What is a R.V.?

How to infer info about A given B? Bayes Rule

How to (efficiently) encode relationships between R.V.s?

How can we determine if measuring one R.V. will reveal info about another?

What does "Markov" mean in "MDP"?

Joint

All info about a collection of R.V.s

Conditional

$P(A|B)$

a function returns a distribution of A given b

Marginal

$P(A)$

without any knowledge of B, distribution of A

A, B binary

3 indep. param

a	b	P(a,b)

$P(A|B)$ 2 indep param

b=0

a	P(a)

b=1

a	P(a)

$P(A)$ 1 indep param

$P(B)$ 1 indep param

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

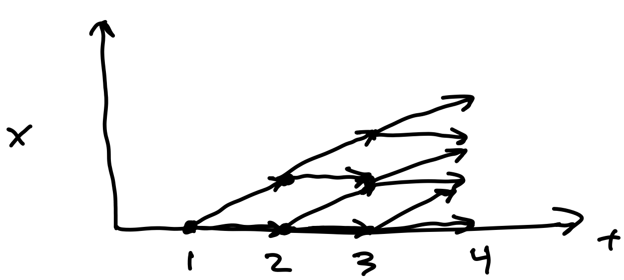
$$\underline{P(A,B)} = \underline{P(A|B)}\,\underline{P(B)}$$

Stochastic Process

$\{x_t\}_{t=1}^{\infty}$

Example: $x_1 = 0$ $\quad x_{t+1} = x_t + v_t$ $\quad v_t$ are i.i.d.

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



Marginal

x_2	$P(x_2)$
0	0.5
1	0.5

Marginal

x_3	$P(x_3)$
0	0.25
1	0.5
2	0.25

Conditional $P(x_3 | x_2)$

$x_2 = 0$

x_3	$P(x_3)$
0	0.5
1	0.5

$x_2 = 1$

x_3	P
1	0.5
2	0.5

$$\sum_{x_2} P(x_3 | x_2) P(x_2)$$

Joint

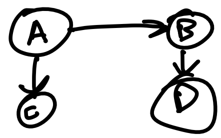
x_1	x_2	x_3	$P(x_1, x_2, x_3)$
0	0	1	
0	1	1	
0	0	0	
0	1	2	

x_4 Joint

x_1	x_2	x_3	x_4	P

Encode Sparsity in Joint Distributions

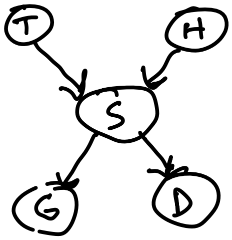
Bayesion Networks



Edge : Direct probabilistic "causal" relationship

$$P(X_i \mid \{X_j\}_{j \neq i}) = P(X_i \mid Pa(X_i))$$

Acyclic



Temp	0	1
Humid	0	1
Snow	0	1
Delay	0	1
~~Powt~~ Gnar	0	1

Joint

$2^5 - 1$ indep params

$= 31$

Chain rule

$$P(X_1, X_2 \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

$$P(T \mid \{X_i\}_i) = P(T)$$

$P(T)$ 1 param

$P(H)$ 1

$P(S \mid T, H)$ 4 — 1 param for each value of T, H

→ $P(G \mid S)$ 2

$P(D \mid S)$ 2

10 indep param — save 21

s=0

G	P(G|S)
1	0.5
0	0.5

s=1

G	P(G|S)
0	0.25
1	0.75

$$A \perp B \iff P(A,B) = P(A)P(B)$$
$$\iff P(A \mid B) = P(A)$$

$G \perp D$? No

$G \perp D \mid S$? Yes

$$P(G \mid D, S) = P(G \mid S)$$

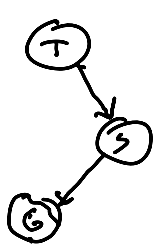
d-separation rules:

A path between A and B is d-separated by $\mathcal{C}$ ← set of nodes if any of the following are true

1. contains "chain" $X \to Y \to Z$ where $Y \in \mathcal{C}$
2. contains "fork" $X \leftarrow Y \to Z$ where $Y \in \mathcal{C}$
3. contains "inverted fork"/"v-structure" $X \to Y \leftarrow Z$ s.t. $Y \notin \mathcal{C}$ and no descendent of Y is in $\mathcal{C}$

All paths between A and B d-sep. by $\mathcal{C} \iff A \perp B \mid \mathcal{C}$

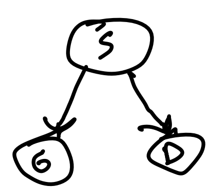
1.



$\mathcal{C} = \{S\}$

$T \perp G \mid S$

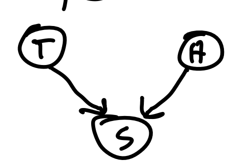
2.

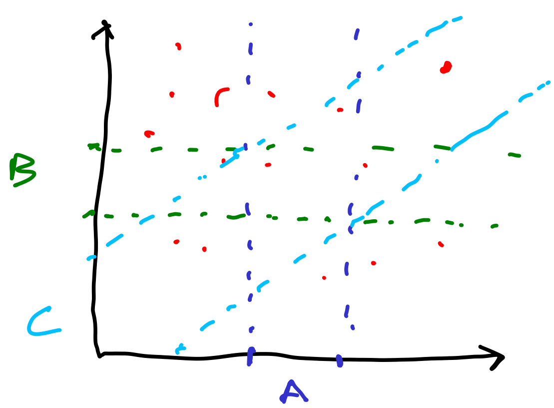


$\mathcal{C} = \{S\}$

$G \perp D \mid S$

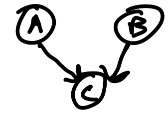
3.





$A \perp B$

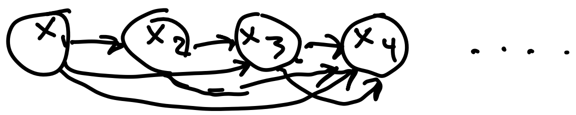
$A \not\perp B \mid C$



know $C = \text{true}$

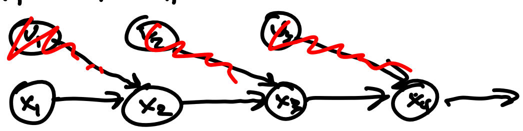
get $A = \text{true}$

conclude B is likely true



Markov Blanket

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



$G = \{x_3\}$

$x_4 \perp x_2 \mid x_3$

Markov process "state"

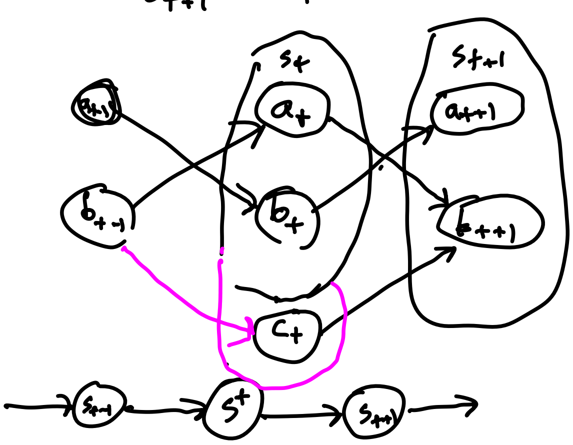
Markov chain

$$P(x_{>t} \mid x_t, x_{t-1}, x_{t-2} \dots x_1) = P(x_{>t} \mid x_t)$$

$a_{t+1} = b_t + v_t$

$b_{t+1} = a_t$

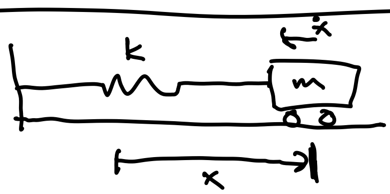
Markov w.r.t. a? No



$G = \{a_t\}$

$s_t = (a_t, b_t)$

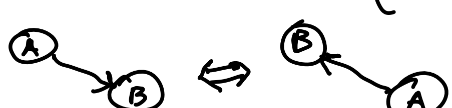
Markov w.r.t. s ✓



$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix}$$

state

Aside: Markov Equivalence Classes in Book



Stationary Markov Process

$$P(s_{t+1} \mid s_t) = P(s_{k+1} \mid s_k) \quad \forall t, k$$

- state changes
- cond. p.d. stay the same

defined by pair

(S, T)

T: Transition "kernel"

S: state space

Explicit

$$T(s' \mid s) = P(s_{t+1} \mid s_t)$$

Generative

$$s' \sim G(s)$$