Shanghai University of Finance and Economics

Entrance Examination for 2024 Interdisciplinary Sciences Elite Program

Date: September 06, 2024
Time: 12:00pm – 3:00pm
There are 7 questions, and total score is 100.

1 Background

Mobile network providers typically offer their customers a variety of different data plans. In this problem, we examine how different pricing methods affect the profit (利润) in various scenarios.

Consider the following two pricing methods: volume-based pricing (按量收费) and monthly subscription fee (包月收费). In the first method, the plan has a single parameter R (RMB/GB), meaning that the customer will be charged $R \times D$ RMB if D GB are used. In the second method, the plan has two parameters F (RMB/month) and C (GB), meaning that the customer will be charged F RMB each month as long as no more than C GB are used.

We assume the cost (成本) for providing data is negligible (可忽略), and so the profit of the company is simply the total amount of money paid by all customers.

2 Customers with fixed demands and fixed budgets

In a market survey, we observe that there are N customers of the same type. They all have the same monthly demands (需求) D GB and are willing to pay up to P RMB for this service. Note that the customers will not pay anything if their demands cannot be fulfilled within their budgets (预算).

Question 1 (10 pts): What are the maximum monthly profits for each of the pricing methods, and the corresponding parameters R, F, C? Which pricing method is better for the company?

Answer 1: The maximum profits are $N \times P$ RMB in both cases. The profit cannot be higher as each customer are willing to pay at most P RMB and there are N of them.

For volume-based pricing, optimal $R = P/D$, so each customer pays P RMB.

For monthly subscription, optimal $F = P$ and $C \geq D$.

Both pricing methods are equally good.

Suppose there are 2 types of customers instead. There are N low-end customers and N high-end customers. The low-end customers are willing to pay P_1 RMB for D_1 GB/month, and the high-end customers are willing to pay P_2 RMB for D_2 GB/month. We assume that $P_1 < P_2 \leq 2P_1$, $D_1 < D_2$, and $P_1/D_1 < P_2/D_2 \leq 2P_1/D_1$.

Question 2 (10+10 pts): What is the profit using volume-based pricing with parameter R? You can consider different ranges of R and give a formula for each of the cases.

What is the maximum profit and the corresponding R?

Answer 2: We consider the following three cases.

Case 1: $0 \leq R \leq P_1/D_1$. The profit is $R \times N \times (D_1 + D_2)$. Maximized at $R = P_1/D_1$.

Case 2: $P_1/D_1 < R \leq P_2/D_2$. The profit is $R \times N \times D_2$. Maximized when $R = P_2/D_2$.

Case 3: $P_2/D_2 < R$. The profit is 0.

So, if $(P_1/D_1) \times N \times (D_1 + D_2) \geq (P_2/D_2) \times N \times D_2 = NP_2$, then the optimal $R = P_1/D_1$ and the profit is $NP_1 + NP_1D_2/D_1$. Indeed, we have $P_1(1 + D_2/D_1) \geq P_1(1 + P_2/(2P_1)) = P_1 + P_2/2 \geq P_2$.

Question 3 (10+5 pts): What is the maximum profit using monthly subscription fee, and the corresponding parameters F, C?

Compared with the volume-based pricing in Question 2, which pricing method is better for the company?

Answer 3: If we want to fulfill high-end customers only, the optimal (F, C) can be chosen to be (P_2, D_2) as the most we can get from each of them is P_2. The corresponding profit is NP_2.

If we want to fulfill all customers, the optimal (F, C) can be chosen to be (P_1, D_2) and the corresponding profit is $2NP_1$.

Since $2P_1 \geq P_2$, the optimal $(F, C) = (P_1, D_2)$ and the maximum profit is $2NP_1$.

Since $D_2/D_1 > 1$, $NP_1 + NP_1D_2/D_1 > 2NP_1$ and hence volume-based pricing is better than monthly subscription.

3 Customers with diminishing marginal utilities

In reality, customers often value (估值) the first few GBs higher than the rest, which means they are willing to pay more for the 1st GB than the 100th GB. This effect is called the law of diminishing marginal utilities (边际效用递减).

We assume that a customer with maximum demand D GB is willing to pay D RMB for the 1st GB, $D-1$ RMB for the 2nd GB, $D-2$ RMB for the 3rd and so on, until 1 RMB for the n-th GB. The total amount is thus the area of Figure 1. For simplicity, we approximate (近似) the figure with a triangle, and hence the customer is willing to pay the area of the trapezoid $Dx - x^2/2$ RMB for x GBs when $0 \le x \le D$, and $D^2/2$ RMB for more than D GB. Formally, when a customer uses x GBs, we define his valuation function (估值函数) as

$$V(x) = \begin{cases} Dx - x^2/2 & \text{if } 0 \le x \le D, \\ D^2/2 & \text{if } x > D. \end{cases}$$

Note that now x can also be fractional (小数).

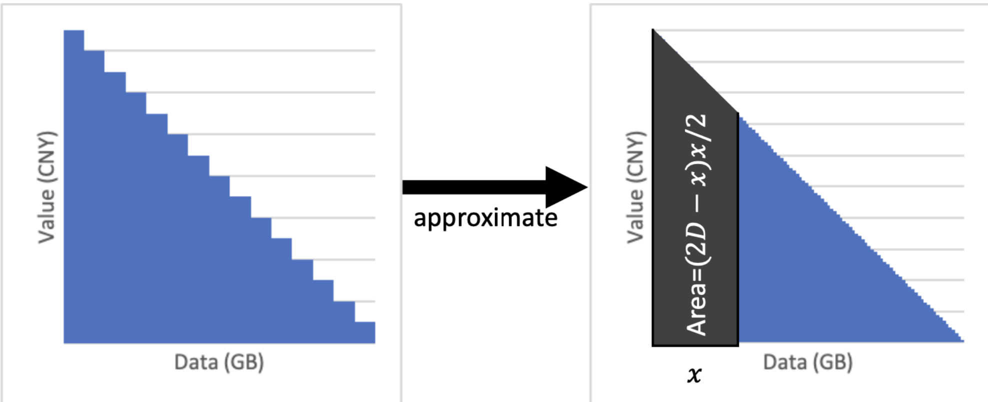

Figure 1: The customer values the first x GB at $(2D-x)x/2$ RMB.

The customer's utility function (效用函数) is defined as his valuation function subtracted by the amount he actually pays. In this scenario, the customer always chooses the x that maximizes (最大化) his utility function. In case of tie (相同效用), we assume the customer is willing to pay the most.

Suppose there are N customers with maximum demands D.

Question 4 (10+5 pts): What is the profit using volume-based pricing with parameter R? Note that the customers may choose to use less than D GB because of the diminishing utility.

What is the maximum profit and the corresponding R?

Answer 4: The utility of each customer for using x GB is $Dx - x^2/2 - Rx$ for $0 \le x \le D$ and $0 \le R \le D$, and it maximizes when $x = D - R$. Hence the profit is $R(D-R)N$. When $R > D$ the profit is 0 as the customers have negative utility for using any data.

The maximum profit is $D^2N/4$ at $R = D/2$.

Question 5 (5+5 pts): What is the maximum profit using monthly subscription fee, and the corresponding parameters F, C? Compared with the volume-based pricing in Question 4, which pricing method is better for the company in this scenario?

Answer 5: The maximum profit is $D^2N/2$ by setting $(F, C) = (D^2/2, D)$. The customers have 0 utility using this plan and negative utility for any plan with higher subscription fee.

The monthly subscription is always better in this scenario.

Suppose there are N customers with maximum demands D_1 and N customers with maximum demands D_2. We assume that $D_1 < D_2 \leq 2D_1$.

Question 6 (10+5 pts): What is the profit using volume-based pricing with parameter R?

What is the maximum profit and the corresponding R?

Answer 6: For $0 \leq R \leq D_1$, the profit is $R \times (D_1 - R) \times N + R \times (D_2 - R) \times N$. For $D_1 < R \leq D_2$, the profit is $R \times (D_2 - R) \times N$. For $D_2 < R$, the profit is 0.

Since the second case maximizes at $R = D_2/2$ which is $\leq D_1$ by assumption, we conclude that cases 2 and 3 are never optimal.

The function $RN(D_1 - R + D_2 - R)$ maximizes at $R = (D_1 + D_2)/4$ and the maximum profit is $(D_1 + D_2)^2N/8$.

Question 7 (10+5 pts): What is the maximum profit using monthly subscription fee, and the corresponding parameters F, C? Compared with the volume-based pricing in Question 6, which pricing method is better for the company in this scenario?

Answer 7: If we want to fulfill high-end customers only, the optimal (F, C) can be chosen to be $(D_2^2/2, D_2)$ and the corresponding profit is $ND_2^2/2$.

If we want to fulfill all customers, the optimal (F, C) can be chosen to be $(D_1^2/2, D_2)$ and the corresponding profit is ND_1^2.

So, the maximum profit is $\max(ND_2^2/2, ND_1^2)$, with corresponding parameters $(D_2^2/2, D_2)$ and $(D_1^2/2, D_2)$. Since $\max(ND_2^2/2, ND_1^2) \geq ND_2^2/2 \geq N(D_1 + D_2)^2/8$ as $D_2 > D_1$, we conclude that monthly subscription is always better in this scenario.

Shanghai University of Finance and Economics

Entrance Examination for 2023 Interdisciplinary Sciences Elite Program

Date: September 01, 2023
Time: 6:00pm – 9:00pm
There are 10 questions, and total score is 100.

1 Background

Inventory management (库存管理) is the process of planning and controlling the flow of goods in and out of a business. It helps to ensure that the business has enough products to meet customer demand, while avoiding the waste and expense of overstocking or understocking. Inventory management involves deciding when to order new products, how many to order, and how to store them efficiently. Suppose you own a bookstore that sells books, magazines, and stationery (文具). You want to keep track of how many items you have in stock (存货), how much they cost, and how fast they sell. You also want to avoid the problems of overstocking or understocking your products. If you store too few products, you may run out of stock (缺货) and lose sales, profits, and competitive edge. If you store too many products, you may waste money and space, and risk having unsold or damaged goods.

One common inventory problem is how to replenish (补充) inventory when it is depleted by customer demand. This problem can be modeled by assuming that the demand rate (单位时间需求量) for a product is constant and known, denoted by R, and that the product is ordered in batches. The costs involved in this problem are: (1). the cost of holding each unit of product in inventory per unit time, C_1; (2). the cost of being unable to meet each unit of customer demand per unit time, C_2; (3). the fixed cost of placing an order for a batch of (一批) products, C_3; (4). the cost of producing or purchasing each unit of product, K.

The goal of inventory management is to find the best ordering policy that minimizes the total cost per unit time, while satisfying customer demand. This policy specifies when to order a new batch of products, and how many units to order each time.

2 Inventory replenishment with zero shortage and instant delivery

To simplify the problem, we first assume that we never run out of stock. In other words, we set the shortage cost, C_2, to be very high. We also assume that we can order a batch of products and receive them instantly when our inventory level reaches zero. This way, we can avoid any delays or uncertainties in the supply chain.

Let t be the time interval between two consecutive orders, and Q be the quantity of products ordered each time. Figure 1 shows how the inventory level changes over time under this policy. We start with an inventory level of zero, and order Q units at time zero. Then, we sell the products at a constant rate of R units per unit time, until the inventory level drops to zero again. At this point, we order another batch of Q units, and repeat the cycle.

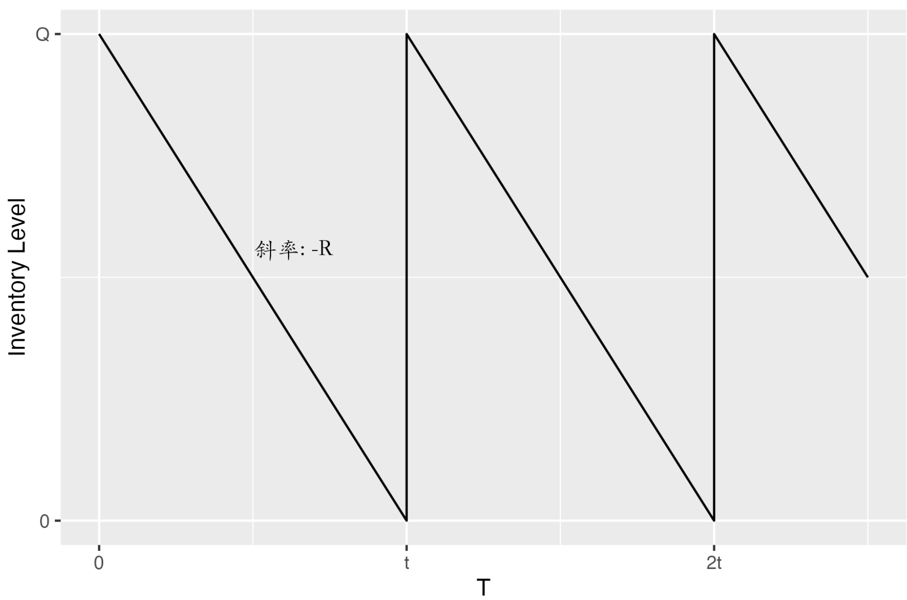

图 1: Inventory level over time when we order and receive products immediately.

We know the values of R, C_1, C_3 and K, which are the demand rate, the holding cost per unit product per unit time, the fixed cost per order, and the unit cost per product, respectively. Our goal is to find the optimal value of Q, which minimizes the total cost per unit time.

To help you understand the notations more clearly, we give you an example of how to calculate the total cost per cycle. The production or ordering cost per cycle is $C_3 + KQ$. The holding cost per cycle is $C_1Qt/2 = C_1Q^2/2/R$. Therefore, the total cost per cycle is

$C_3 + KQ + C_1Q^2/2/R$.

Question 1. What is the expression for the total cost per **unit time**, denoted by $C(Q)$, as a function of Q?

Answer 1: The total cost per unit time is

$$C(Q) = \frac{C_3 + KQ}{t} + C_1Q/2 = \frac{C_3 + KQ}{Q/R} + \frac{C_1Q}{2} = \frac{C_3R}{Q} + KR + \frac{C_1Q}{2}.$$

Now we determine when and by how much to replenish inventory through minimizing the total cost per **unit time**.

Question 2. What is the optimal value of Q, denoted by Q^*, that minimizes $C(Q)$? What is the corresponding value of t, denoted by t^*?

Answer 2: The optimal value of Q is given by

$$Q^* = \sqrt{\frac{2C_3R}{C_1}}.$$

The corresponding t^* is $t^* = Q^*/R = \sqrt{\frac{2C_3}{C_1R}}$.

3 Inventory replenishment with stockouts and backorders

Sometimes, it may be beneficial to allow some stockouts (缺货), or planned shortages, in inventory management. This means that we are willing to accept a delay in fulfilling some customer orders, if the holding cost of inventory (C_1) is too high compared to the shortage cost (C_2). The shortage cost is the cost of losing customer satisfaction or loyalty due to unavailability of products.

When we have stockouts, we keep track of the customer orders that are not met, and we call them backorders. Backorders are orders that are placed by customers but cannot be delivered immediately because the product is out of stock. The business promises to deliver the product as soon as possible, but the customer has to wait until then. This can affect the customer's perception of the business and its service quality. We fill the backorders as soon as we receive a new batch of products. In this case, the inventory level over time looks like Figure 2. It shows that when the inventory level reaches zero, we do not order immediately, but wait for some time, t_1, until we order a batch of Q

units. During this time, we accumulate backorders, which are filled when the new batch arrives. Then we sell the remaining at a constant rate of R units per unit time. **Note that there is no holding cost during the period** $(0, t_1)$**.**

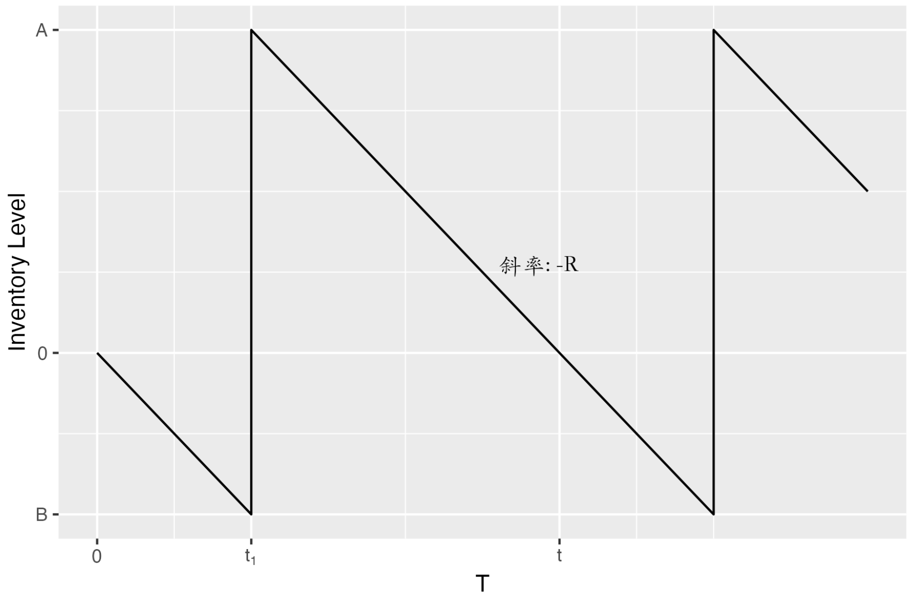

图 2: Inventory level over time when we allow some stockouts and backorders.

We know the values of R, C_1, C_2, C_3 and K, which are the demand rate, the holding cost per unit product per unit time, the shortage cost per unit product per unit time, the fixed cost per order, and the unit cost per product, respectively. Our goal is to find the optimal values of t_1 and Q, which minimize the total cost per unit time.

Question 3. What is the expression for the total cost per **unit time**, denoted by $C(t_1, Q)$, as a function of t_1 and Q?

Answer 3: The production or ordering cost per cycle $C_3 + KQ$. The holding cost per cycle is $C_1 * A/2 * (t - t_1) = C_1 * R * (t - t_1)^2/2 = C_1 * R * (Q/R - t_1)^2/2$. The shortage cost per cycle is $C_2 * B/2 * t_1 = C_2 * R * t_1^2/2$. Therefore, the total cost per unit time is

$$\begin{aligned} C(t_1, Q) &= \frac{C_3 + KQ}{t} + \frac{C_1 R(Q/R - t_1)^2}{2t} + \frac{C_2 R t_1^2}{2t} \\ &= \frac{C_3 R}{Q} + KR + \frac{C_1 (Q - t_1 R)^2}{2Q} + \frac{C_2 R^2 t_1^2}{2Q}. \end{aligned}$$

Now we minimize the total cost per **unit time** to determine when and by how much to replenish inventory.

Question 4. For each fixed Q, what is the optimal value of t_1, denoted by $t_1^*(Q)$, that minimizes $C(t_1, Q)$?

Answer 4: For each Q, the optimal t_1 is given by

$$t_1^*(Q) = \frac{C_1 Q}{(C_1 + C_2)R}.$$

Question 5. What are the optimal values of t_1 and Q, denoted by t^* and Q^*, that minimize $C(t_1, Q)$?

Answer 5: Substituting $t_1^*(Q) = \frac{C_1 Q}{(C_1+C_2)R}$ into $C(t_1^*(Q), Q)$, we have,

$$\begin{aligned} C(t_1^*(Q), Q) &= \frac{C_3 R}{Q} + KR + \frac{C_1(Q - \alpha Q)^2}{2Q} + \frac{C_2 \alpha^2 Q^2}{2Q} \\ &= \frac{C_3 R}{Q} + KR + \frac{C_1(1-\alpha)^2 Q}{2} + \frac{C_2 \alpha^2 Q}{2}, \end{aligned}$$

where $\alpha = C_1/(C_1 + C_2)$. Then the optimal Q is given by

$$Q^* = \sqrt{\frac{2C_3 R(C_1 + C_2)}{C_1 C_2}}.$$

The corresponding t^* is

$$t^* = \sqrt{\frac{2C_1 C_3}{RC_2(C_1 + C_2)}}.$$

4 Inventory replenishment with finite production rate

Sometimes, we cannot order a batch of products and receive them instantly. Instead, we need to produce the products ourselves, or wait for the supplier to produce them for us. We assume that we can start the production process whenever we want, but it takes some time to complete. We also assume that we can store the products in our inventory as soon as they are produced, without any delay in transportation. Let P be the constant production rate, which is higher than R. In this case, the inventory level over time looks like Figure 3.

We know the values of R, P, C_1, C_2, C_3 and K, which are the demand rate, the production rate, the holding cost per unit product per unit time, the shortage cost per

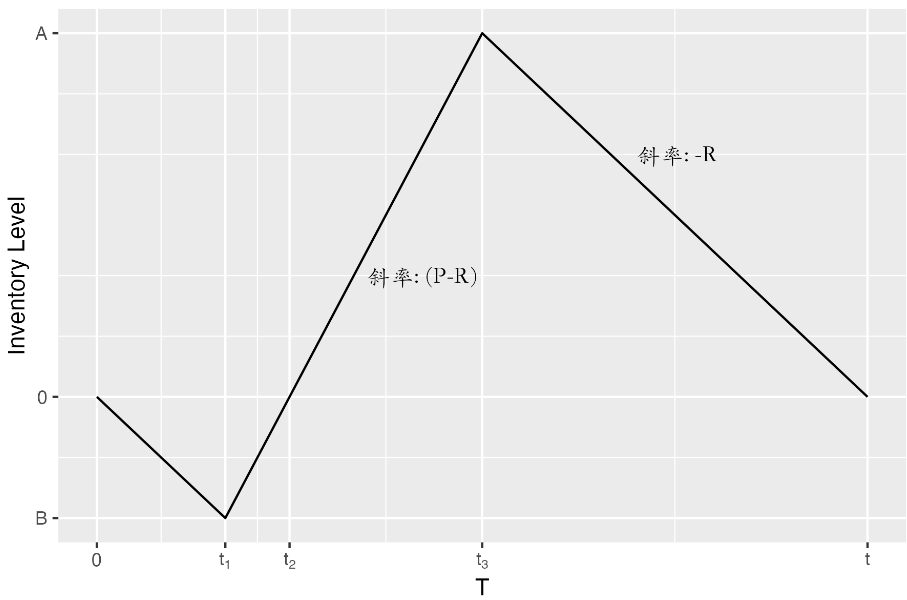

图 3: Inventory level over time when we have a delay in producing or receiving products.

unit product per unit time, the fixed cost per order, and the unit cost per product, respectively. Our goal is to find the optimal values of t_1 and Q, which minimize the total cost per unit time.

Question 6. What is the expression for the total cost per **unit time**, denoted by $C(t_1, Q)$, as a function of t_1 and Q?

Answer 6: First of all, $Q = Rt$, $B = Rt_1 = (P-R)(t_2-t_1)$, $A = R(t-t_3) = (P-R)(t_3-t_2)$. Then we have $t = Q/R$, $t_2 = Pt_1/(P-R)$, and $t_3 = t_1 + R/Pt = t_1 + Q/P$.

The production or ordering cost per cycle $C_3 + KQ$. The holding cost per cycle is $C_1 * A/2 * (t - t_2) = C_1 * (Q - Rt_1 - RQ/P) * (Q/R - Pt_1/(P-R))/2$. The shortage cost per cycle is $C_2 * B/2 * t_2 = C_2 * P * R * t_1^2/(P-R)/2$. Therefore, the total cost per unit time is

$$\begin{aligned} C(t_1, Q) &= \frac{C_3 + KQ}{t} + \frac{C_1(Q - Rt_1 - RQ/P)(Q/R - Pt_1/(P-R))}{2t} + \frac{C_2PRt_1^2}{2t(P-R)} \\ &= \frac{C_3R}{Q} + KR + \frac{C_1(Q-\beta Rt_1)^2}{2\beta Q} + \frac{C_2\beta R^2t_1^2}{2Q}, \end{aligned}$$

where $\beta = P/(P-R)$.

Now we minimize the total cost per **unit time** to determine when and by how much to replenish inventory.

Question 7. For each fixed Q, what is the optimal value of t_1, denoted by $t_1^*(Q)$, that minimizes $C(t_1, Q)$?

Answer 7: For each Q, the optimal t_1 is given by

$$t_1^*(Q) = \frac{C_1 Q}{\beta(C_1 + C_2)R}.$$

Question 8. What are the optimal values of t_1 and Q, denoted by t^* and Q^*, that minimize $C(t_1, Q)$?

Answer 8: Substituting $t_1^*(Q) = \frac{C_1 Q}{\beta(C_1+C_2)R}$ into $C(t_1^*(Q), Q)$, we have,

$$\begin{aligned} C(t_1^*(Q), Q) &= \frac{C_3 R}{Q} + KR + \frac{C_1(Q - \alpha Q)^2}{2\beta Q} + \frac{C_2 \alpha^2 Q}{2Q\beta} \\ &= \frac{C_3 R}{Q} + KR + \frac{C_1(1-\alpha)^2 Q}{2\beta} + \frac{C_2 \alpha^2 Q}{2\beta}, \end{aligned}$$

where $\alpha = C_1/(C_1 + C_2)$. Then the optimal Q is given by

$$Q^* = \sqrt{\frac{2C_3 R(C_1 + C_2)}{C_1 C_2}} \sqrt{\frac{P}{P - R}}.$$

The corresponding t^* is

$$t^* = \sqrt{\frac{2C_1 C_3}{RC_2(C_1 + C_2)}} \sqrt{\frac{P}{P - R}}.$$

5 Inventory management with stochastic demand

So far, we have assumed that the demand for the product is constant and known in advance. However, in reality, the demand may vary from day to day, and we may not be able to predict it accurately. In this case, we say that the demand is stochastic, meaning that it is random and follows a certain probability distribution.

For simplicity, we only consider a single-period inventory problem, such as a newsvendor who sells newspapers every day. The newsvendor needs to decide how many newspapers to order before knowing the actual demand for that day, and then sell them at a fixed price per unit. At the end of the day, the newsvendor earns a profit of k dollars for each newspaper sold, and incurs a cost of h dollars for each newspaper left unsold.

The newsvendor faces a trade-off between ordering too many newspapers and having excess inventory, or ordering too few newspapers and missing potential sales. The goal is to find the optimal order quantity, Q^*, that maximizes the expected profit per day.

Let R be the random variable that represents the demand for newspapers per day. Let $\mathrm{P}(r)$ be the probability that the demand is equal to r, where $r = 0, 1, \ldots$. We assume that $\sum_{r=0}^{\infty} \mathrm{P}(r) = 1$.

Question 9. What is the expected revenue if the order quantity is Q? (Hint: Use the formula for the expected value of a function of a random variable: $E[f(R)] = \sum_{r=0}^{\infty} f(r)\mathrm{P}(r)$.)

When the demand is r, the revenue is $kr - h(Q-r)$ if $r \le Q$ and kQ if $r > Q$. Therefore, the expected revenue is

$$\begin{aligned}
&\sum_{r=0}^{Q}\{kr - h(Q-r)\}\mathrm{P}(r) + \sum_{r=Q+1}^{\infty} kQ\mathrm{P}(r) \\
= \;&\sum_{r=0}^{Q}\{kr - h(Q-r)\}\mathrm{P}(r) + \sum_{r=Q+1}^{\infty} \{k(Q-r)\}\mathrm{P}(r) + \sum_{r=Q+1}^{\infty} kr\mathrm{P}(r) \\
= \;& kE(R) - \sum_{r=Q+1}^{\infty} \{k(r-Q)\}\mathrm{P}(r) - \sum_{r=0}^{Q}\{h(Q-r)\}\mathrm{P}(r).
\end{aligned}$$

Question 10. Define $F(Q) = \sum_{r=0}^{Q} \mathrm{P}(r)$. This is the probability that the demand is less than or equal to Q. Show that the optimal order quantity, Q^*, can be determined by finding the value of Q such that

$$F(Q-1) < \frac{k}{k+h} \le F(Q).$$

Let $C(Q)$ be the difference between $kE(R)$ and the expected revenue when the order quantity is Q. That is,

$$C(Q) = \sum_{r=Q+1}^{\infty} \{k(r-Q)\}\mathrm{P}(r) + \sum_{r=0}^{Q}\{h(Q-r)\}\mathrm{P}(r).$$

Then we have

$$C(Q+1) - C(Q) = (k+h)\left[F(Q) - \frac{k}{k+h}\right].$$

This implies that when $F(Q) < \frac{k}{k+h}$, $C(Q+1) < C(Q)$, while when $F(Q) \ge \frac{k}{k+h}$, $C(Q+1) \ge C(Q)$. Therefore, when $F(Q^*-1) < \frac{k}{k+h} \le F(Q^*)$, $C(Q^*)$ attains the minimum of $C(Q)$, or equivalently, the revenue attains the maximum.

Shanghai University of Finance and Economics

Entrance Examination for 2022 Interdisciplinary Sciences Elite Program

Date: September 02, 2022

Time: 6:00pm – 9:00pm

There are 10 questions, and total score is 100.

1 Background

The coronavirus disease 2019 (COVID-19) is a contagious disease caused by a particular virus. It has caused a global epidemic (疫情) up to the present year of 2022. Besides studying the biological attributes of the virus, it is also critical to understand the evolution of the epidemic in the population. There exists several classical models that describe how certain types of diseases spread among people. Such epidemiological models are useful tools to predict the future development of an epidemic.

2 A two-segment model

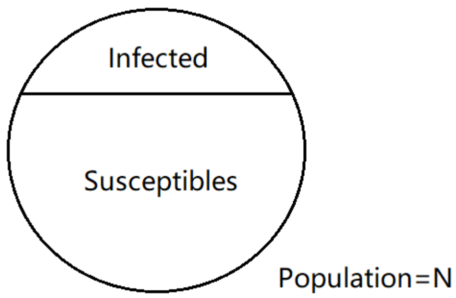

Figure 1: Division of population.

Assume there is an epidemic progressing in a population consisting of a fixed number of N people. Suppose once an individual gets the disease, he/she becomes infectious (有传染性的), and will not recover from the disease in the foreseeable future. However, the disease is not vital, meaning no people will die from it. In light of such facts, we divide the population into two disjoint (不相交的) groups: the *susceptibles* (待感染者) and the *infected* (已感染者). See Figure 1 for an illustration of the division. We keep track of the number of individuals from each group at the end of each day. In particular, at the end of day t ($t = 1, 2, \ldots$), the susceptible group includes S_t people, and the infected group includes I_t people. We also use S_0 and I_0 to denote the number from the two groups at the beginning of day 1. We assume $0 < I_0 < N$. It can be seen easily that $S_t + I_t = N$ for all $t = 0, 1, 2, \ldots$, where the population N is a constant that does not depend on t.

Now assume on day $t \geq 1$, every susceptible individual has the same probability $\beta I_{t-1}/N$ of getting infected due to contacts with people from the infected group. Here $\beta \in (0, 1)$ is a constant. Also, the event whether a susceptible individual gets infected is independent (独立于) of the event whether any other susceptible individual gets infected. ***On the average sense, the number of newly infected people on day t is counted as $\beta I_{t-1}S_{t-1}/N$.** Figure 2 below demonstrates the transition of the two groups. Thus we have the following recursive formula (递推式) for I_t:

$$I_t - I_{t-1} = \beta I_{t-1}S_{t-1}/N, \quad t = 1, 2, \ldots. \tag{1}$$

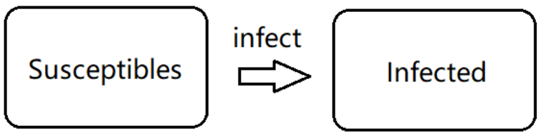

Figure 2: Transition of two groups.

As it turns out, it is more convenient to record the proportions of the susceptibles and the infected to the whole population, instead of recording their actual headcount. To this end, we define the two proportions: $s_t = S_t/N$, $i_t = I_t/N$. In order to understand how i_t and s_t changes day by day, we walk through some basic analysis.

Question 1 (10 pts): Using your knowledge of probability (概率知识), prove the sentence marked with star. That is, prove that the average/expected number (平均数或者期望数) of newly infected people on day t is $\beta I_{t-1}S_{t-1}/N$.

Answer 1: For each susceptible individual, whether he/she gets infected on day t is a Bernoulli random variable with success probability $\beta I_{t-1}/N$. The total number of newly infected people is then a binomial random variable with S_{t-1} trials and success probability $\beta I_{t-1}/N$. Its expectation is $\beta I_{t-1}S_{t-1}/N$.

Question 2 (10 pts): Write out two recursive formulas similar to (1), one for i_t and one for s_t. The quantities I_{t-1}, I_t, S_{t-1}, S_t should disappear in both formulas. Further show that, $\{i_t\}$ is an non-decreasing sequence and $\{s_t\}$ is a non-increasing sequence.

Answer 2: The two recursive formulas are:

$$i_t - i_{t-1} = \beta i_{t-1} s_{t-1}$$
$$s_t - s_{t-1} = -\beta i_{t-1} s_{t-1}.$$

Since we start from $i_0 \in (0, 1)$, we are guaranteed that $i_t - i_{t-1} \geq 0$ and $s_t - s_{t-1} \leq 0$ for all t.

Question 3 (10 pts): Suppose the epidemic starts with $i_0 \in (0, 1/2)$. We count the number of days until i_t exceeds $1 - i_0$. Let t^* be the largest t such that $i_t \leq 1 - i_0$. Prove that

$$t^* \leq \frac{1 - 2i_0}{i_0(1 - i_0)\beta}.$$

(Hint: First try to find a lower bound (下界) for the daily increase $i_t - i_{t-1}$, then find an upper bound (上界) for the total increase up to day t^*.)

Answer 3: The daily increase of $\{i_t\}$ is $\beta i_{t-1} s_{t-1} = \beta i_{t-1}(1 - i_{t-1})$, which is a quadratic function of i_{t-1}. When $i_{t-1} \in [i_0, 1 - i_0]$, we have that $\beta i_{t-1}(1 - i_{t-1}) \geq \beta i_0(1 - i_0)$. This holds true for the first t^* days. The total increment is then $\geq t^* \beta i_0(1 - i_0)$. On the other hand, the total increment is $\leq 1 - i_0 - i_0 = 1 - 2i_0$ by the definition of t^*. We must then have $t^* \beta i_0(1 - i_0) \leq 1 - 2i_0$, which leads to the conclusion.

Question 4 (10 pts): Prove by contradiction (反证法) that, as t grows larger and larger, i_t gets arbitrarily close to 1. The meaning of this result is, all people will eventually get infected. You can start the proof by assuming $i_t \leq 1 - \epsilon_0$ for all t with some small constant $\epsilon_0 > 0$. A contradiction can be reached by an argument similar to Question 3.

Answer 4: Assume $i_t \leq 1 - \epsilon_0$ for all t with some constant $\epsilon_0 > 0$. Then $i_t \in [i_0, 1 - \epsilon_0]$ for all t. The daily increase $\beta i_{t-1}(1 - i_{t-1})$ is then $\geq \min\{\beta i_0(1 - i_0), \beta \epsilon_0(1 - \epsilon_0)\} := c_0$. Since the total increment should be $\leq 1 - \epsilon_0 - i_0$, this amount of increase can last no more than $(1 - \epsilon_0 - i_0)/c_0$ days. This leads to a contradiction to the fact that $i_t \leq 1 - \epsilon_0$ for all large enough t.

3 Model with recovery

The model in the previous section ignores the fact that infected people may recover from the disease. Now assume that an infected individual may recover from the disease, and once recovered, he/she is no longer infectious. However, a recovered individual may later catch the disease again. For each infected individual, we assume that he/she recovers with probability $\alpha \in (0,1)$ independently on any given day. A more detailed explanation is the following: if John belongs to the infected group at the beginning of day t, then he recovers with probability α on day t. If he does recover, then he becomes a member of the susceptible group at the end of day t. If he does not recover on day t, then he still belongs to the infected group, and recovers with probability α on day $t+1$. The events whether he recovers on any particular day are mutually independent. On average, the proportion (to the whole population) of newly recovered people on day t is just αi_{t-1}. The transition of the two groups is illustrated in Figure 3.

We then have the recursive formulas

$$i_t - i_{t-1} = \beta i_{t-1} s_{t-1} - \alpha i_{t-1} \tag{2}$$

$$s_t - s_{t-1} = -\beta i_{t-1} s_{t-1} + \alpha i_{t-1}. \tag{3}$$

As usual, we assume $i_0 \in (0,1)$.

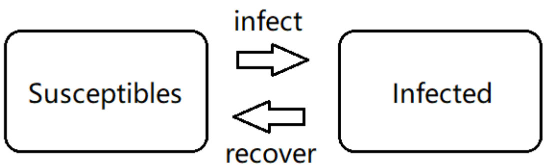

Figure 3: Transition of two groups with recovery.

One critical parameter in this system is $R_0 = \beta/\alpha$, which basically represents how contagious the disease is. The future progression of the epidemic largely depends on whether $R_0 < 1$ or $R_0 > 1$.

Question 5 (10 pts): Suppose $R_0 < 1$. Explain why $\{i_t\}$ is a non-increasing sequence.
Answer 5: Since $\beta < \alpha$ (from $R_0 < 1$) and $s_{t-1} \leq 1$, it must be that $\beta s_{t-1} - \alpha < 0$ for all t. From (2), we know that $i_t \leq i_{t-1}$.

Question 6 (10 pts): Suppose $R_0 > 1$. Assume that the two variables i_t and s_t approach their respective (各自的) *steady states* (平稳状态) i^* and s^* after a long enough period. In plain words, the *steady states* i^* and s^* are two constants such that $i_t \approx i^*$ and $s_t \approx s^*$ for all $t \geq T$ (T is some big integer). If we know $i^* \in (0,1)$, try to find the values of i^* and s^*.
Answer 6: For all $t \geq T$, we have $i_t \approx i^*, s_t \approx s^*$. Plugging these into (2) and (3), we get $0 = \beta i^* s^* - \alpha i^*$. Since $i^* > 0$, we have $s^* = \alpha/\beta = 1/R_0$. The complement is $i^* = 1 - 1/R_0$.

A more popular understanding of R_0 is the average number of people who will get the disease directly from the first infected individual. Now suppose there is a large population of N people who are completely healthy (susceptible). At the beginning of day 1, there comes from outside an extra "patient zero", who is infected by the disease. By our previous assumption, every susceptible individual has probability β/N of getting infected directly by "patient zero" on a given day, as long as "patient zero" has not recovered. Suppose N is so large that, for a long long time, the infected only account for an infinitesimal (极微小的) faction of the population. In other words, you can admit that $N = S_0 \approx S_1 \approx S_2 \approx \dots$. Also remember that "patient zero" recovers with probability α on each day. We count the total number of people infected directly by "patient zero" until he/she recovers.

Question 7 (10 pts): Show that the total average number of people who get infected directly from "patient zero" is approximately R_0.

Answer 7: On day t, "patient zero" is still infected with probability $(1-\alpha)^{t-1}$. Then the average number of people infected by "patient zero" is

$$\beta/N \cdot S_{t-1} \cdot (1-\alpha)^{t-1} \approx \beta/N \cdot N \cdot (1-\alpha)^{t-1} = \beta(1-\alpha)^{t-1}.$$

The cumulative average number is then approximately

$$\sum_{t=1}^{\infty} \beta(1-\alpha)^{t-1} = \beta/\alpha = R_0.$$

4 A three-segment model

Consider another scenario where people recovered from the disease get lifetime immunity. That is to say, recovered people will never get the disease again. They are *not* infectious either. We then need to divide the population into three disjoint groups: the susceptibles, the infected, and the *recovered* (已康复者). The proportion of people from each group are denoted s_t, i_t and r_t respectively. Remember $s_t + i_t + r_t = 1$ for all t. Figure 4 describes the transition between groups in this scenario.

Based on previous assumptions, we have the recursive formulas

$$i_t - i_{t-1} = \beta i_{t-1} s_{t-1} - \alpha i_{t-1} \quad (4)$$

$$s_t - s_{t-1} = -\beta i_{t-1} s_{t-1} \quad (5)$$

$$r_t - r_{t-1} = \alpha i_{t-1}. \quad (6)$$

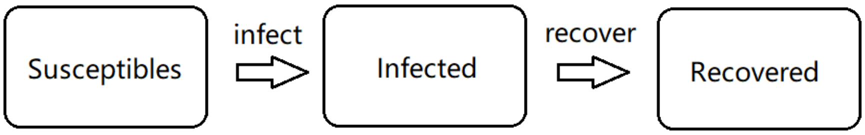

Figure 4: Transition between three groups.

We assume $i_0 > 0$, $r_0 = 0$. We define the parameter $R_0 = \beta/\alpha$ exactly the same as before, and assume $R_0 > 1$.

Question 8 (10 pts): Assume for the moment that the approximation $(b-a)/a \approx \ln(b/a)$ holds for $a > 0, b > 0$. Use this approximation and recursive formulas (4)–(6) to prove $s_t \approx s_0 e^{-R_0 r_t}$.

Answer 8: Combining (5) and (6), we get

$$\frac{s_t - s_{t-1}}{s_{t-1}} = -R_0(r_t - r_{t-1}).$$

By the given approximation, we have

$$\ln s_t - \ln s_{t-1} \approx -R_0(r_t - r_{t-1}).$$

Taking the sum, we get

$$\ln s_t - \ln s_0 = \sum_{\tau=1}^{t}(\ln s_\tau - \ln s_{\tau-1}) \approx -R_0 \sum_{\tau=1}^{t}(r_\tau - r_{\tau-1}) = -R_0(r_t - r_0) = -R_0 r_t.$$

This leads to the desired result.

Question 9 (10 pts): Recall the definition of steady states in Question 6. Assume that the three variables (i_t, s_t, r_t) approach their respective steady states (i^*, s^*, r^*) after a long enough period. Use the result of Question 8 and the three recursive formulas (4)–(6) to prove: $i^* = 0$, $s^* = 1 - r^*$, and r^* satisfies the approximate equation

$$1 - r^* - s_0 e^{-R_0 r^*} \approx 0.$$

Answer 9: Plugging the steady states into (6), we get $0 = \alpha i^*$. Thus $i^* = 0$, and $s^* = 1 - r^*$. Also from the result of Question 8, we have $s^* \approx s_0 e^{-R_0 r^*}$. Therefore $1 - r^* \approx s_0 e^{-R_0 r^*}$

5 A four-segment model

Consider the same scenario as Section 4, except for an additional feature. Let us assume that the disease has an incubation period (潜伏期). The population is divided into four disjoint groups: the susceptibles, the infected, the recovered, and the *exposed* (潜伏者). The proportion of people from each group are denoted s_t, i_t, r_t and e_t respectively. Once a susceptible individual gets the disease, he/she becomes one of the exposed at first. The exposed people are *not* infectious. When the incubation period ends for an exposed individual, he/she becomes one of the infected, who are infectious. Each exposed individual has probability $\delta \in (0,1)$ of becoming infected on any given day, so that the average proportion of newly infected people on day t is δe_{t-1}. See Figure 5 for a description of such transition.

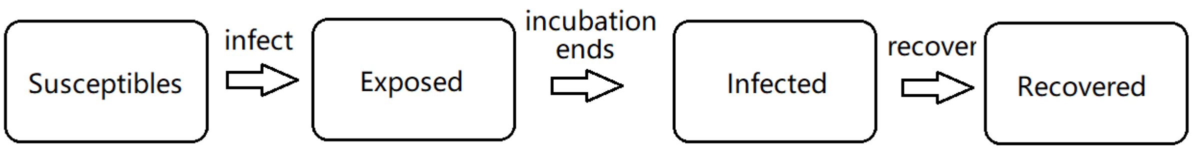

Figure 5: Transition between four groups.

Question 10 (10 pts): Write out the four recursive formulas for s_t, i_t, r_t and e_t.

Answer 10: The formulas are

$$
\begin{aligned}
s_t - s_{t-1} &= -\beta i_{t-1} s_{t-1} \\
i_t - i_{t-1} &= \delta e_{t-1} - \alpha i_{t-1} \\
r_t - r_{t-1} &= \alpha i_{t-1} \\
e_t - e_{t-1} &= \beta i_{t-1} s_{t-1} - \delta e_{t-1}.
\end{aligned}
$$