Project introduction and outline.

The goal of this project is many fold: for one we want to see if we can use branching processes specifically the Galton-Watson process to model the spread of contagious diseases namely COVID-19, Ebola, and AIDS. We know [citation] that the early stages of a pandemic or epidemic can be well modeled by the Galton-Watson process, but then after a certain period of time the spread of disease is better modeled by an ODE called the SIR-model because at a certain point the random effect is lost [citation needed].

We are also interested in parameter estimation. Specifically in the Galton-Watson process for modeling the spread of disease we know that the offspring distribution is Poisson with parameter Lambda the goal is to see if we can estimate Lambda from data that we have from various countries. We want to see if we can determine the efficacy of lockdowns during a pandemic. To do this we want to determine when lockdowns were established in various countries and then do parameter estimation on either side of that date and time to see if the parameter changes, we can use Bayesian inference [citation of Hackers] To see if there is actually a change in parameter after lockdowns.

Finally, the very last thing that we're going to be interested in in this, is if we can estimate fatality rates of diseases using Bayesian neural networks. The goal of this is to see if certain facets of the population are more susceptible to death from disease namely COVID-19. This could lay the foundation for another analysis where we determine the number of highly susceptible people in the United States and other countries and what outbreaks of certain diseases could do to these populations. In the United Stata for example we have an endemic problem with heart conditions and heart disease brought on by obesity if we can determine the number of people susceptible to a lethal case of a disease such as COVID-19 then we can make a model that would estimate the number of expected deaths.

Using these three seemingly disparate models in conjunction, we want to see if we can come up with an aggregate model that can be used to estimate mortality in certain countries along with the efficacy of lockdowns mass vaccination and other protocols.

What is R0?

If you are at all familiar with epidemiology or the study of disease in general, then you've probably heard of the value R0. Simply stated R0 is the average number of people an infected person will infect during their infectious period. The value varies by disease, and it depends on many factors, but in general the higher the R0 value the more likely the disease is to spread through a population.

In the context of this paper, we can think of R0 as the expected number of children in the Galton-Watson process we will use to describe the spread of an epidemic. It is widely accepted that the number of infectious interactions a carrier will have is Poisson distributed

Parameter Estimation Using Regression.

Perhaps a better way to estimate our parameter of interest, R0, it's just simply take advantage of the fact that the expected value of Y is in fact Lambda. We also know from Theorem 7.2 in Gut’s text that the expected value of each generation in a galton Watson branching process is simply the expected value of Y to the NTH power, where n is the generation.

COVID-19 By County

Gives us the total number of confirmed cases of COVID-19 and deaths by COVID-19 as a function of date. The data starts on January 22nd 2020 and ends on July 27th 2020. During this timespan the CDC tracked these numbers by county for every county in the United States.

For the purposes of this analysis we'll choose several individual counties from throughout the country to work with individually. Those counties will be Berks County, Pennsylvania (my home county), Manhattan county New York, Los Angeles County California, and Philadelphia county, Pennsylvania. The charts below show the cumulative cases and deaths for each of these counties:

COVID-19 worldwide.

The World Health Organization (WHO) track the number of cumulative cases and cumulative deaths by day similarly to how the CDC did for each county in the United States. From this data, using a similar method as before, we will again determine R0-- but this time, for the entire world. Again, the data is visualized below:

Notes on the data:

What's interesting about each graph, is that even though three of them have very similar shapes the scales are drastically different. Not to mention one county, that being Los Angeles, actually has a very different shape than the other three counties in terms of XN versus N. This was a little surprising to see, since we would have expected similar R0 values for each county considering we're dealing with the same disease in each. But there are some rational reasons why this could have occurred. For one, access to testing resources could have artificially deflated the number of detected cases in certain areas. Of the four counties that we've chosen, Berks County is by far the smallest and least densely populated. R0 Can be thought of as a random variable in its own right. Really, it represents the number of interactions that people have each day, which itself can be modeled probabilistically, multiplied by the probability of infection of each interaction. So in densely populated counties like Los Angeles or Manhattan, you would expect to see more interactions per day therefore driving up the effective R0, or Re. This could explain why those two counties have many more cases in pure numbers than places like Berks county

Early Stages: The Galton- Watson Process

The spread of disease, especially the early stages of epidemics, is easily modeled using branching processes—namely, the Galton-Watson process. A branching process is a stochastic process, and the Galton-Watson has two distinct features: a) all individuals give birth according to the same probability law independently of each other, and b) the number of offspring produced by an individual is independent of the number of individuals in that generation (GIP).

SIR Model

Where R0 is the number of contacts per day that are sufficient to spread the disease, and k is the fraction of the infected group that will recover during a given generation. For our purposes, K will equal one. Again, intuitively, these equations make sense. Of course the number of individuals leaving susceptible status is proportional to -R0, the number of susceptible people and the number of infected people, at time T. And then, of course, the number of infected increases by the same, and decreases by the number of infected at time T, when K = 1. And of course the number of recovered is just the number of people leaving infected status at time T.

R0 vs. Re

R0 is the basic reproductive number and is the reproductive ratio that most people are familiar with, but there is also the effective reproductive number, RE, that changes as the immunity of the population changes. Where you might consider r0 to be the “true” reproductive ratio, RE can be thought of as describing how the disease acts in reality. For example, R0 is affected purely by the infectiousness of the Organism and the rate of recovery and death during an outbreak. While, RE is affected by herd immunity, either through natural immunity of significant proportions of the population contracting the disease or through immunization efforts. Obviously, such efforts will affect the rate at which the disease spreads, but we don't describe this using R0, we described this using RE.

Within the confines of this paper, we will be attempting to estimate R0 values of diseases in various counties and locations. We will also be seeing how certain efforts, namely lockdown efforts, affect the spread of disease. In these cases, we will be estimating RE and the effect that mitigation efforts have on re.

Discussion: Ebola Spread

As previously mentioned in our discussion of R0 versus RE, the way the public attempts to mitigate the spread of disease has major implications on the resultant RE value. An interesting note, in the case of ebola, is that unlike swine flu or COVID-19, ebola is spread only through contact with infected fluids. This means one can't contract the disease through coughing or other aerosolized methods of transmission. This means that all of the cases of ebola in western Africa from 2014 to 2016 came directly from physical contact with an infected person’s viscera.