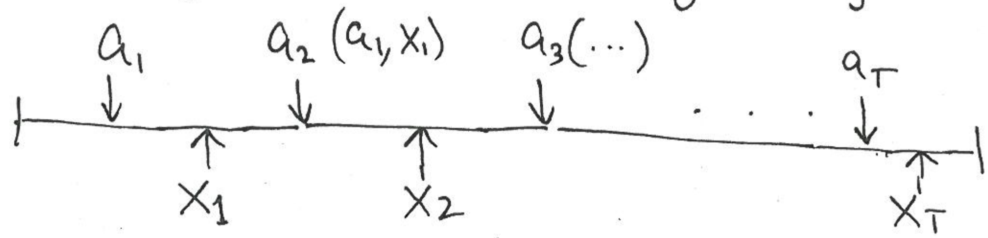
Dynamic Programming

- Sequential decision making
  - Stochastic problem: $\max\limits_{a} \mathbb{E}_X [f(\overset{\text{action}}{a}, \underset{\text{randomness}}{X})]$
  - If actions interleave realization of random vars



Later actions can depend on 'realized' randomness

- Stochastic DP - 5 components

i) Horizon $\boxed{T}$ $\Rightarrow$ discrete periods $1, 2, \ldots, T$

ii) State - $\boxed{s_t \in S}$ $\equiv$ (concise) summary of history of process
(finite)

iii) Action - $\boxed{a_t \in A(s_t)}$ $\equiv$ control in each period

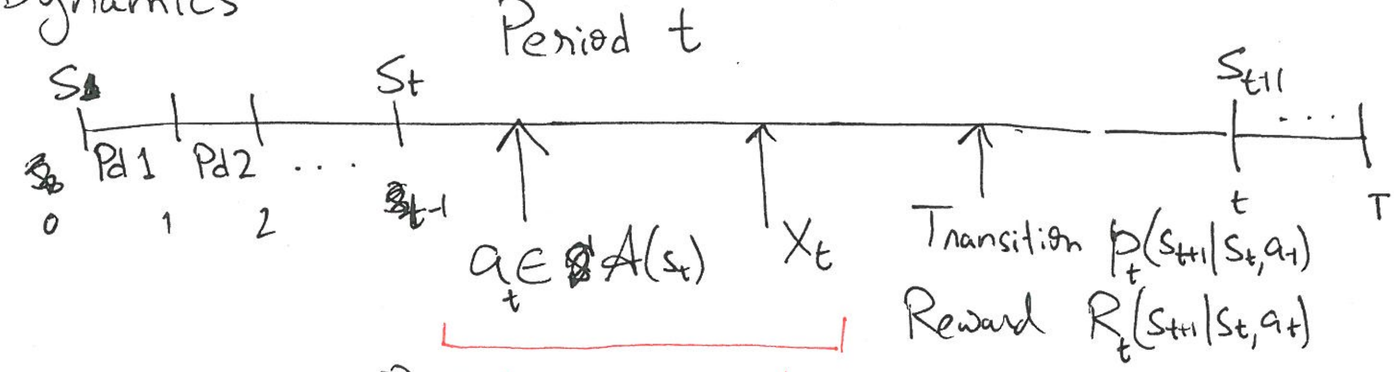
iv) Randomness | Disturbance | Noise $\equiv$ X_t in period t

- determines transition $\boxed{p(s_{t+1} | s_t, a_t)}$
  (or $s_{t+1} = f(s_t, a_t, X_t)$)

v) Reward / Cost $\equiv$ $R_t(s_t, a_t, X_t)$ or $\boxed{R_t(s_{t+1} | s_t, a_t)}$

• Backward induction

– Dynamics



(*) order is crucial

– $\underbrace{V_t(s)}_{\text{'Value function'}} \triangleq$ maximum possible total expected reward over periods $t, t+1, \ldots, T$

$$V_T(s) = \max_{a \in A(s)} \sum_{x \in S} p(x|s,a)\, R_t(x|s,a)$$

Bellman Eqn

$$\boxed{V_t(s) = \max_{a \in A(s)} \sum_{x \in S} p_t(x|s,a)\Big(R(x|s,a) + V_{t+1}(x)\Big)}$$

$$\left(V_t(s) = \max_{a} \mathbb{E}_{X_t}\Big[R_t(s,a,X_t) + V_{t+1}\big(S_{t+1}(s,a,X_t)\big)\Big]\right)$$

(*) Note – What if X_t realized BEFORE a_t is executed?

Jensen's Ineq $\Rightarrow$ $\max_a \mathbb{E}_{X_t}[f(a,X)] \leq \mathbb{E}_{X_t}[\max_a f(a,X)]$

$\Rightarrow$ More info $\rightarrow$ higher rewards