

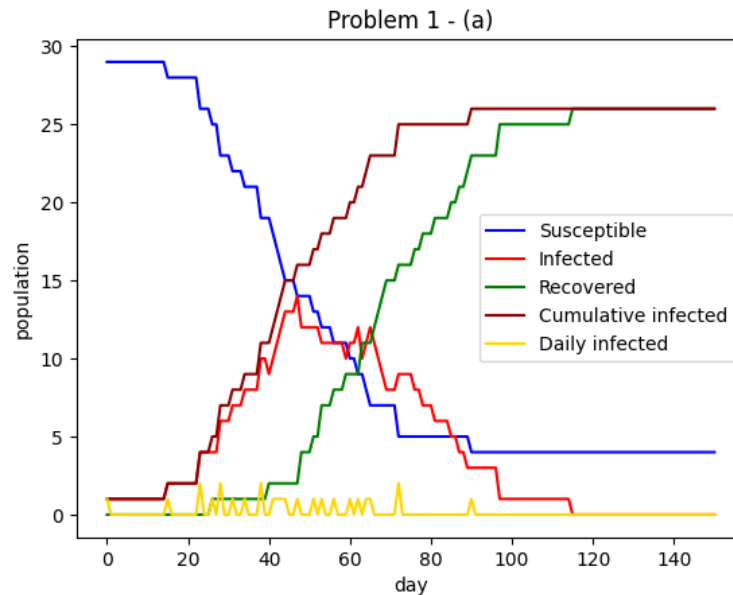
Mathematical Data Science HW3

20180617 You SeungWoo

October 6, 2023

Problem 1

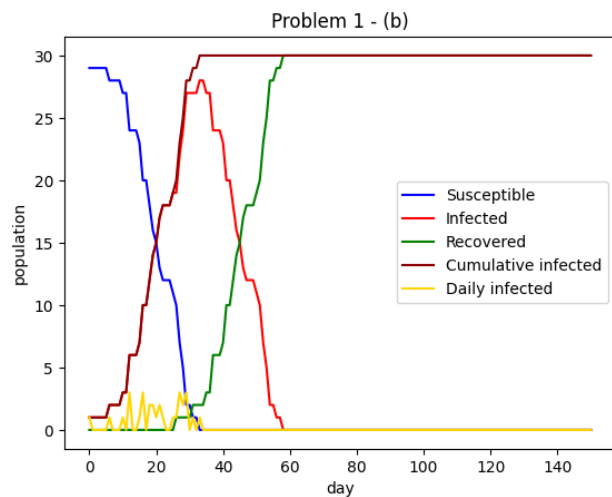
- a) *Solution.* Use the variables: Total population = 30, initial infected = 1, day = 150, speed = 0.01, recovery = 25, distance = 0.03.



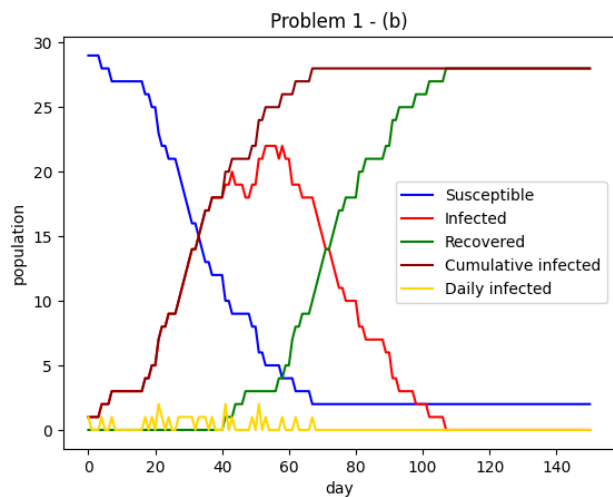
□

- b) *Solution.* Control the 3 variables: speed, recovery, and distance. Before running the model, we can expect the behavior of each population.

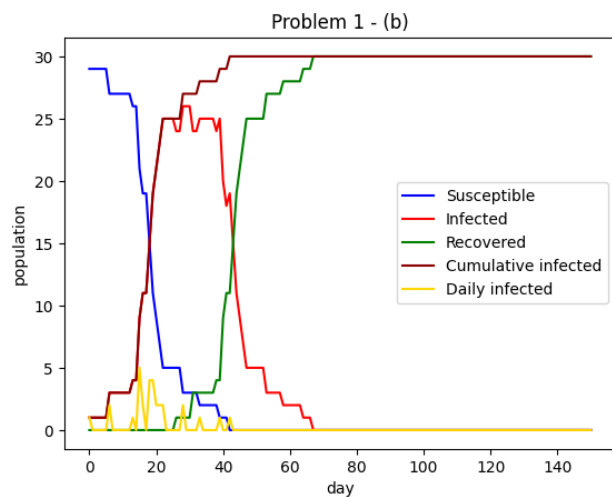
- If **speed** increases, then the chances of infected people coming into contact with susceptible increases. Additionally, as the distance traveled per day increases, the number of infected can increase significantly in an instant.
- If **recovery** increases, then the number of infected people is likely to increase due to increased opportunities for infected people to encounter susceptible. **recovery** is mainly influenced by biology, so it can be difficult to control socially.
- If **distance** increases, then it becomes easier to spread to others. It means that although the number of infected is small, the possibility of spread is high. This suggests that it is important to maintain distance in daily life.



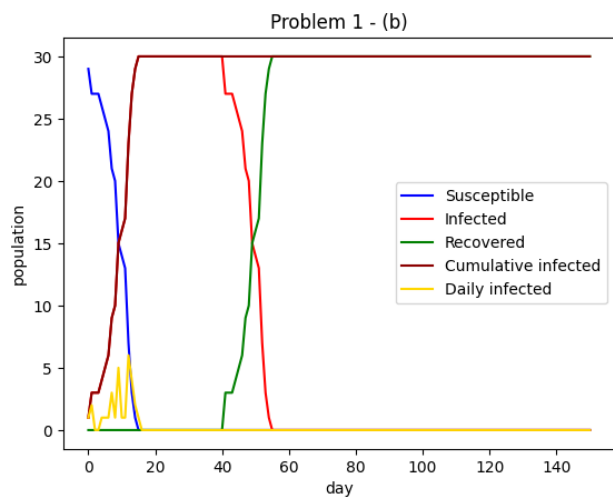
(a) speed = 0.03



(b) recovery = 40



(c) distance = 0.06



(d) speed = 0.03, recovery = 40, distance = 0.06

Differential equation models are based on theoretical formulas. Infection and recovery occur numerically according to the social distance β and recovery rate γ . This is advantageous in explaining the long-term spread trend of infectious diseases, but may not properly reflect local reality.

The particle model takes into account social movement: distance and speed. Therefore, the number of infected and recovered people each day is not consistent. This is advantageous for simulated observations based on reality, but it may be difficult to calculate long-term trends due to the complexity of calculations and various variables.

□

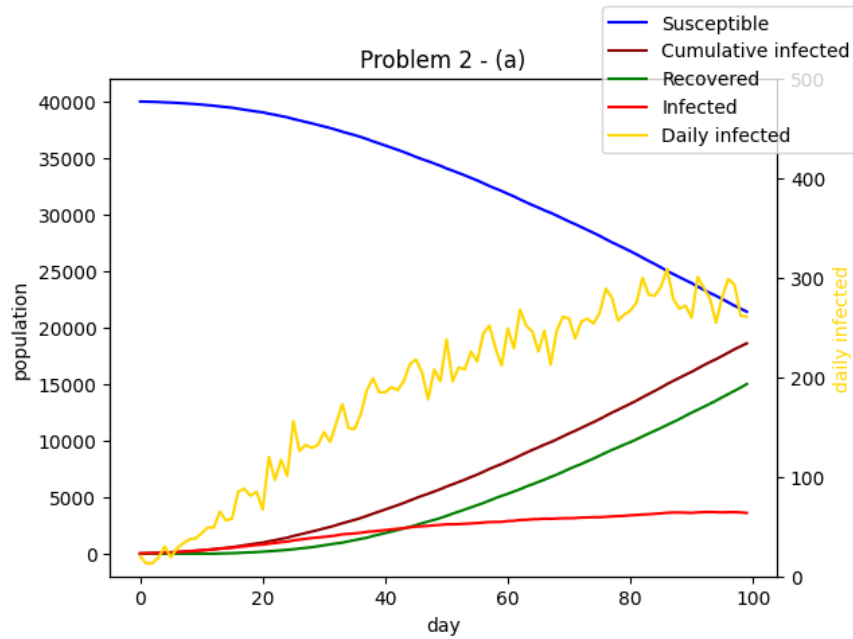
- c) *Solution.* In model a), since the total number of people is only 30, it is a little difficult to compare the number of daily infections. Nevertheless, the model in a) can be seen as similar to the Seoul and Gyeonggi regions. Compared to other regions, the infection spreads slowly over a long period of time after it begins. On the other hand, there is a big difference from the Daegu and Gyeongbuk regions. This can rather be seen as similar to model b). In the early stages of the spread, the number of infected people increases rapidly. According to our particle model, the **speed**(social movement distance) and **distance**(degree of crowding) of early infected people were high.

Let's also look at foreign cases. Like the model in a), Rome shows a long-term trend of gradual spread after the initial infection. Italy and Germany increase rapidly in a short period of time, which is similar to the changes in the Daegu and Gyeongbuk regions. Sweden has maintained a consistently low number of infected people despite early infections, perhaps due to good control of variables. However, given that a rapid increase of nearly three times occurred at the end, it appears that it will follow Germany's pattern in the future.

□

Problem 2

- a) *Solution.* I use the given CA model. Note that the daily infected population is plotted in another scale of second y axis.



□

- b) *Solution.* Infection begins at a specific point and spreads randomly from that point. And recovery also follows the flow at certain time intervals. In this model, infection does not occur through other routes, and recovered people are no longer infected. Due to randomness and the impossibility of reinfection after recovery, there may still be uninfected areas even if surrounded by a large yellow area. While it did not spread, a barrier was created for those who have recovered. If both variables were modified, we would expect there to be no such regions.

Domestic infections show similar patterns to the CA model. It can be interpreted that it spread around Seoul and Daegu. However, it does not spread out in a perfectly spherical shape like the model. This is presumed to be due to the existence of additional variables such as transmission distance and infection speed. Therefore, adding these variables to the CA model will allow it to be implemented similar to reality.

□

Problem 3

「대량살상수학무기 Chapter 6. 디지털 골상학: 당신은 우리가 원하는 직원이 아닙니다」

요약: 6장은 많은 기업이 채용 과정에서 활용하는 인적성검사를 설명한다. 최근 인적성검사는 컴퓨터를 통해 실시하고 평가한다. 이는 정해진 답이 있는 것처럼 문항에 따라 평가 지표가 만들어진다. 물론 사람의 평가보다는 공정할 수 있다. 하지만 평가 알고리즘 또한 사람이 만든 것이기에, 편견이나 오류가 발생할 수 있다. 게다가 불합격한 사람은 정확한 이유를 알 수 없기에 차별이 발생해도 이를 눈치챌 수 없다. 인적성검사 과정이 매우 불투명하고 수정 불가능한 모델에 기반하기 때문이다. 그럼에도 이러한 절차가 여전히 사용되는 이유는 선발 과정의 편리함 덕분이다. 기업 입장에서는 인재를 선발하는 과정에 소요되는 시간과 비용을 최소화하는 것이 중요하다. 기업에 입사를 희망하는 지원자는 여전히 많으며, 컴퓨터 프로그램을 이용한다면 몇 명의 인재를 놓치더라도 그들을 편리하게 선별할 수 있다. 사람이 하나하나 선별한다고 하더라도 모든 인재를 선별할 수 있을지는 미지수이며, 투자 대비 결과가 그렇게 좋지 않을 수 있다. 따라서 이러한 시스템은 여전히 유지되고 있으며, 프로그램에 대한 정보와 이를 쟁취할 부를 가진 사람들이 경쟁에서 우위를 차지할 가능성이 높아진다.

해결 방안: 기업에서 평가 기준을 투명하게 공개하고, 알고리즘을 개선하고 검증하는 과정이 지속해서 이루어져야 할 것이다. 또한 컴퓨터 평가 결과를 주로 삼지 말고, 보조 용도로 사용하면 좋을 것이다. 애매한 경우에 대해 다시 한 번 사람이 보고 판단해 불합리한 탈락이 발생하지 않도록 해야 한다. 또한 이러한 프로그램을 기업에 부적합한 사람을 배제하기 위한 용도가 아니라 필요한 자원을 제공하는 용도로 사용한다면 더 훌륭한 인재를 발굴할 수 있을 것이다.