

# The Reed-Frost Model and its Application to Real Life

Mark: 72

## 1 Preface

The Reed-Frost model is one of the earliest models used to study the behaviour of infectious diseases. It was created by Lowell Reed and Wade Hampton Frost during the 1920s. Nowadays, the model is used more as a template for more complex models. However, it is still worth looking at because it gives a good introduction to stochastic epidemiology theory.

Epidemiological models are used in modern life to provide predictions and data for possible outbreaks of infections. This information can be useful for governments and public health organisations to implement measures to combat current epidemics and to make plans for possible future outbreaks. Whilst epidemiology covers a wide range of infectious diseases, the Reed-Frost model is used primarily for acute infectious diseases. An acute infectious disease is an infection that appears suddenly and may be of brief or prolonged duration [1] such as colds, flus and STIs.

We will explore the Reed-Frost model and a couple of example simulations, before analysing how applicable the model is to real life.

## 2 Introduction of Concepts

Before looking at the Reed-Frost model a few introductory concepts need to be explained, namely the different types of epidemiological and statistical models.

Firstly, from an epidemiology viewpoint, the Reed-Frost model is an S-I-R epidemic model.

**Definition 1** (S-I-R Model). An S-I-R (Susceptible-Infection-Removal) Model is an epidemiological model that calculates the number of people infected with a contagious illness in a closed population over time. [2]

The S-I-R format provides the basis of many epidemic models and is used mainly to model diseases transmitted by humans, where individuals pass through the following progression of states:

Susceptible to the disease  $\rightarrow$  Infected  $\rightarrow$  Removed (e.g. immunity or death)

The infection can only be passed between two individuals through “adequate contact”. The concept of “adequate contact” is relative as different infections are passed in different ways. For example, close proximity to an infected individual may be sufficient for a common cold to spread but isn’t for an infection such as an STI. This makes the term “adequate contact” difficult to define universally and it can be hard to define precisely for a specific infection (e.g requirements to catch a cold).

For the Reed-Frost model, we consider an acute, infectious disease that is spread only by “adequate contact”. The model itself comes in two forms: deterministic form and stochastic form.

**Definition 2** (Deterministic Model). A deterministic model is one in which a given input into the model always produces the same output. [3]

Equivalently, a deterministic model is one where the output is fully determined by the parameters and initial conditions. For example, in population dynamics a model that takes some initial population size as its input and outputs a predicted population size after some interval is a deterministic model. The lack of randomness within these models makes them relatively simple to work with, but limits its application to real life. For example, real populations have random influences affecting growth such as varying birth and death rates. Therefore, we need to use stochastic processes in order to create a more realistic model.

**Definition 3** (Stochastic Model). A stochastic model is one in which there is a random element such that for a given input to the model, the outcome takes a range of possible values. [3]

This means that the outcome is not uniquely determined by the given input. Therefore, a stochastic model needs to be run multiple times to generate trends in behaviour.

The stochastic version of the Reed-Frost model that we will be considering is called an epidemic chain binomial.

**Definition 4** (Epidemic Chain Binomial Model). An epidemic chain binomial model is one in which the number of infected individuals to appear in the next unit interval of time follows a binomial distribution, with the probability of infection dependent on the number of infected individuals in the current time unit. [4]

It is important to note that an epidemic chain binomial model is a discrete time model and are used to describe the spread of infection within closed populations, such as households. Therefore, we need to make some basic assumptions to base the model on.

## 3 Set-Up for the Model

### 3.1 Assumptions

There are many factors that can influence the spread of an infectious disease. Factors can range from simple influences like the movement of individuals and changes in living standards, to changes as extreme as a natural disaster or the outbreak of war.” Since there are numerous influences on the spread of an infection, it is essential that some assumptions are made in order to create a coherent model.

The Reed-Frost model is based on 5 assumptions [5] :

1. The infection can only be spread via “adequate contact” with infected individuals.
2. Susceptible individuals become infected after “adequate contact” with an infected individual in a given time period. They will then only be infectious for the subsequent time period, before becoming fully immune.
3. The infection is introduced to a closed population (i.e. the total population is fixed with no one entering or leaving the pool of individuals).
4. Individuals have a fixed probability of “adequate contact” with any other individual in a time period
5. The above conditions remain constant for the duration of the epidemic.

It is important to note that the final assumption also includes the requirement of external factors remaining constant. For example, a war or a famine could still impact a closed population and thus influence the spread of disease.

Furthermore, the discrete time periods are defined to correspond to the time between an individual becoming infected and the point at which the individual is most infectious.

### 3.2 Notation

In a given time interval  $t$ , the number of susceptible individuals in the population is denoted by  $S_t$  and the number of infected individuals are denoted by  $I_t$ . For simplicity, immune individuals are often ignored as they stay immune for the remainder of the epidemic. The (fixed) probability that any two individuals come into “adequate contact” in one time frame is  $p$ , which is usually expressed as  $p = (1 - q)$ .

## 4 Deterministic Approach

Given the set-up as defined above, the probability that, in a given time period  $t$ , a susceptible individual *doesn't* come into contact with a given individual is  $1 - p = q$ . Therefore, the probability that an individual *doesn't* come into contact with an infectious individual during time period  $t$  is  $(1 - p)^{I_t} = q^{I_t}$ . From this, we obtain the probability of a susceptible individual becoming infected during a time period  $t$ :

$$1 - q^{I_t}$$

Now, consider the number of infectious individuals in the time period  $t + 1$ . It is given in the model's set-up that infected individuals become wholly immune in the subsequent time period. This implies that  $I_{t+1}$  is independent of  $I_t$  and is comprised of the number of susceptible individuals who came into contact with infectious individuals in time period  $t$ . From this, we can calculate the expected number of infectious individuals in time period  $t + 1$  as:

$$I_{t+1} = S_t(1 - q^{I_t}) \tag{1}$$

The number of susceptible individuals in time period  $t + 1$  is then simply the remaining susceptible individuals from the previous time period, which is clearly:

$$S_{t+1} = S_t - I_{t+1} \tag{2}$$

These two equations form the deterministic Reed-Frost model for initial conditions  $(S_t, I_t, p)$ . The entire epidemic is then modelled using iterations of these formulae.

### 4.1 Examples

**Example 4.1.** Take a total population made up of 100 susceptible individuals and 1 infected individual. Let the probability of “adequate contact” be 0.02.

As seen in figure 1, the number of infected individuals peaks at 16 individuals at time  $t = 6$ . The number of infected individuals never surpasses the number of susceptible individuals, leaving 19 susceptible individuals remaining at the end.

**Example 4.2.** Take a total population made up of 400 susceptible individuals and 1 infected individual. Let the probability of “adequate contact” to be 0.04.

From figure 2, we can see that the number of infected individuals exceeds the number of susceptible people and peaks at a value of 199 at time  $t = 4$ . At the end of the epidemic, there are no more susceptible individuals remaining.

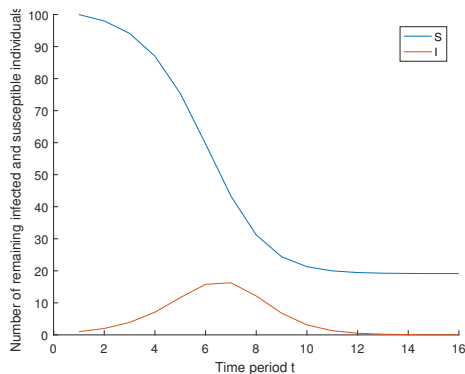


Figure 1: Deterministic model with initial conditions (100, 1, 0.02)

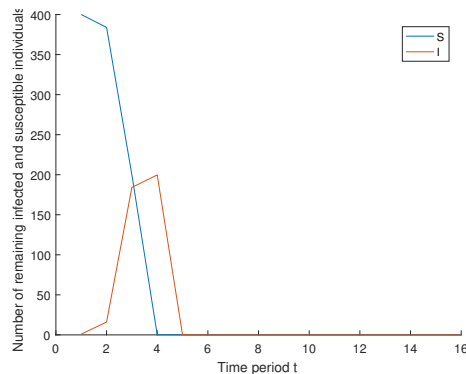


Figure 2: Deterministic model with initial conditions (400, 1, 0.04)

## 5 Stochastic Approach

The stochastic approach to the Reed-Frost model uses the same set-up defined previously and takes a series of binomial trials to model the epidemic.

As before, the probability of a susceptible individual becoming infected is:

$$1 - q^{I_t}$$

However, now the number of infected individuals in the time period  $t + 1$  is modelled using a binomial distribution with parameters  $S_t$  and  $1 - q^{I_t}$  [5]. We denote this as:

$$I_{t+1} \sim \text{Binomial}(S_t, 1 - q^{I_t})$$

Therefore, in a given time period  $t + 1$ , the probability that there are  $r$  infected individuals is:

$$\mathbb{P}(I_{t+1} = r) = \binom{S_t}{r} (1 - q^{I_t})^r (q^{I_t})^{S_t - r} \quad (3)$$

The number of susceptible individuals in time period  $t + 1$  is simply the difference between the number of susceptible individuals in time  $t$  and the number of infected individuals in time  $t + 1$ . So we have that:

$$S_{t+1} = S_t - I_{t+1}$$

These two equations form the stochastic Reed-Frost model for initial conditions  $(S_t, I_t, p)$ .

The sequence  $\{I_0, I_1, \dots, I_\tau\}$  is called the epidemic chain, where we define  $\tau = \min\{t : I_{t+1} = 0\}$  to be the time when the infection dies out (i.e. when there are no more infected individuals). Using this set up, can also calculate the (expected) final size of the epidemic. [6]

**Definition 5** (Final Size of an Epidemic). The final size of an epidemic,  $T$ , is the total number of individuals who became infected during the epidemic.

Therefore, we can see that:

$$T = S_0 - S_\infty = \sum_{k=1}^{\tau} I_k$$

where  $\tau = \min\{t : I_{t+1} = 0\}$

It should be noted that the final size of the epidemic excludes the number of individuals who were initially infected.

Whereas  $I_t$  gives the value of the number of infected individuals in one time period,  $T$  provides a clearer indication of the total number of individuals who are affected by the infection. Using the above formula, the probability mass function and expectation for the variable  $T$  can be calculated in order to obtain estimations for the size of an epidemic. <sup>1</sup>

## 5.1 Example

**Example 5.1.** Take a total population made up of 100 susceptible individuals and 1 infected individual. Let the probability of “adequate contact” to be 0.02. Then our initial conditions are (100, 1, 0.02). Figures 3 and 4, seen below, depict the number of susceptible and infected individuals respectively when the model was run 100 times, each for 10 time periods.

As seen in figure 3, general trend is for the number of susceptibles to decrease over time with the time interval of sharpest decline being [3, 6] and the majority of final values of susceptible individuals lying in the range [10, 30].

The number of infected individuals tends to be more random. There is a slight trend in figure 4 of plots increasing when  $t \in [0, 6]$  and decreasing when  $t \in [6, 11]$ . However, the number of infected individuals doesn’t exceed 35 in these epidemic chains, with most plots peaking below 25 individuals.

Since the model was only run for 10 time steps, the value of  $\tau$  for many of the epidemic chains wasn’t reached. However, figure 5 shows that the size of epidemics lay in the range [0, 93] with an average infection size of 64.3.

---

<sup>1</sup>The formulation of the probability mass function for  $T$ , and thus the expectation of  $T$ , is complex. Therefore, this has been omitted here but can be found in [6].

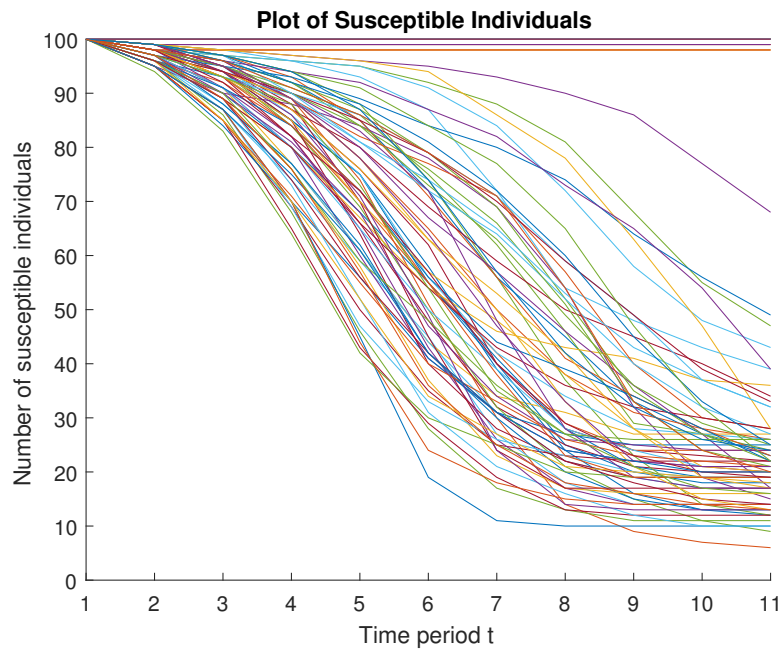


Figure 3: the number of susceptible individuals

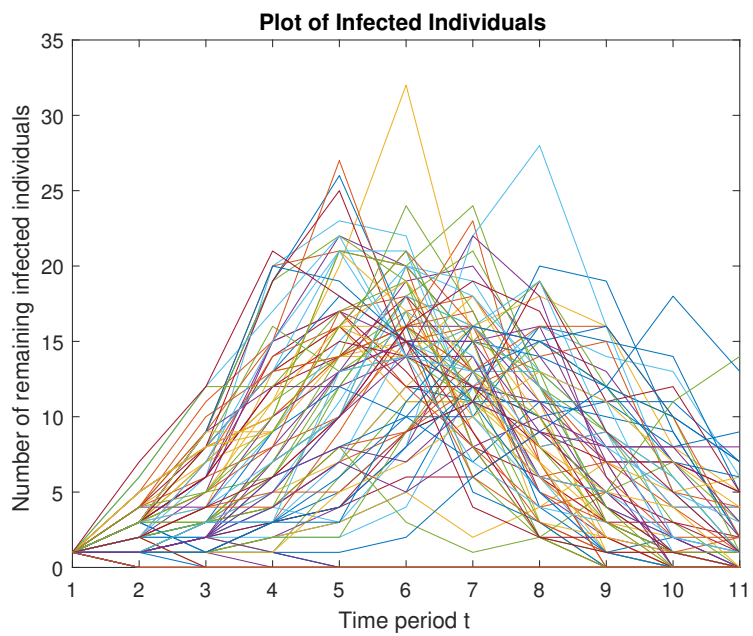


Figure 4: the number of infected individuals

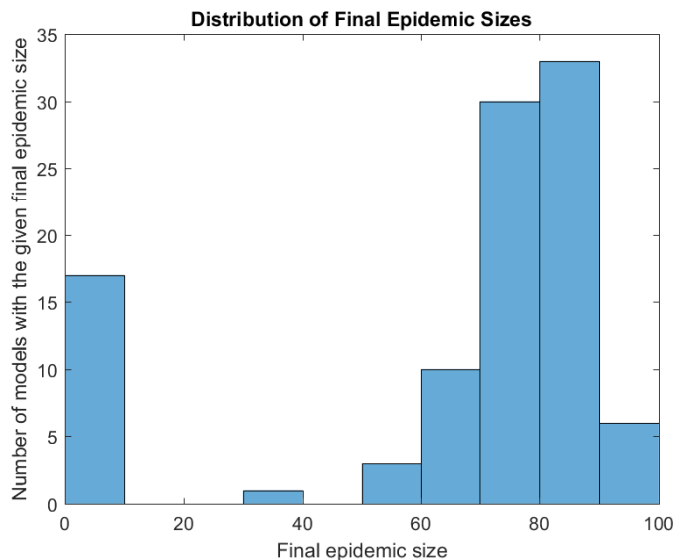


Figure 5: Histogram of the final epidemic sizes generated

## 6 S-I-S Variation

Consider a modified version of the model that simulates epidemics where there is no immunity, so infected people can become re-infected after recovering. An example of this is the common cold. This would put the Reed-Frost model into SIS form (Susceptible-Infective-Susceptible).

This implies that the number of susceptible individuals increases by the number of recovered individuals:

$$S_{t+1} = S_t - I_{t+1} + I_t$$

However, this assumes that all infectious individuals recover and that they recover at the same rate.

Instead, let  $r$  be the probability an infected individual recovers during one time interval. Then we have  $I_t$  infected individuals, each with probability  $r$  of recovering during by time  $t + 1$ . Let  $R_t$  denote the number of newly recovered individuals time  $t$ . Then  $R_t \sim \text{Binomial}(I_t, r)$ .

As before, the number of newly infected individuals is binomially distributed. Let this be denoted by  $N_t$ . Then  $N_t \sim \text{Binomial}(S_t, 1 - q^{I_t})$ .

Therefore, we have:

$$I_{t+1} = I_t + N_t - R_t \tag{4}$$

$$S_{t+1} = S_t + R_t - N_t \tag{5}$$

These equations form the SIS variation on Reed-Frost model. <sup>2</sup>

---

<sup>2</sup>This section was heavily influenced by [7]



## 7 Analysis of the Model

The main strength of the Reed-Frost model is its simplicity and adaptability.

The simplicity of the model's formulation means that it can be easily converted from a deterministic model to a stochastic model. This has meant the model has been used as an introduction to stochastic epidemic theory. Yet, despite its simplicity, the model is still relatively complex so requires approximations in order to be understood and used [8]. Furthermore, the method "fails to convince those untrained to probability and statistics" [9], which could cause problems when the model is being used by those from an epidemiological (rather than statistical) background within public health bodies.

Although flawed in some areas, the model can be seen as a reasonable model for outbreaks of contagious illnesses within institutions such as schools. This is because the assumptions for the model to hold are reasonable for such situations. For example schools naturally have a closed population and regular contact. However, it can be argued that the assumptions made are unrealistic and that the model would be better suited to simulate the spread of highly contagious diseases.

One of the main strengths of the model is its adaptability to encompass more realistic assumptions in order to become a more suitable model for real life applications. This has resulted in the Reed-Frost model being a template for more complex models, such as the aforementioned SIS model.

We've seen that the SIS version of the model assumes that individuals return to a susceptible state after being infected. Another possible scenario is one where immunisation is only short term. In this case, immune individuals are reintroduced into the model as susceptible individuals, thus the model needs to incorporate the rate at which new susceptible individuals are being introduced.

Another realistic assumption would be that during the course of an epidemic, public health measures may be put in place to reduce the rate of infection, such as restriction of movement [5]. As a result, it would be expected that  $p$  would be reduced and may vary depending on how such measures change over the duration of the epidemic.

A further assumption which could be considered is that an immunisation program would be introduced in order to reduce the number of susceptible individuals and prevent the spread of infection. This would result in some susceptible individuals moving to an immune state without becoming infected.

All these adaptations would be complex to incorporate within the model but could improve the accuracy of the predictions generated.

In addition, the model can be extended further to calculate other random variables such as the duration of infection (previously denoted  $\tau$ ), the final size of the epidemic ( $T$ ) and the remaining number of susceptibles. These variables would be useful to calculate in order for public health measures to

be enforced efficiently and effectively. For example, the expectation of the final size of an epidemic can be used to predict the number of immunisations that will be needed.

In conclusion, the Reed-Frost model is a reliable epidemic model for a specific epidemic scenario that can be easily adapted. Therefore, with the some aforementioned improvements and extensions, the model can be applicable to real life.

## References

- [1] acute infection. (n.d.), Medical Dictionary for the Health Professions and Nursing, 2012, <https://medical-dictionary.thefreedictionary.com/acute+infection>
- [2] Weisstein, Eric W., *SIR Model*, MathWorld-A Wolfram Wed Resource, <http://mathworld.wolfram.com/SIRModel.html>
- [3] D. Brown and P. Rothery, *Models in Biology: Mathematics, Statistics and Computing*, Part 1 Chapter 1, Pages 13-14, 1993.
- [4] Nath, D.C., Das, K.K. and Chakraborty, T., *A Modified Epidemic Chain Binomial Model and Its 2,3-Introductory Probabilities.*, Open Journal of Statistics, Volume 7, pages 225-239.
- [5] Paul E.M. Fine, *A commentary on the mechanical analogue to the Reed-Frost epidemic model*, American Journal of Epidemiology, volume 106, pages 88-100, 1977.
- [6] Philip O'Neill, *Perfect Simulation for Reed-Frost Epidemic Models*, Statistics and Computing journal, Volume 13, Issue 1, pages 37-44 2003.
- [7] *Chain binomial epidemic models*, 2013, accessed from <http://lalashan.mcmaster.ca/theobio/mmed/images/8/8f/Reed-frost-5jun2013.pdf>
- [8] A.D. Barbour, Sergey Utev, *Approximating the Reed-Frost epidemic process*, Stochastic Processes and their Applications 113, pages 173-197, 2004.
- [9] *Reed-Frost Epidemic Model*, May 2012, accessed from <https://www.osc.edu/education/si/projects/epidemic>

## Appendix A Matlab Code for Deterministic Simulation

```
1 I=zeros(15,1);  
2 S=zeros(15,1);  
3 I(1)=1;  
4 S(1)=100;
```

```

5 p=0.02;
6 q=1-p;
7
8 for t=1:15
9     I(t+1)=S(t)*(1-q^I(t));
10    S(t+1)=S(t)-I(t+1);
11 end
12
13 I
14 S
15 hold on
16 plot(S)
17 plot(I)
18 xlabel('Time period t')
19 ylabel('Number of remaining infected and susceptible individuals')
20 legend('S','I')

```

## Appendix B Matlab Code for Stochastic Simulation

```

1 % simulates the decrease of susceptibles/infected in a population
2
3 p = 0.02;      % p = prob of contact
4 q = 1-p;      % q = probability of no contact
5 S(1) = 100;   % Initial population
6 I(1) = 1;     % Initial number of infected
7
8 for j=1:100
9     A=0;      %resets the value of A for each run of the model
10    for k=1:10
11        I(k+1) = binornd(S(k),1-q^(I(k)));
12        S(k+1) = S(k) - I(k+1);
13        A=A+I(k+1);
14        %running count of the number of individuals who become infected
15    end
16
17    plot(I)
18    hold on
19    plot(S)
20    %Note, in practice, these plots are split into 2 graphs
21
22    xlabel('Time period t')
23    ylabel('Number of remaining infected and susceptible individuals')
24    legend('S','I')
25
26    T(j)=A;
27    j=j+1;
28 end
29
30 T;
31 max(T)
32 min(T)
33 mean(T)
34 histogram(T,10) %plots histogram of T split into 10 bins

```