IE361 – Stochastic Models in Operations Research

Case Study – Airline Pricing Policy

GROUP NUMBER 12

"Academic integrity is expected of all students of METU at all times, whether in the presence or absence of members of the faculty. Understanding this, I declare that I shall not give, use, or receive unauthorized aid in this study."

Abderrahmane Harkat	2415297
Alperen Oktay Şahin	2305373
Youssef Nsouli	2487494

Table of Contents

Introduction

In this case study, the sales policy of an airline firm will be studied, managing its revenue using a dynamic programming approach; price discrimination is a very common approach for firms that want to maximize the revenue from a certain flight. The firm wants to maximize its total revenue by setting the prices over time according to a couple key factors: time and empty seats. These temporal differences make the flight company decide the prices for every selling period dynamically. These decisions are centered about the understanding of the customers' willingness to pay for the ticket, and actually boarding the plane, in different situations.

Main Assumptions and Restrictions:

The system in question is a single period system. As in, only one flight is considered. Also, it is a single-category problem which means, for example, only economy tickets are sold to the customers rather than different types of tickets, including business class, first class, and the like. Moreover, this company is working as a monopolistic company, where competition is not taken into account. Any service cost is ignored and not accounted for. Ticket cancellations are not a possibility. Therefore, when a ticket is bought by a customer, the firm takes the money for sure.

In addition to these assumptions, batch booking is not allowed, meaning tickets are sold individually. For the situation given in the system, there is a trade-off between lower and higher prices. Lower prices can be chosen to early customer arrivals. On the other hand, with higher prices, more money can be acquired from late customer arrivals.

Tradeoff Considerations

1 seat remains, 5 periods left	vs	1 seat remains, 1 period left
5 seats remain, 5 periods left	vs	5 seats remain, 1 period left

Table 1. Different Situations for Price Determination

In table 1, two different situations are given to understand the problem easily. When the number of remaining seats is constant and the number of periods left increases, the posted price is expected to increase. In addition to this, when the time left is constant and the number of seats remaining increases, the posted price is again expected to increase.

The sales horizon is divided into very small parts. So, at most, one customer can arrive at the given time interval. Therefore, a customer arrives under a probability denoted as ρ_t.

With these stated assumptions and restrictions, the company will try to maximize the expected revenue by manipulating ticket prices at, possibly, every time stage over the entire sales horizon. The dynamicity of the problem allows the airline to utilize dynamic programming to model an optimal pricing policy. As such, a maximization dynamic programming formulation is to be considered.

Question 1

Part a)

There are a total of S seats, and the total time until the flight is given as T. Also, overbookings are not to be considered. Meaning, the seat limit, S, cannot be exceeded.

Parameters:

- Time until departure, t:

  t is the smallest time interval in T so that only one customer can arrive. t is a decreasing value, starting from whenever seats can be booked, denoted as t_{start}, and decreasing all the way to 0, denoting flight departure. Since it is best to consider how the formulation changes as t decreases, the DP formulation will be backwards recursive. In the recursive formulation, f_t will be given in terms of $f_{t-1}$and f_t

- Number of seats booked, s:

  s corresponds to the number of seats booked starting from t_{start} until current time, t. s increases from a value 0 (meaning no seats have been booked) and cannot exceed the seat limit S. If a booking occurs, s is revised as $s + 1$ for the next period. Otherwise, it remains as s.

- Customer Arrival Rate, ρ_t:

Under a probability, ρ_t, a customer comes and faces a posted price. In a given time interval, a customer either arrives or not, and if they do arrive, then they either book a seat or reject it. The probability is independent of how many seats there are and independent of how many customers have already arrived, it only depends on t.

- Reservation Price, P_t:

The reservation price is the maximum price a customer is willing to pay for the airplane seat. Each customer has a different level of willingness to pay. If the posted price by the company is less than or equal to the reservation price, the customer goes for the seat and books it, otherwise, the customer does not buy a ticket. If a booking occurs, the number of seats booked, s, advanced to $s+1$.

In addition, reservation price is a time-dependent continuous random variable; meaning that, as time progresses, the distribution of the random variable changes accordingly, perhaps increasing to simulate the rush to buying tickets days before a flight. Therefore, the distribution of P_t is very crucial to the decision-making process of the company.

With all the given information, assumptions, and restrictions, the model is as follows:

Dynamic Programming Model

- Stage: The time interval, t
- State: The number of booked seats, s
- Function: $f_t(s)$: Maximum (optimal) expected revenue from period $t, t-1, \ldots 2, 1, 0$ when s seats have been booked
- Decision Variable: The price per seat at s and t, x_{ts}
- Recursive Function:

$$f_t(s) = \max_{x_{ts}} \{\rho_t[P(x_{ts} < P_t)\big(x_{ts} + f_{t-1}(s+1)\big) + P(x_{ts} \geq P_t)f_{t-1}(s)] + (1-\rho_t)f_{t-1}(s)\} \ \forall\, s < S$$

$$f_t(S) = f_{t-1}(S) \qquad\qquad for\ s = S$$

- Boundary Condition: $f_0(s) = 0\ \forall\, s \in [0, S]$

- Solution: $f_T(0)$

As a stage definition, the time to flight departure, t, is chosen so transactions between states are in the T time domain. As time goes on, stages are changing and decrease one by one.

As a state definition, the number of seats that have been booked, s, is considered instead of the remaining seat inventory, but the former is preferred.

When the seat booking limit S is reached, the following periods will remain at a state S since overbooking is not allowed.

The probability of accepting the posted price is $P(x_{ts} \leq P_t)$, an x_{ts} amount of revenue will be gained and the system enter the next stage as $f_{t-1}(s+1)$. If the posted price is rejected with probability $P(x_{ts} > P_t)$, then no revenue will be earned and the system will go to the next stage as $f_{t-1}(s)$. Moreover, there is a probability of a customer not coming at that t stage. That probability is $(1 - \rho_t)$ and the next stage will be $f_{t-1}(s)$ because of no possibility to sell a ticket.

Part b)

Under the assumption that overbooking is allowed under a limit η and d as the amount of penalty incurred for each denied boarding, it is assumed that the penalty will be incurred at the end of the sales horizon, i.e.: during flight. So, the DP model will be modified so that:

- $f_0(s) = -\pi(s)\ \forall\ s \in [0, S+\eta]$
- $\pi(s) = \begin{cases} \left(\sum_{k=S}^{s} \binom{s}{k} (1-\beta)^k \beta^{s-k}\right) d & if\ s > S \\ 0 & if\ s \leq S \end{cases}$

where β is the probability that a customer will not show up to the flight.

Basically, a random variable follows the binomial probability distribution where it is defined as the number of people who bought an airplane ticket and are going to show up, the experssion $\binom{s}{k}(1-\beta)^k\beta^{s-k}$ defines the probability that k number of people more than or equal to the plane's capacity will show up to the flight. This probability is then multiplied by the penalty amount for each denied boarding, giving the expected penalty for the denied boardings, if any.

Part c)

The parametrized problem now can be solved using an appropriate software package. The problem can be solved in multiple ways, and a heuristic approach (through simulation) can be done. After that, a more precise approach can be taken.

Heuristic Approach

The heuristic approach involves simulating a customer arriving (or not) and simulating a reservation price, which is uniformly distributed between \$41 and \$710. The heuristic approach involves selling a seat to the arriving customer at exactly their reservation price. This was simulated through Java, in the *"SimulatorSubOptimal.class"* file under the *"Simulation"* Folder, and the expected revenue, after 100,000 simulations, is \$20,349.30. Graphs and further explanation can be found in Appendix A. So, the DP formulation must beat and get a higher expected revenue.

Dynamic Programming Model

Since the reservation price is a uniform distribution that is independent of time, it is safe to substitute for P_t while disregarding time. Meaning that prices should only range from \$41 to \$710. If the price is at the lower limit, then that means all arriving customers will purchase and book a seat. A price of \$710, the upper limit, means that the seat pricing is too high for all customers and no one will be able to afford a seat, which effectively means that the airline is refusing to sell that particular seat. This is the equivalent of having $P(41 \leq P_t) = 1$ and $P(710 \leq P_t) = 0$. As such, we will limit our price per seat, x_t, to the range $[41{,}710]$.

In addition, the probability of the customer's reservation price being higher than the price can now be formulated:

$$P(x_{ts} \leq P_t) = 1 - F_{P_t}(x_{ts}) = 1 - \frac{x_{ts} - 41}{710 - 41} = \frac{710 - x_{ts}}{669}$$

The arrival rate is, too, time-stationary, where $\rho_t = (0.9)$. So, after substituting the parameters, the formulation is now as follows:

- $f_t(s) = \max_{x_{st} \in [41,710]} \left\{ (0.9) \left[\left(\frac{710 - x_{ts}}{669} \right) \left(x_{ts} + f_{t-1}(s+1) \right) + \left(\frac{x_{ts} - 41}{669} \right) f_{t-1}(s) \right] + (0.1)[f_{t-1}(s)] \right\} \; \forall \; s < 100$
- $f_t(100) = f_{t-1}(100) \; \forall \; t > 0$
- Boundary Condition: $f_0(s) = -\pi(s) \; \forall \; s \leq 100$
- Solution: $f_{144}(0)$

Where $\pi(s) = \begin{cases} (200) \left(\sum_{k=90}^{s} \binom{s}{k} (0.95)^k (0.05)^{s-k} (k - 90) \right) & if \; s > 90 \\ 0 & if \; s \leq 90 \end{cases}$

After implementing the usual DP formulation solving technique that consists of trying many, many values of x and finding the maximum in each iteration, the maximum expected attainable revenue is $f_{144}(0)$ =\$24,413.81. The solutions were found in the python file, *"Discrete_Prices.py"*, and are presented in the excel file, *"Optimal_Prices_Discrete.xlsx"*. It can be observed that most of the prices are \$355, which is half of the upper limit of the uniformly distributed reservation price. However, the price seems to increase the more the number of the booked seats reaches the maximum seating limit. For example, the optimal price at $t = 47$ for the values of s from zero to 62 is \$355, but slowly increases to $x_{47,97}$ =\$635 at $s = 97$. It is worth noting that since a maximum of one customer may arrive for each stage t, there are some values of t where certain s values cannot be reached, and a staircase of possible prices can be seen in the excel file.

Also, nowhere in any stage or state did the price hit one of its boundaries, whether it is the upper or lower one. The maximum optimal price over all is for $t = 44, s = 100$, at a price of $x_{44,100} = \$681$. So, the airline will always keep offering seats at different price levels until the overbooking limit is reached.

The discrete way of solving the DP Model takes a lot of time and computational resources, so it may be worthwhile looking into a more rigid and general way to approach the problem. Since the Optimality function, $f_t(s)$ does not have the price variable, x_{ts}, in its argument and the reservation price distribution is fairly simple, its derivative with respect to x_{ts} will be taken and equated to zero. It is also worth noting that the second derivative of $f_t(s)$ is negative for all values of x_{ts}, meaning it is a concave function that admits a global maximum.

$$\frac{d}{dx_{ts}} f_t(s) = (0.9)[(-\frac{1}{669})(x_{ts} + f_{t-1}(s+1) + \left(\frac{710 - x_{ts}}{669}\right)(1) + \left(\frac{1}{669}\right) f_{t-1}(s) = 0$$

$$710 - 2x_{ts} + f_{t-1}(s+1) + f_{t-1}(s) = 0$$

$$x_{ts} = \frac{710 + f_{t-1}(s+1) + f_{t-1}(s)}{2}$$

The respective formulation is formulated and solve in a Python package, *"Continuous_Price.py"*, and the result comes out to a maximum expected attainable revenue of $f_{144}(0) = \$24{,}413.81$. The optimal prices for each stage and state are provided in the file *"Optimal_Prices_Continuous.xlsx"*. The expected revenue is not different than the discrete prices case (with the exception of a difference in the magnitude of 10^{-5}), but the continuous case is more accurate and much faster to calculate. As expected, the same features and behaviors presented in the excel file of the optimal pricing in the discrete case can also be found in the continuous case. As such, the continuous prices will be used.

Simulation using the Optimal Policy

To mostly confirm and determine whether the DP model is valid, and to find, on average, how many seats are booked and sold, a similar simulation is formed as the heuristic approach, but instead, the seat is sold if a customer arrives and has a reservation price higher than or equal to the optimal price for that seat, according to the current stage, t, and state, s. After running 100,000 simulations on Java, through the *"SimulatorOptimal.class"* file, the average revenue is $24,396.70 and the average amount of seats that were sold is ~69 seats. Graphs and further explanations can be found in Appendix B.

Part c-i)

Seat inventory levels of -5 and 20 correspond to $s = 95$ and $s = 70$, respectively. The graph of optimal prices over time for each of the two specified inventory levels can be found in figure 1: *Note: Since the time value is decreasing from T to zero, the t-axis is inverted.*

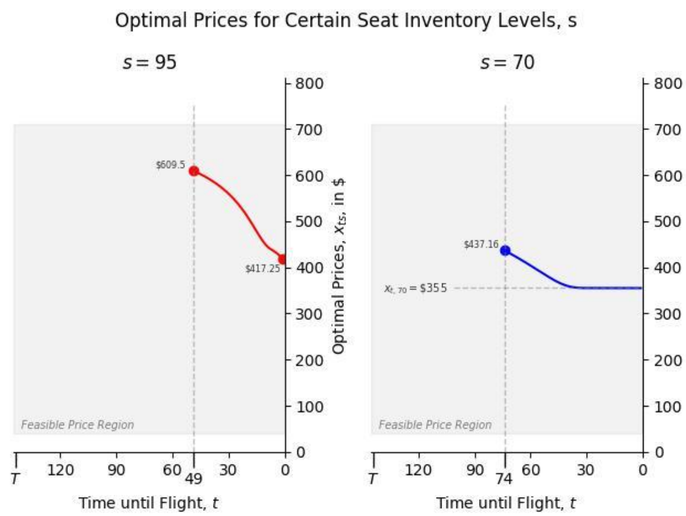

Figure 1: Optimal Prices at Certain Seat Inventory Levels

The highlighted vertical line indicates the time at which it is possible to have that much inventory level s. It seems overall that the optimal price is decreasing as the time until flight decreases, perhaps indicating a sense of urgency to sell the tickets. The prices for $s = 70$ converge to $355, whereas the prices at $s = 95$ decrease but never cross the $400 mark.

Part c-ii)

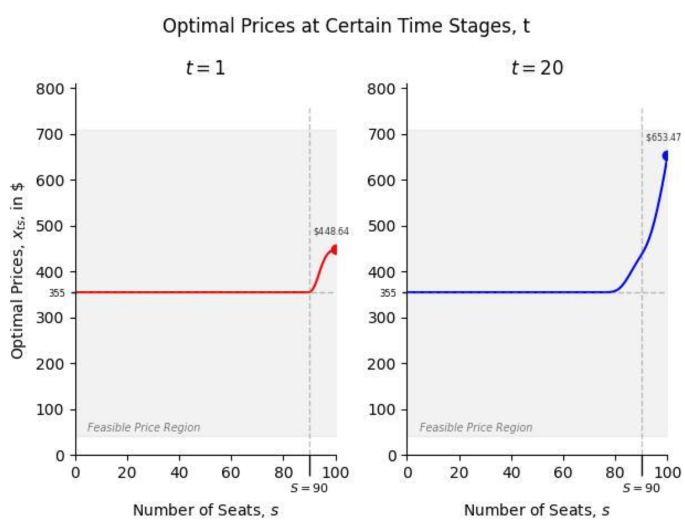

Figure 2: Optimal Prices at Certain Time Stages

As expected at $t = 1$, as number of seats approaches the maximum seating capacity, the price of the final seat increases to compensate for the penalty that may result from that one extra seat booking, as observed in figure 2. In the $t = 20$ case, the prices start soaring before that, possibly to prevent people from booking too many seats too early, in which the resulting penalty might be high if too many people show up.

Question 2

Part a)

To choose the optimal number of seats to be sold, the optimal revenue should be determined, the parameters are as follows:

- j: Price level; $j \in \{1,2, \dots m\}$
- μ_j: Expected Demand (number of seats sold) at a price level, j
- p_j: The price according to the price level, j
- π_j: The expected penalty incurred at a price level, j, which determines the expected demand

Mathematical Model

$$\textbf{Max}\, z = \{p_j \mu_j - \pi_j\}$$

Where: $\pi_j = \begin{cases} d(\mu_j - S) & if\ \mu_j > S \\ 0 & if\ \mu_j \leq S \end{cases}$

Constraints:

- $\mu_j \leq S + \eta$ (Maximum overbooking constraint)

Since there isn't much information about how the price level affects the demand, the mathematical model cannot be detailed much farther. The expected number of seats to be sold can be found from the value of μ_j that maximizes the respective revenue.

Part b)

The effect of price on demand is now known, and price $p_j = 20j$. If the value of j is chosen such that the revenue, $R_j = 20\mu_j j$ is maximized, then the price and expected demand can be determined as well. The relationship between expected aggregate demand and price is given as follows:

$$\mu_j = \begin{cases} 90 - \dfrac{20(j-1)}{20} & for\ j \in [1,10] \\ 80 - \dfrac{20(j-11)}{10} & for\ j \in [11,20] \\ 60 - \dfrac{60j - 1260}{20} & for\ j \in [21,30] \\ 30 - \dfrac{20j - 620}{5} & for\ j \in [31,36] \end{cases}$$

Since there is no information about the standard deviation of this *expected* aggregated demand parameter, it is not possible to determine a value for the penalty if more than 90 people show up. So, the incurred penalty will be ignored and/or deemed negligible. In addition, the model is of a non-linear nature, involving a piece-wise linear function. As such, the model is as follows:

Mathematical Model

Sets and Indices:

- $d \in D = \{1,2, \ldots 5\}$, domain d that bounds the piecewise function ranges
- $r \in R = \{1,2,3,4\}$, range r, the amount of ranges the piece-wise function is split into.

Parameters:

- μ_j: the expected demand for a chosen price level, j

Decision Variables:

- j: is the price level, which in turn determines the respective price per seat
- y_r: a binary variable that corresponds to the range, r, of the piece-wise function
- z_d: corresponds to the domain, d, of the piece-wise function

Objective Function:

$$\text{Max}\, z = (\mu_1 z_1 + \mu_{11} z_2 + \mu_{21} z_3 + \mu_{31} z_4 + \mu_{36} z_5)(20j)$$

$$\text{Max}\, \text{z} = [\text{Expected Seats to be Sold}] \times [\text{Price Per Seat}]$$

Note that the price, $p_j = 20j$, for a certain price level j

Constraints:

- $\sum_{d \in D} z_d = 1$ (All z_d must sum to 1)
- $\sum_{r \in R} y_r = 1$ (j can only fall within one range)
- $\begin{cases} z_1 \leq y_1 \\ z_d \leq y_d - y_{d-1} \ \forall \ 1 < d < 5 \\ z_5 \leq y_4 \end{cases}$ (z_d can take a value if the respective y_r is 1)
- $j = z_1 + 11z_2 + 21z_3 + 31z_4 + 36z_5$ (Decides the value of j, based on z_d)

Sign and Domain Restrictions:

- $j \in \{1,2, \dots 36\}$
- $z_d \in [0,1] \ \forall \ d \in D$
- $y_r \in \{0,1\} \ \forall \ r \in R$

The expected demand function is piece-wise liner, and the expected revenue function is concave, considering that it is also piece-wise quadratic (multiplication of two linear functions), and its second derivative is strictly negative. The constraints are linear, so the NLP admits a feasible, optimal solution.

After calculating the maximum expected attainable revenue using a Python package, *Pyomo*, in the file, *"NLP Model.py"*, the optimal value that maximizes the revenue of j is $j = 21$, which corresponds to $p_{21} = \$420$ and a revenue $R_j = \$25{,}200$. This is more than the expected revenue from the probabilistic case and dynamic price-changing case, but that is due to the parameter changes. The corresponding seats – on average – sold under this price level is $\mu_{21} = 60$, also comparable with the average seats sold in the Dynamic case. The expected revenue concaves downwards, as expected (Figure 3):

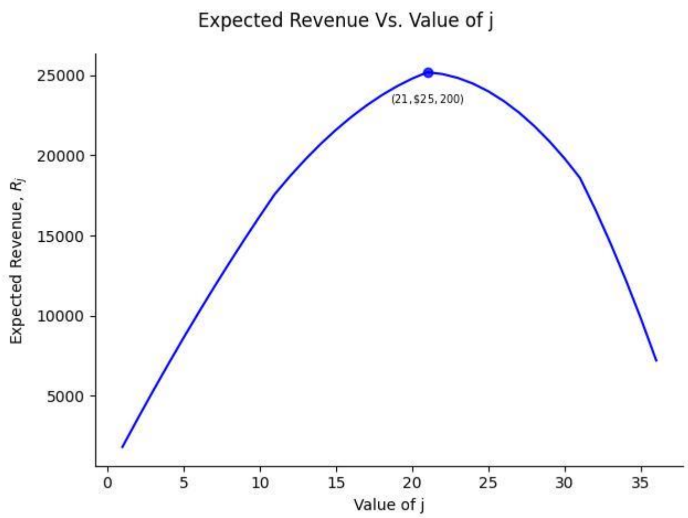

Figure 3: Expected Revenue Against Price Level, j

Part c)

The DP and MP (or NLP) model each approach the problem in a different manner. The DP Model-constructed prices are *dynamic* – as suggested by the title of the formulation – checking for optimality stage by stage. With each passing stage, a new price may be considered in order to optimize between high revenue per customer, more incentivized customers, or trying to cash out before the flight. Meaning that it is in the airline's best interest to update the system governing the prices at every stage. On the other hand, in the NLP model, the philosophy is different. The MP model optimizes the price at the beginning, and never changes it again. This price is determined by the expected demand and its interaction with the price level. Even if the actual seats sold are lower or higher than expected at a certain time, the NLP Model falls short in trying to accommodate for these aforementioned edge cases.

In order for the MP model to achieve a comparable result as the DP model, the NLP Model would be studied and run every so often, optimally every period, to determine how prices should vary depending on how many seats have been sold up to that point and according to how much time is left. After observing and updating the expected demand to fit with the current state of the system, the MP model would be run again and a new optimal price would be found. These series of steps should be done over and over again over the whole sales horizon so that the MP model can be comparable to the efficacy of the DP model.

Conclusion

In conclusion, an airline firm has considered maximizing their profit using various optimization models that each utilize different philosophies in approaching a problem. The pricing decisions were affected by multiple factors, including how much time is left until flight and how seats have been booked so far. Firstly, a dynamic programming approach is used to tackle the problem. The DP Model featured dynamic prices to optimize between the different decision-affecting factors. Secondly, a mathematical programming model is used to work out the problem. In this model, the prices were stable, unlike the former model.

At first, in the DP model, the prices range between \$355 and \$680.83 per seat. When the number of seats sold, s, increases, prices also increase, and as time till flight, t, decreases, prices also decrease. On the other hand, in the MP model, the price is stable and valid for all t values. When the problem is solved optimally with the given parameters, the price for the MP model is given at \$420 over the entire sales horizon without changing or varying. Despite the slightly varying parameters between the two situations, for some of the t and s values, the MP model gives a higher price than the DP Model; however, for other values, the DP model gives a higher price.

As the mathematical programming model approach is different than the dynamic programming approach, the way to find the optimal solution was different. Also, as both questions have some different informations and parameters, the optimal values differed. At the end of the case study, these two different approaches are compared to each other and the necessary comments are done.

Appendix A

Heuristic Simulation

The simulation involved simulation a customer arrival at a 90% chance and choosing a random number from 41 to 710 as their respective reservation price. Then, the seats were sold to **each** customer at their reservation price, meaning no customer (that has already arrived) backed out. In addition, to counter the penalty effect, after 90 seats are sold, the seats can only be booked by the customers if their reservation prices happen to be more than 200.

This simulation assumes that everyone who buys and books a seat shows up, and assumes that the algorithm governing the pricing policy knows exactly the reservation price of each customer. So, in almost all simulations, the entire plane was booked in addition to the overbooking limit. The mean revenue was $20,349.90, as discussed in the main body of this report, and 100 seats were booked in 100% of the simulations. The histogram of the data is below (Figure A1):

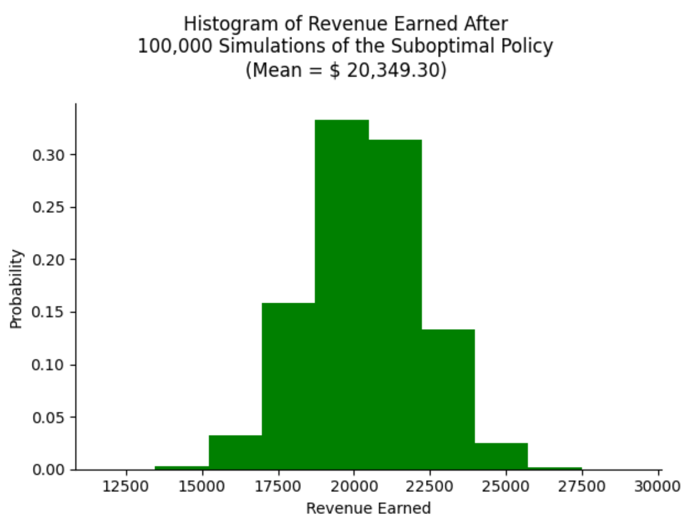

Figure A1: Histogram of Revenue Earned after 100,000 simulations of the suboptimal policy

Appendix B

The simulation is pretty much similar to previous one, with the exception that the optimal price has already been calculated and stored, so the algorithm (or the decision maker) waits until a customer with a high enough reservation price walks in and sells a seat. The penalty system assumes everyone shows up.

The average revenue was $24,396.70, and 68.74 seats were sold, on average. Less than 0.52% of simulations reached the overbooking region. The histograms are provided below:

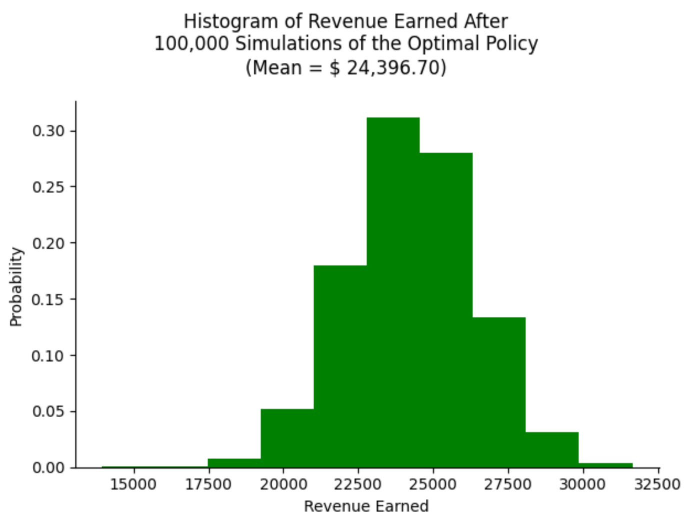

Figure B1: Histogram of Earnt Revenue after 100,000 simulations of the Optimal Policy

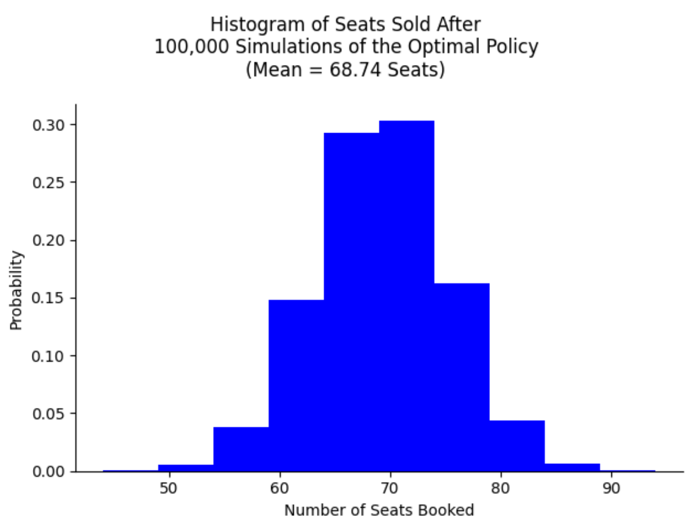

Figure B2: Histogram of Seats Sold after 100,000 simulations of the Optimal Policy