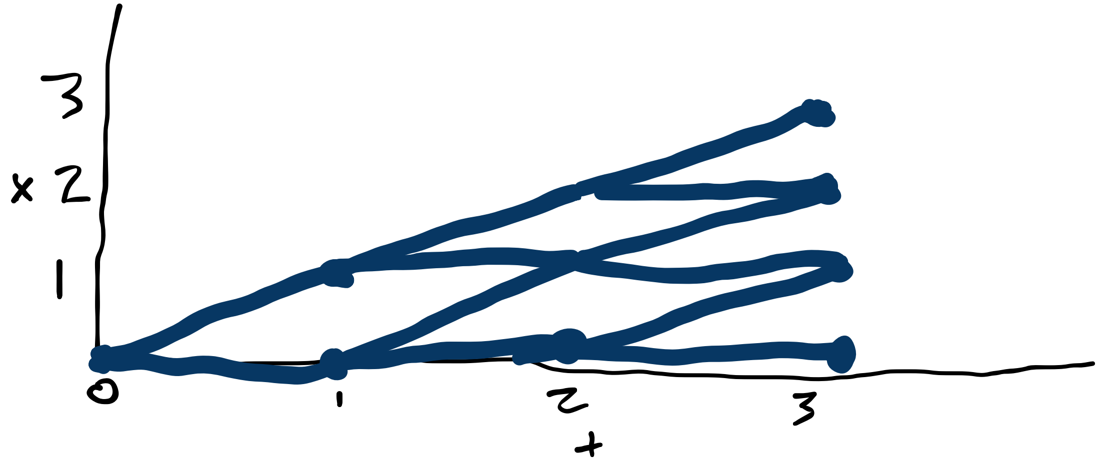
$v_t \sim \mathcal{U}(\{0, 1\})$ (i.i.d.)

Shorthand:

$x' = x + v$

Conditional

x_{t+1}	$P(x_{t+1} \mid x_t)$
x_t	0.5
$x_t + 1$	0.5



Joint

x0	x1	x2	P(x1, x2, x3)
0	0	0	0.25
0	0	1	0.25
0	1	1	0.25
0	1	2	0.25

Marginal

x_2	
0	0.5
1	0.5

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

Example:

$$x_0 = 0 \qquad x_{t+1} = x_t + v_t$$

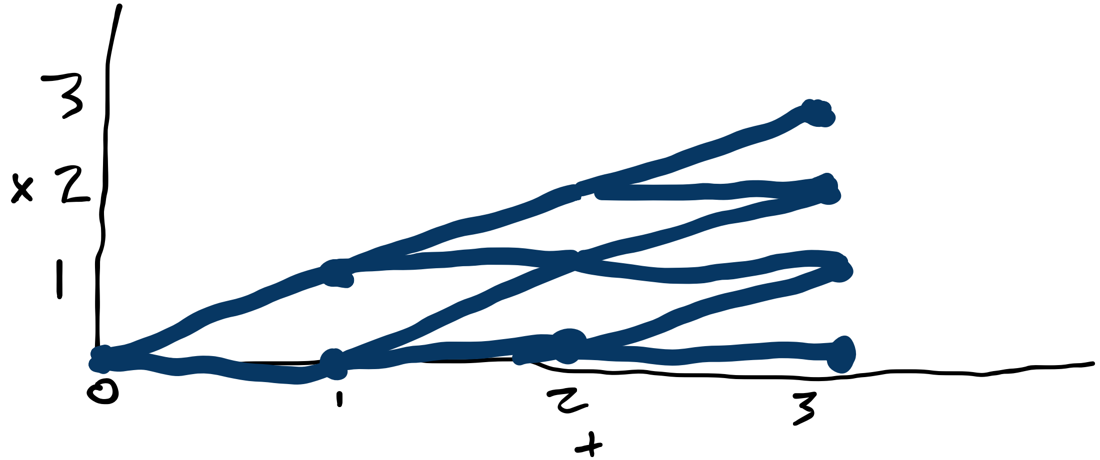
$$v_t \sim \mathcal{U}(\{0,1\}) \text{ (i.i.d.)}$$

Shorthand:

$$x' = x + v$$

Conditional

x_{t+1}	$P(x_{t+1} \mid x_t)$
x_t	0.5
x_t+1	0.5



Joint

x0	x1	x2	P(x1, x2, x3)
0	0	0	0.25
0	0	1	0.25
0	1	1	0.25
0	1	2	0.25

Marginal

x_2	
0	0.5
1	0.5

x_3	
0	0.25
1	0.5
2	0.25

Stochastic Process

- A stochastic process is a collection of R.V.s indexed by time.
- $\{x_t\}_{t=1}^{\infty}$ or just $\{x_t\}$ (shorthand for $\{x_1, x_2, x_3, \ldots\}$)

$P(x_{t+1} \mid x_t, x_{t-1}, \ldots) = P(x_{t+1} \mid x_t)$

In a *stationary* stochastic process (all in this class), this relationship does not change with time

Example:

$x_0 = 0$

$$\boxed{x_{t+1} = x_t + v_t}$$

$v_t \sim \mathcal{U}(\{0, 1\})$ (i.i.d.)

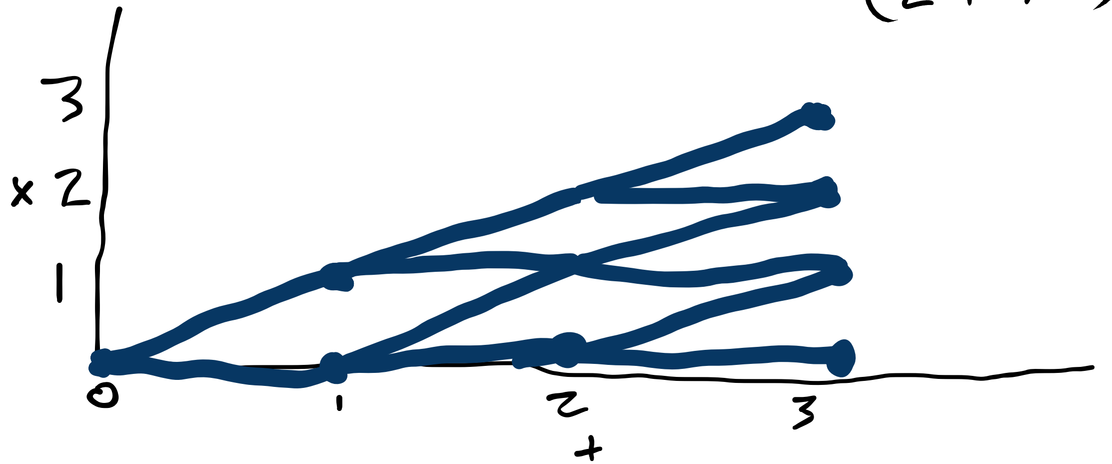
$U(0,1)$ $U(\{1,2,3\})$

Shorthand:

$x' = x + v$

Conditional

x_{t+1}	$P(x_{t+1} \mid x_t)$
x_t	0.5
$x_t + 1$	0.5



Joint

x0	x1	x2	P(x1, x2, x3)
0	0	0	0.25
0	0	1	0.25
0	1	1	0.25
0	1	2	0.25

Marginal

x_2	
0	0.5
1	0.5

x_3	
0	0.25
1	0.5
2	0.25

Simulating a Stochastic Process

030-Stochastic-Processes.ipynb

Markov Process

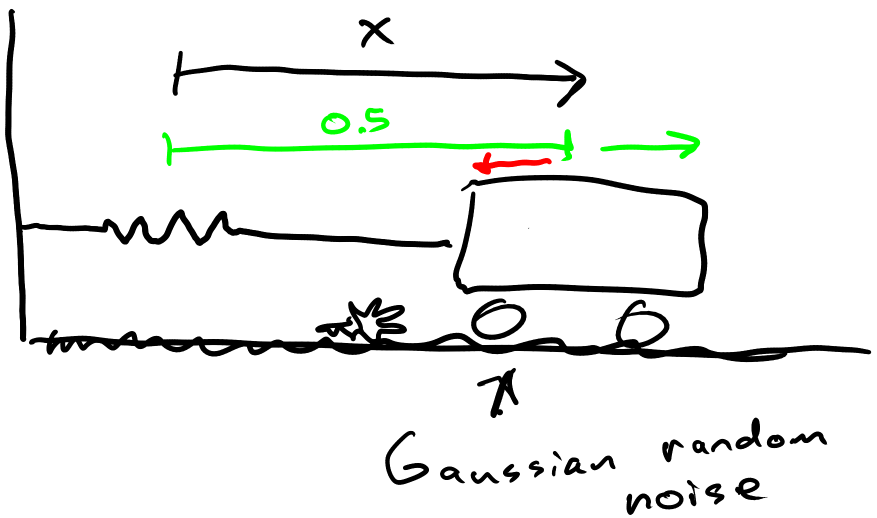
Markov Process

- A stochastic process $\{x_t\}$ is *Markov* if

$$P(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_0) = P(x_t \mid x_{t-1})$$

Markov Process

- A stochastic process $\{x_t\}$ is *Markov* if
  $P(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_0) = P(x_t \mid x_{t-1})$
- x_t is called the "state" of the process



Is $\{x_t\}$ a Markov Process? No

$$\underbrace{\begin{bmatrix} x_{t+1} \\ \dot{x}_{t+1} \end{bmatrix}}_{y_{t+1}} = \begin{bmatrix} 1 & \Delta t \\ -\frac{k}{m}\Delta t & 1 \end{bmatrix} \underbrace{\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix}}_{y_t} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v_t$$

$v_t \sim N(0, \sigma)$

Is $\{y_t\}$ a Markov Process? Yes

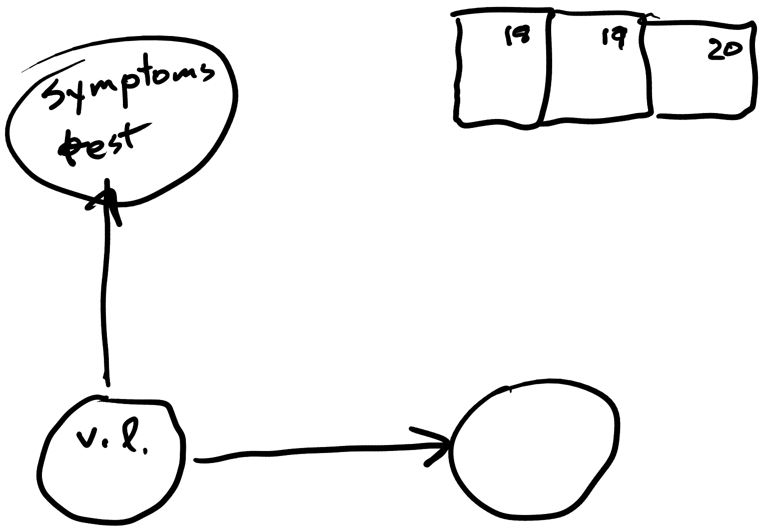
Breakout Rooms

Breakout Rooms

- Name

Breakout Rooms

- Name
- Suppose you want to create a Markov model that describes whether you will test positive for COVID on a given day. What information should be included the state of that model?



state: #cases today
\# exposure in past 5 days
prior test result
days since vaccination
symptoms
observations

Breakout Rooms

Breakout Rooms

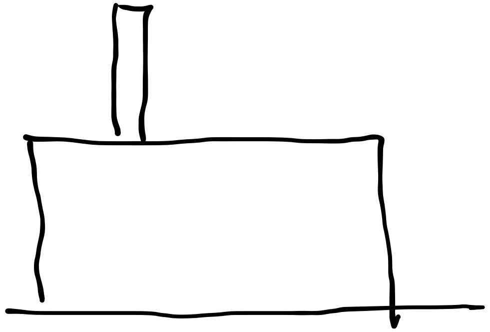
- Name

Breakout Rooms

- Name
- Suppose you have a factory with an entrance/exit road, and you want to define a Markov process to model when trucks will reach the intersection. What should be in the state?

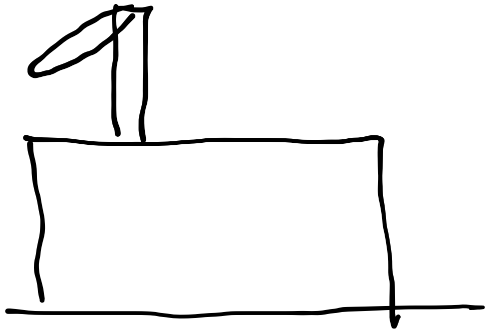
Breakout Rooms

- Name
- Suppose you have a factory with an entrance/exit road, and you want to define a Markov process to model when trucks will reach the intersection. What should be in the state?



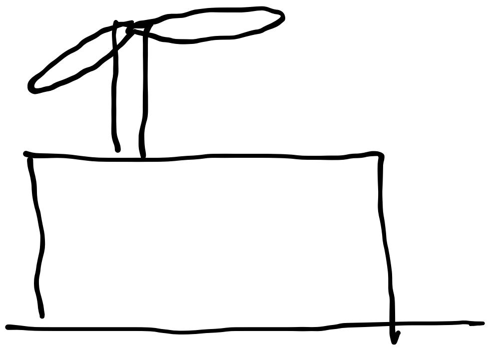
Breakout Rooms

- Name
- Suppose you have a factory with an entrance/exit road, and you want to define a Markov process to model when trucks will reach the intersection. What should be in the state?



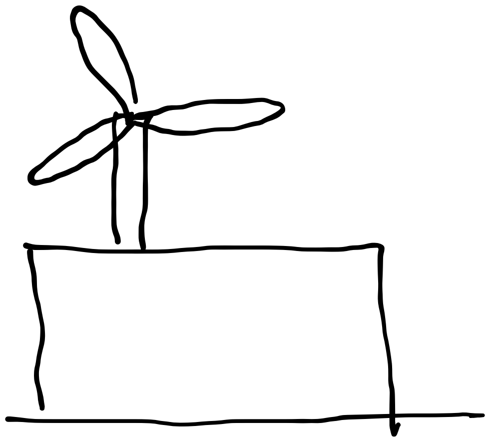
Breakout Rooms

- Name
- Suppose you have a factory with an entrance/exit road, and you want to define a Markov process to model when trucks will reach the intersection. What should be in the state?



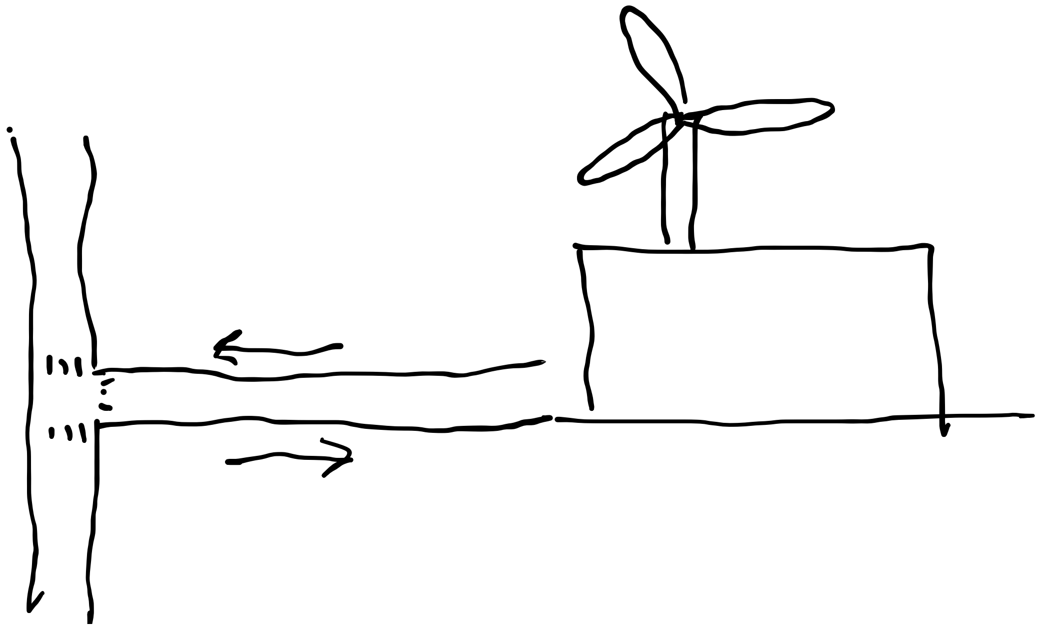
Breakout Rooms

- Name
- Suppose you have a factory with an entrance/exit road, and you want to define a Markov process to model when trucks will reach the intersection. What should be in the state?



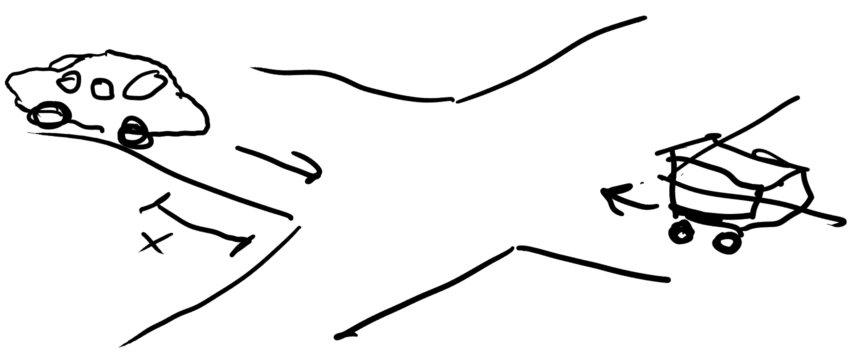
Breakout Rooms

- Name
- Suppose you have a factory with an entrance/exit road, and you want to define a Markov process to model when trucks will reach the intersection. What should be in the state?



Hidden Markov Model

(Often you can't measure the whole state)

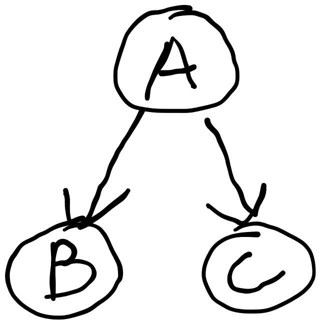


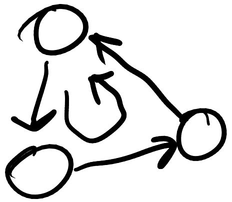
state: $(x, \dot{x}, \text{intention})$

observation: $(x, \dot{x})$

Bayesian Networks

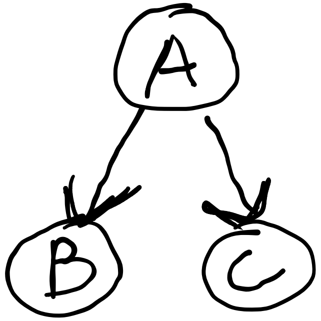
Bayesian Networks



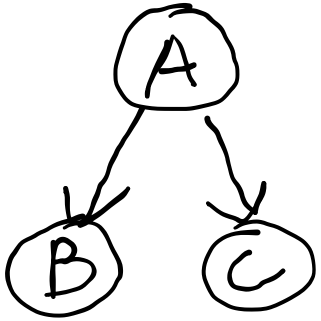


Bayesian Networks

A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s



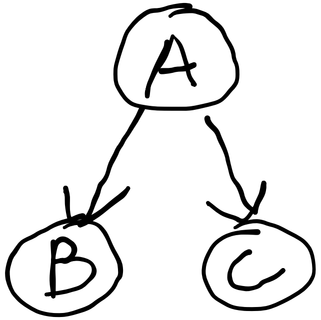
Bayesian Networks



A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s

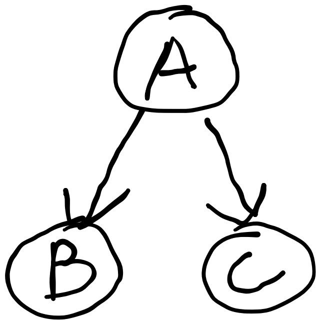
Bayesian Networks



A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Bayesian Networks

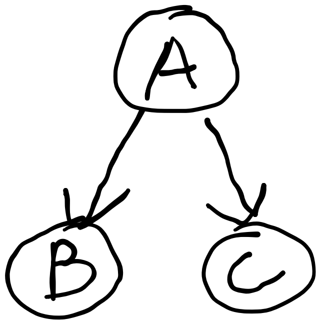


A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Concretely:

Bayesian Networks



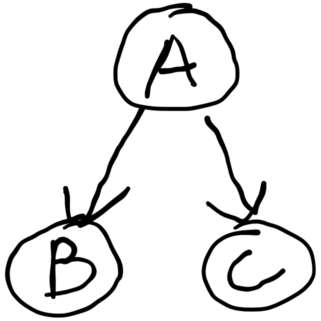
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

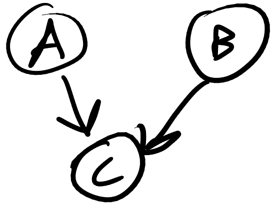
x_1
x_2
$\vdots$
x_n

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Concretely: $$P(x_i \mid \underline{x_{1:n \setminus i}}) = P(x_i \mid Pa(x_i))$$

Bayesian Networks





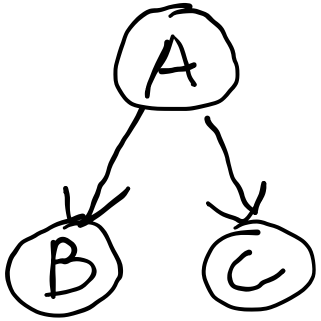
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Concretely: $$P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$$

Parents

Bayesian Networks



$$P(B \mid A, C) = P(B \mid A)$$

$B \perp C$? No
$B \perp C \mid A$ Yes

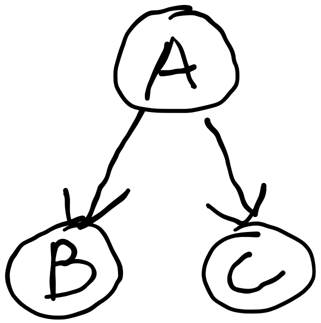
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Concretely: $\underline{P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))}$.

Parents

Bayesian Networks



$$P(B \mid A, C) = P(B \mid A)$$

A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

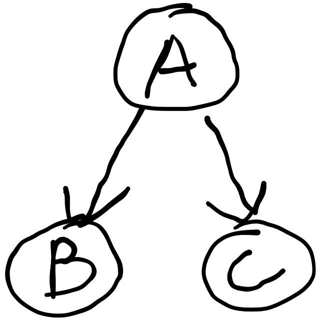
- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Concretely: $$P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$$

Parents

Markov Process

Bayesian Networks



A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

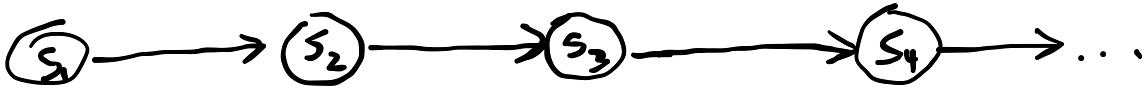
$P(B \mid A, C) = P(B \mid A)$

Concretely: $P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$

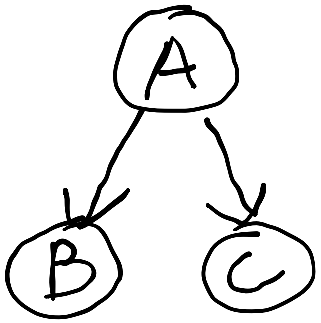
Parents

$P(s_3 \mid s_1, s_2) = P(s_3 \mid s_2)$

Markov Process



Bayesian Networks



$$P(B \mid A, C) = P(B \mid A)$$

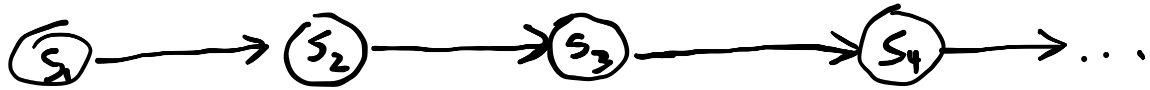
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Concretely: $$P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$$

Parents

Markov Process

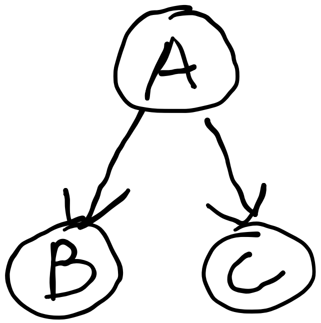


Hidden Markov Model

Bayesian Networks

A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

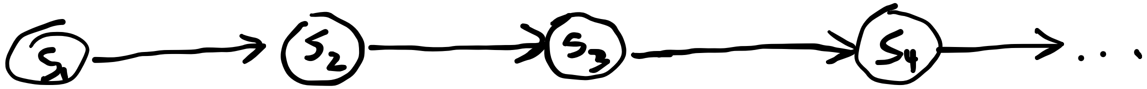


$$P(B \mid A, C) = P(B \mid A)$$

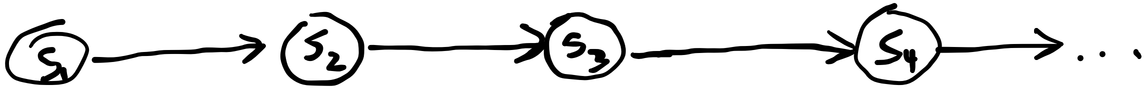
Concretely: $$P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$$

Parents

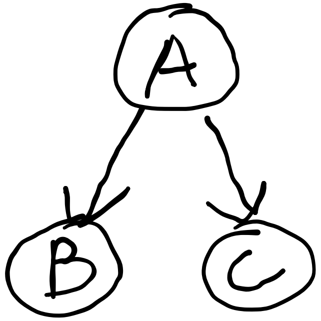
Markov Process



Hidden Markov Model



Bayesian Networks



$P(B \mid A, C) = P(B \mid A)$

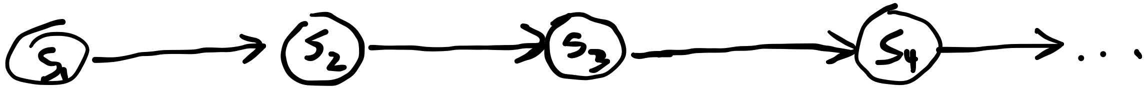
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

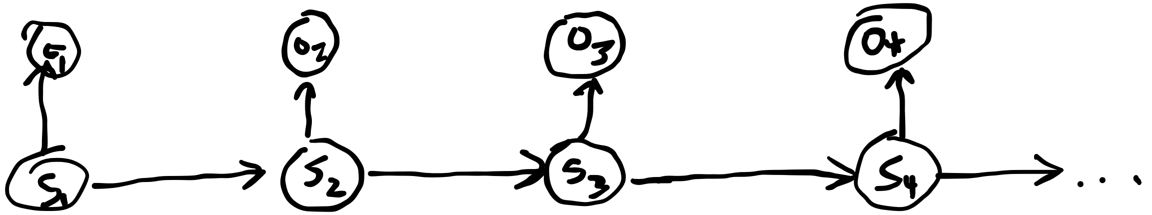
Concretely: $P(x_i \mid x_{1:n\setminus i}) = P(x_i \mid Pa(x_i))$

Parents

Markov Process



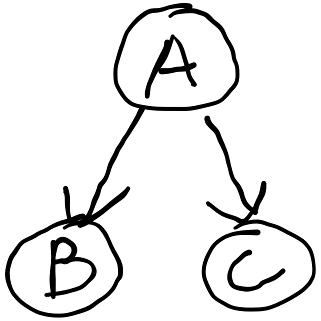
Hidden Markov Model



Bayesian Networks

A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

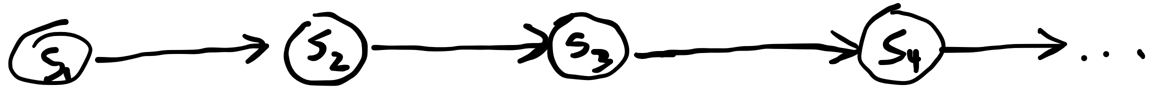


$$P(B \mid A, C) = P(B \mid A)$$

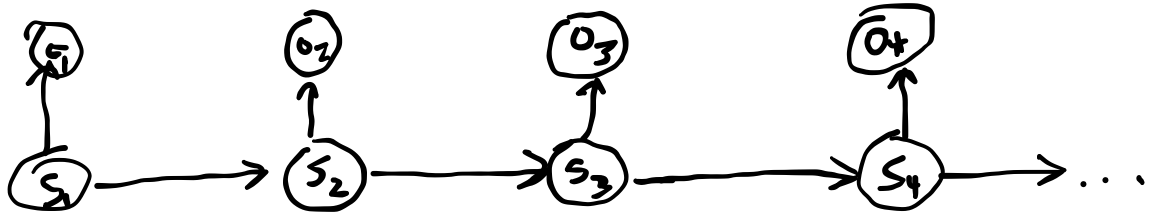
Concretely: $$P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$$

Parents

Markov Process

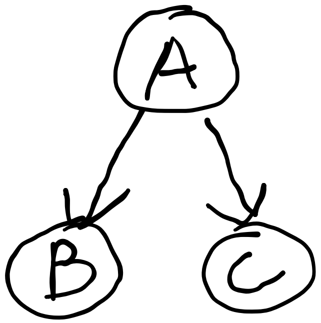


Hidden Markov Model



Dynamic Bayesian Network

Bayesian Networks



$$P(B \mid A, C) = P(B \mid A)$$

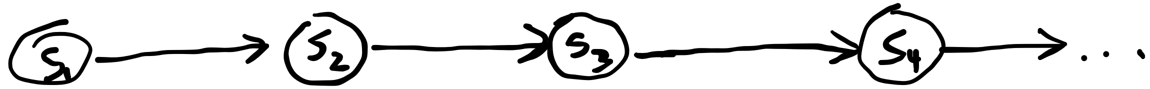
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

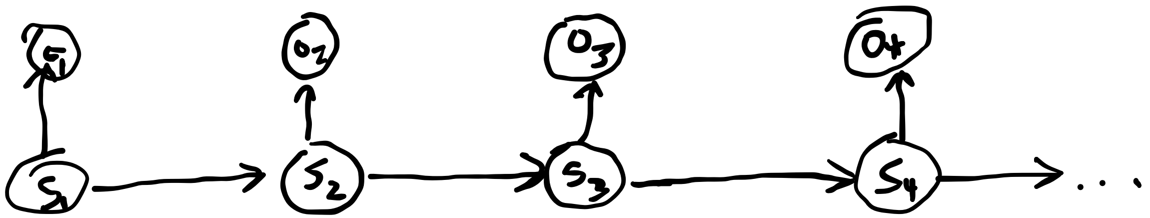
Concretely: $$P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$$

Parents

Markov Process

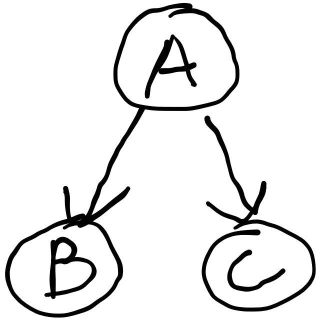


Hidden Markov Model



Dynamic Bayesian Network (One step)

Bayesian Networks



$$P(B \mid A, C) = P(B \mid A)$$

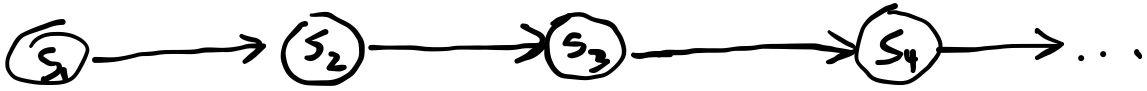
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

Concretely: $$P(x_i \mid x_{1:n \setminus i}) = P(x_i \mid Pa(x_i))$$

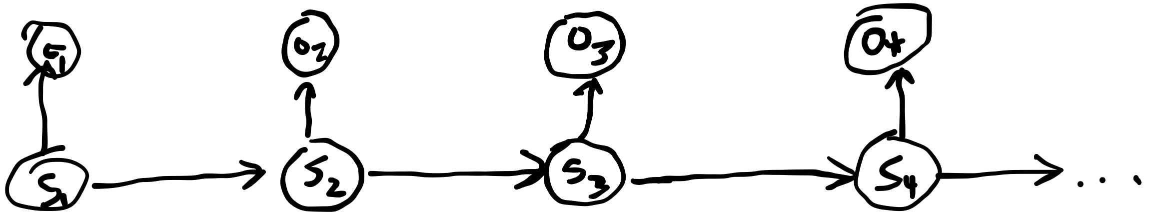
Parents

Markov Process



Dynamic Bayesian Network (One step)

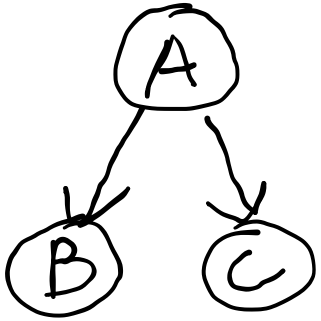
Hidden Markov Model



Bayesian Networks

A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

- Nodes: R.V.s
- Edges: Direct probabilistic relationships

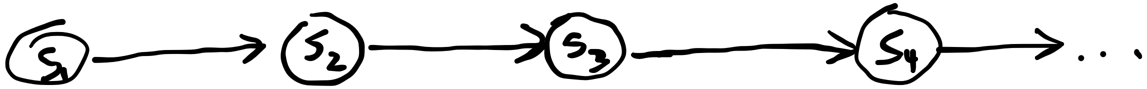


$$P(B \mid A, C) = P(B \mid A)$$

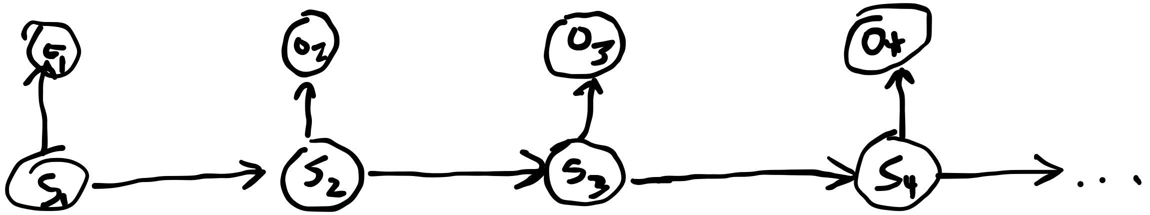
Concretely: $P(x_i \mid x_{1:n\setminus i}) = P(x_i \mid Pa(x_i))$

Parents

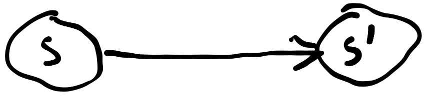
Markov Process

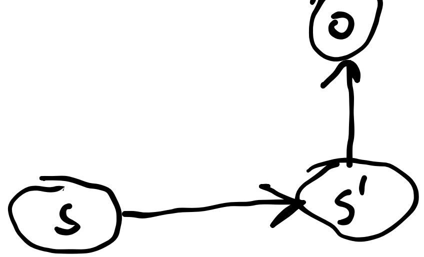


Hidden Markov Model



Dynamic Bayesian Network (One step)





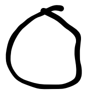
Decision Networks and MDPs

Decision Networks and MDPs

Decision Network

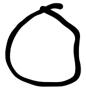
Decision Networks and MDPs

Decision Network



Decision Networks and MDPs

Decision Network

 Chance node

Decision Networks and MDPs

Decision Network

Chance node

Decision Networks and MDPs

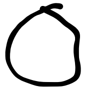
Decision Network

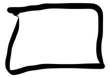
Chance node

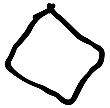
Decision node

Decision Networks and MDPs

Decision Network

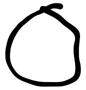
 Chance node

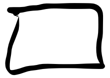
 Decision node

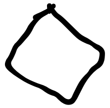


Decision Networks and MDPs

Decision Network

 Chance node

 Decision node

 Utility node

Decision Networks and MDPs

Decision Network

Chance node

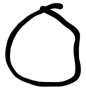
Decision node

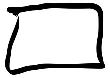
Utility node

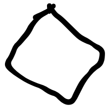
MDP Dynamic Decision Network

Decision Networks and MDPs

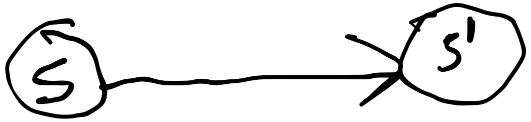
Decision Network


Chance node


Decision node

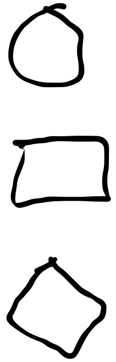

Utility node

MDP Dynamic Decision Network



Decision Networks and MDPs

Decision Network

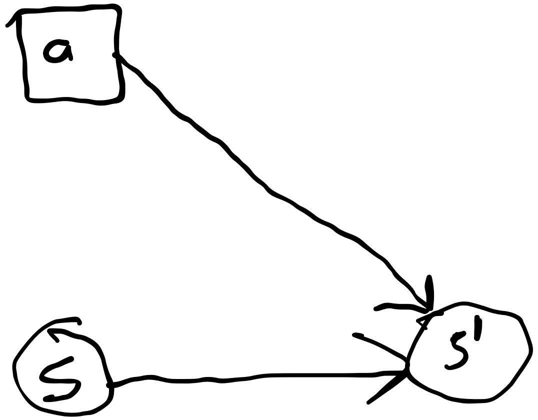


Chance node

Decision node

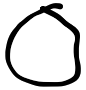
Utility node

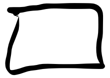
MDP Dynamic Decision Network

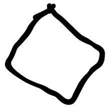


Decision Networks and MDPs

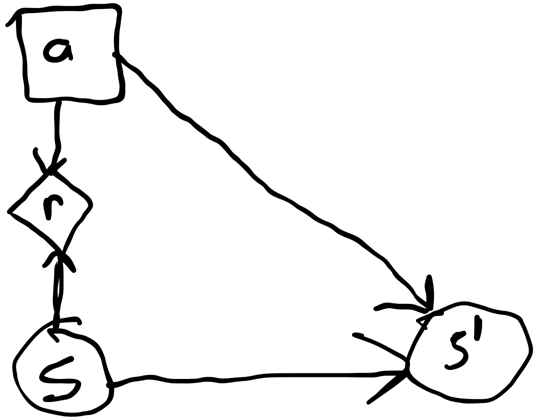
Decision Network

 Chance node

 Decision node

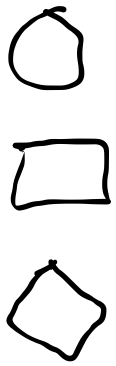
 Utility node

MDP Dynamic Decision Network



Decision Networks and MDPs

Decision Network

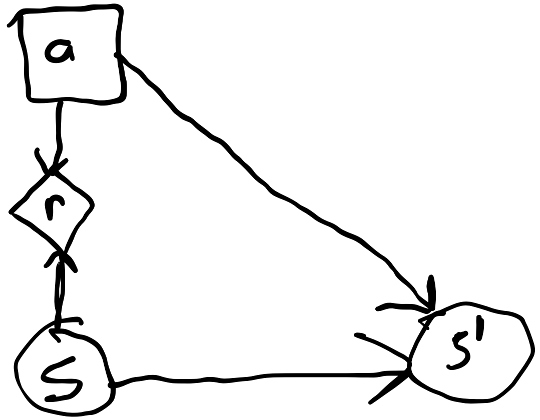


Chance node

Decision node

Utility node

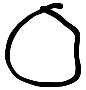
MDP Dynamic Decision Network

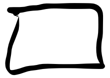


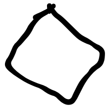
MDP Optimization problem

Decision Networks and MDPs

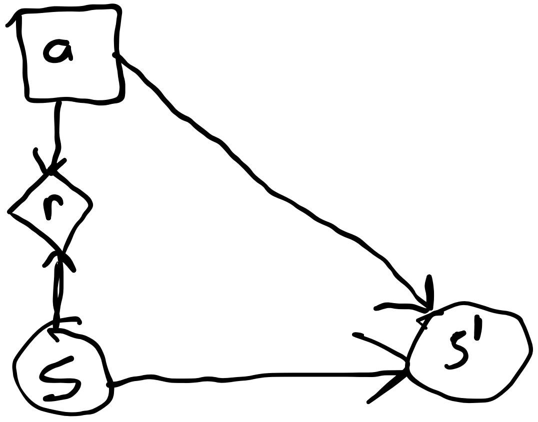
Decision Network

 Chance node

 Decision node

 Utility node

MDP Dynamic Decision Network

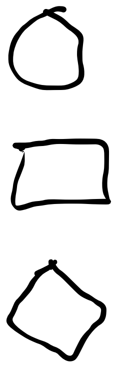


MDP Optimization problem

$$\text{maximize} \quad \mathrm{E}\left[\sum_{t=1}^{\infty} r_t\right]$$

Decision Networks and MDPs

Decision Network

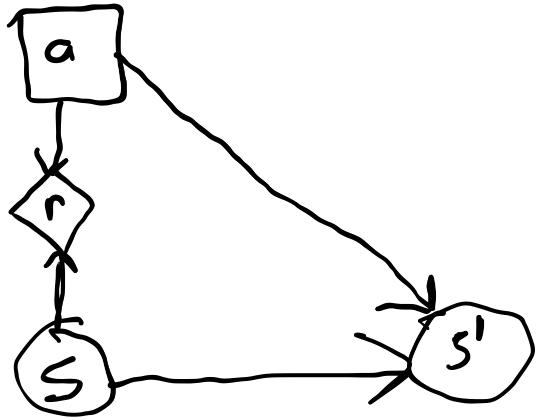


Chance node

Decision node

Utility node

MDP Dynamic Decision Network



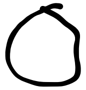
MDP Optimization problem

$$\text{maximize} \quad \mathrm{E}\left[\sum_{t=1}^{\infty} r_t\right]$$

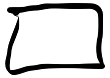
Not well formulated!

Decision Networks and MDPs

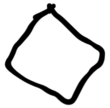
Decision Network



Chance node

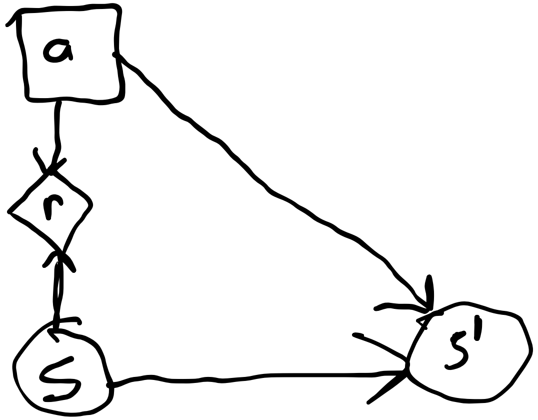


Decision node



Utility node

MDP Dynamic Decision Network



MDP Optimization problem

$$\text{maximize} \quad \mathrm{E}\left[\sum_{t=1}^{\infty} r_t\right]$$

Not well formulated!
Infinite

Finite MDP Objectives

Finite MDP Objectives

1. Finite time

Finite MDP Objectives

1. Finite time $$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty} \mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty} \mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty}\mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

$$\mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty} \mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

$$\mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

discount $\gamma \in [0, 1)$

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty} \mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

$$\mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

discount $\gamma \in [0,1)$

typically 0.9, 0.95, 0.99

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty} \mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

$$\mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

discount $\gamma \in [0, 1)$

typically 0.9, 0.95, 0.99

if $\underline{r} \leq r_t \leq \bar{r}$

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty} \mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

$$\mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

discount $\gamma \in [0, 1)$

typically 0.9, 0.95, 0.99

if $\underline{r} \leq r_t \leq \bar{r}$

then

$$\frac{\underline{r}}{1-\gamma} \leq \sum_{t=0}^{\infty} \gamma^t r_t \leq \frac{\bar{r}}{1-\gamma}$$

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty} \mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

$$\mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

discount $\gamma \in [0,1)$

typically 0.9, 0.95, 0.99

if $\underline{r} \le r_t \le \bar{r}$

then

$$\frac{\underline{r}}{1-\gamma} \le \sum_{t=0}^{\infty} \gamma^t r_t \le \frac{\bar{r}}{1-\gamma}$$

4. Terminal States

Finite MDP Objectives

1. Finite time

$$\mathbf{E}\left[\sum_{t=0}^{T} r_t\right]$$

2. Average reward

$$\lim_{n\to\infty}\mathbf{E}\left[\sum_{t=0}^{n} r_t\right]$$

3. Discounting

$$\mathbf{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

discount $\gamma \in [0,1)$

typically 0.9, 0.95, 0.99

if $\underline{r} \leq r_t \leq \bar{r}$

then

$$\frac{\underline{r}}{1-\gamma} \leq \sum_{t=0}^{\infty} \gamma^t r_t \leq \frac{\bar{r}}{1-\gamma}$$

4. Terminal States

Infinite time, but a terminal state (no reward, no leaving) is always reached with probability 1.

Guiding Question

- What does "Markov" mean in "Markov Decision Process"?

$$P(x_t \mid x_{t-1} \ldots x_0) = P(x_t \mid x_{t-1})$$