Risky Communal Bahavior in an Epidemic

Brendan Wallace

February 20, 2022

1 Motivation

I plan to model how individual variation in proclivity to engage in "risky" behavior (i.e. going to a bar, working in an indoor area together) affects the dynamics of an epidemic, and in particular the impact of the concentration of risky behavior among a subset of connected individuals.

Things to cover in an intro:
Early theory of super spreading events (SSEs) focuses on variation in the number of secondary infections due to a single individual, but remains largely agnostic as to whether that variation is due to biological differences or behavior differences.[1]

There are some similar papers, in particular [2] showed that SSEs are an important factor in COVID outbreaks, showed some theoretical difference between different distributions of number of secondary infections, and layed out a qualitative framework for SSEs: behavioral, biological, high-risk facilities, and ""opportunistic" scenarios". Another paper: [3] went further to develop these ideas within the context of COVID-19 and modeled transmission in fairly complicated network structures.

On the other extreme, compartment models have been extremely developed by [4]. But these papers are deterministic, don't engage with theory of SSEs and variation.

I want to take a middle-ground approach. Try to answer questions of variation, diffusion (an SSE term...), etc, but keep things simple and idealized.

(Note: this is not a fully developed intro but as per advice from that economics-modeler, I want to not read too much else about the approach taken by all these folks and just try to make my own model for a while!).

2 General Model Description

The model will be agent based, consisting of a fixed size, finite population of discrete agents. Disease spreads through the population as per a standard compartmental (SIR) epidemic model (agents can be "susceptible", "infected" or "recovered"). Agents are further characterized by their riskyness p_r (a value in [0, 1].

At the outset one agent is "infected" and all other agents are "susceptible" (to represent evolution or introduction of a new disease into a subpopulation).

Each time unit (i.e. day), disease spreads in two ways. First: every agent makes contact with every other agent and "infected" agents pass the disease to "susceptible" agents with very low probability α_c: this is community spread.

Simultaneously, each agent engages in "risky behavior" with probability p_r: and all agent so doing make contact with each other and "infected" agents pass the disease to "susceptible" agents with larger probability α_r.

Finally, each "infected" agent that has been infected for β time steps recovers and is moved to the inert category "recovered".

3 Pseudo code

```
function simulate:
        initialize array of agents
        set one agent to infected

        while number of infected > 0:
                # community spread
                for infected_agent in agents:
                        for susceptible_agent in agents:
                                spread to susceptible_agent with p=alpha_c

                # risky behavior spread
                populate risk_takers from agents according to p_r
                for infected_risk_taker in risk_taker:
                        for susceptible_risk_taker in risk_taker:
                                spread to susceptible_risk_taker with p=alpha_r

                # agents recover
                all agents in infected_agents increment timer
                for agent in infected_agents:
                        if agent timer > beta:
                                agent is recovered
```

4 Description of planned experiment

I'm interested in the following questions:

1. How is the likilihood of an epidemic (rather than a disease failing to take hold) affected simultaneously by the p_r (riskiness) of the initial diseased agent and the distribution of p_r across the population?

   My hypothesis is that distributions with overall higher values of p_r will have much lower chances of an epidemic taking off. This is the basic theory/result of SSE. I think I will be able to find a lot of these values analytically and then use the model to verify.

2. How does the timing and overall size of the epidemic vary as a function of the distribution of p_r across the population?

   My hypothesis is that compensatorily with my hypothesis for 1., distributions with overall higher values of p_r will have faster-occurring and higher-peaking epidemics when they do occur.

3. What is the expected number of secondary infections (S_2) caused by primary infection S_1 of an agent with riskiness p_r? And since we can compute this simultaneously for all values of p_r, how does this curve (i.e. the function from p_r to $\mathrm{E}[S_2]$) vary in response to the distribution of p_r in the population and the timing of when in the course of the epidemic the primary infection occurs?

   I think this is the most novel question of my project!

I hypothesize that, because initial outbreaks associated with "risky behavior" will be necessary for an initial outbreak in the subpopulation, that the disease will burn its way through risk takers early; and resultantly "risky behavior" becomes less risky later on in a pandemic.

This would potentially have management implications for the control of disease which has high diffusion (variation in R_t between individuals) due to this kind of behavioral SSEs (i.e. bars with frequent repeat customers, high risk work places, etc.): interventions are most crucial ***before*** *an initial outbreak in a community and that by the time an outbreak has begun, targetting control at high risk behavior of this type is too late.*

To answer these questions I will run lots of simulations with different parameters, but keep effective R_0 fixed (the expected number of S_2 when only one agent is infected) by balancing α_c, α_r and the distibution of p_r.

For 1., all that's necessary to measure is whether the epidemic went extinct before infecting some large fraction of the population. For the most part, the threshold of being "some large fraction" should be pretty easy to compute. I can also approximate the early phase of the system as a branching process and find the likelihood of extinction.

For 2., I will plot the standard graphs in an SIR model: the numbers of "infected", "susceptible" and "recovered" agents over time, and compare these curves (and possible the average of them?) for different distributions of p_r while holding effective R_0 fixed.

For 3.; for one distribution of p_r I need to estimate $E[S_2]$ for different realized values of p_r and different timing – for the definiton of "timing" I can use the number of timesteps that have occurred or the number of individuals that have already been infected.
I can still estimate this via simulation: any time an individual is infected I count how many secondary infections they cause, and save this 3-d point (p_r, time, S_2); then I can approximate a surface in this space using some kind of smoothing method.

5 Statement of work

I worked alone on this proposal.

References

[1] Superspreading and the effect of individual variation on disease emergence - 2005

[2] Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and contro - 2020 Althouse et al

[3] Heterogeneity matters: Contact structure and individual variation shape epidemic dynamics - 2021 Grossmann et al

[4] [1980s papers from laurant]