AMSI VACATION RESEARCH SCHOLARSHIPS 2020–21

Get a Thirst for Research this Summer

Control Strategies for a Superspreading Virus at Low Prevalence

Kristian Caracciolo

Supervised by Joel Miller

La Trobe University

Vacation Research Scholarships are funded jointly by the Department of Education, Skills and Employment and the Australian Mathematical Sciences Institute.

Abstract

This project sought to investigate the super-spreading dynamics displayed by viruses such as SARS-CoV-2 in low prevalence settings like those that exist in Australia. In near extinction settings, the importance of random events to the transmission of the virus are greater as opposed to populations where the virus is better established and the behaviour becomes more deterministic. Super-spreading events are the manifestation of the stochastic nature of disease transmission wherein the number of infections caused by a single person is occasionally well above average. Mathematically this behaviour can be modelled using probability distributions such as the negative binomial distribution. To study this behaviour, we develop a generation-based stochastic SIR model to simulate the propagation of a super-spreading virus beginning from a single infected person. The developed model is subsequently used to gain some insight into optimal resource allocation by studying the effect that varying parameters such as tracing efficiency, tracing capacity and background testing rate have on the probability of an epidemic emerging.

Keywords: SARS-Cov-2, Super-spreading, Negative Binomial Distribution, Stochastic SIR Model, Resource Allocation

1 Introduction

The emergence of the SARS-CoV-2 in late 2019 from Wuhan China (Shereen et al. 2020) has had a profound impact on the world. Beyond the obvious consequences to the health of those who develop COVID-19 which has resulted in more than 2 million deaths (Dong, Du, and Gardner 2020), the hazard to mental health posed by the disease is emerging with high rates of anxiety, depression and psychological distress being reported in many countries experiencing outbreaks (Xiong et al. 2020). Travel restrictions, lock-downs, and unpredictable spending behaviour by the public have slowed supply chains, trade and created a number of market anomalies resulting in a reduction of global GDP (Baldwin and Mauro 2020).

A great number of these deleterious effects resulting from the spread of the virus can be mitigated or avoided entirely if the virus is prevented from being establishing itself. Thus, it is important to understand how viruses like SARS-Cov-2 propagate in populations where it is near extinction and how effective containment responses should be to eliminate the chance of pandemic spread. The transmission dynamics of SARs-CoV-2 are characterised by super-spreading events in which a small number of infected are responsible for a disproportionately large number of infections. The random nature of these super-spreading events means the impact on low prevalence communities differs from high prevalence ones. For instance, a super-spreading event in which 30 people are infected will be more impactful to the propagation of the virus in a community in which only a handful of infected exist versus one in which many thousand exist.

To investigate the effects of interventions on the probability that a super-spreading virus reaches extinction, we develop a stochastic SIR model to simulate the propagation of a virus through a community beginning with a single infection. We subsequently vary parameters relating to the containment response like the tracing efficiency, tracing capacity and background testing rates to gain some insight on where limited resources would be best allocated to combat a super-spreading virus. We conclude that within the context of preventing outbreaks, high tracing capacities are least effective whilst high tracing efficiencies and background testing rates are highly effective albeit potentially difficult to achieve.

2 Statement of Authorship

This report, the SIR model at the core of the project along with all supporting functions were written by Caracciolo with input and supervision from Miller. All figures contained in the report were generated by Caracciolo in Python 3.7.4 with input from Miller. Interpretation and discussion of all results produced by the project are once again the product of Caracciolo and Miller.

3 Virus Modelling

A commonly employed model for the transmission of infectious diseases is the SIR (susceptible, infected, removed or recovered) model. The model emerged in the early twentieth century in part from the work of Ronald Ross and William Hamer (Weiss 2013). The model begins with a population divided into three groups; those who are susceptible to infection, those who are infected and those who are removed following recovery from the infection. The SIR model employs a number of assumptions which are present in our implementation such as a closed population, the absence of any natural deaths or births, individuals becoming infectious immediately upon being infected, interaction of groups being proportional to the group size and once recovered, individuals are unable to be infected again (Weiss 2013).

Traditionally, an SIR model is deterministic, using differential equations to describe the rate at which individuals transition between each of the groups of the model. For this project we instead opt for a stochastic SIR model which simulates the propagation of the virus through the population according to discrete probability distributions called offspring distributions. Stochastic SIR models allow us to account for the random nature of super-spreading events. Additionally, the model we employ is generational which for the purpose of our simulation means that the unit of time represented is the period of time that infected individuals remain infected. This means we are fundamentally unable to account for heterogeneity in the period of time that people are infectious and cannot account for any incubation time, heterogeneous or otherwise.

Our model begins with a list of susceptible people of a size defined by user inputs; 10000 for all simulations performed here. One of these people is then infected and added to a list of infected persons. For each person in the list of infected, the simulation samples from an offspring distribution to determine the number of infections that would be caused. If the output of the offspring distribution is n, then n people are selected at random from the total population list and added to the list of infected if susceptible whilst the persons causing infections in the current generation have their status set to recovered. This process repeats until the number of infected reaches zero.

Prior to causing any new infections there is a chance that infected people are essentially sequestered by having the outcome of the offspring set to 0. This probability is referred to as the tracing efficiency and was made to decay as a function of the current number of infections to represent a tracing system becoming overwhelmed. The rate of decay is based on a Holling's type II functional response predator prey model where the rate at which predators can process new prey is not proportional to the rate at which the prey population increases (Real 1977). In the context of our SIR model, as the number of infected increases, it becomes less likely they are identified by tracing. The rate of decay is determined by the simulation input named tracing capacity as shown below in equation 2 where $E(i)$ is the current tracing efficiency, E_{max} is the input maximum efficiency, i is the number of infected and C is the input tracing capacity. The rate of decay is also shown in

Figure 1. Our model loops over a number of tracing efficiencies and capacities.

$$E(i) = E_{max} - E_{max}\left(\frac{\frac{i}{c}}{1-\frac{i}{c}}\right) \tag{1}$$

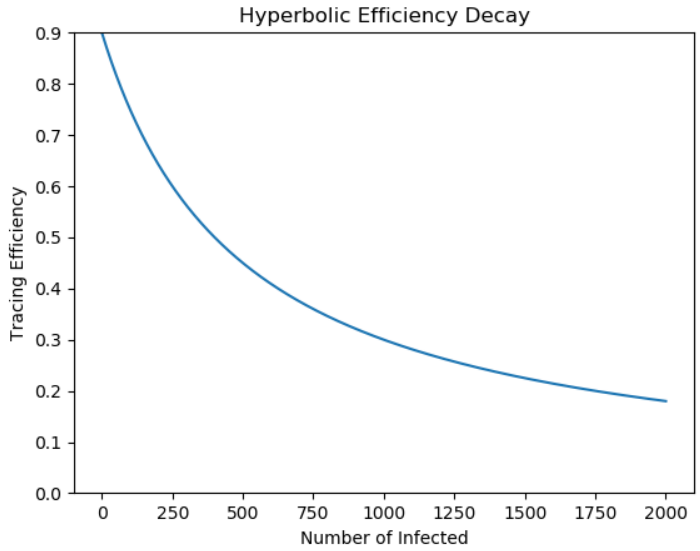

Figure 1: Hyperbolic decay of tracing efficiency as a function of number of infected. Curve based on Holling's type II functional response. Input parameters are $E_{max} = 0.9$ and c = 500.

As mentioned earlier, offspring distributions are those used to describe the manner in which a virus propagates in models like SIR. The distributions we consider fall into one of two categories, those describing heterogeneous and homogeneous transmission. The latter of these modes is often modelled using the Poisson distribution shown in Figure 2. The Poisson distribution is characterised by a mean value, R, such that on average the number of new infections caused by an infected person is R. This mean value R is often referred to as the 'reproduction number'. This distribution bears a relatively small variance compared to that used to model heterogeneous transmission; the negative binomial distribution. The negative binomial distribution shown in Figure 3 is characterised by a dispersion parameter, k, in addition to the mean value, R, that dictates the variance of the distribution. It is this dispersion parameter which allows for super-spreading to be accounted for by means of introducing a large variance. Values of $k < 0.5$ are considered to represent extreme super-spreading whilst $0.5 < k < 1.0$ indicate moderate super-spreading. When modelling the spread of SARS-CoV-2, the Poisson distribution is often used where reproduction numbers alone are of interest (Chintalapudi et al. 2020). Such investigations are typically considering the rate at which the virus propagates in higher prevalence settings. Studies of lower prevalence setting which investigate the emergence of outbreaks tend to utilise the negative binomial distribution (Kucharski et al. 2020) for that fact that it can account for the stochastic super-spreading events which influence the probability the virus reaches extinction. The use of the negative binomial distribution doesn't appear to have any motivation beyond its ability to fit observed offspring distributions.

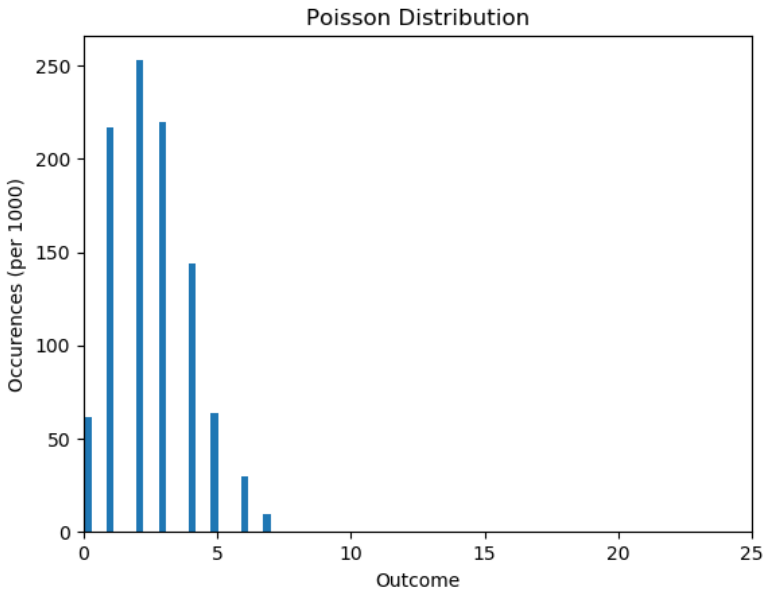

Figure 2: Histogram plot showing the results of sampling the Poisson distribution with a mean value of 2.5 a thousand times. All outcomes are clustered relatively closely around the mean value.

A key output of our simulation was the extinction probability which we define as the probability that the virus dies out before infecting a significant portion of the population. Within the context of our SIR model, we consider the virus as having failed to propagate if less then 100 of our 10000 population are infected before extinction. The probability of this occurring emerges as the proportion of a large number of simulations that fall in this category of fewer than 100 infections. The extinction probability depends on a number of the simulations inputs such as tracing efficiency, background testing rates and offspring distribution. In the absence of a limited tracing capacity, it is also possible to extract the extinction probability using probability generating functions.

Probability Generating Functions (PGFs) are a useful tool in modelling the stochastic behaviour of pandemic outbreaks. Despite the name, PGFs do not in fact generate probabilities from a given input value. Instead, their name indicates that the function is generated from a sequence of numbers in the form shown in equation (2) where r_i is the probability of selecting the value i from a given discrete probability distribution (Miller 2018). In the case of a PGF, this sequence of numbers represents a probability distribution of integers.

$$f(x) = \Sigma_i r_i x^i \tag{2}$$

Using PGFs, we may derive the probability of a virus reaching extinction at some generation of its spread. Though useful tools, it can be difficult if not impossible to account for the effect that interventions like contact tracing, the delay with which tracing is implemented after an outbreak is discovered and the background testing rate have on the probability of extinction using PGFs. Part of this difficultly arises from the fact that PGFs rely on two events being independent and aspects of our simulation like tracing capacity introduces a dependence

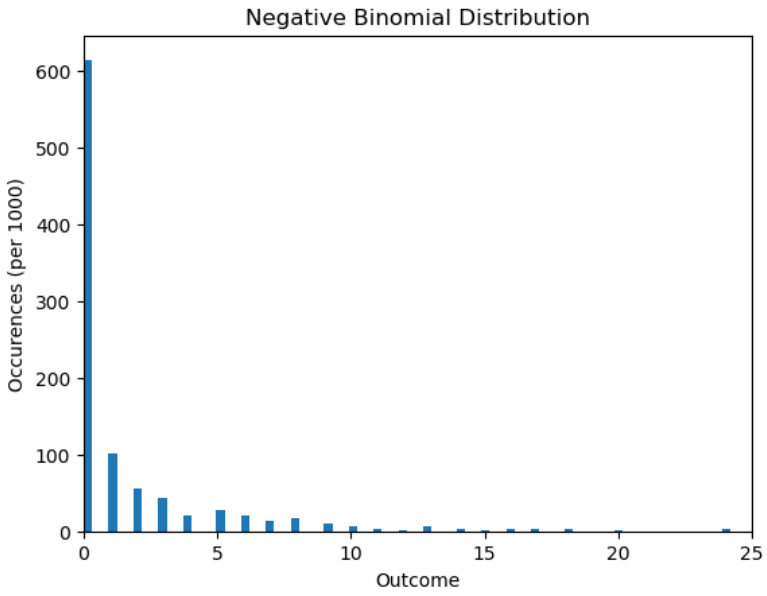

Figure 3: Histogram plot showing the results of sampling the Negative Binomial distribution with a mean value of 2.5 and dispersion parameter 0.16 a thousand times. Compared to the Poisson distribution (Figure 2) a much greater variance of observed. The occurrences where the outcome is greater than 10 could be considered super-spreading events given that they are well above the mean. To maintain the same mean value as the Poisson distribution, these super-spreading events are accompanied by a greater number of 0 outcomes.

between the probability of extinction and the number of infections in previous generations. The only aspect of the interventions that we account for using PGFs is a constant tracing efficiency.

4 Method

The stochastic SIR model used was written in Python 3.7.4. The parameters of the simulation are population size, offspring distribution type, reproduction factor, dispersion coefficient, tracing efficiency, tracing capacity, self-reporting rate, tracing delay and stopping threshold. Reproduction factor and dispersion coefficient have been explained in the earlier paragraph on distribution types. Unless stated otherwise, all simulations presented with this report utilise a reproduction factor of 2.5 and in the case of the negative binomial distribution a dispersion parameter of 0.16. All simulations shown use a population size of 10000. Tracing efficiency and tracing capacity were explored in the earlier discussion on the SIR model and Holling's type II functional response alongside Figure 1. Our simulations sweep over a range of tracing efficiencies between 0 and 1 and tracing capacities between 0 and 300.

Self-reporting rate and tracing delay dictate the generation at which tracing begins. Self-reporting rate represents the proportion of infected individuals that test and isolate themselves. Tracing can only begin after the first successful self-reported case either beginning at that moment if tracing delay is zero or where tracing delay equals some positive integer n, beginning n generations after the first self-reported case. The stopping

threshold is a parameter used to speed up the computation time of simulations. It does so by terminating the simulation prematurely if the number of infected reaches the stopping threshold under the assumption that the virus would have reached pandemic levels. This shortcut arose from our early observation that simulations produced results in one of two distinct categories; extinction by failure to propagate or extinction via depletion of the susceptible population. Failure to propagate can be seen in Figure 4 where the number of recovered individuals and the number of times the virus failed to infect a person who had either recovered or was currently infected is shown for 1000 repetitions of the simulation. The clustering around 0 recovered indicates the virus failing to propagate whilst the cluster around 10000 failed infections indicates a depletion of the susceptible population. The clear segregation of results into these two groups served as justification for categorising any simulation resulting in greater than 100 infections as a pandemic.

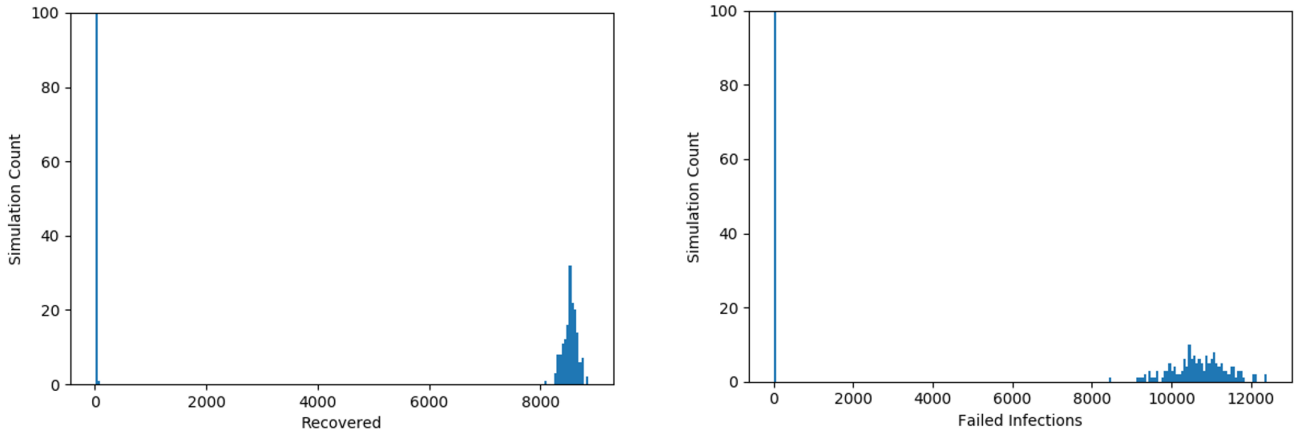

Figure 4: Histogram plot showing the results of 1000 stochastic SIR model simulations in terms of the number of recovered individuals and infections that failed due to the recipient already having been infected for the negative binomial distribution. Note, the y-axis has been limited to 100 to better resolve the clusters at high recovered and failed reinfection counts.

5 Results

Given that we swept over a range of tracing efficiencies and capacities, we found our data set to be best represented using contour plots seen in Figures 5 and 6. In these contour plots we use the colour bar to define the probability of an pandemic occurring (1 - extinction probability) to gain a visually intuitive depiction of the trends with respect to tracing capacity and efficiency. Our early results are shown in Figure 5 which utilised a finite capacity tracing that begins immediately upon the first infection. It is apparent that for the same reproduction factor, outbreaks are much more likely to occur when modelling with the Poisson distribution over the entire range of capacities and efficiencies. This is certainly a result of the greater likelihood for 0 new infections using the negative binomial distribution resulting in instant extinction of the single infection our SIR model begins with. The trend which emerges with respect to tracing efficiency is an unremarkable one; the probability of extinction increases rapidly with increasing tracing efficiency. The trend emerging with respect to tracing capacity was notable in that the extinction probability quickly approaches a limit above a

capacity of approximately 60. At especially low tracing capacities, a significant increase in the probability of a pandemic is observed even for high tracing efficiencies above 0.9. Both trends seem to suggest that with regard to the likelihood of extinction, it is only necessary to have a tracing capacity high enough to prevent immediate collapse of the tracing system as infection counts begin to increase.

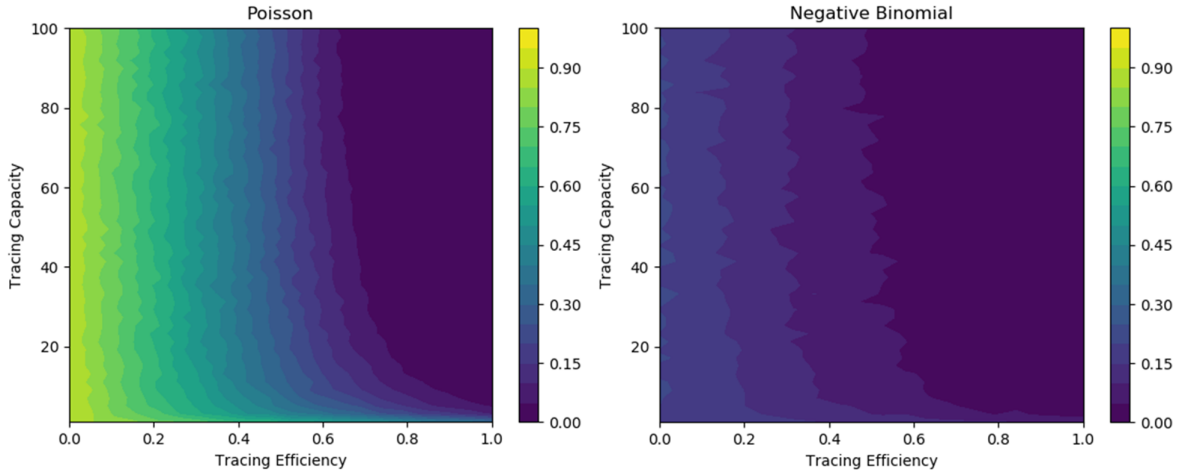

Figure 5: Contour plots showing the probability that an outbreak reaches pandemic levels for both the Poisson and Negative Binomial offspring distributions. For high tracing capacities, extinction probabilities presented here are approximately equivalent to those attained from PGFs.

Incorporating self-reporting rates and a delayed tracing system as detailed in the methods section, allowed the investigation of new trends. Comparing the results of these simulations to those in Figure 5 helps to realise the threshold at which the delay to the onset of tracing cannot be overcome by the aid of the persistent background testing rates. The left plot of Figure 6, representing a self-reporting rate of 10%, is clearly below this threshold as pandemics appear more likely across the range of capacities and efficiencies. To better visualise the comparison being made we produced contour plots that display the difference in extinction probability between the delayed tracing case in which capacity is finite and tracing does not begin until an infected person successfully self-reports, and the PGF case wherein tracing is both immediate and has infinite capacity. We refer to these as difference plots (see Figures 7 - 9). It should be noted that in these plots, colour represents the difference in extinction probability as opposed to the probability of a pandemic like in Figures 5 and 6.

The difference calculated in these plots is defined as simulation probability minus PGF probability. Thus, bluer hues in the difference plots indicate regions where the simulation is performing more poorly than the PGF. With this knowledge, two regions separated by a minimum at a tracing efficiency of 0.6 emerge. For tracing efficiencies less than 0.6, the simulation performs better than the PGF. This is most likely due to the self-reporting rate being comparable or greater than the tracing efficiency in this regime leading to more rapid extinction. Above the minimum, the simulation seems unable to produce the same extinction probabilities achieved by the PGF. In this regime, the same contours seen in the earlier plots of pandemic probability appear to return. We believe this to be due to the PGF having an extinction probability of zero for tracing efficiencies above 0.6, which by our definition of difference, means the difference is equal to the simulation value of extinction. We return to this point with Figures 10 and 11.

Experimenting with beginning tracing in the generation following detection of the virus, extinction in the

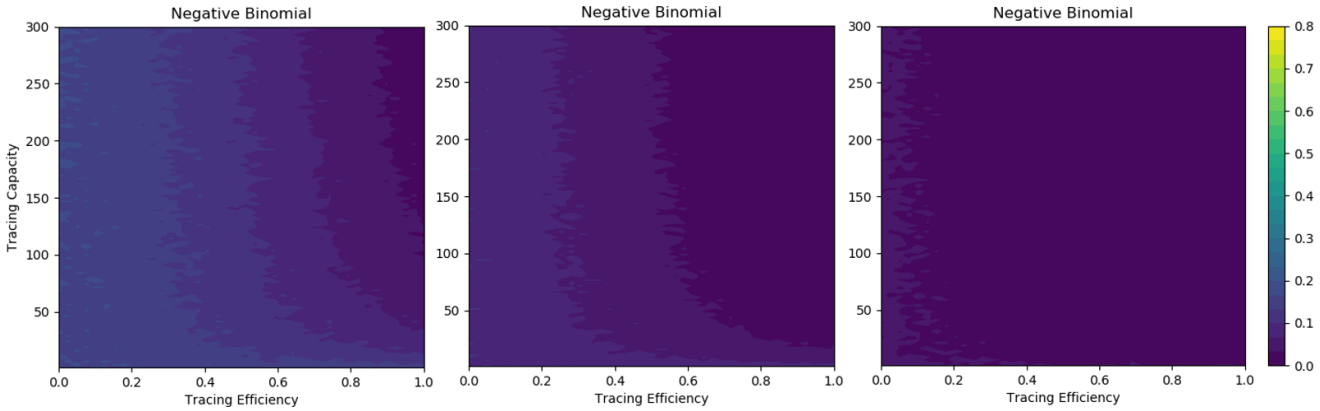

Figure 6: Contour plots showing the probability of a pandemic in a simulation with a delayed tracing onset using the negative binomial distribution. Plots from left to right correspond to self-reporting rates of 10% , 30% and 50% . Outcomes for 10% plot poorer than those presented in Figure 5 which incorporated no tracing delay.

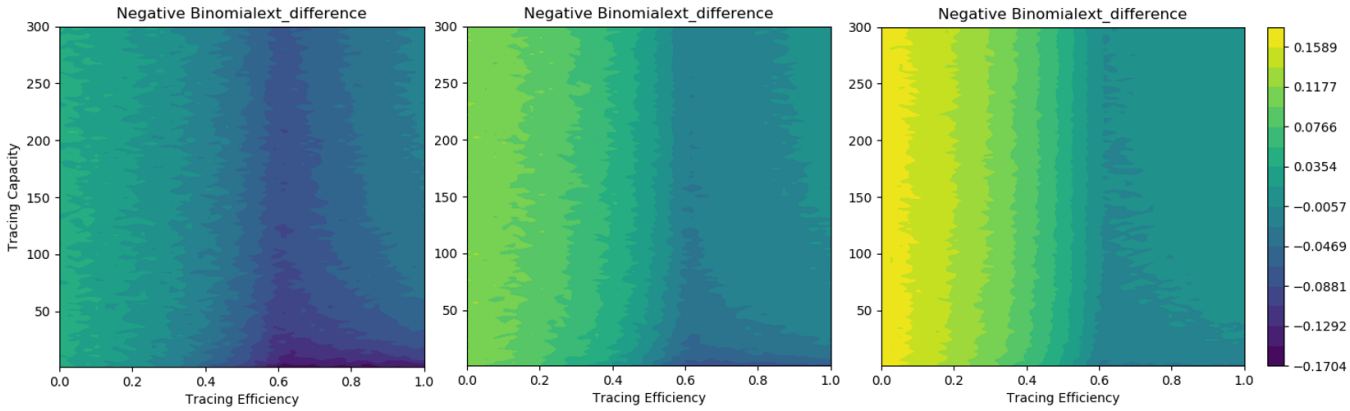

Figure 7: Contour plots showing the difference between PGF and simulated extinction probability in the manner described on page 7 for the same simulation seen in Figure 6. Simulation results attained with a tracing delay of zero using the negative binomial distribution for tracing efficiencies from left to right of 10% , 30% and 50%. Notably, a minimum in the difference appears at 0.6 tracing efficiency.

high tracing efficiency region are notably poorer (see in Figure 8). Conversely, for low tracing efficiencies below the minimum difference, the additional delay does not appear to produce substantially poorer results. Once again this can be explained by tracing having less relative importance in this region. We see almost identical trends in Figure 9 where the Poisson distribution is used with all other parameters remaining unchanged. However, the magnitude of the difference is drastically larger as seen by the colour bar that covers a range of roughly $\pm$ 70% .

The minimum value which emerged in the difference plots was of interest to us. Figures 10 to 13 were produced to investigate the behaviour and origins of this minimum. the tracing efficiency at which it manifests appears to depend only on the reproduction factor of the offspring distribution used whilst its magnitude

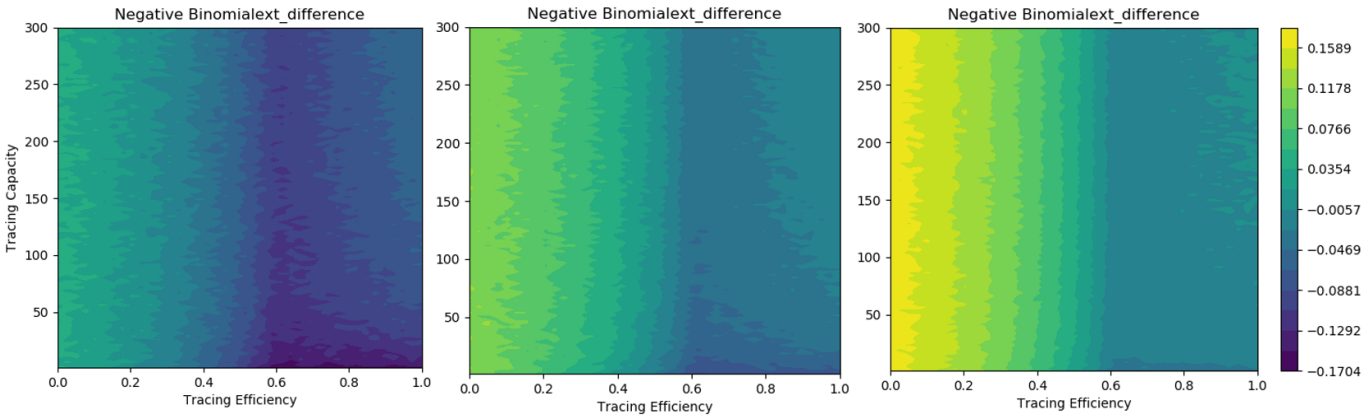

Figure 8: Contour plots showing the difference between PGF and simulated extinction probability. Input parameters are the same as those used in Figure 7 barring the tracing delay which here is 1. A tracing delay of 1 indicates that tracing begins at the generation following that in which the first successfully self-reported infection.

depends on the reproduction factor, reporting rate and in the case of the negative binomial distribution, on the dispersion coefficient as well. From the trends observed in Figures 10 and 11, it becomes clear that the minimum corresponds to the tracing efficiency at which we have an effective reproduction factor of 1; where the product of tracing efficiency and reproduction number are 1. As the majority of our simulations were performed with a reproduction factor of 2.5 then we have an effective reproduction number of 1 if 60 % of cases are removed and outbreaks essentially become impossible. This is confirmed by noting that Figure 10 shows the minimum at a tracing efficiency of 0.5 and 0.66 for reproduction factors of 2 and 3, respectively.

The position of the minimum doesn't appear to have any clear dependence on reporting rate. The minimum appears insensitive up to reporting rates between 0.5-0.6 although the magnitude of that minimum approaches zero over this range. Beyond 0.6, position becomes extremely erratic. The reporting rate of 0.6 is significant for the reason that it corresponds to an effective reproduction factor of 1, similarly to why the minimum appeared at 0.6 tracing efficiency in earlier plots. Beyond this point there cannot be outbreaks for any tracing capacity or efficiency given that our reporting rate does not decay like our tracing efficiency does. Thus, we believe the erratic behaviour of minimum position above 0.6 to be a manifestation of the noise inherent to the simulation involving the negative binomial distribution. The trend with respect to the dispersion parameter is effectively the opposite that was observed for self-reporting rate. Here, the position of our minimum becomes volatile at extremely low dispersion parameter values in the range below 0.2 whilst being almost entirely insensitive to the parameter above this value.

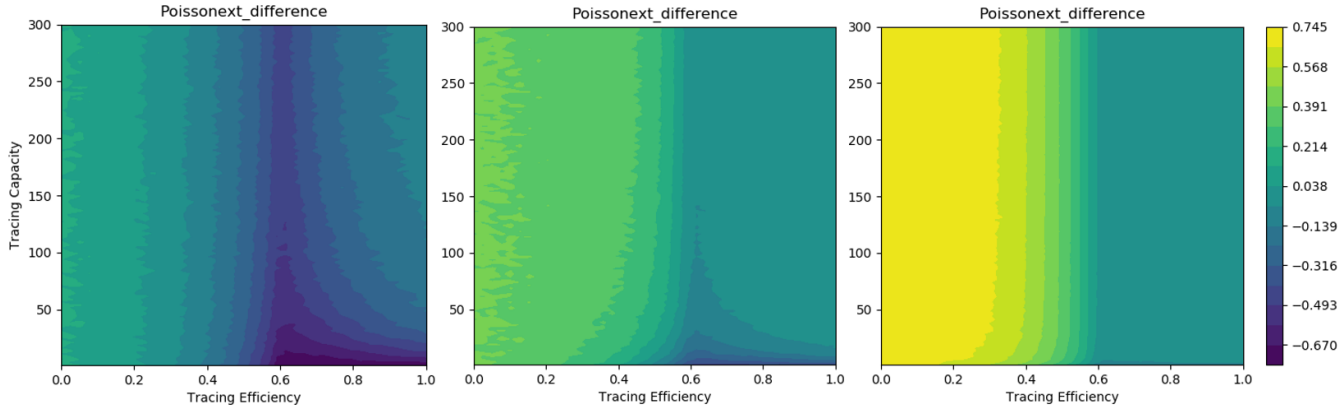

Figure 9: Contour plots showing the difference between PGF and simulated extinction probability using the the Poisson distribution and a tracing delay of 1. All other parameters are identical to those employed in Figure 7. Note the larger range of the contours as indicated by the colourbar relative to those used in Figures 7 and 8.

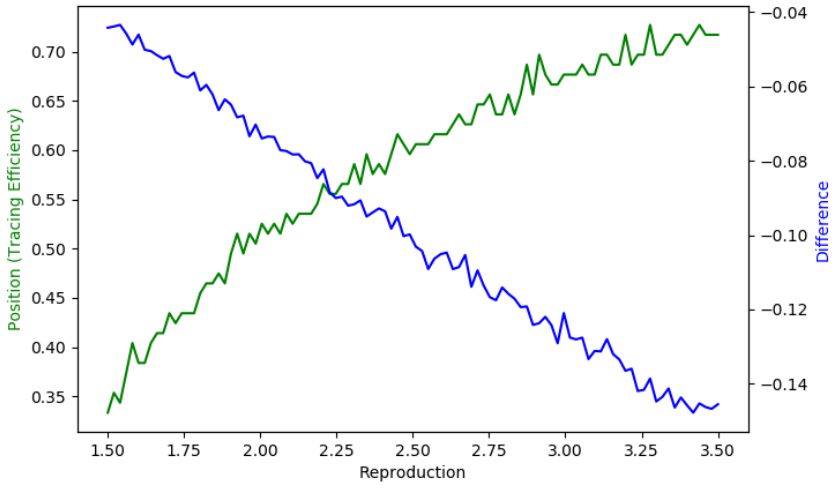

Figure 10: Double line plot showing the behaviour of the minimum observed in Figures 7 to 9 for the negative binomial distribution as a function of reproduction factor. Position of the minimum plotted in green indicates the tracing efficiency at which the minimum occurred. Difference plotted in blue indicates the minimum value of the difference between the PGF and simulation extinction probability. Position and difference both demonstrate a strong dependence on the reproduction number.

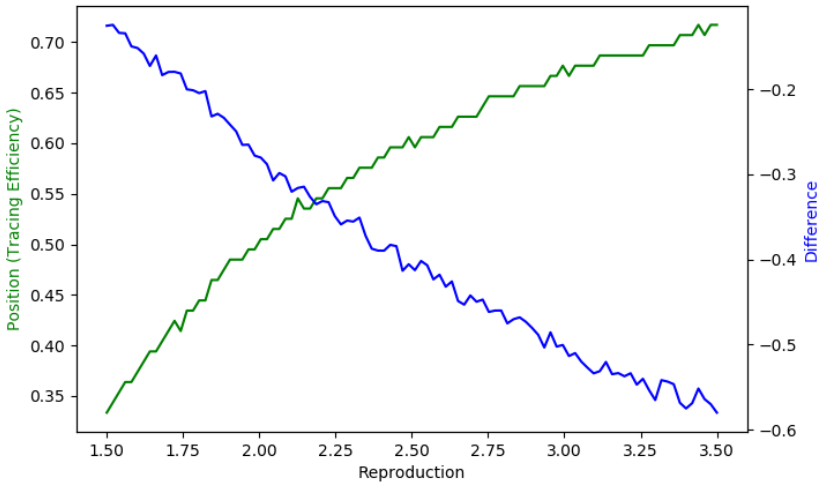

Figure 11: Double line plot showing the behaviour of the minimum observed in Figures 7 to 9 for the Poisson distribution as a function of reproduction factor. Plot is colour coded in the same manner as Figure 10 and demonstrates the same trend.

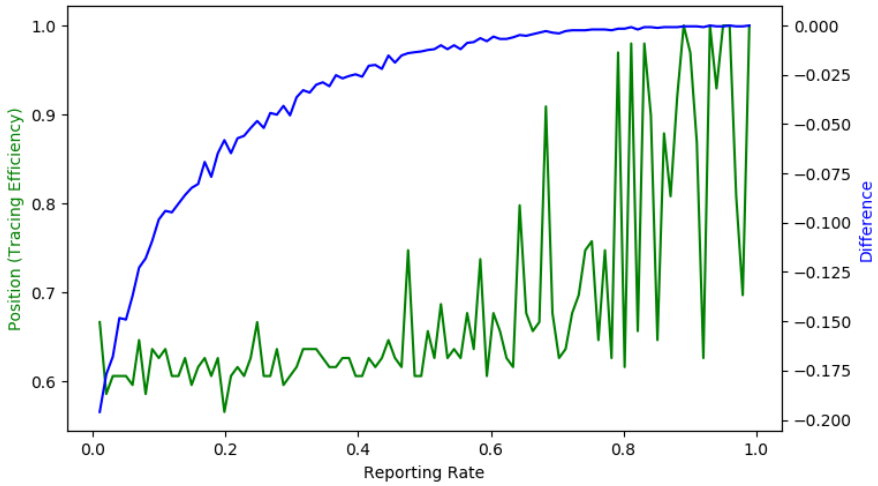

Figure 12: Double line plot showing the behaviour of the minimum observed in Figures 7 to 9 for the negative binomial distribution as a function of self-reporting rate. Position of the minimum appears insensitive to reporting rate up until approximately 0.5 to 0.6 at which point position becomes extremely erratic. At this point, the minimum is so shallow that it is likely a minimum of noise in the simulation.

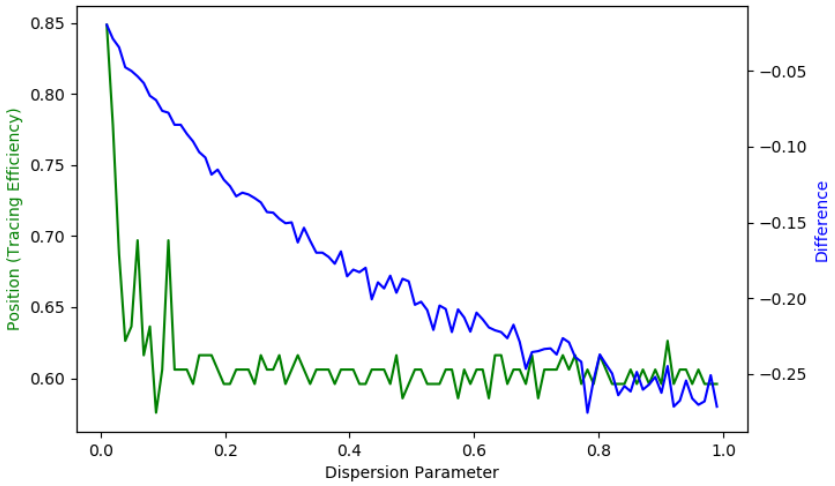

Figure 13: Double line plot showing the behaviour of the minimum observed in Figures 7 to 9 for the negative binomial distribution as a function of dispersion parameter. The minimum appears almost entirely insensitive to dispersion parameter with the exception of some values far into the extreme super-spreading regime less than 0.2.

6 Discussion and Conclusion

A stochastic SIR model was produced and used to investigate questions on control strategies for super-spreading viruses. Namely, we hoped to gain some insight on whether resources would be best allocated in increasing the efficiency of tracing systems, the capacity of tracing systems or on background testing rates. Additionally we studied whether optimal resource allocation may differ for super-spreading versus non-super-spreading viruses. The probability of extinction as attained from our simulation and PGFs was used as a figure of merit to evaluate the relative impact of each of these modes of resource allocation. We conclude that in the context of preventing a pandemic from emerging, tracing capacity need not be much greater than 100 and is the least impactful of the three interventions. Tracing efficiency and background testing rates are both highly effective in preventing outbreaks, particularly when their rates bring the effective reproduction down to 1. Reporting rate seems especially effective where tracing efficiency is low or tracing may not have begun. However we acknowledge that it may be difficult to achieve high reporting rates or tracing efficiencies depending on the cooperation of the population. We also conclude that this is true for both super-spreading and non-super-spreading viruses alike, with the same trends emerging in both cases albeit with a greater likelihood of extinction for the super-spreading case.

There are a number of avenues along which the simulation could be expanded to produce more realistic results and investigate a greater number of trends. The stochastic SIR model created did not implement a real time structure, instead only operating on generations. This lack of time structure means that heterogeneity in the period of time that individuals are infected, incubation times or period before recovered individuals are moved into the susceptible pool could not be implemented. Additional aspects that would have been incorporated into the simulation given more time would be a dynamic self-reporting rate which varies in response to the number of infections and a true contact tracing system in which a network of infections is recorded and tracing would work through this network in reverse rather than simply tracing people on chance alone.

References

Baldwin, Richard and B Weder di Mauro (2020). "Economics in the time of COVID-19: A new eBook". In: *VOX CEPR Policy Portal*.

Chintalapudi, Nalini et al. (2020). "COVID-19 outbreak reproduction number estimations and forecasting in Marche, Italy". In: *International Journal of Infectious Diseases*.

Dong, Ensheng, Hongru Du, and Lauren Gardner (2020). "An interactive web-based dashboard to track COVID-19 in real time". In: *The Lancet infectious diseases* 20.5, pp. 533–534.

Kucharski, Adam J et al. (2020). "Early dynamics of transmission and control of COVID-19: a mathematical modelling study". In: *The lancet infectious diseases*.

Miller, Joel C (2018). "A primer on the use of probability generating functions in infectious disease modeling". In: *Infectious Disease Modelling* 3, pp. 192–248.

Real, Leslie A (1977). "The kinetics of functional response". In: *The American Naturalist* 111.978, pp. 289–300.

Shereen, Muhammad Adnan et al. (2020). "COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses". In: *Journal of Advanced Research*.

Weiss, Howard Howie (2013). "The SIR model and the foundations of public health". In: *Materials matematics*, pp. 0001–17.

Xiong, Jiaqi et al. (2020). "Impact of COVID-19 pandemic on mental health in the general population: A systematic review". In: *Journal of affective disorders*.