Assignment 2 DPRL

Group 71: Gordon Gao (2802034) and Vishwamitra Mishra (2720410)

November 2024

Intro

In this assignment, we solved this problem with MRC, used Poisson Equations to calculate avg. costs in the long run, and found optimal policy for an inventory problem. In the beginning, we calculated the stationary distribution and found long-run average restock costs. Then, we solved Poisson Equations to find an unique solution for this problem. Finally, we used preventive policy for each state to find the Optimal policy.

A. State Space $\mathcal{X}$ and Action Space $\mathcal{A}$

The state can be defined as a tuple representing the inventory levels of both products.

$$\mathcal{X} = (I_1, I_2)$$

. Where I_1 and I_2 are the inventory levels of product 1 and product 2, respectively, such that $1 \leq I_1, I_2 \leq 20$

Actions depend on the state. Given the fixed policy, the only action occurs when the inventory level of either product drops to 1 and possible actions are:

$$\mathcal{A} = \{(O_1, O_2)\}$$

where O_1 and O_2 are the order quantities for product 1 and product 2, respectively, ensuring $I_1 = 5$ and $I_2 = 5$ after the order. For all other states, no action is taken.

B. Possible transitions and their probabilities

At each timestep, demand for each product (D_1, D_2) can be 0 or 1, each with a probability 0.5.

Let the current state be $S = (I_1, I_2)$. After the demand and potential orders:

- If $I_1 > 1$ and $I_2 > 1$, the new state is:

  $$\mathcal{X}' = (I_1 - D_1, I_2 - D_2)$$

  Transition probability:

  $$P(\mathcal{X}'|\mathcal{X}) = P(D_1) \cdot P(D_2) = 0.5 \cdot 0.5 = 0.25$$

  So, we have$P(D_1 = 0, D_2 = 0) = P(D_1 = 0, D_2 = 1) = P(D_1 = 1, D_2 = 0) = P(D_1 = 1, D_2 = 1) = 0.25$

- If $I_1 = 1$ or $I_2 = 1$, an order is placed to restock to 5 units. After the demand:

  $$\mathcal{X}' = (5 - D_1, 5 - D_2)$$

  Transition probabilities remain 0.25 per combination of demands.

C. Simulate the system for a long period under the given policy

Simulation involves the following:
Initializing the inventory state $\mathcal{X} = (5, 5)$. Simulating demand and applying the fixed policy over T timesteps.

Costs to be considered:

- Holding cost: $C_h = I_1 + 2 \cdot I_2$.
- Ordering cost: $C_o = 5$ if an order is placed.

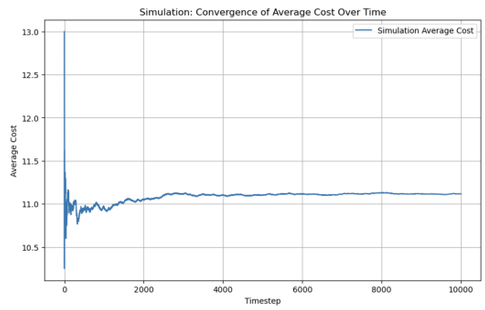

Figure 1: Convergence of average cost over time

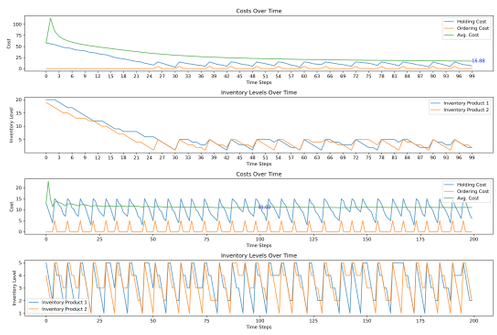

Figure 2: This is the simulation over 100 and 200 step, starting from {20,20} and {5,5}

Using the above simulation, compute the long-run average cost by the following formula:

$$\text{Average Cost} = \frac{\text{Total Cost}}{T}$$

Stopping Criteria: Simulation is run for a fixed number of steps (10000 timesteps). The Long-run average cost (Simulation-based) is: **11.1168**. When I increase the number of simulations, the long-run average converges.

D. Compute the limiting distribution using iteration

A Markov Chain analyzes The system's behavior over an infinite time horizon. We compute the stationary distribution $\pi(\mathcal{X})$ of the inventory states and use it to calculate the average cost.

Using the transition probabilities from the point above in section b, compute the limiting distribution $\pi(\mathcal{X})$ by solving:

$$\pi(\mathcal{X}) = \sum_{\mathcal{X}'} \pi(\mathcal{X}') \cdot P(\mathcal{X}' \to \mathcal{X})$$

Iterate until $\pi(\mathcal{X})$ converges. Use $\pi(\mathcal{X})$ to compute the long-run average cost:

$$\text{Average Cost} = \sum_{\mathcal{X}} \pi(\mathcal{X}) \cdot \text{Cost}(\mathcal{X})$$

Define a matrix P where each entry $P(\mathcal{X}'|\mathcal{X})$ represents the probability of transitioning from state $\mathcal{X}$ to state $\mathcal{X}'$. For example:

$P(\mathcal{X}'|\mathcal{X}) = P(D_1) \cdot P(D_2) = 0.25$ for all valid transitions.

With Limiting Distribution solve:

$$\pi = \pi P, \quad \text{where} \quad \sum_{\mathcal{X}} \pi(\mathcal{X}) = 1$$

to find the stationary distribution $\pi(\mathcal{X})$ (probabilities of being in each state). Using $\pi(\mathcal{X})$, calculate the average cost:

$$\text{Average Cost} = \sum_{\mathcal{X}} \pi(\mathcal{X}) \cdot C(\mathcal{X})$$

Iterative computation of π is stopped when:

$$\max_{\mathcal{X}} |\pi_{new}(\mathcal{X}) - \pi_{old}(\mathcal{X})| < \epsilon$$

where $\epsilon = 10^{-6}$ ensures convergence.

Result: Long-run average cost (Limiting distribution)is: **11.17196412789927**

E. Average-cost Poisson equation

The Poisson equation solves for the differential costs $J(\mathcal{X})$ and extracts the long-run average cost g. The fixed policy simplifies the Bellman equation to:

$$J(\mathcal{X}) + g = C(\mathcal{X}) + \sum_{S'} P(\mathcal{X}'|\mathcal{X}) J(\mathcal{X}')$$

Initialize $J(\mathcal{X}) = 0$ for all states. At each step, let's update:

$$J_{new}(\mathcal{X}) = C(\mathcal{X}) + \sum_{\mathcal{X}'} P(\mathcal{X}'|\mathcal{X}) J(\mathcal{X}')$$

Normalize $J(\mathcal{X})$ by subtracting the average cost g:

$$g = \frac{\sum_{\mathcal{X}} J_{new} - J_{old}}{\text{Number of States}}$$

We stop the iteration when:

$$\max_{\mathcal{X}} |J_{new} - J_{old}| < \epsilon$$

Result: The Long-run average cost (Poisson equation): **11.171945701357458**

F. Bellman equation by value iteration

The Bellman equation for minimizing the long-run average cost is:

$$J(\mathcal{X}) + g = \min_{A \in \{0,1\}} \left[C(\mathcal{X}, A) + \sum_{\mathcal{X}'} P(\mathcal{X}'|\mathcal{X}, A) J(\mathcal{X}') \right]$$

Where: The differential cost of being in the state $\mathcal{X}$. g: Long-run average cost. $A \in \{0, 1\}$: Possible actions:$A = 0$: Do not order. $A = 1$: Place an order.

$C(\mathcal{X}, A)$: Cost incurred in state $\mathcal{X}$ under action A.

$C(\mathcal{X}, 0)$: Holding cost (no order). $C(\mathcal{X}, 1)$: Fixed ordering cost + holding cost after restocking. $P(\mathcal{X}'|\mathcal{X}, A)$: Transition probability from state $\mathcal{X}$ to state $\mathcal{X}'$ under action A.

Solution:
Start with $J(\mathcal{X}) = 0$ for all states. Initialize the long-run average cost $g = 0$. Define transition probabilities $P(\mathcal{X}'|\mathcal{X}, A)$ for each state-action pair.

For each state $\mathcal{X}$, compute the cost-to-go for each action:

$$Q(\mathcal{X}, A) = C(\mathcal{X}, A) + \sum_{\mathcal{X}'} P(\mathcal{X}'|\mathcal{X}, A) J(\mathcal{X}')$$

Update $J(\mathcal{X})$ with the minimum cost-to-go:

$$J_{new}(\mathcal{X}) = \min_{A} Q(\mathcal{X}, A)$$

Update the optimal action:

$$A^*(\mathcal{X}) = \arg\min_{A} Q(\mathcal{X}, A)$$

Update the long-run average cost g as:

$$g = \frac{1}{|\mathcal{X}|} \sum_{\mathcal{X}} (J_{new}(\mathcal{X}) - J(\mathcal{X}))$$

Stop when the change in $J(\mathcal{X})$ across all states is less than a threshold ϵ.

Result: Long-run average cost (Bellman equation): **11.171945701357464**

G. Optimal Policy - Bellman Equation

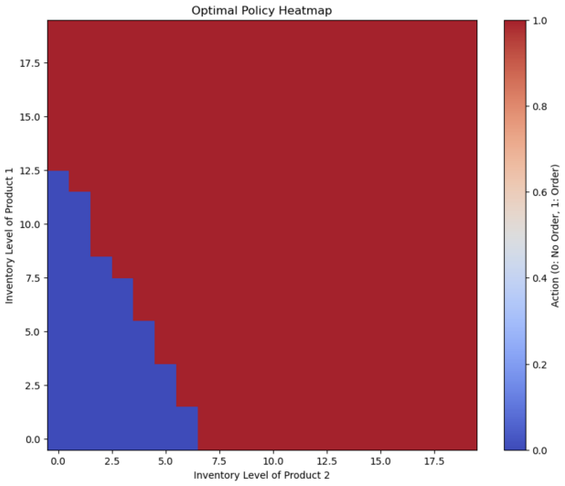

Figure 3: Optimal Policy Heatmap

The x-axis represents the inventory level of Product 2. The y-axis represents the inventory level of Product 1. The colors indicate the optimal action:

- 0 (No Order): Do not place an order.
- 1 (Order): Place an order to restock inventory.

The policy recommends "Order" in regions of low inventory for both products. As inventory levels increase, the policy shifts to "No Order," reducing unnecessary holding costs.