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:

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

$v_t \sim \mathcal{U}(\{0, 1\})$ (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:

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

$v_t \sim \mathcal{U}(\{0, 1\})$ (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 $\{x_1, x_2, x_3, \ldots\}$)

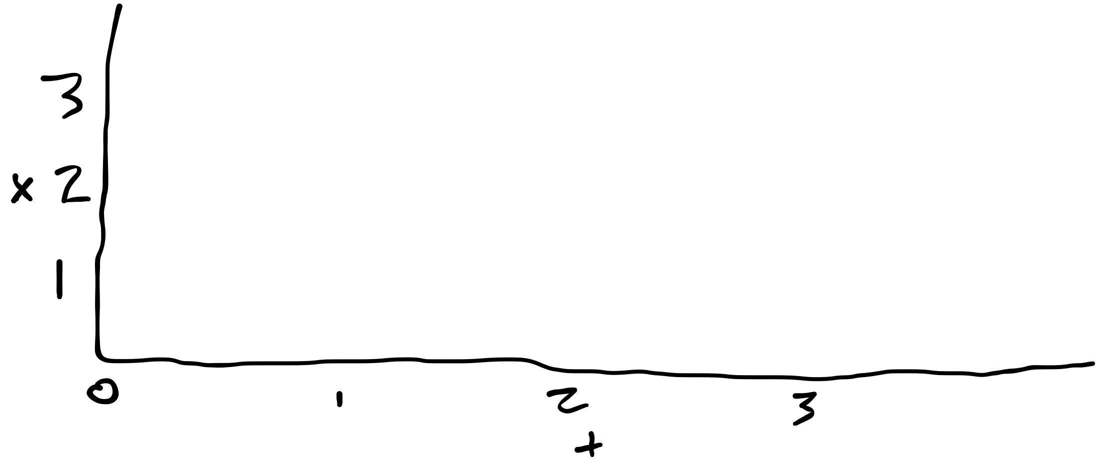
Example:

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

$v_t \sim \mathcal{U}(\{0,1\})$ (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 $\{x_1, x_2, x_3, \ldots\}$)

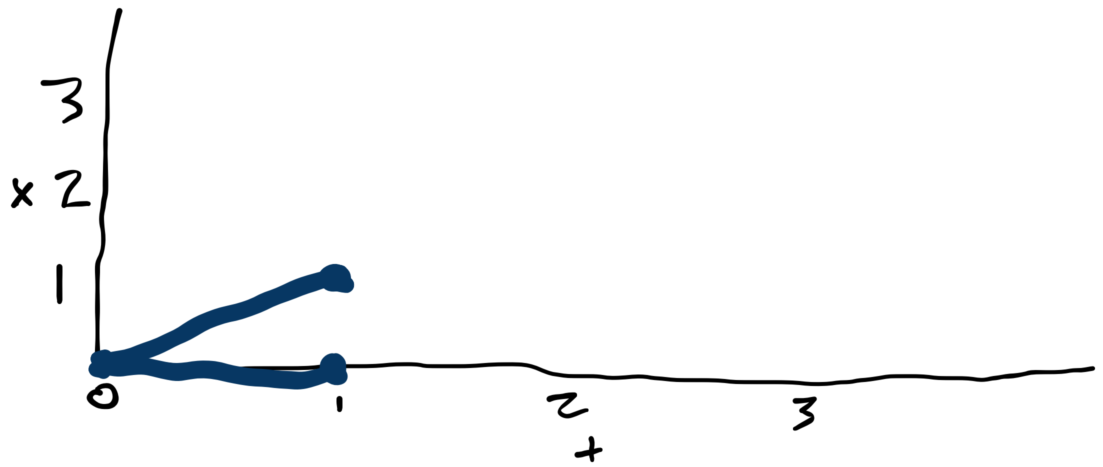
Example:

$x_0 = 0$

$$x_{t+1} = x_t + v_t$$
$$v_t \sim \mathcal{U}(\{0, 1\}) \text{ (i.i.d.)}$$

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:

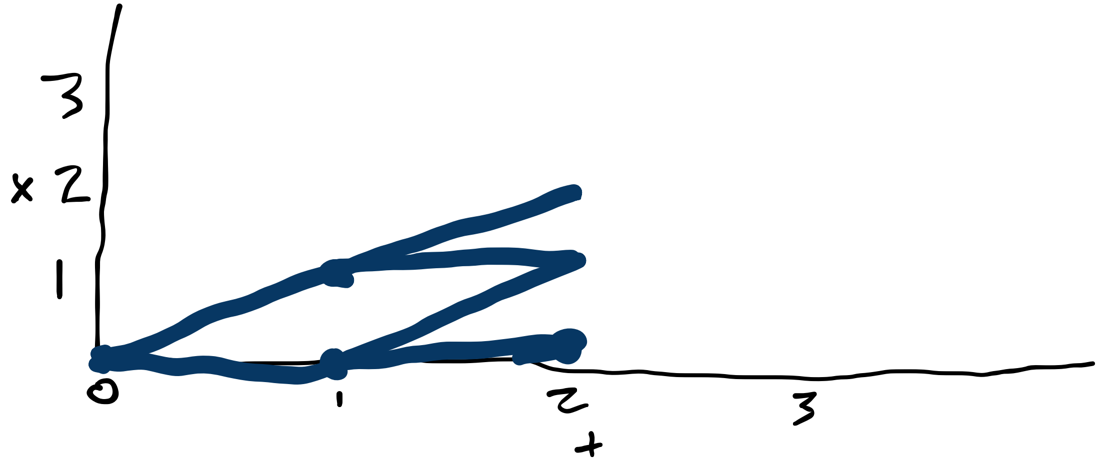
$x_0 = 0$

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

$v_t \sim \mathcal{U}(\{0, 1\})$ (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:

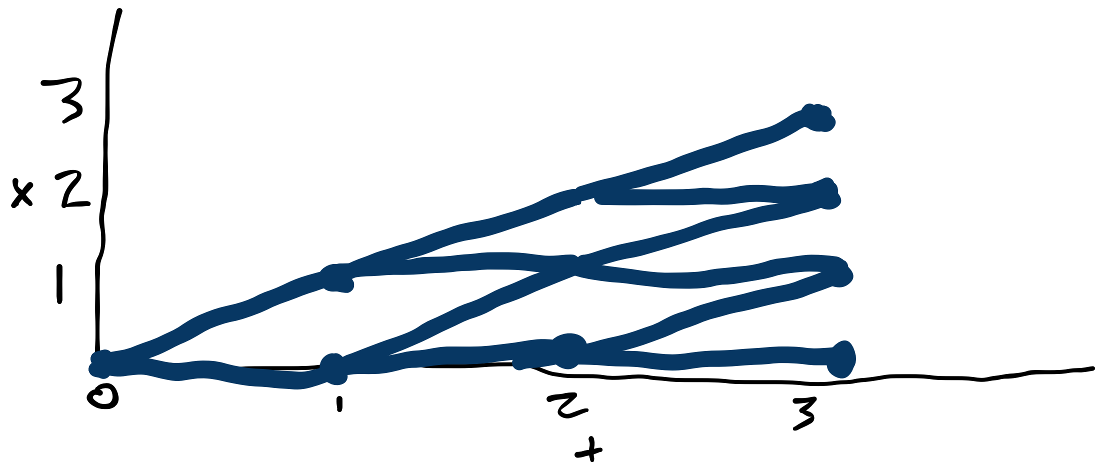
$x_0 = 0$

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

$v_t \sim \mathcal{U}(\{0, 1\})$ (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:

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

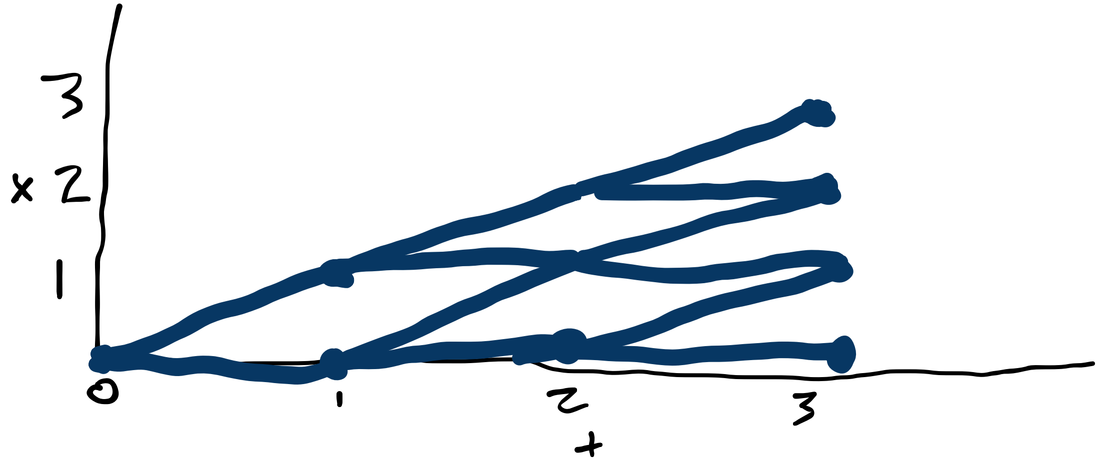
$$x_0 = 0$$

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

$v_t \sim \mathcal{U}(\{0, 1\})$ (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:

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

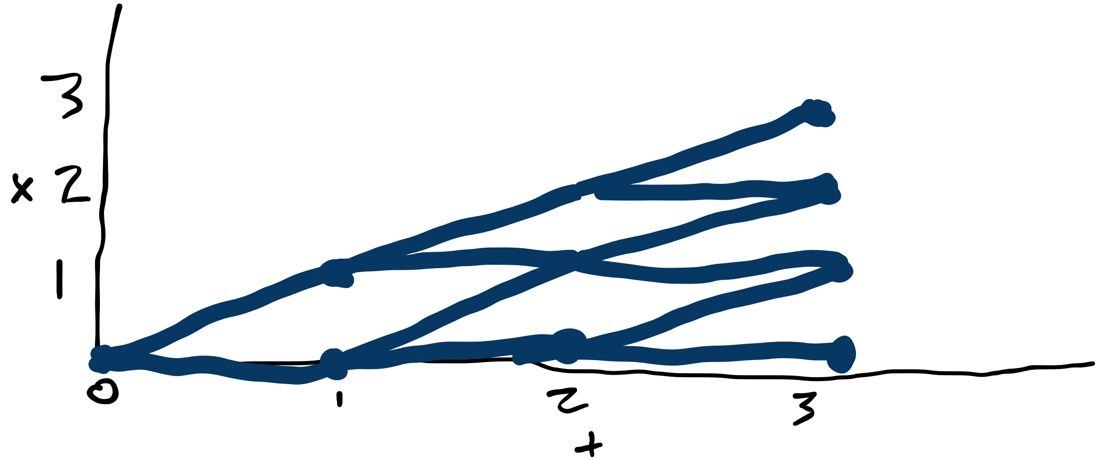
$x_0 = 0$

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

$v_t \sim \mathcal{U}(\{0,1\})$ (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 $\{x_1, x_2, x_3, \ldots\}$)

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

Example:

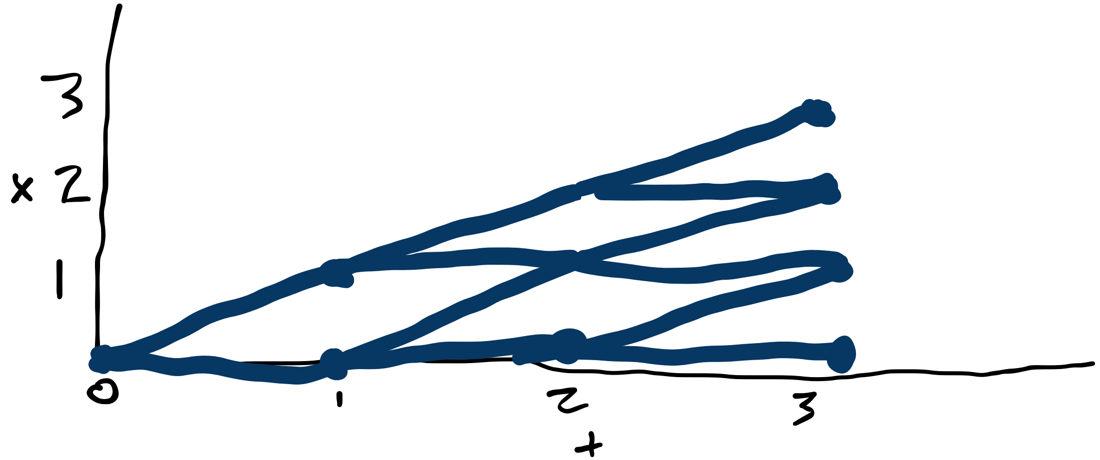
$$x_0 = 0 \qquad x_{t+1} = x_t + v_t$$

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

Shorthand:

$$x' = x + v$$

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:

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

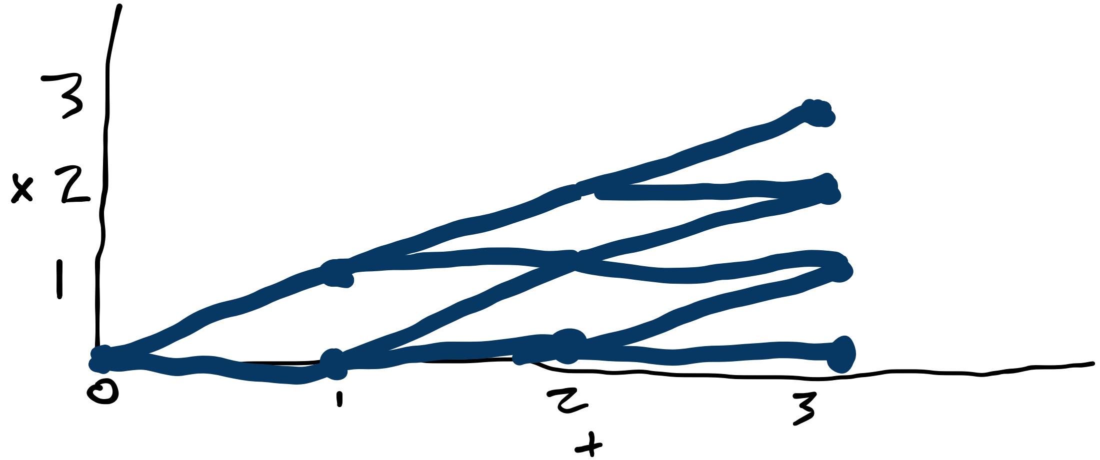
$x_0 = 0$

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

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

Shorthand:

$$x' = x + v$$



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

Simulating a Stochastic Process

030-Stochastic-Processes.ipynb

Markov Process

Markov Process

- A stochastic process $\{s_t\}$ is *Markov* if
  $P(s_{t+1} \mid s_t, s_{t-1}, \ldots, s_0) = P(s_{t+1} \mid s_t)$

Markov Process

- A stochastic process $\{s_t\}$ is *Markov* if
  $P(s_{t+1} \mid s_t, s_{t-1}, \ldots, s_0) = P(s_{t+1} \mid s_t)$
- s_t is called the "state" of the process

Break

Break

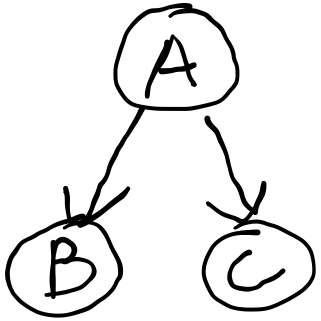
- Suppose you want to create a Markov model that describes how many new COVID cases will be detected on a particular day. What information should be in the state of the model?

Hidden Markov Model

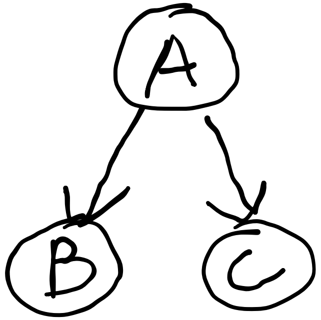
(Often you can't measure the whole state)

Bayesian Networks

Bayesian Networks

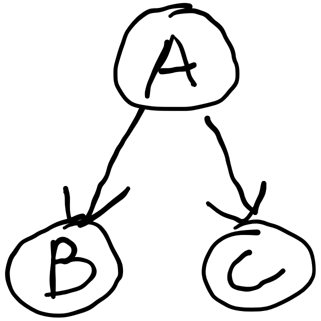


Bayesian Networks



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

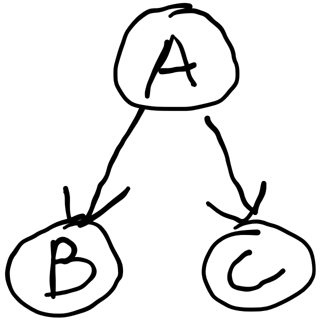
Bayesian Networks



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

- Nodes: R.V.s

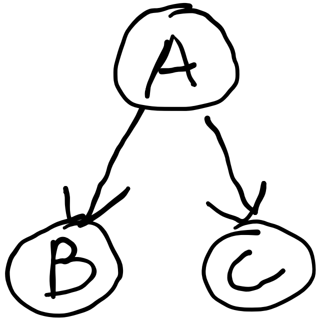
Bayesian Networks



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

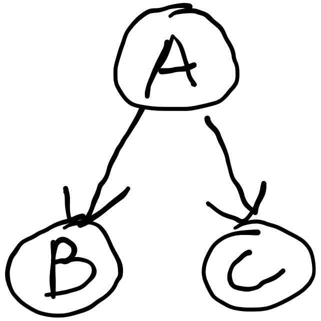


A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

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

Concretely:

Bayesian Networks

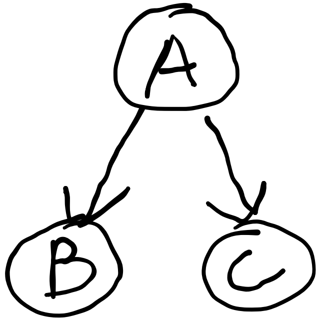


A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

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

Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Bayesian Networks



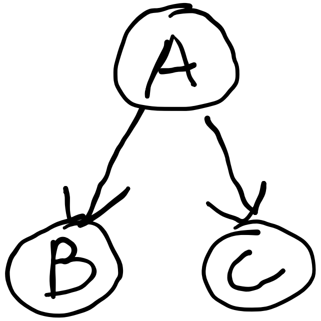
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

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

Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

Bayesian Networks



$P(A, B, C)$

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

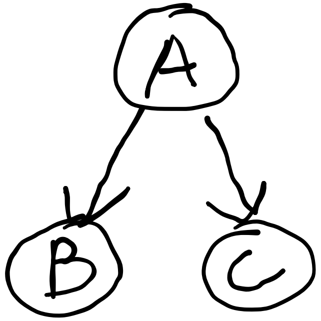
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

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

Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

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

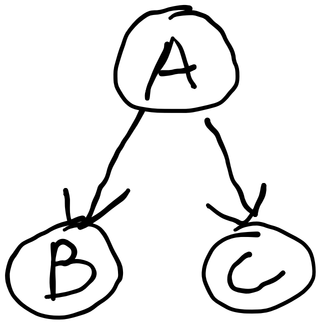
$$P(A, B, C)$$
$$= P(A)P(B \mid A)P(C \mid A)$$

Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

Markov Process

Bayesian Networks



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

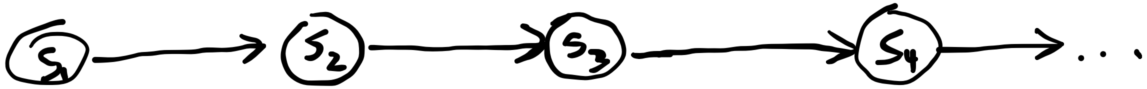
- Nodes: R.V.s
- Edges: Direct probabilistic relationships

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

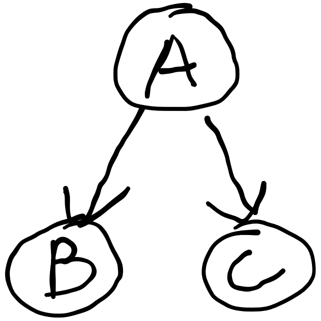
Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

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

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

Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

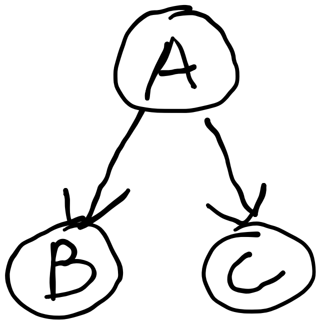
Parents

Markov Process

$s_1 \to s_2 \to s_3 \to s_4 \to \cdots$

Hidden Markov Model

Bayesian Networks



$P(A, B, C)$

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

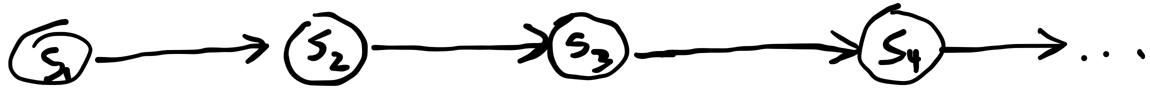
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

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

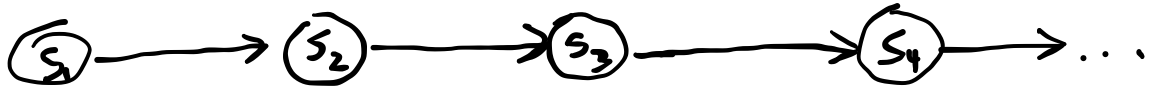
Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

Markov Process



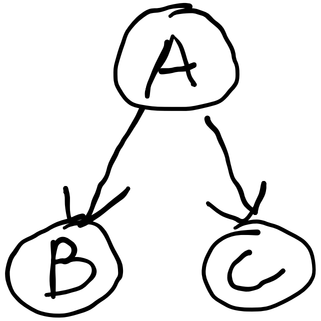
Hidden Markov Model



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

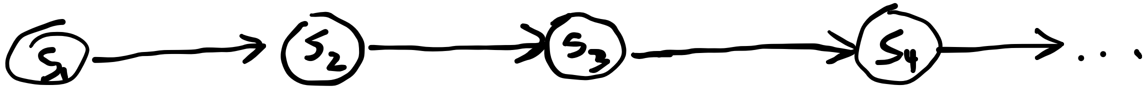


$$P(A, B, C)$$
$$= P(A)P(B \mid A)P(C \mid A)$$

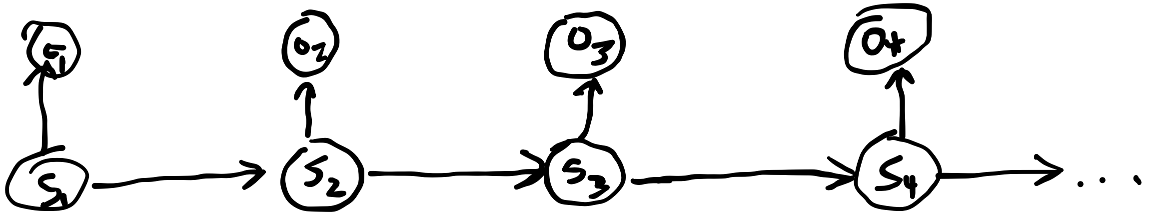
Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

Markov Process



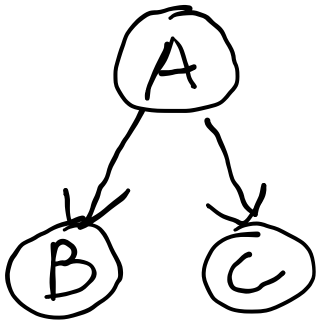
Hidden Markov Model



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

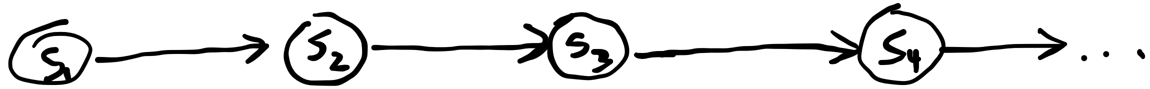


$$P(A, B, C)$$
$$= P(A)P(B \mid A)P(C \mid A)$$

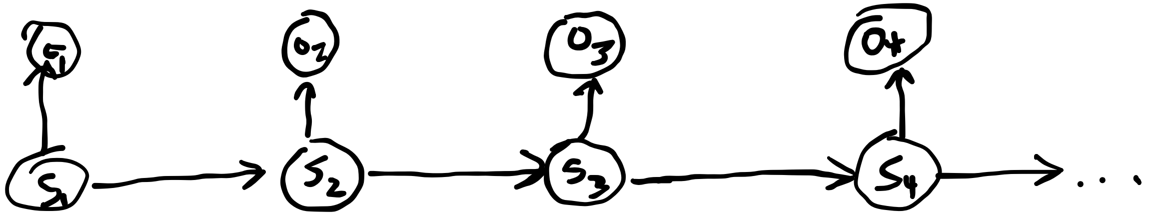
Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

Markov Process



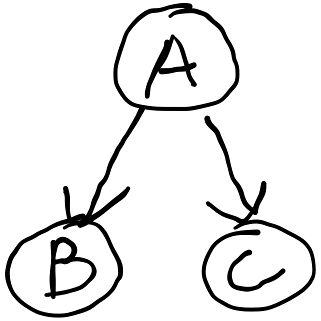
Hidden Markov Model



Dynamic Bayesian Network

Bayesian Networks

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



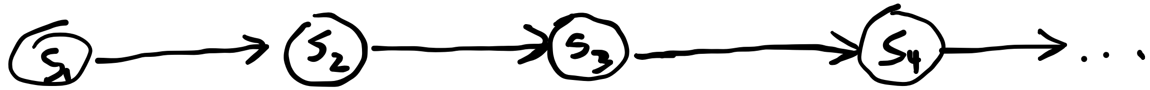
- Nodes: R.V.s
- Edges: Direct probabilistic relationships

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

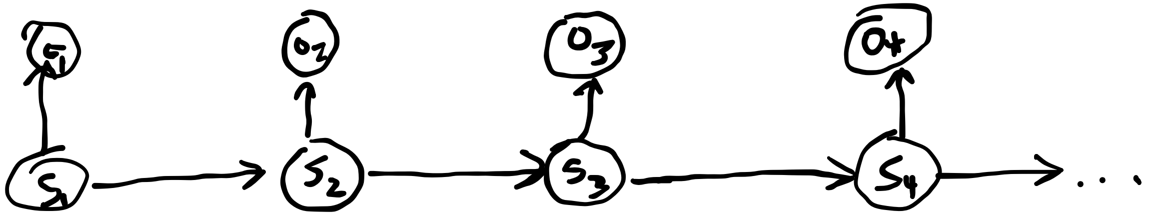
Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

Markov Process

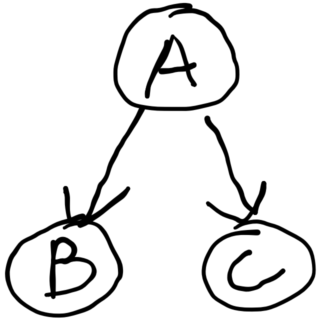


Hidden Markov Model



Dynamic Bayesian Network (One step)

Bayesian Networks



$P(A, B, C)$

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

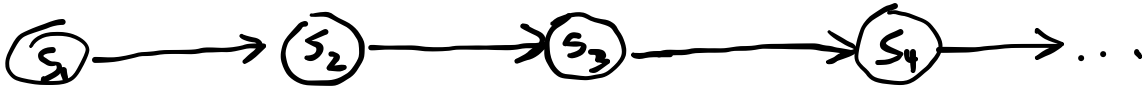
A *Bayesian Network* is a directed acyclic graph (DAG) that encodes probabilistic relationships between R.V.s

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

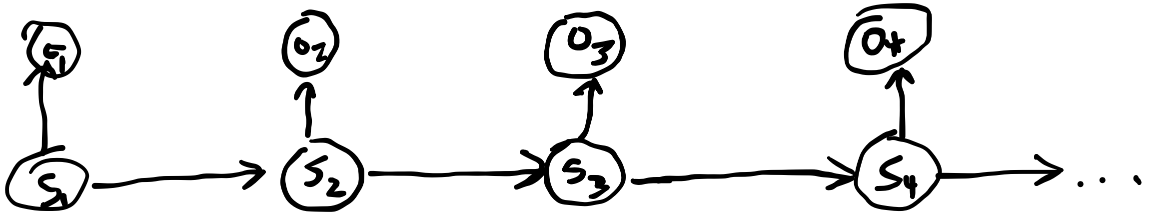
Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

Markov Process



Hidden Markov Model

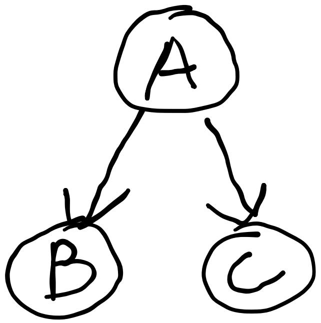


Dynamic Bayesian Network (One step)

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

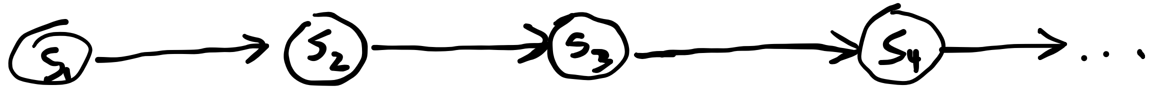


$$P(A, B, C)$$
$$= P(A)P(B \mid A)P(C \mid A)$$

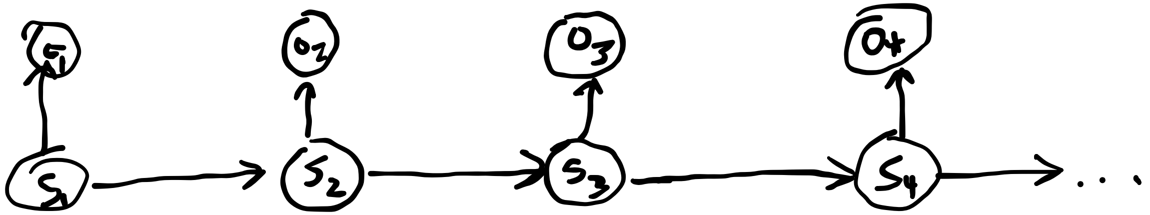
Concretely: $P(x_{1:n}) = \prod_i P(x_i \mid pa(x_i))$

Parents

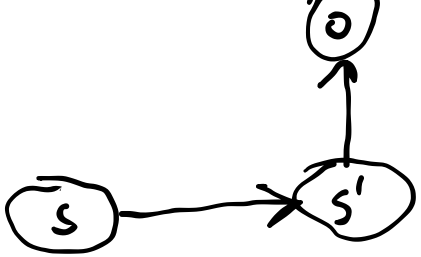
Markov Process



Hidden Markov Model



Dynamic Bayesian Network (One step)



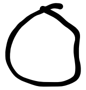
Decision Networks and MDPs

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

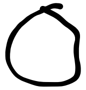
Decision Network

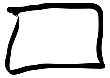
Chance node

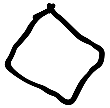
Decision node

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

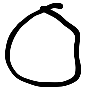
Decision node

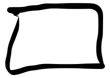
Utility node

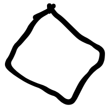
MDP Dynamic Decision Network

Decision Networks and MDPs

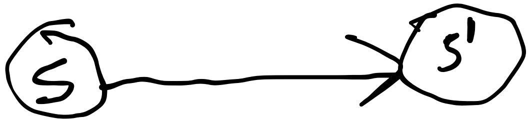
Decision Network

 Chance node

 Decision node

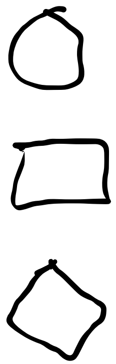
 Utility node

MDP Dynamic Decision Network



Decision Networks and MDPs

Decision Network

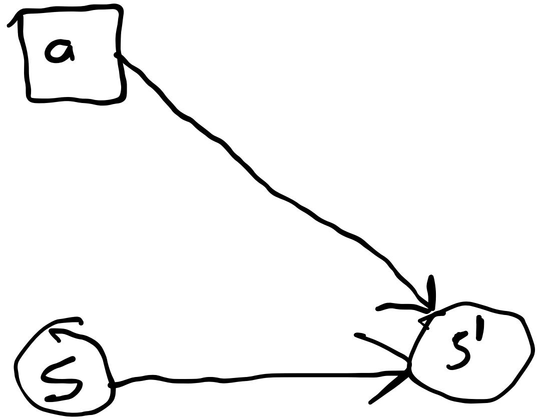


Chance node

Decision node

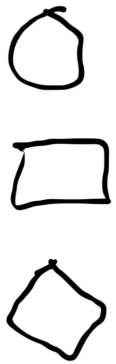
Utility node

MDP Dynamic Decision Network



Decision Networks and MDPs

Decision Network

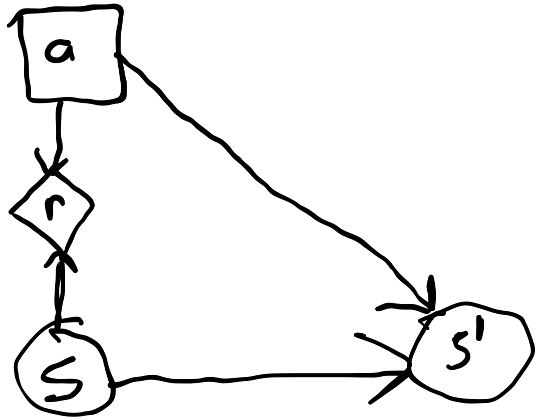


Chance node

Decision node

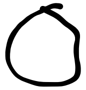
Utility node

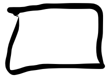
MDP Dynamic Decision Network

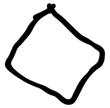


Decision Networks and MDPs

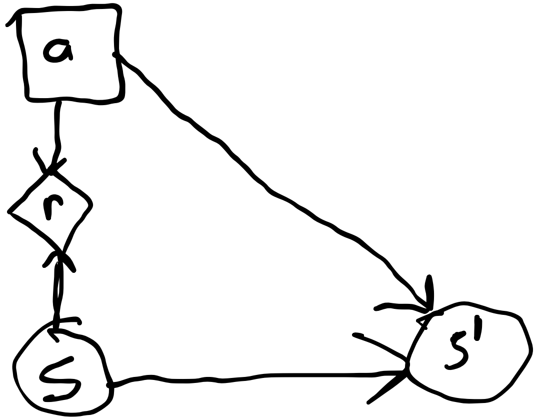
Decision Network

 Chance node

 Decision node

 Utility node

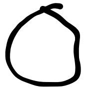
MDP Dynamic Decision Network

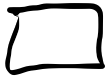


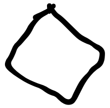
MDP Optimization problem

Decision Networks and MDPs

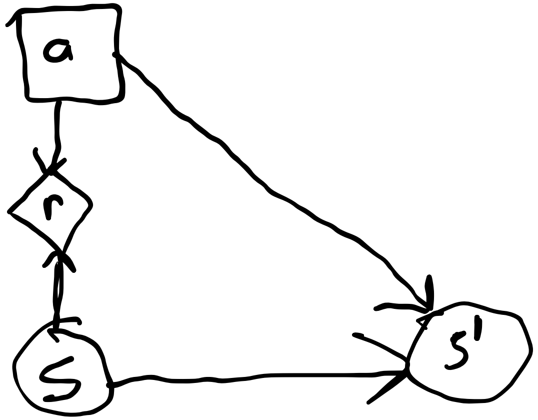
Decision Network

 Chance node

 Decision node

 Utility node

MDP Dynamic Decision Network

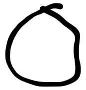


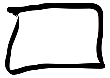
MDP Optimization problem

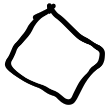
$$\text{maximize} \quad \mathrm{E}\left[\sum_{t=1}^{\infty} r_t\right]$$

Decision Networks and MDPs

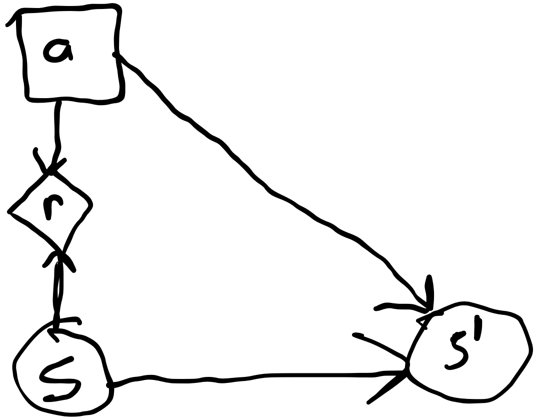
Decision Network

 Chance node

 Decision node

 Utility node

MDP Dynamic Decision Network



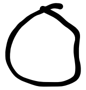
MDP Optimization problem

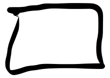
$$\text{maximize} \quad \mathrm{E}\left[\sum_{t=1}^{\infty} r_t\right]$$

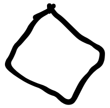
Not well formulated!

Decision Networks and MDPs

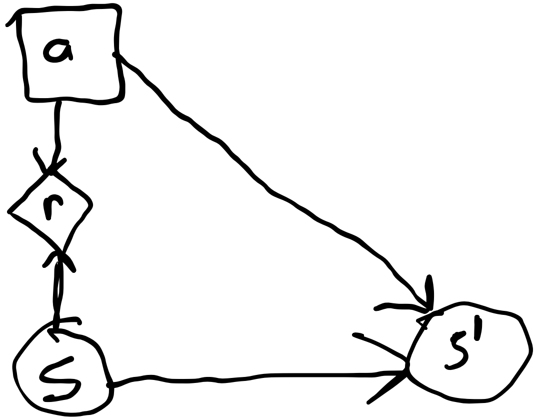
Decision Network

 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

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

2. Average reward

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

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

then

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

4. Terminal States

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"?