Proposing a new probabilistic approach in modelling epidemics

IB Mathematics HL Internal Assessment

Introduction

This investigation aims to propose a new probabilistic approach in modelling epidemics. This topic is of significance as though humanity continues to progress in terms of technology, the threat of epidemics to our survival is still imminent. A recent example would be our inability to control the outbreak of the Zika virus in South America.

My interest in this topic grew after I played a game called Plague Inc. Plague Inc is a game in which you can create different types of virus and adjust their properties, such as the virus' probability of infection, lethality etc. The goal of the game is to kill the population of the world using the virus you created. This, adding to the recent outbreak of the Zika virus, led me to researching epidemic models. The SIR Model stood out as it was used before by the Hong Kong Department of Health in predicting the SARS outbreak in Hong Kong in 2003, a time when all of us living in Hong Kong lived in fear.

However, I noticed that some of the underlying problems of the SIR Model, such as the number of people infected would be exactly equal to the expected value of the number of people infected resulting from a certain probability of infection. Thus this investigation would propose a new probabilistic approach in modelling epidemics using numerical methods, inverse transform sampling and the Monte-Carlo method via a spreadsheet software and graphs.

Background research

Epidemic Models

An epidemic model is a mathematical model that predicts population-level epidemic dynamics (Keeling 7). In other words, it predicts how different groups within a population reacts to the invasion of an epidemic. The SIR Model is a type of simple epidemic model.

The SIR Model

The SIR Model is used in epidemiology to compute the amount of susceptible, infected, recovered people in a population (Johnson 1). In other words, the model divides the population into three groups of people: the people that are susceptible to infection by the epidemic (S), the people infected by the epidemic (I), and the people that recovered from the epidemic with immunity (R).

Five key assumptions are made in this model to simplify the calculations.

1. A fixed number of population. The number of people within the population is constant. There are no net birth rates and death rates, as well as immigration to and emigration from the population.

2. Immunity upon recovery. A person is immune from the epidemic upon recovery. In other words, if a person from the I population has recovered, he or she would be put into the R population group. He or she would be immune to infection by the epidemic and thus unable to be classed as the S population again.
3. Constant probability of infection. There is a constant probability of infection, and the probability of infection of the epidemic would not change for people of different ages, genders, races etc.
4. No inheritance of disease. There is no inheritance of the epidemic, which also means that immunity of the disease is not inheritable as well.
5. Homogenous mixing. There is homogenous mixing of the population, meaning that all members of the population have an equal number of interactions with each other over time.

The Deterministic SIR Model

In a Deterministic SIR Model, there is a fixed infection and recovery rate. In other words, the number of people infected and recovered would be exactly equal to the expected number of people to be infected and recovered. This leads to far more straightforward equations and exponentially distributed infectious periods (Keeling 2). In this report, the accuracy and the realism of the Deterministic SIR Model would be evaluated.

Inverse of the Cumulative Binomial Probability Distribution Function

The binomial probability distribution function of a discrete random variable X is as follows.

$$P(X=x)=\binom{n}{x}p^{x}(1-p)^{n-x}, x=0,1,2,\dots n \text{ (Fabio 530)}$$

Equation 1 The Binomial Probability Distribution Function

(Refer to Equation 1) Where X denotes the number of successes in n trials such that the probability of a success on any one trial is $p,\ 0 \le p \le 1$ (Fabio 530). In other words, the binomial probability distribution function provides the probability of x number of successes in n trials given the value of x. Thus the inverse of this function would provide the value of x given the probability of x number of successes in n trials. In this report, the inverse of the cumulative binomial probability distribution function would be used in Microsoft Excel to formulate a new Probabilistic SIR Model.

Monte-Carlo Methods

Monte-Carlo methods are methods which solve a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties ("Monte-Carlo Method"). The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically ("Monte-Carlo Method"). In this report, Monte-Carlo Methods would be used in Microsoft Excel to randomly generate samples from a uniform probability distribution from 0 to 1 in the inverse of binomial probability distribution function to simulate the random probability of one being infected by an epidemic.

Data Collection and Analysis

Equations of the Deterministic SIR Model

The independent variable in SIR Models is the time step or the unit time, given by t. On the other hand, there are 3 dependent variables that gives the number of people as a function of unit time.

Group	Dependent variables as functions of unit time	Explanation
S	$S(t)$	Represents the number of people S population group.
I	$I(t)$	Represents the number of people I population group.
R	$R(t)$	Represents the number of people R population group.

Table 1 The dependent variables in a SIR Model.

Furthermore, there are other variables that are needed for the creation of a SIR Model.

Variables	Explanation
N	Total number of people in the whole population.
α	The average number of interactions between people in the population per capita per unit time.
i	Probability of (Rate of) infection of the epidemic.
r	Probability of (Rate of) recovery from the epidemic.

Table 2 Other Variables needed for the creation of the SIR Model.

From the above we can create 3 differential equations that represent the rates of change of each dependent variable in a Deterministic SIR Model (DSIRM).

$$\frac{dS(t)}{dt} = -\frac{S(t)}{N} \times \frac{I(t)}{N} \times i \times N \times \alpha$$

Equation 2 The rate of change of the number of people in the S population group over time in a DSIRM.

(Refer to Equation 2) The probability that a person from the S population group meeting a person from the I population group is $\frac{S(t)}{N} \times \frac{I(t)}{N}$. The probability that the interaction between the S and I person is infectious is i. The number of interactions per day in the population is $(N \times \alpha)$. Thus all of the variables multiplied results in the rate of change of the number of people in the S population group over time. Note that a negative sign is put in front of the equation as the number of people in the S population group decreases over time as people are infected.

$$\frac{dI(t)}{dt} = \frac{S(t)}{N} \times \frac{I(t)}{N} \times i \times N \times \alpha - I(t) \times r$$

Equation 3 The rate of change of the number of people in the I population group over time in a DSIRM.

(Refer to Equation 3) Contrary to the previous equation, the first part of the equation is the negative of $\frac{dS(t)}{dt}$, as the rate of decrease of the number of people being infected in the S population group over time is the negative of the rate of increase of the number of people in the I population group over time. However, as the people in in the I population group recover, the number of people in the I population group decreases. The rate at which the number of people in the I population group decreasing over time can be given by $-I(t) \times r$, thus forming the second part of the equation.

$$\frac{dR(t)}{dt} = I(t) \times r$$

Equation 4 The rate of change of the number of people in the R population group over time in a DSIRM.

(Refer to Equation 4) The number of people that recover over time is equal to the number of people in the I population multiplied by the probability of recovery from the epidemic, which is the negative of the second part of the previous equation.

The DSIRM that will be created in this report will simulate the situation when a hypothetical non-lethal epidemic is introduced into a school of 1200 students. The students in the school are assumed to have 40 interactions with other students per unit time. The epidemic is assumed to have a probability of infection 0.001 and a probability of recovery of 0.01. In other words, $N = 1200, \alpha = 40, i = 0.001, r = 0.01$.

Euler's Method

Euler's Method is a numerical method used to find numerical approximations to the solution of the differential equation $\frac{dy}{dx} = f(x, y)$ (Blythe 232).

The fundamental approximation in Euler's Method is as follows.

$$\frac{dy}{dx} = f(x, y) \approx \frac{\Delta y}{\Delta x}$$

Equation 5 The fundamental approximation in Euler's method.

From equation 5, we can obtain the following (See Figure 1 for a graphical illustration).

If x_0 increases by h from x_0 to x_1,

$$f(x_0, y_0) \approx \frac{\Delta y}{h} = \frac{y_1 - y_0}{x_1 - x_0}$$

$$h \times f(x_0, y_0) \approx y_1 - y_0$$

$$y_1 \approx y_0 + h \times f(x_0, y_0)$$

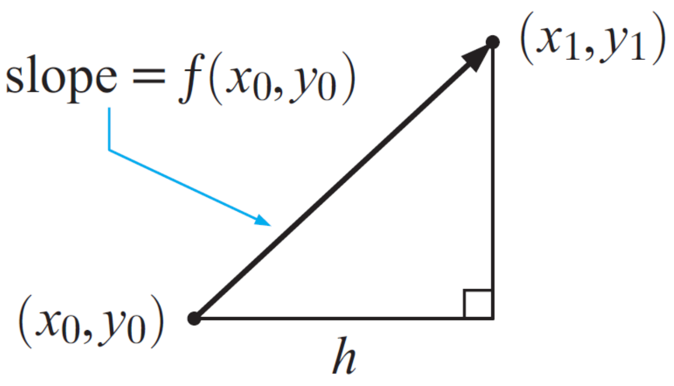


Figure 1 A graphical illustration of Euler's method.

(Source: Blythe, Peter. *Mathematics for the International Student: International Baccalaureate Mathematics HL (options), International Baccalaureate Diploma Programme*. Adelaide Airport, S. Aust.: Haese & Harris Publications, 2005, p. 233)

In other words, the slope of function at (x_0, y_0) is used to approximate the position of (x_1, y_1). Clearly, Euler's Method only gives an approximate solution to an initial-value problem (Blythe 233). However, as h decreases in value and as the number of course corrections increase, the accuracy of the approximation using Eurler's Method would increase.

In a DSIRM, Eurler's Method will be used to approximate how the number of people in the S, I, and R change over time.

The follwing equation can then be derived for the DSIRM.

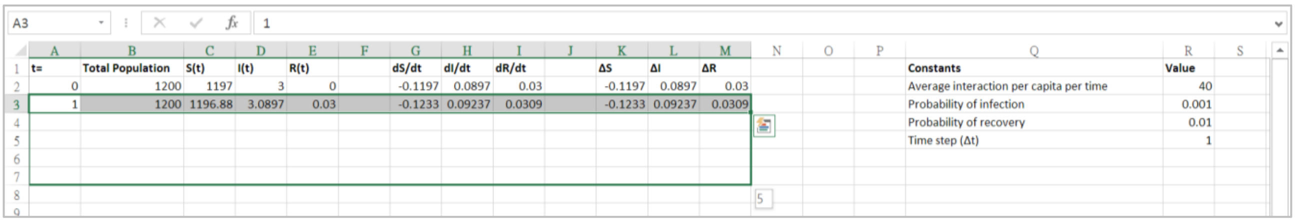
$$\begin{cases} S_1 = S_0 + \Delta t \times \frac{dS(t)}{dt} = S_0 + \Delta t \times \left(-\frac{S_0}{N} \times \frac{I_0}{N} \times i \times N \times \alpha\right) \\ I_1 = I_0 + \Delta t \times \frac{dI(t)}{dt} = I_0 + \Delta t \times \left(\frac{S_0}{N} \times \frac{I_0}{N} \times i \times N \times \alpha - I_0 \times r\right) \\ R_1 = R_0 + \Delta t \times \frac{dR(t)}{dt} = R_0 + \Delta t \times (I_0 \times r) \end{cases}$$

Equation 6 Equation for the DSIRM.

However, errors do still persist when numerical methods are used since they are numerical approximations. The most prominent of which is the truncation error, which is the error caused by using simple approximations to represent exact formulas (Ledder 2). Truncation errors persist as long as exact calculations are not used. Thus truncation errors will always persist in DSIRMs as DSIRMs cannot be solved analytically since we cannot obtain an exact analytical expression for the dynamics of S and I through time (Keeling 5). Yet truncation error can be reduced by applying the same approximation to a larger number of smaller intervals (Ledder 2). Therefore in this report, Δt is set to equal to 1 and the model is created for the time period from $t = 0$ to $t = 1000$.

Procedures of creating the Deterministic SIR Model in Microsoft Excel

The spreadsheet software Microsoft Excel is used to create the DSIRM.



1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	t=	Total Population	S(t)	I(t)	R(t)		dS/dt	dI/dt	dR/dt		ΔS	ΔI	ΔR				Constants	Value	
2	0	1200	1197	3	0		-0.1197	0.0897	0.03		-0.1197	0.0897	0.03				Average interaction per capita per time	40	
3	1	1200	1196.88	3.0897	0.03		-0.1233	0.09237	0.0309		-0.1233	0.09237	0.0309				Probability of infection	0.001	
4																	Probability of recovery	0.01	
5																	Time step (Δt)	1	
6																			
7																			
8														5					
9																			

Figure 2 Microsoft Excel layout of creating the DSIRM.

The step-by-step procedures for creating this DSIRM are shown in Appendix 1.

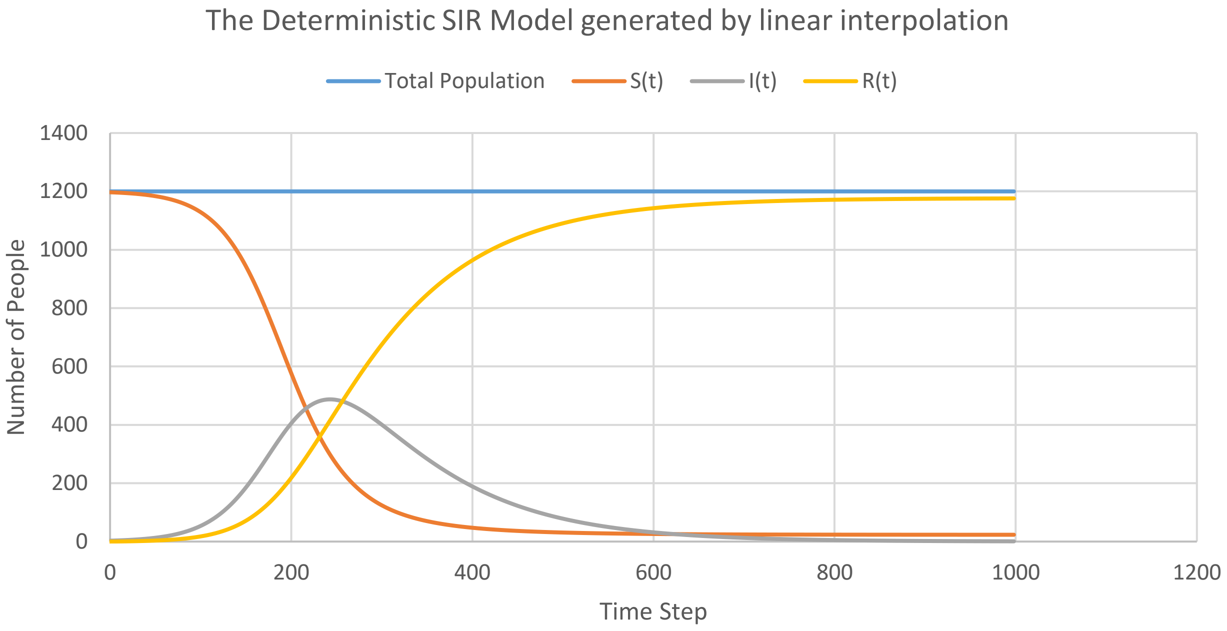


Figure 3 The DSIRM generated (See a sample of the data in Appendix 2).

Graph Analysis

From Figure 3, $S(t)$ decreases with time with an initially increasing and then decreasing rate. $I(t)$ increases and then decreases with time. $R(t)$ increases with time with an initially increasing and then decreasing rate. The functions are expected to behave this way. The total population remains constant, which fits the assumption that the epidemic introduced into the population is non-lethal. When more and more people get infected initially, the number of people in the S population group should decrease and the number of people in the I population group should increase. As the model progresses further over time, the infected people start to recover, thus the number of people in the I population group should decrease and the number of people in the R population group should increase.

The DSIRM generated in Figure 3 closely resembles other DSIRMs in other researches, meaning that the model is accurately generated and its error is not large.

Evaluation of the Deterministic SIR Model

Even though the DSIRM provided a rather thorough simulation of the introduction of the epidemic into a closed population, there are two major underlying errors with the model.

1. It assumed that the number of people infected and recovered would be exactly equal to the expected value of number of people being infected and recovered, which is not true, as in reality there would still be an underlying random probability on whether the person would be infected or not, or recovered or not.

2. The DSIRM doesn't account for the fact that it is impossible for a non-integer amount of people to be infected or recovered.

Take $t = 1$ as an example, the number of infected people increased from 3 to 3.0897, while the number of recovered people increased from 0 to 0.0309. In reality, it is impossible for the epidemic to infect an amount exactly according to the expected value of the probability, and it is also impossible to infect 0.0897 persons with a disease or recover 0.0309 persons from a disease as there exists only positive integer numbers of people.

Even though one might argue that rounding off the numbers would make the model accurate enough for the governments to use, it is not known whether rounding off the numbers would cause Round-off errors in the model. Round-off error is caused by using a discrete number of significant digits to represent real numbers (Ledder 2). This is another reason why the Probabilistic SIR Model is worthwhile to investigate, as there will be no round-off error in it.

Formulation of the Probabilistic SIR Model

To solve the underlying problems the DSIRM, the new SIR model has to have the following properties.

1. It is probabilistic. It takes into account that there would still be an underlying random probability on whether the person would be infected or not in an interaction between a member of the S population and a member of the I population.
2. It only infects a positive integral number of people. The number of people in the S, I and R population must be positive integers.

The number of people being infected by the epidemic can be considered to follow a binomial distribution, as the epidemic can be considered as performing "Bernoulli trials" on members of the S population when they came into contact members of the I population. The outcomes of this situation are a certain member of the S population either being "infected" or "not infected", but not both. The infections can also be assumed to be independent and the probability of infection is assumed to be identical in each interaction between S and I individuals. Furthermore, the number of interactions between S and I individuals per time period can be estimated based on current $S(t)$ and $I(t)$ value and the interaction rate. Thus on these conditions, a new Probabilistic SIR Model could be proposed.

The new Probabilistic SIR Model (PSIRM) will utilize the inverse of the binomial distribution function find the ΔS, ΔI and ΔR. Monte-Carlo Methods would be used to generate this PSIRM. Microsoft Excel will be used to randomly generate samples from a uniform probability distribution from 0 to 1 in the inverse of the cumulative binomial probability distribution function (ICBPDF) to simulate the random probability of one being infected by an epidemic.

Equations of the Probabilistic SIR Model

The ICBPDF is represented in this report as follows.

$$s = \mathrm{InvBi}(n, p, x)$$

Equation 7 The ICBPDF.

(Refer to Equation 7) Where s denotes the number of successes, n denotes the number of Bernoulli Trials conducted, p denotes the probability of success in the Bernoulli Trials, and x denotes the probability of a s number of successes in n Bernoulli Trials with a probability of success of p.

From the above equation, the equations for the dependent variable in a PSIRM can be represented as follows.

$$\Delta S = -\mathrm{InvBi}\left(\frac{S(t)}{N} \times \frac{I(t)}{N} \times N \times \alpha, i, X\right)$$

Equation 8 The change of the number of people in the S population group over time in a PSIRM.

(Refer to Equation 8) Similar to Equation 2, the first part of the ICBPDF represents the number of interactions between members of the S population group and members of the I population group per unit time, i.e. the number of trials performed on members of the S population group. The probability of success of the infection is i. X is a continuous random variable that is uniformly distributed between 0 and 1, simulating the underlying random probability of a person being infected by an epidemic. A negative sign is used here as the number of people in the S population group must decrease as people get infected.

$$\Delta I = \mathrm{InvBi}\left(\frac{S(t)}{N} \times \frac{I(t)}{N} \times N \times \alpha, i, X\right) - \mathrm{InvBi}(I(t), r, Y)$$

Equation 9 The change of the number of people in the I population group over time in a PSIRM.

(Refer to Equation 9) The first part of this equation is the negative of Equation 8, as the decrease of the number of people in the S population group should be equal to the negative of the increase of the number of people in the I population group. However, as the people in in the I population group recover, the number of people in the I population group decreases. The number of trials performed on the people in the I population group is $I(t)$. The probability of success of the recovery is r. Y is a continuous random variable that is uniformly distributed between 0 and 1, simulating the underlying random probability of a person recovering from an epidemic, thus forming the second part of the equation.

$$\Delta R = \mathrm{InvBi}(I(t), r, Y)$$

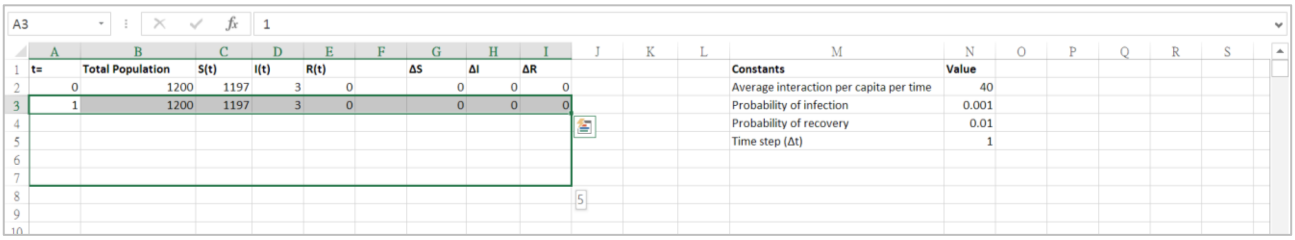
Equation 10 The change of the number of people in the R population group over time in a PSIRM.

(Refer to Equation 10) The number of people that recover over time is equal to the number of people that leave the I population due to recovery from the epidemic over time, which is the negative of the second part of the previous equation.

To allow a comparison between the DSIRM and the PSIRM, a virus with the same specifications will be introduced into a population identical to that when used to construct the DSIRM previously in this report.

Procedures of creating the Probabilistic SIR Model

The spreadsheet software Microsoft Excel is used to create the PSIRM.



t=	Total Population	S(t)	I(t)	R(t)		ΔS	ΔI	ΔR
0	1200	1197	3	0		0	0	0
1	1200	1197	3	0		0	0	0

Constants	Value
Average interaction per capita per time	40
Probability of infection	0.001
Probability of recovery	0.01
Time step (Δt)	1

Figure 4 Microsoft Excel layout of creating the PSIRM.

The step-by-step procedures for creating this PSIRM are shown in Appendix 3.

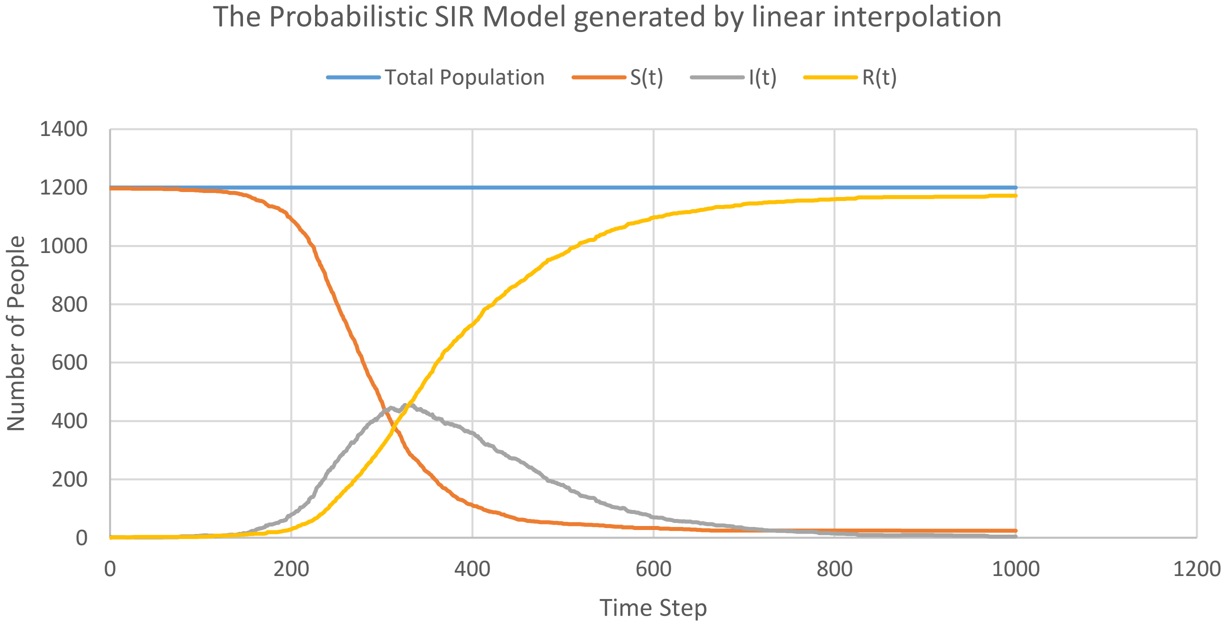


Figure 5 A sample of the PSIRM generated (See a sample of the data in Appendix 4).

However, since this SIR Model is probabilistic, it means that for almost every time this model is reproduced, the resulting model yielded would be different. This is illustrated in Figure 6, as the same model as above is reproduced for 20 times.

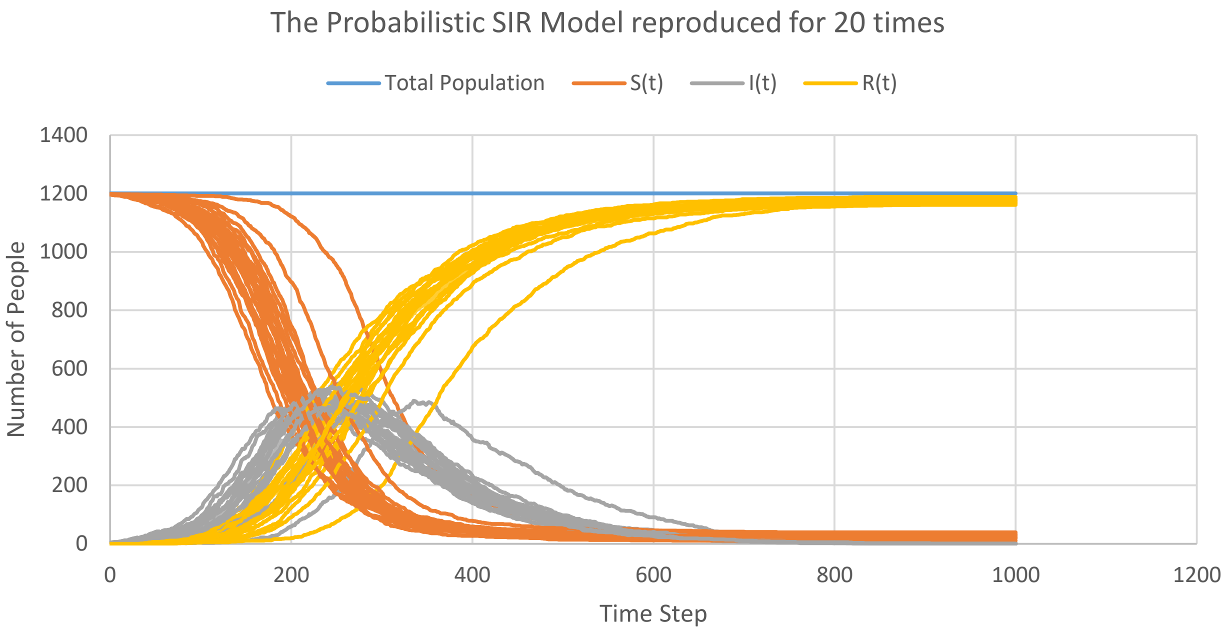


Figure 6 The Probabilistic SIR Model reproduced for 20 times.

Graph Analysis

(Refer to Figure 3 and Figure 4) As we can see, compared to the smooth lines the dependent variables formed over time in the DSIRM, the dependent variables form rough lines over time in the PSIRMs. It is due to the fact the outputs of the functions of time of the dependent variables in the PSIRM are all discrete values. The total population remains constant, which fits the assumption that the epidemic introduced into the population is non-lethal. In Figure 6, the areas formed the functions of time of the dependent variables represents the trends of the dependent variables over time. $S(t)$ decreases with time with an initially increasing and then decreasing rate. $I(t)$ increases and then decreases with time. $R(t)$ increases with time with an initially increasing and then decreasing rate.

Comparison between the DSIRM and the PSIRM

By comparing Figure 3 with Figure 5 and Figure 6, we can compare the properties of the DSIRM and the PSIRM.

The major similarity of the DSIRM and the PSIRM is that overall, the yield similar outputs, i.e. the shapes formed by the functions of time of the dependent variables of both models are similar, as well as the number of people in the S, I, and R population groups as t approaches a really large value. This implies that in a long term, the DSIRM can indeed give a rather accurate estimate as to how an epidemic interacts in a population group, and thus justifies the use of the DSIRM in predicting the long term trends of epidemics.

However, a major difference between the DSIRM and the PSIRM are the short term results. For example, $I(t)$ in Figure 3 has a much higher rate of increase initially compared to the rate of increase of $I(t)$ in Figure 5. This implies that in a short term, the PSIRM can give a much better estimate as to how an epidemic interacts in a population group, and thus the PSIRM should be used in predicting the short term trends of epidemics instead of the DSIRM.

Evaluation of the Probabilistic SIR Model

The advantages of using the PSIRM is that addresses the problem of the number of people infected and recovering over time being exactly equal to the expected number of people being infected or recovering over time, as well as the number of people in each population group not being discrete, that seen in the DSIRM. It is also a more realistic representation of how an epidemic interacts in a population in the short term.

However, there are still some disadvantages to the PSIRM. One major disadvantage it is still not an accurate estimation of how a disease interacts within a population. It still doesn’t resolve the core assumptions in the model that would make the model inaccurate, such as a fixed number of population, immunity upon recovery, constant probability of infection for everyone in the population in different time periods etc.

Conclusion and Further Investigations

Conclusion

This investigation aims to propose a new probabilistic approach in modelling epidemics. The reason to propose a new model is that the Deterministic SIR Model has the problems of the number of people infected and recovering over time being exactly equal to the expected number of people being infected or recovering over time, as well as the number of people in each population group not being discrete.

A Probabilistic SIR Model is constructed using the Inverse of the Cumulative Binomial Distribution Function. Monte-Carlo Methods is used in the Inverse of the Cumulative Binomial Distribution Function to obtain samples from an interval of 0 to 1 to obtain the number of people being infected and recovering over time.

The resulting Probabilistic SIR Model addresses the problems mentioned above of the Deterministic SIR Model, but still doesn't resolve some of the assumptions of a SIR Model that would make the model inaccurate, such as a fixed number of population, immunity upon recovery, constant probability of infection for everyone in the population in different time periods etc.

It can be observed that the Probabilistic SIR Model would yield results different from the Deterministic SIR Model in a short term but similar results in a long term. This implies that in a long term, the DSIRM can indeed give a rather accurate estimate as to how an epidemic interacts in a population group, and thus justifies the use of the DSIRM in predicting the long term trends of epidemics. But in a short term, the PSIRM can give a much better estimate as to how an epidemic interacts in a population group, and thus the PSIRM should be used in predicting the short term trends of epidemics instead of the DSIRM.

Further Investigations

In Figure 6, we can see that there are areas at which the functions of time of the dependent variables converge upon the reproduction of the SIR Model. This means that there is a high probability for the functions of time of the dependent variables to be in that area. Thus finding the probability distribution of the position of the functions in their respective area could be the subject of further investigation, as it would give an area where the functions would probably be in a certain period of time, which is more useful than an one-line estimate the DSIRM provides.

Furthermore, a PSIRM can be developed for larger, city-scale communities, such as Hong Kong. In order for such a model to be created, factors such as death rate of infection, natural death rate in the society, virus changing nature over time due to demographics and seasons etc. The model created can then be used as a basis to model real-life epidemics, such as the Zika Virus.

Bibliography

Blythe, Peter. *Mathematics for the International Student: International Baccalaureate Mathematics HL (options), International Baccalaureate Diploma Programme*. Adelaide Airport, S. Aust.: Haese & Harris Publications, 2005. Print.

Buckle, Nigel, Fabio Cirrito, and Iain Dunbar. *Mathematics Higher Level (core)*. Camberwell, Vic: IBID, 2012. Print.

Johnson, Teri. "Mathematical Modeling of Diseases: Susceptible-Infected-Recovered (SIR) Model." *Op12no2.me*. N.p., 2009. Web. 7 Oct. 2016.

Keeling, Matthew James, and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton: Princeton UP, 2008. Print.

Ledder, Glenn. "Error in Eurler's Method." University of Nebraska-Lincoln, n.d. Web. 31 Dec. 2016.

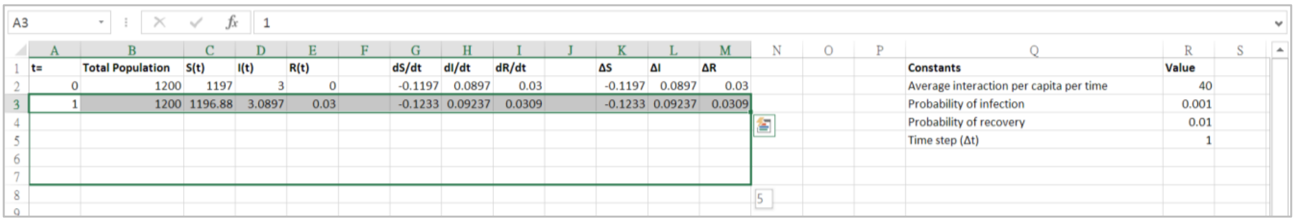
Mackay, D. J. C. "Introduction to Monte-Calo Methods." *Stat.duke.edu*. Duke University, n.d. Web. 8 Oct. 2016.

"Monte Carlo Method." *Monte Carlo Method*. Wolfram MathWorld, n.d. Web. 31 Dec. 2016.

Appendix 1

Procedures of creating the Deterministic SIR Model in Microsoft Excel

The spreadsheet software Microsoft Excel is used to create the DSIRM.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	t=	Total Population	S(t)	I(t)	R(t)		dS/dt	dI/dt	dR/dt		ΔS	ΔI	ΔR				Constants	Value
2	0	1200	1197	3	0		-0.1197	0.0897	0.03		-0.1197	0.0897	0.03				Average interaction per capita per time	40
3	1	1200	1196.88	3.0897	0.03		-0.1233	0.09237	0.0309		-0.1233	0.09237	0.0309				Probability of infection	0.001
4																	Probability of recovery	0.01
5																	Time step (Δt)	1

Figure 1 Microsoft Excel layout of creating the DSIRM.

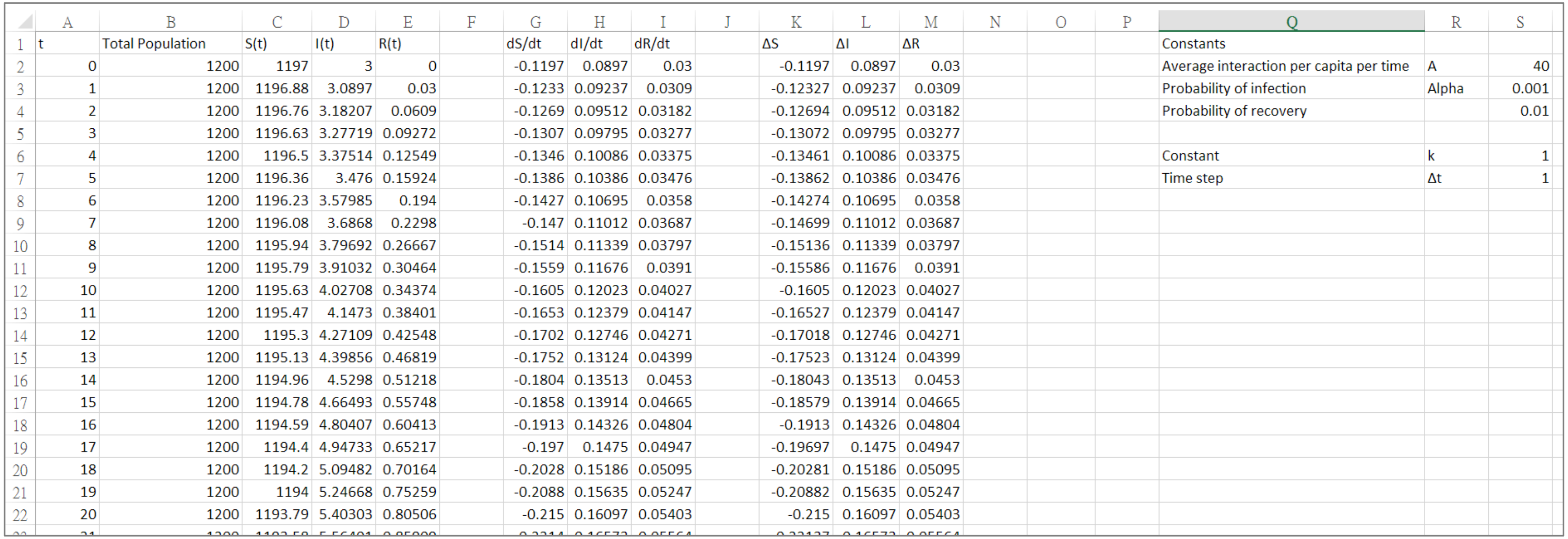
The procedures for creating this DSIRM is as follows.

1. Enter the constants and their corresponding values in columns Q and R as shown in Figure 2.
2. Enter the column headers as shown in boxes A1, B1, C1, D1, E1, G1, H1, I1, K1, L1 and M1 as shown in Figure 2.
3. Assuming that the total number of students is 1200, in which 1197, 3 and 0 are initially ($t = 0$) susceptible, infected and recovered respectively, enter the respective values into Boxes A2, B2, C2, D2 and E2 as shown in Figure 2.
4. Enter Equation 2 into Box G2 in Excel code (“=-C2/B2*D2/B2*R3*R2*B2”).
5. Enter Equation 3 into Box H2 in Excel code (“=-G2-D2*R4”).
6. Enter Equation 4 into Box I2 in Excel code (“=-G2-H2).
7. Obtain ΔS for $t = 0$ through multiplying $\frac{dS(t)}{dt}$ with Time step (Δt) by entering “=G2*R5” into Box K2.
8. Obtain ΔI for $t = 0$ through multiplying $\frac{dI(t)}{dt}$ with Time step (Δt) by entering “=H2*R5” into Box L2.
9. Obtain ΔR for $t = 0$ through multiplying $\frac{dR(t)}{dt}$ with Time step (Δt) by entering “=I2*R5” into Box M2.
10. For $t = 1$, enter “1” into Box A3.
11. To monitor the total population, enter “=C2+D2+E2” into Box B2.
12. To see the value of $S(t)$ at $t = 1$, enter “=C2+K2” into Box C3.
13. To see the value of $I(t)$ at $t = 1$, enter “=D2+L2” into Box D3.
14. To see the value of $R(t)$ at $t = 1$, enter “=E2+M2” into Box E3.
15. Highlight Boxes horizontally from Box G2 to Box M2 and drag it down to Row 3.
16. Highlight Boxes horizontally from Box A3 to Box M3 and drag it down to Row 1000.

Appendix 2

Sample of the data in the Deterministic SIR Model

The following is a sample of the Microsoft Excel data that forms the DSIRM in Figure 2 of this paper.

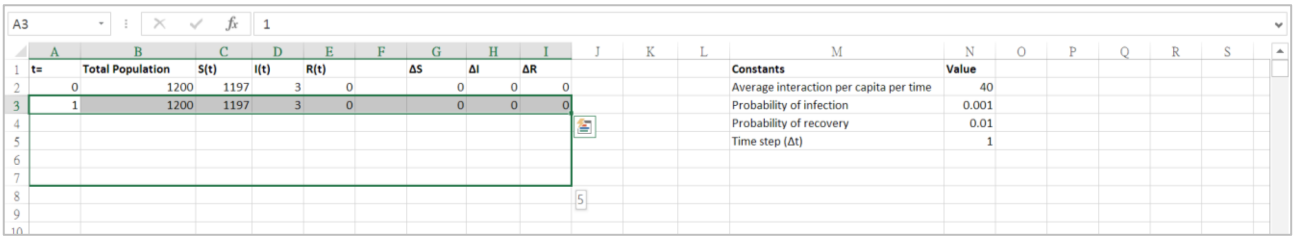


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	t	Total Population	S(t)	I(t)	R(t)		dS/dt	dI/dt	dR/dt		ΔS	ΔI	ΔR				Constants		
2	0	1200	1197	3	0		-0.1197	0.0897	0.03		-0.1197	0.0897	0.03				Average interaction per capita per time	A	40
3	1	1200	1196.88	3.0897	0.03		-0.1233	0.09237	0.0309		-0.12327	0.09237	0.0309				Probability of infection	Alpha	0.001
4	2	1200	1196.76	3.18207	0.0609		-0.1269	0.09512	0.03182		-0.12694	0.09512	0.03182				Probability of recovery		0.01
5	3	1200	1196.63	3.27719	0.09272		-0.1307	0.09795	0.03277		-0.13072	0.09795	0.03277						
6	4	1200	1196.5	3.37514	0.12549		-0.1346	0.10086	0.03375		-0.13461	0.10086	0.03375				Constant	k	1
7	5	1200	1196.36	3.476	0.15924		-0.1386	0.10386	0.03476		-0.13862	0.10386	0.03476				Time step	Δt	1
8	6	1200	1196.23	3.57985	0.194		-0.1427	0.10695	0.0358		-0.14274	0.10695	0.0358						
9	7	1200	1196.08	3.6868	0.2298		-0.147	0.11012	0.03687		-0.14699	0.11012	0.03687						
10	8	1200	1195.94	3.79692	0.26667		-0.1514	0.11339	0.03797		-0.15136	0.11339	0.03797						
11	9	1200	1195.79	3.91032	0.30464		-0.1559	0.11676	0.0391		-0.15586	0.11676	0.0391						
12	10	1200	1195.63	4.02708	0.34374		-0.1605	0.12023	0.04027		-0.1605	0.12023	0.04027						
13	11	1200	1195.47	4.1473	0.38401		-0.1653	0.12379	0.04147		-0.16527	0.12379	0.04147						
14	12	1200	1195.3	4.27109	0.42548		-0.1702	0.12746	0.04271		-0.17018	0.12746	0.04271						
15	13	1200	1195.13	4.39856	0.46819		-0.1752	0.13124	0.04399		-0.17523	0.13124	0.04399						
16	14	1200	1194.96	4.5298	0.51218		-0.1804	0.13513	0.0453		-0.18043	0.13513	0.0453						
17	15	1200	1194.78	4.66493	0.55748		-0.1858	0.13914	0.04665		-0.18579	0.13914	0.04665						
18	16	1200	1194.59	4.80407	0.60413		-0.1913	0.14326	0.04804		-0.1913	0.14326	0.04804						
19	17	1200	1194.4	4.94733	0.65217		-0.197	0.1475	0.04947		-0.19697	0.1475	0.04947						
20	18	1200	1194.2	5.09482	0.70164		-0.2028	0.15186	0.05095		-0.20281	0.15186	0.05095						
21	19	1200	1194	5.24668	0.75259		-0.2088	0.15635	0.05247		-0.20882	0.15635	0.05247						
22	20	1200	1193.79	5.40303	0.80506		-0.215	0.16097	0.05403		-0.215	0.16097	0.05403						

Appendix 3

Procedures of creating the Probabilistic SIR Model

The spreadsheet software Microsoft Excel is used to create the PSIRM.



t=	Total Population	S(t)	I(t)	R(t)		ΔS	ΔI	ΔR				Constants	Value
0	1200	1197	3	0		0	0	0				Average interaction per capita per time	40
1	1200	1197	3	0		0	0	0				Probability of infection	0.001
												Probability of recovery	0.01
												Time step (Δt)	1

Figure 1 Microsoft Excel layout of creating the PSIRM.

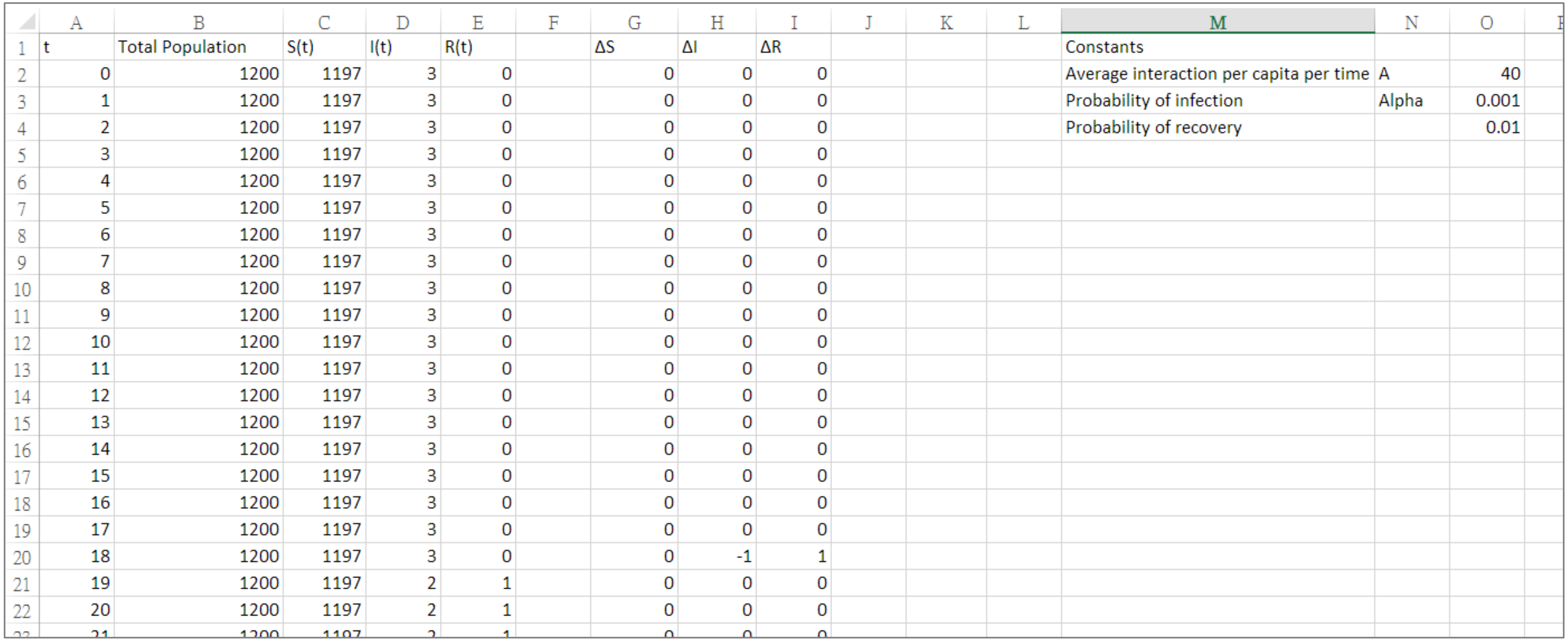
The procedures for creating this PSIRM is as follows.

1. Enter the constants and their corresponding values in columns M and N as shown in Figure 4.
2. Enter the column headers as shown in boxes A1, B1, C1, D1, E1, G1, H1 and I1 as shown in Figure 4.
3. Assuming that the total number of students is 1200, in which 1197, 3 and 0 are initially ($t = 0$) susceptible, infected and recovered respectively, enter the respective values into Boxes A2, B2, C2, D2 and E2 as shown in Figure 4.
4. Obtain ΔS for $t = 0$ by entering Equation 8 into Box G2 in Excel code ("=-BINOM.INV(ROUND(C2*D2/B2*N2,0), N3, RAND())"). Note that a "ROUND(number, num_digits)" function is needed for the first part of Equation 8, as a positive integral number of trials is needed for the ICBPDF. Monte-Carlo methods is used here as a "Rand()" function would give random values from 0 to 1, simulating the underlying random probability of a person being infected by an epidemic.
5. Obtain ΔI for $t = 0$ by entering Equation 9 into Box H2 in Excel code ("=-G2-BINOM.INV(D2, N4, RAND())").
6. Obtain ΔR for $t = 0$by entering Equation 10 into Box I2 in Excel code ("=-G2-H2").
7. For $t = 1$, enter "1" into Box A3.
8. To monitor the total population, enter "=C2+D2+E2" into Box B2.
9. To see the value of $S(t)$ at $t = 1$, enter "=C2+K2" into Box C3.
10. To see the value of $I(t)$ at $t = 1$, enter "=D2+L2" into Box D3.
11. To see the value of $R(t)$ at $t = 1$, enter "=E2+M2" into Box E3.
12. Highlight Boxes horizontally from Box G2 to Box I2 and drag it down to Row 3.
13. Highlight Boxes horizontally from Box A3 to Box I3 and drag it down to Row 1000.

Appendix 4

Sample of the data in the Probabilistic SIR Model

The following is a sample of the Microsoft Excel data that forms the PSIRM in Figure 3 of this paper.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	t	Total Population	S(t)	I(t)	R(t)		ΔS	ΔI	ΔR				Constants		
2	0	1200	1197	3	0		0	0	0				Average interaction per capita per time	A	40
3	1	1200	1197	3	0		0	0	0				Probability of infection	Alpha	0.001
4	2	1200	1197	3	0		0	0	0				Probability of recovery		0.01
5	3	1200	1197	3	0		0	0	0						
6	4	1200	1197	3	0		0	0	0						
7	5	1200	1197	3	0		0	0	0						
8	6	1200	1197	3	0		0	0	0						
9	7	1200	1197	3	0		0	0	0						
10	8	1200	1197	3	0		0	0	0						
11	9	1200	1197	3	0		0	0	0						
12	10	1200	1197	3	0		0	0	0						
13	11	1200	1197	3	0		0	0	0						
14	12	1200	1197	3	0		0	0	0						
15	13	1200	1197	3	0		0	0	0						
16	14	1200	1197	3	0		0	0	0						
17	15	1200	1197	3	0		0	0	0						
18	16	1200	1197	3	0		0	0	0						
19	17	1200	1197	3	0		0	0	0						
20	18	1200	1197	3	0		0	-1	1						
21	19	1200	1197	2	1		0	0	0						
22	20	1200	1197	2	1		0	0	0						
23	21	1200	1197	2	1		0	0	0						

Appendix 5

Summary of the Equations, Tables and Figures

The following is a summary of the Equations, Tables and Figures shown in the paper.

Equations

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}, x=0,1,2,\dots n$$ (Fabio 530)

Equation 1 The Binomial Probability Distribution Function

$$\frac{dS(t)}{dt}=-\frac{S(t)}{N}\times\frac{I(t)}{N}\times i\times N\times\alpha$$

Equation 2 The rate of change of the number of people in the S population group over time in a DSIRM.

$$\frac{dI(t)}{dt}=\frac{S(t)}{N}\times\frac{I(t)}{N}\times i\times N\times\alpha-I(t)\times r$$

Equation 3 The rate of change of the number of people in the I population group over time in a DSIRM.

$$\frac{dR(t)}{dt}=I(t)\times r$$

Equation 4 The rate of change of the number of people in the R population group over time in a DSIRM.

$$\frac{dy}{dx}=f(x,y)\approx\frac{\Delta y}{\Delta x}$$

Equation 5 The fundamental approximation in Euler's method.

$$\begin{cases} S_1=S_0+\Delta t\times\frac{dS(t)}{dt}=S_0+\Delta t\times\left(-\frac{S_0}{N}\times\frac{I_0}{N}\times i\times N\times\alpha\right) \\ I_1=I_0+\Delta t\times\frac{dI(t)}{dt}=I_0+\Delta t\times\left(\frac{S_0}{N}\times\frac{I_0}{N}\times i\times N\times\alpha-I_0\times r\right) \\ R_1=R_0+\Delta t\times\frac{dR(t)}{dt}=R_0+\Delta t\times(I_0\times r) \end{cases}$$

Equation 6 Equation for the DSIRM.

$$s=\mathrm{InvBi}(n,p,x)$$

Equation 7 The ICBPDF.

$$\Delta S=-\mathrm{InvBi}\left(\frac{S(t)}{N}\times\frac{I(t)}{N}\times N\times\alpha,i,X\right)$$

Equation 8 The change of the number of people in the S population group over time in a PSIRM.

$$\Delta I = \text{InvBi}\left(\frac{S(t)}{N} \times \frac{I(t)}{N} \times N \times \alpha, i, X\right) - \text{InvBi}(I(t), r, Y)$$

Equation 9 The change of the number of people in the I population group over time in a PSIRM.

$$\Delta R = \text{InvBi}(I(t), r, Y)$$

Equation 10 The change of the number of people in the R population group over time in a PSIRM.

Tables

Group	Dependent variables as functions of unit time	Explanation
S	$S(t)$	Represents the number of people S population group.
I	$I(t)$	Represents the number of people I population group.
R	$R(t)$	Represents the number of people R population group.

Table 1 The dependent variables in a SIR Model.

Variables	Explanation
N	Total number of people in the whole population.
α	The average number of interactions between people in the population per capita per unit time.
i	Probability of (Rate of) infection of the epidemic.
r	Probability of (Rate of) recovery from the epidemic.

Table 2 Other Variables needed for the creation of the SIR Model.

Figures

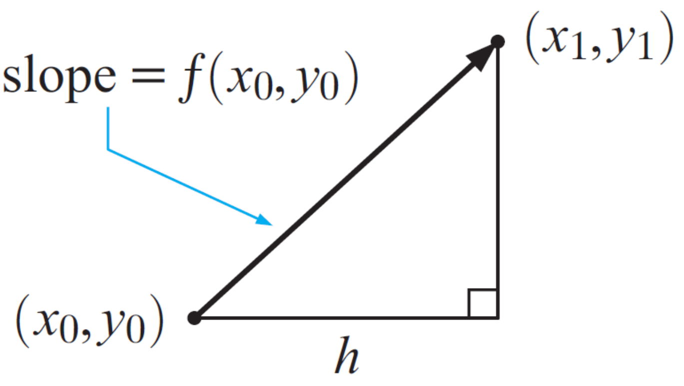
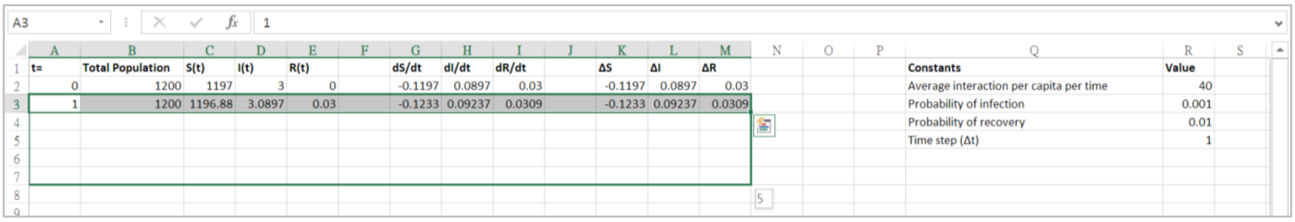


Figure 1 A graphical illustration of Euler's method.

(Source: Blythe, Peter. *Mathematics for the International Student: International Baccalaureate Mathematics HL (options), International Baccalaureate Diploma Programme*. Adelaide Airport, S. Aust.: Haese & Harris Publications, 2005, p. 233)



t=	Total Population	S(t)	I(t)	R(t)		dS/dt	dI/dt	dR/dt		ΔS	ΔI	ΔR
0	1200	1197	3	0		-0.1197	0.0897	0.03		-0.1197	0.0897	0.03
1	1200	1196.88	3.0897	0.03		-0.1233	0.09237	0.0309		-0.1233	0.09237	0.0309

Constants	Value
Average interaction per capita per time	40
Probability of infection	0.001
Probability of recovery	0.01
Time step (Δt)	1

Figure 2 Microsoft Excel layout of creating the DSIRM.

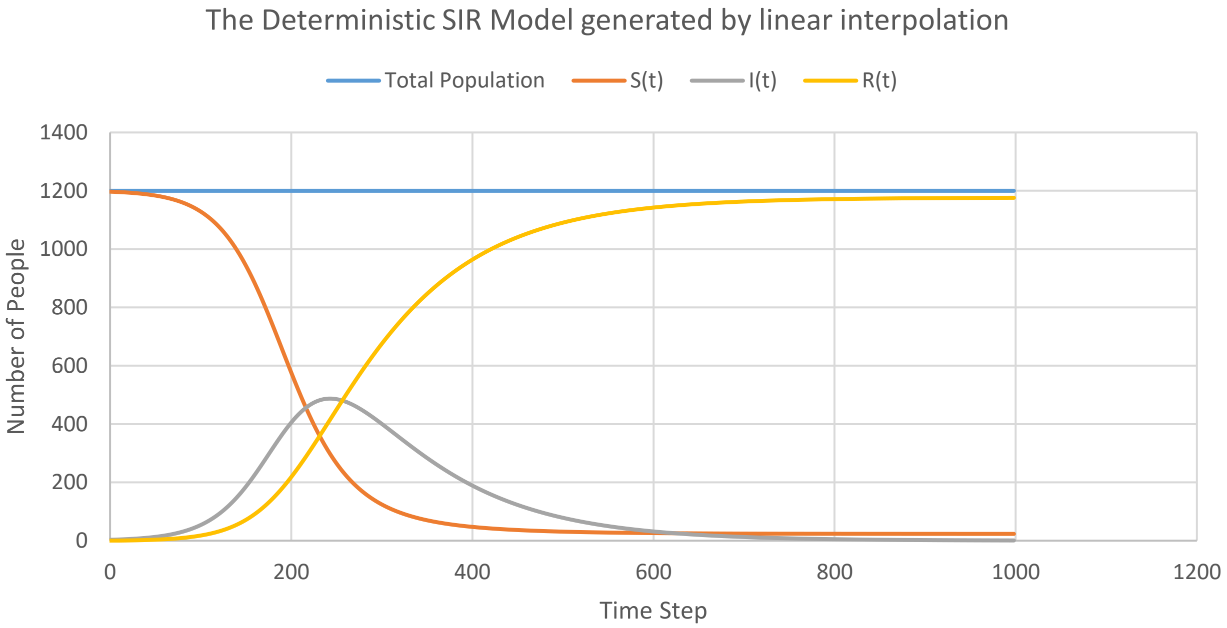
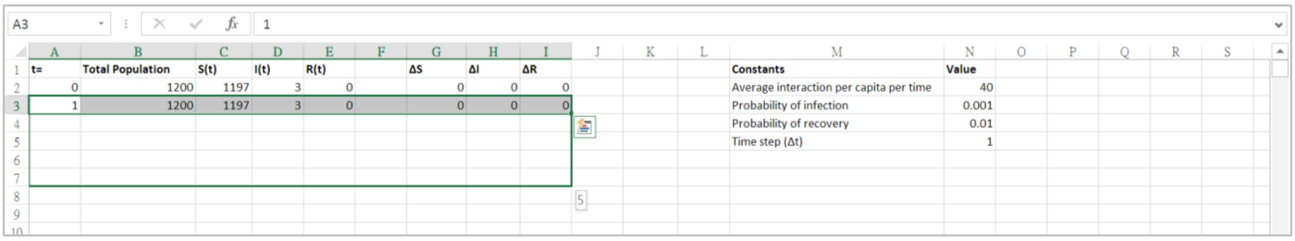


Figure 3 The DSIRM generated (See a sample of the data in Appendix 2).



t=	Total Population	S(t)	I(t)	R(t)		ΔS	ΔI	ΔR
0	1200	1197	3	0		0	0	0
1	1200	1197	3	0		0	0	0

Constants	Value
Average interaction per capita per time	40
Probability of infection	0.001
Probability of recovery	0.01
Time step (Δt)	1

Figure 4 Microsoft Excel layout of creating the PSIRM.

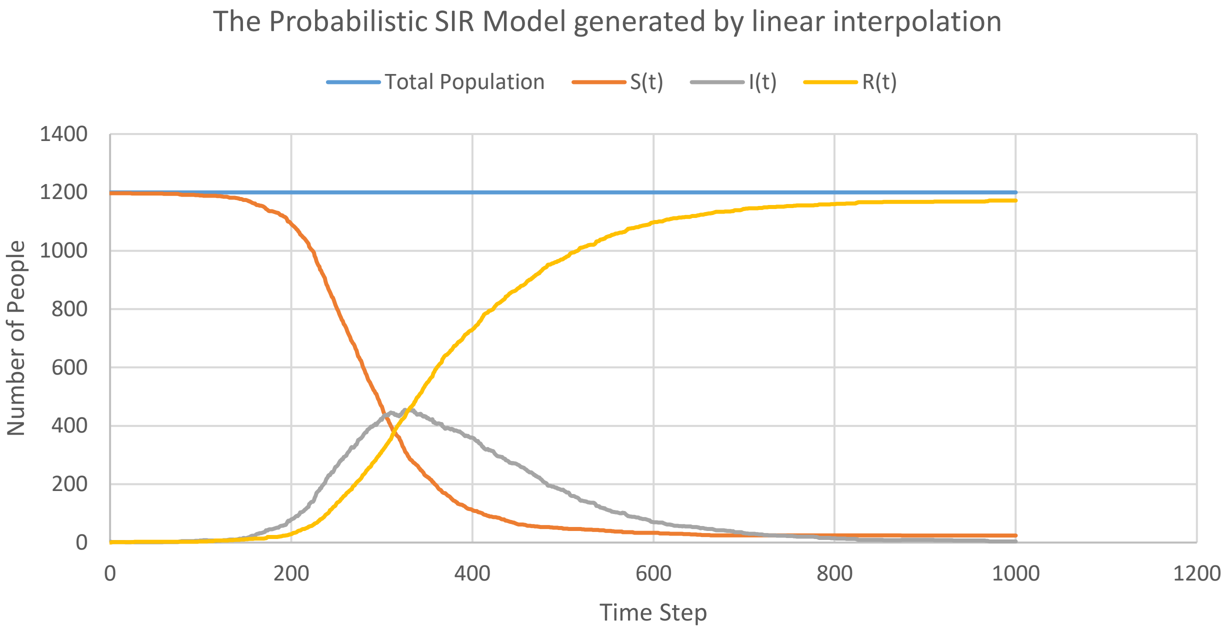


Figure 5 A sample of the PSIRM generated (See a sample of the data in Appendix 4).

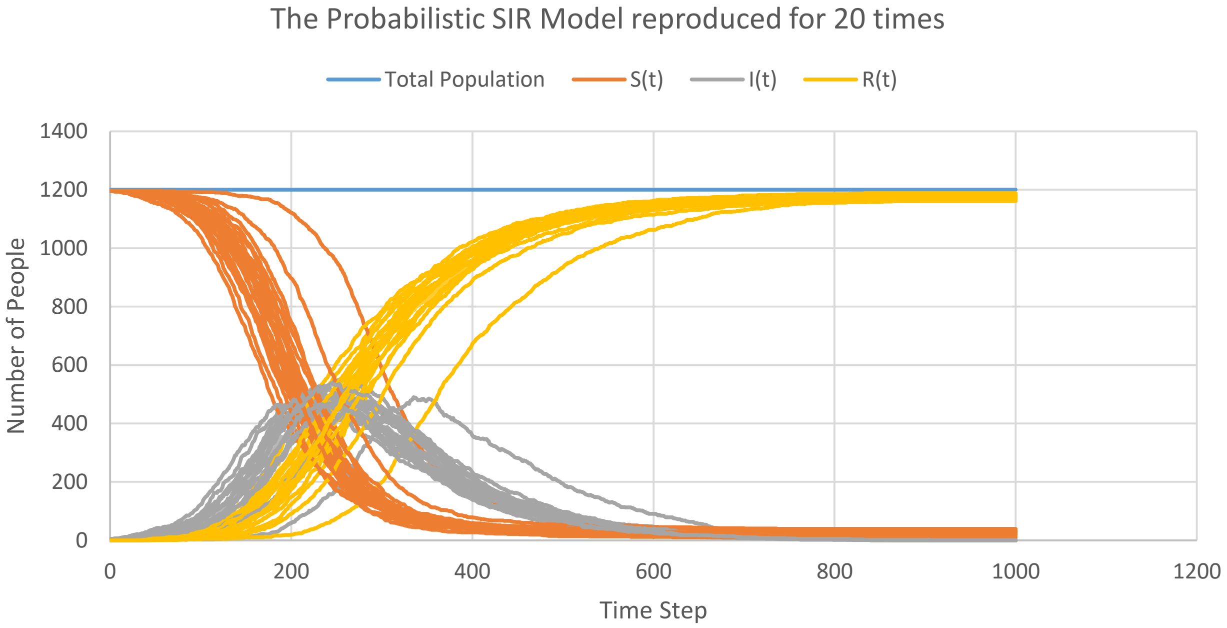


Figure 6 The Probabilistic SIR Model reproduced for 20 times.