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

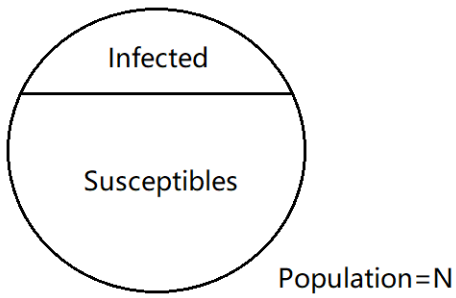

Figure 1: Division of population.

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 ($t = 1, 2, \dots$), 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 $t = 0, 1, 2, \dots$, 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:

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

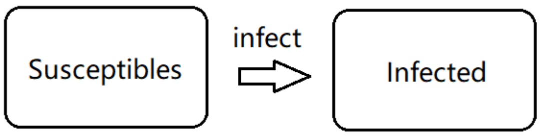

Figure 2: Transition of two groups.

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:

$$
\begin{aligned}
i_t - i_{t-1} &= \beta i_{t-1} s_{t-1} \\
s_t - s_{t-1} &= -\beta i_{t-1} s_{t-1}.
\end{aligned}
$$

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 $i_0 \in (0, 1)$.

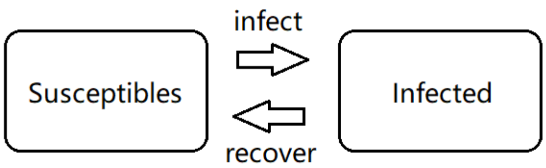

Figure 3: Transition of two groups with recovery.

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

$$i_t - i_{t-1} = \beta i_{t-1} s_{t-1} - \alpha i_{t-1} \tag{4}$$

$$s_t - s_{t-1} = -\beta i_{t-1} s_{t-1} \tag{5}$$

$$r_t - r_{t-1} = \alpha i_{t-1}. \tag{6}$$

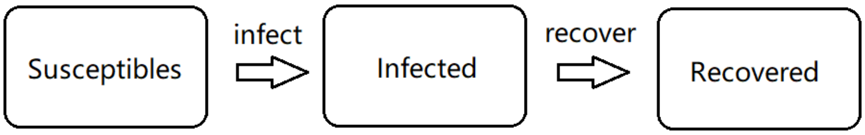

Figure 4: Transition between three groups.

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.

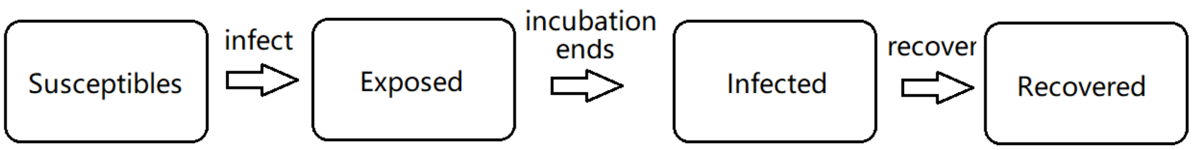

Figure 5: Transition between four groups.

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

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