ما لىرنامدرىيى يوياى فطعى ١٠

م منال ما کشر دن فروس العادد افل به ول

· Ling	3	مہ 2	6	1	
and enoutable à de suis à	4	2	5	2	وروسته
acécemi in la cut il and nature.	5	4		3	

- A policy determines actions based on history and observation period.
- Define history sets:

$$\underbrace{H_n} = \Gamma^{n-1} \times S \text{ for } n > 0.$$

$$\underbrace{H_0} = S.$$

- A policy $\pi = (\pi_0, \pi_1, ...) \in \Pi$:
 - For any $n \ge 0$ and history $h_n = (i_0, a_0, \dots, i_n) \in H_n$:
 - $\pi_n(h_n)$ is a probability distribution on $A(i_n)$.

Question 1 What is a deterministic policy?

Question 2 What is a Markov policy?

ewlobiside with coloring : New . I cause -
$$T(t_{+} = (T_{+}, t_{+}, 0) \in TI_{m} : P$$
 in .

- For $n \ge 0$:
 - X_n : State at period n.
 - Δ_n : Action chosen at period n.
- The process $\{X_n, \Delta_n, n \geq 0\}$ is well-defined under any policy $\pi \in \Pi$.
- Under a Markov policy $\pi \in \Pi_M$ Forms a discrete-time Markov chain.
- For each $\pi \in \Pi$ and $i \in S$:
 - $P_{\pi,i}$: Probability under policy π with initial state i.
 - $E_{\pi,i}$: Expectation under policy π with initial state i.
- Reward structure:
 - Reward $r(X_n, \Delta_n)$ at period n is random.

Question How to compare different policies?

D Average reward: V_(TC) ijzlim in f 1 V_{19N} (TC))

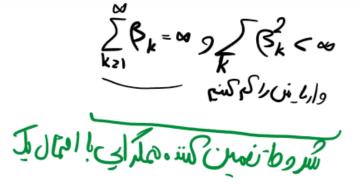
N→∞

Discounted criterion/total reward criterion

$$V_{\beta}(\pi,i) = \sum_{n=0}^{\infty} \beta^n \mathbf{E}_{\pi,i}(r(X_n,\Delta_n)), \quad i \in S, \pi \in \Pi$$

In the literature, the discount rate $\beta \in [0,1]$ is often assumed. Why? The optimal value function for this criterion is defined by:

$$V_{\beta,N}(i) = \sup_{\pi \in \Pi} V_{\beta,N}(\pi,i), \quad i \in S$$



مستوانيم كارا لرصب زمان تعريف كنيم.

A Gambling Problem [Ross, 2014]

At each play of the game, a gambler can bet any nonnegative amount up to his present fortune and will either win or lose that amount with probabilities p and q=1-p, respectively. The gambler is allowed to make n bets and his objective is to maximize the expectations of the logarithm of his final fortune. What strategy achieves this end?

$$\begin{array}{l}
\chi \longrightarrow V_{n}(X) = \max \left\{ P V_{n+1}(X_{+\alpha X}) + 9 V_{n+1}(X_{-\alpha X}) \right\} \\
V_{n}(X) = \log X \longrightarrow V_{n}(X) = \max \left\{ P \log (X_{+\alpha X}) + 9 \log (X_{-\alpha X}) \right\} \longrightarrow V_{n}(X) = \max \left\{ P \log (X_{+\alpha X}) + 9 \log (X_{-\alpha X}) \right\} \longrightarrow V_{n}(X) = C + \log X \longrightarrow V_{n}(X) = C + \log X \longrightarrow V_{n}(X) = 2 C + \log X \longrightarrow V_{n}(X) = n C + \log X \\
\frac{dV}{d\alpha} = \frac{P}{X_{+\alpha X}} - \frac{9}{X_{-\alpha X}} = 0 \longrightarrow \alpha_{\frac{1}{2}} \left(1 \longrightarrow P - 9 = \alpha \cup 2P \right) = \alpha
\end{array}$$

Sequential Investment Problem

Suppose one has an amount M of money and considers investing this money over N future periods. However, the opportunity for investment is not deterministic. At each period, an investment opportunity occurs with probability p, which is independent of the past and the amount of remaining money. When an investment opportunity occurs, if he invests x, he will earn a revenue r(x), including his investment. Assume that both his investment and his return at any period cannot be reinvested in the future. What is the optimal strategy for this problem?

Let $V_n(X)$ be the maximal expected profit when there are n periods remaining, X money available for future investment, and an investment opportunity occurs.

- 1. Write the optimality equation.
- 2. Assume that r(x) is nondecreasing, concave, and satisfies r(0) = 0. Show that $V_n(X)$ is also concave in X.

