Optimizing Inventory Pricing Strategies Using Dynamic Programming

Dynamic Programming and Reinforcement Learning (DPRL) – Assignment 1

November 10, 2024

Group 57

Sara Abesová (2735679)
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
`s.abesova@student.vu.nl`

Matúš Halák (2724858)
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
`m.halak@student.vu.nl`

1 INTRODUCTION

Maximizing revenue through optimal pricing strategies is a crucial aspect of inventory management, especially for seasonal products with limited sales season.

This report addresses an inventory pricing optimization problem in which a product must be sold over 500 time periods with an initial inventory of 100 items. At each time period, the company has the option to set one of three price levels, each with a specific probability of sale. The objective is to maximize total expected revenue by determining the optimal price to set at each period, accounting for both remaining inventory and the number of periods left until end of the sales season.

To achieve this goal, we model the problem as a finite-horizon dynamic programming (DP) problem. By solving it backward from the final period to the initial state, we calculate the expected maximal reward and derive an optimal pricing policy, which specifies the best price for each state, defined by the inventory level and time remaining. Additionally, we explore a variant of the problem with an added constraint where prices cannot increase over time. This constraint simulates realistic scenarios in which raising prices over time may be impractical and deter customers from buying the product. To incorporate this constraint, we expand the DP state space to include the previous price level, allowing the model to track price history and ensure that the no-increase rule is followed.

Thus, our research question is: **What is the optimal flexible pricing policy to maximize revenue over a finite sales horizon, and how does it differ from a constrained pricing policy?**

We hypothesize that a flexible pricing strategy, where prices can be increased or decreased over time, will yield a higher total expected revenue than a constrained strategy. Specifically, we expect that allowing flexible pricing increases maximizes the revenue, because the product price can immediately adapt to successful or unsuccessful sales in the stochastic environment. In contrast, the no-increase constraint may limit the adaptability of the value function, potentially resulting in a lower expected revenue due to the restricted pricing options available as inventory and time diminish.

2 METHODS

For both flexible and constrained pricing strategies[1], we define time as a vector of time periods during a season when the seasonal product is sold, ranging from the beginning until the end of the season, after which the product becomes worthless.

$$t \in \mathcal{T} = (1, 2, \dots, 500)$$

2.1 Flexible Pricing

2.1.1 Part A: State Space and Action Space. With flexible pricing, we define the state space $\mathcal{X}$ as the set of all possible inventory levels left at any given time t ranging from empty to full inventory.

$$x_t \in \mathcal{X} = \{0, 1, \dots, 100\}$$

We define the action space as a vector of possible product prices that can be set in any given state at any given time x_t. With flexible pricing, the actions are independent of x_t.

$$a \in \mathcal{A} = (50, 100, 200)$$

2.1.2 Part B: Stochastic Finite-horizon DP Definition & Solution. The transition probabilities representing probability of sale at a given x_t given an action a are defined as follows:

$$p \in P = (0.8, 0.5, 0.1)$$

$$P(\text{sale} \mid a_i) = P(x_{t+1} = x_t - 1 \mid a_i) = p_i, \quad \text{for } i = 1, 2, 3$$

Using flexible pricing our immediate reward is both time and state-homogeneous and depends only on the product price set with the chosen action. However, for the DP algorithm, we use the expected immediate reward $\mathbb{E}r_t(x, a)$.

$$\begin{aligned} r_t(x, a_i) &= r(a_i) = a_i \cdot \mathbb{I}(U < p_i) \\ \mathbb{E}r_t(x, a_i) &= a_i \cdot p_i \end{aligned} \qquad \text{for } i = 1, 2, 3$$

Finally, to determine the maximal revenue $V_t(x)$ at each state and time we employ backward recursion using the Bellman equation, starting from the time horizon $t = 500$. For every state and time, this equation considers all the possible actions, and chooses one that maximizes revenue. To do this, the Bellman equation considers the expected immediate reward of taking an action in addition to a weighed sum of future rewards weighed by probability that a sale will or will not happen when taking that action. The boundary conditions are that at $t = 500$ the seasonal product becomes worthless, and therefore the maximal revenue for any remaining inventory level at that time is 0. An equally crucial boundary condition is that at depleted inventory $x = 0$ a sale cannot be made (meaning the weighed sum of future rewards in undefined at that point), and thus regardless of the action taken, the revenue for depleted inventory at any time is 0. Using this approach, $V_1(100)$ gives the maximum expected revenue possible to obtain in 500 time periods when starting at inventory level 100[2].

$$V_t(x) = \begin{cases} 0 & \text{if } t = 500 \text{ or } x = 0, \\ \max\limits_{a \in (50,100,200)} \left\{ \begin{array}{c} \mathbb{E}r_t(x, a) + P(\text{sale} \mid a) \cdot V_{t+1}(x - 1) \\ + (1 - P(\text{sale} \mid a)) \cdot V_{t+1}(x) \end{array} \right\} \end{cases}$$

We obtain optimal policy $\alpha_t(x)$, within the same backward recursion loop, for only those times and inventory levels where we can make sales $t \in (1, \dots, 499)$ and $x \in (1, \dots, 100)$, as those are the only points where we need to and can make a decision about the product price. $\alpha_t(x)$ determines the optimal action to take at each x_t to maximize expected revenue earned from that x_t onward.

$$\alpha_t(x) = \arg \max_{a \in (50,100,200)} \left\{ \begin{array}{c} \mathbb{E}r_t(x, a) + P(\text{sale} \mid a) \cdot V_{t+1}(x - 1) \\ + (1 - P(\text{sale} \mid a)) \cdot V_{t+1}(x) \end{array} \right\}$$

2.1.3 Part C: Simulation of seasonal product sales under optimal policy. To verify that the maximum expected revenue is truly the expected revenue when following the optimal policy, we simulated the stochastic process of seasonal product sales 1000 times and computed the revenue earned after 500 time periods, when starting with 100 inventory, 0 revenue and following $\alpha_t(x)$ at every x_t. For the simulation, we used the stochastic definition of immediate reward $r(a) = a \cdot \mathbb{I}(U < p_a)$ for 500 time periods. At each x_t we sampled

[1] Note that while the provided definitions are specific to the assignment, our implementation is generalizable and works for season of any length, inventory of any size and any number of actions with corresponding prices and sale probabilities.

[2] A powerful feature of the DP approach here is that in general, $V_t(x)$ gives the maximum expected revenue possible to obtain in $500 - t$ time periods starting at time t with x inventory level.

from $U(0,1)$ and if sampled value was lower than p_a corresponding to a from $\alpha_t(x)$, the immediate reward was added to the revenue and 1 was subtracted from the inventory level. We computed the average and plotted a histogram of revenues earned across the 1000 simulations. We also plotted the changes in inventory level and revenue over time for an example simulation run.

2.2 Constrained Pricing

2.2.1 Part D: Constrained State Space and Constrained Action Space. Under constrained pricing, we need to adjust the state space $\mathcal{X}$, so that any x_t captures not only the remaining inventory, but also reflects the action taken and price set at time t. The action space $\mathcal{A}$ now depends on both x and t and contains the conditions that the prices allowed to be set at any x_t must be greater or equal than prices set in the future, taking into account both the case that a sale will and will not be made in the future.

$$\begin{aligned} x_t \in \mathcal{X}_{\text{constrained}} &= \{(0, a_{t,x}), (1, a_{t,x}), \ldots, (100, a_{t,x})\}, \\ a_{t,x} \in \mathcal{A}_{\text{constrained}} &= \{50, 100, 200\} \mid \forall t < 500, \forall x > 0 : \\ & \qquad a_{t,x} \geq a_{t+1,x} \land a_{t,x} \geq a_{t+1,x-1} \end{aligned}$$

2.2.2 Part E: Constrained Stochastic Finite-horizon DP Definition & Solution. Having defined the new state space $\mathcal{X}$ as a set of (inventory, price) tuples, and the new actions space $\mathcal{A}$ with the constraints of possible prices to be set at any x_t, the rewards and algorithm described in 2.1.2 works exactly the same, with no necessary adjustments. However, for implementation, it was most efficient to implement the changes in $\mathcal{X}$ and $\mathcal{A}$ in the backward recursion algorithm itself. Therefore, while the logic and interpretation remains the same, the values of $V_t(x)$ are actually (revenue, price) tuples, and the price element of each tuple is the optimal policy $\alpha_t(x)$ at that x_t. The revenue element of $V_1(100)$ gives the maximum expected revenue possible to obtain in 500 time periods when starting at inventory 100 under the constrained pricing strategy.

$$\alpha_t(x) = \arg \max_{a_{t,x} \in \mathcal{A}_{\text{constrained}}} \left\{ \begin{array}{l} \mathbb{E} r_t(x, a_{t,x}) \\ +P(\text{sale} \mid a_{t,x}) \cdot V_{t+1}(x-1) \\ + \left(1 - P(\text{sale} \mid a_{t,x})\right) \cdot V_{t+1}(x) \end{array} \right\}$$

$$V_t(x) = \begin{cases} (0,0) & \text{if } t = 500 \text{ or } x = 0, \\ \max\limits_{a_{t,x} \in \mathcal{A}_{\text{constrained}}} \left\{ \begin{array}{l} \mathbb{E} r_t(x, a_{t,x}) \\ +P(\text{sale} \mid a_{t,x}) \cdot V_{t+1}(x-1) \\ + \left(1 - P(\text{sale} \mid a_{t,x})\right) \cdot V_{t+1}(x) \end{array} \right\}, \alpha_t(x) \end{cases}$$

3 RESULTS & DISCUSSION

3.1 Flexible Pricing Policy

The results show that the optimal policy suggests starting with the highest price (200) when inventory levels are high and there are many time periods remaining. This way, the revenue per item sold is maximized, as there is no need to sell quickly. If sales are going well, it is possible to sell out all inventory for the higher price. However, because of the lower sale probability with the maximum price, this is generally not the case and the optimal policy makes use of the full time horizon to sell out the stock at as high a price as possible, see example simulation run in Figure 3. As time progresses towards the horizon, the policy shifts to a lower, mid-level price (100) to increase the probability of selling items, helping ensure that all stock is sold by the end of the time horizon. Interestingly, even though there's no restriction on increasing prices, once prices decrease, they do not increase again under the optimal policy. This behavior occurs because, with a limited number of time periods, raising prices again could reduce the probability of selling the remaining inventory in time, resulting in potential losses for the company (as unsold inventory generates no revenue).

The expected maximal reward obtained using the flexible pricing strategy is 13,715.79, representing the expected revenue achievable by following this policy starting from $t = 0$ and full inventory $x = 100$. In 1,000 simulations described in 2.1.3, the average revenue was 13,714.5, see Figure 2. The close match between the expected revenue and average revenue earned by following the optimal policy $\alpha_t(x)$ at every x_t indicates that the optimal policy reliably, effectively and consistently achieves results close to the highest expected revenue.

3.2 Constrained Pricing Policy

Figure 4 4 illustrates the optimal pricing strategy under the constrained policy, where price increases are not allowed over the sales horizon. The policy begins with the highest price (200) and maintains this level for most of the time. Since the constraint prevents any future price increases, the policy strategically holds onto the highest price as long as possible, taking into account that it cannot be used again once a lower price is chosen. In the final portion of the sales period (approximately the last 100 time periods), as the time horizon approaches and some inventory still remains, the policy shifts to a lower price (100). The price reduction increases the likelihood of selling the remaining items, helping to ensure that all stock is cleared before time runs out. By holding high prices at first, this strategy aims to maximize revenue early on and only lowers prices when needed to sell out inventory by the end. The restriction on price increases limits the policy's flexibility, which, as expected, leads to a slightly lower expected maximum reward of 11,733.33 compared to the flexible pricing approach.

4 CONCLUSION

In this report, we compared flexible and constrained pricing policies for a seasonal product using DP techniques. The results showed that a flexible policy, which does not limit changes in price, generates more revenue. On the other hand, the constrained policy, which only allows price decreases, limits the company's responsiveness to demand changes, resulting in a slightly lower revenue.

Our findings highlight the importance of adaptable pricing strategies for maximizing revenue in a limited sales horizon.

APPENDIX

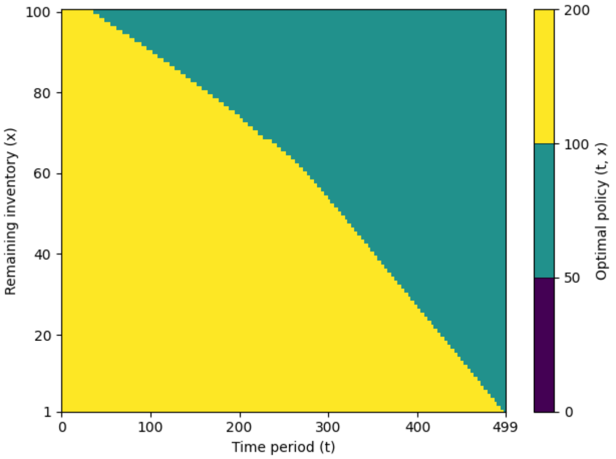

Figure 1: Optimal Policy $\alpha_t(x)$ (Flexible pricing)
Optimal policy $\alpha_t(x)$ when using flexible pricing across 500 time periods. The color-coded regions indicate the optimal price levels for each combination of remaining inventory and time period x_t.

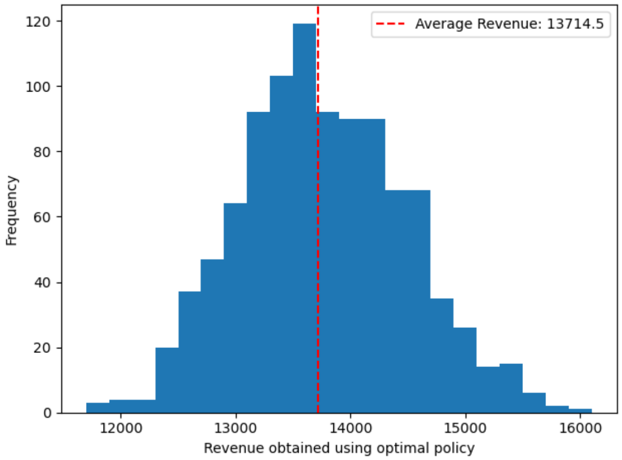

Figure 2: Simulated Revenue Distribution following $\alpha_t(x)$
Frequency distribution of revenue outcomes obtained from 1000 simulations following the optimal policy $\alpha_t(x)$ with flexible pricing. The average revenue is indicated by the red dashed line.

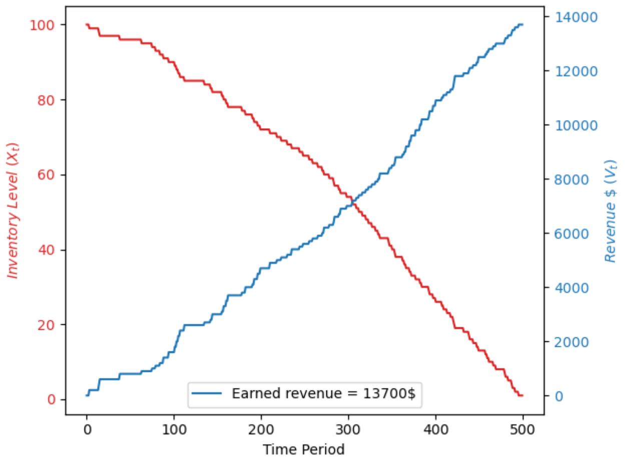

Figure 3: Sample Run following $\alpha_t(x)$
Evolution of inventory level (left y-axis) and accumulated revenue (right y-axis) over the 500 time periods in a single simulation run, following the optimal policy $\alpha_t(x)$ with flexible pricing.

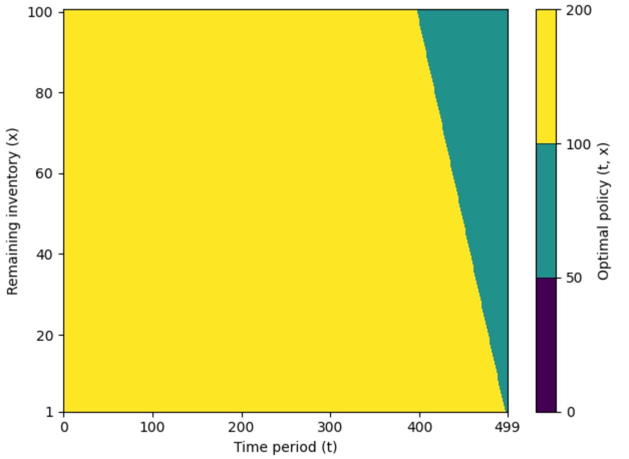

Figure 4: Optimal Policy $\alpha_t(x)$ (Constrained pricing)
Optimal policy $\alpha_t(x)$ under constrained pricing conditions over 500 time periods. The color-coded areas represent the optimal price levels to be set for each combination of remaining inventory and time x_t, with prices restricted from increasing over time.