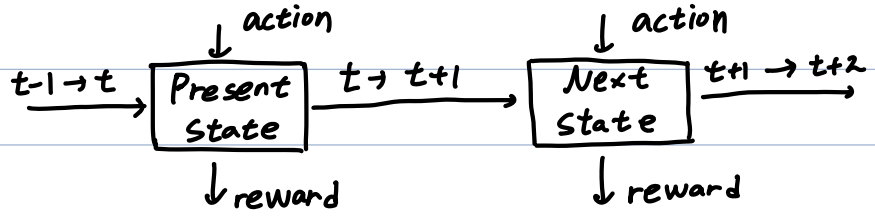
<5. Introduction to Markov Decision Process>

- 목차
  - Sequential Decision Process
  - Markov Decision Process
    - 정의
    - 요소 5가지
  - Decision Rule 4가지 종류
  - Policy (π)
  - Stochastic processes in MDP

* Sequential Decision Process

- 정의
  - 불확실성이 직면하여 어떤 목표를 향해 일련의 행동들을 수반하는 activity
  - 

* Markov Decision Process

- 정의
  - decision epoch을 기반으로 하는 stochastic process
  - 선택한 action과 process 내 state에 따라 일련의 reward를 받음.
- Elements of MDP
  - T - set of decision epochs, discrete ($T = \{1, 2, \ldots, N\}$)
    - MDP는 T의 상한선에 따라 finite horizon 혹은 infinite horizon을 가짐.
    - 마지막 N시점에는 no decision
  - S - the set of states that can be assumed by the process.
  - A_s - 상태가 s 일 때 취할 수 있는 action 집합, $A = \bigcup_{s \in S} A_s$
    - action은

* decision rule (d_t) category

① MD (Markovian + Deterministic)

- $d_t \in D_t^{MD} \sim d_t : S \to A$
- 현재 시점만 고려.
- 어떤 액션 취할 건지 정해져 있음.

② MR (Markovian + Randomized)

- $d_t \in D_t^{MR} \sim d_t : S \to P(A)$
- 현재 시점만 고려
- a probability distribution on the action set A_s
- action is chosen at random using the distribution

③ HD (History-dependent + Deterministic)

- $d_t \in D_t^{HD} \sim d_t : H_t \to A$
- T=1 부터 현 시점까지 모든 state와 action을 고려.
- $h_t = (s_1, a_1, \dots, s_{t-1}, a_{t-1}, s_t)$
- 어떤 액션 취할 건지 정해져 있음.

④ HR (History-dependent + Randomized)

- $d_t \in D_t^{HR} \sim d_t : H_t \to P(A)$
- T=1 부터 현 시점까지 모든 state와 action을 고려.
- $h_t = (s_1, a_1, \dots, s_{t-1}, a_{t-1}, s_t)$
- a probability distribution on the action set A_s
- action is chosen at random using the distribution

- D_t^K 는 다음과 같은 포함관계를 가지며, MD가 가장 specific 함.
  - $D_t^{MD} \subseteq D_t^{MR} \subseteq D_t^{HR}$
  - $D_t^{MD} \subseteq D_t^{HD} \subseteq D_t^{HR}$

* Policy (π)

- π는 MD or MR or HD or HR
- $\pi^K = D_1^K \times D_2^K \times \cdots \times D_{N-1}^K$ ($T=N$일 땐 no decision)
- 매 epoch 마다 동일한 π 사용하면 정책은 고정되어 있음).

* Stochastic processes in MDP

- $X_t \in S$ 가 t 시기에 시스템에 의해 발생하는 state 이고,

  $Y_t \in S$ 가 t 시기에 취해지는 액션일 때,

  discrete-time process $\{X_1, Y_1, X_2, Y_2, \cdots\}$ 를 말함.

Q. 둘 차이가 뭐지?

- $Z_1 = S_1$ 이고, $Z_t = (S_1, a_1, \ldots, S_{t-1}, a_{t-1}, S_t) = h_t$ $\forall t \in T$ 일 때

  history process 군 $= \{Z_1, Z_2, \ldots\}$
- 일반적으로 MDP는 $P^\pi(X_{t+1}=s \mid X_t=S_t, Y_t=a_t) = P_t(s' \mid S_t, a_t)$로 표현됨.
- $\pi \in D_t^{MD}$ 인 경우, $X=\{X_t ; t \in T\}$ 가 discrete time Markovian chain이 됨.