

Hot-spot model of risk structured disease spread

Outline

- 0. Abstract
- 1. Introduction
- 2. Model
 - 2.1 agent based model, introduced textually
 - 2.2 branching process
 - 2.3 integro differential
 - 2.4 difference
- 3. Results
 - 3.1 Extinction Probability
 - 3.2 Max I, Total R
 - 3.3 Disease Evolution
- 4. Discussion
- 5. Appendix
 - A.1 branching process derivations
 - A.2 differential equations - moments
 - A.3 basic reproduction number
 - A.4 differential equations - mean riskyness results
 - A.5 choice of γ and N
 - A.6 model comparisons

0. Abstract

Disease transmission is often not homogeneous: a small number of infected individuals are responsible for a disproportionately large number of subsequent infections.

This is certainly the case in COVID-19, in which 20% of infections may cause as many as 80% of secondary infections[citation?]. However, our mechanistic understanding of this phenomenon – “super-spreading” – is limited[citations?].

Here we introduce and analyze a model of an outbreak with super-spreading via “problem-place” locations: disease spreads homogeneously throughout the population at a low rate, but spreads rampantly at certain locations (bars, restaurants, churches, etc.) that individuals visit daily with their own fixed probabilities.

Compared to homogeneous dynamics, outbreaks are less likely to happen but are accelerated when they do occur; causing larger outbreaks in some cases, but paradoxically smaller

and less severe outbreaks in moderately to highly infectious diseases.

1. Introduction

After the SARS pandemic in 2005, researchers noted that disease transmission was not homogeneous: some small number of infected individuals – “super spreaders” – were responsible for a disproportionately large number of total infections.[2005 nature citation] Subsequent research noted a similar pattern in other diseases: between 15% to 20% of infected individuals caused 75% to 85% of infections in diseases as diverse as (*add examples from 2006 paper*). [2006 super spreading]. The COVID-19 pandemic follows a similar pattern: 15% of super-spreading- infected individuals cause 80% of later infections.

However, research into the dynamics of disease outbreaks with superspreading has been mostly limited to models in which some individuals are simply more infectious – either because of greater biological infectiousness (higher viral load, larger lung capacity, etc.) or greater network connectivity. But this doesn’t match what we understand about the COVID-19 pandemic. Firstly, individual infectiousness doesn’t vary all that much [citation of this]. And while much can be learned from a model that simply assumes some individuals have many more social connections than others, this misses the crucial and salient fact that COVID transmission is centered around certain “hot spots” (bars, restaurants, churches, etc) [problem place paper]. We’d like to understand how the picture changes when disease spreads largely (or even primarily) through certain “hot spots” and the most important predictive variable of an individual’s superspreading capability is their likelihood to visit such a hot spot over any given time interval.

To address this missing piece, we introduce the “hot-spot” SIR model. This model is risk-structured: each individual has a fixed probability (riskiness) of visiting a “hot spot” per unit time, where disease spreads with very high transmission rate between all present individuals, and simultaneously is subject to “community spread” which acts homogeneously between all individuals with much low transmission rate. We find that with these dynamics, initial outbreaks in a fixed population where one individual becomes infected happen much less readily (for comparable overall transmission rates). Similar to previous work, [heterogeneous spread citation, SSE papers] we find that when there is an outbreak it occurs much faster or “explosively”. However, we also see the outbreak die out much more rapidly – due to the average riskiness in the infected and susceptible populations both declining after a point. Surprisingly, this leads to explosive outbreaks that nevertheless infect fewer people both overall and at their peak over a fairly wide range of average disease infectiousness (defined below).

2. Methods

We first consider an agent-based model.

2.1 Agent-based (simulation) model

Setup

We initialize a fixed-size population of N agents. Each agent i ($i \in [0, N]$) is characterized by a fixed “riskiness” parameter $\rho_i \in [0, 1]$ that never changes, and a disease state of Susceptible (S), Infected (I) or Recovered (R). We set the following parameters:

- $\beta_c \in [0, 1], << 1$ community spread rate – the probability of disease spreading from an I to an S individual through a single community contact,
- $\beta_p \in [0, 1], << 1$ hot-spot spread rate - the probability of disease spreading from an I to an S individual through contact in the problem-place; this is typically larger than β_c .
- $\gamma \in (0, 1]$ the recovery rate - the rate at which I individuals move to R (the disease lasts on average $1/\gamma$ time units).

All agents i are initially S, and each’s ρ_i is drawn iid. from a specific fixed distribution P over $[0, 1]$. One agent is chosen at random and set to I to start the simulation.

Dynamics

Risk Taking Every timestep, each agent i visits the problem place with probability ρ_i .

Disease Spread All agents in the problem place make a contact with all other agents in the problem place; I individuals in this subset spread the disease to S individuals with probability β_p (problem place spread). Simultaneously, all agents (in and out of the problem place) make contact with all other agents and (I) individuals spread to (S) individuals with probability β_c (homogeneous community spread).

Recovery

(I) agents recover with probability γ . If there are no (I) agents, the simulation ends.

2.2 Investigation Procedure

To investigate how disease dynamics and outcomes differ, we compare simulations with different levels of “hot-spot” dynamics and different shapes of risk distribution P while controlling for initial basic reproduction number R_0 .

Basic Reproduction Number

The first infected agent has expected infectiousness $E[P] := \bar{\rho}$ and will recover in $\frac{1}{\gamma}$ timesteps (mean of the exponential distribution). Each time step, this agent goes to the problem place with probability $\bar{\rho}$, encounters $\bar{\rho}N$ susceptible agents, and infects each of them with probability β_p . At the same time, the agent infects all other agents with probability β_c each. This leads to¹:

¹This development is intuitive but somewhat imprecise. We develop it more rigorously in the Appendix.

$$R_0 = \frac{1}{\gamma}(\bar{\rho}^2 \beta_p N) + \frac{1}{\gamma}(\beta_c N) = \frac{(\bar{\rho}^2 \beta_p + \beta_c)N}{\gamma}$$

Hot-Spot weight, Risk Distributions

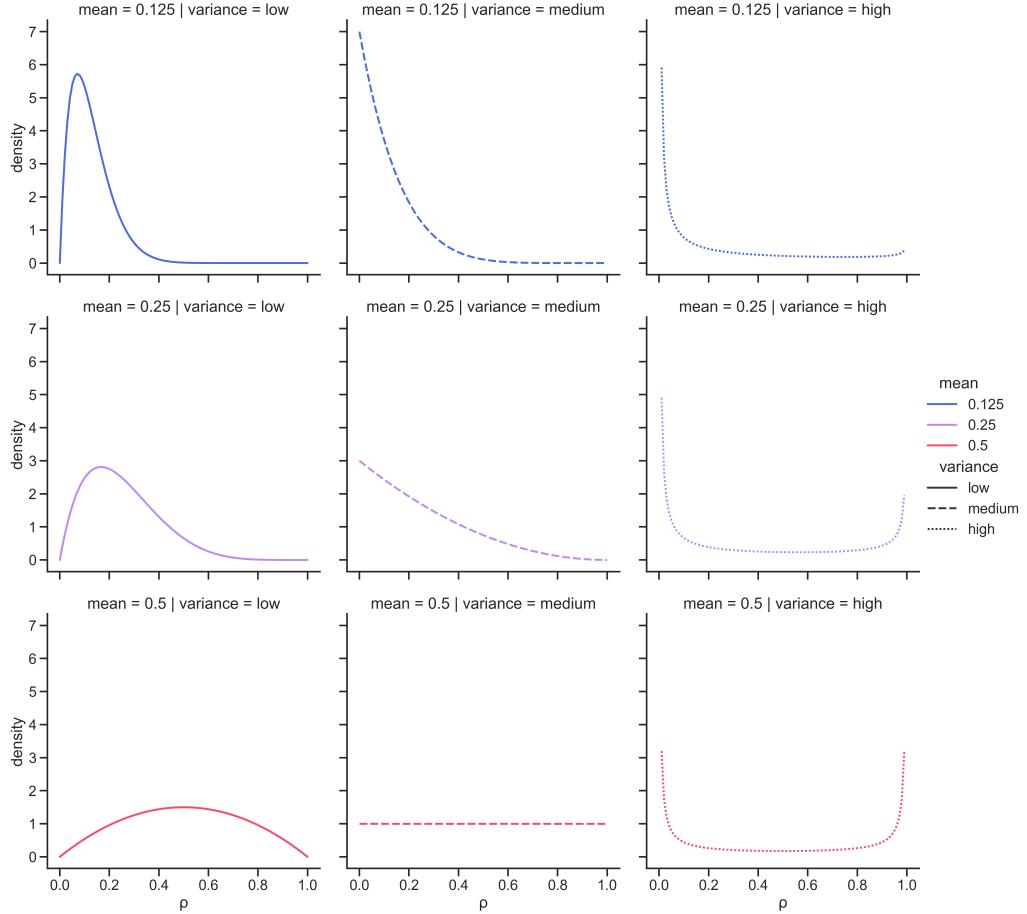


Figure 1: **Risk distributions of the initial population:** the top row shows distributions in which *on average* 1/8th of the population goes to the problem-place per week which increase in variance from left to right. The second row shows the same with mean 1/4, and the last row mean 1/2. We don't consider higher means – if more than half of the population spends time in the problem place per week, we start to approach the homogeneous case.

Using this definition, we can compare scenarios where hot-spot spread contributes 1/4, 1/2 or 3/4 of the expected disease spread by varying β_c and β_p , for the same R_0 .

Further, we can consider different shapes of the risk distribution P .

In particular, we use nine different Beta distributions to consider Ps with low, medium and high variability; and low (1/8), medium (1/4), and high (1/4) means. These are shown in detail in [Beta Distributions Figure].

Procedure

Our procedure is then the following:

We consider these three proportions of hot-spot spread; and within each, nine distributions of risk taking; then we run simulations for values of R_0 ranging from 0 to 8 and compare the outcomes of the outbreaks to those of a similar setup with no risk taking (the homogeneous case).

2.3 Other Models

To aid in investigating the dynamics of the model in certain cases, we consider and analytically analyze the following additional models.

Branching process model

When the number of infected agents is small; we use a branching process model.

[this is developed more in the text]

2.1.3 Integrodifferential model

Conversely, when the number of infected agents is large (either N is very large or the number of infected agents is just large relative to N), we consider the following integrodifferential equations:

$$\begin{aligned}\frac{\partial S(\rho, t)}{\partial t} &= -\beta_c S(\rho, t) \int_0^1 I(u, t) du - \beta_r S(\rho, t) \rho \int_0^1 I(u, t) u du \\ \frac{\partial I(\rho, t)}{\partial t} &= \beta_c S(\rho, t) \int_0^1 I(u, t) du + \beta_r S(\rho, t) \rho \int_0^1 I(u, t) u du - \gamma I(p, t)\end{aligned}$$

We analyze these equations and find several useful analytic results which help to understand the general or expected dynamics of the system; especially when N is large.

2.1.4 Difference model

Finally, to diagnose differences between the integrodifferential model and the agent-based model we consider a difference equation model which combines the deterministic behavior of the former with the discrete time steps of the latter.

3. Results

3.1 Extinction Probability

Homogeneous case Given a disease with \mathcal{R}_0 of 2, the standard SIR model predicts an outbreak to infect 79.681% of the population before running its course. When we simulate

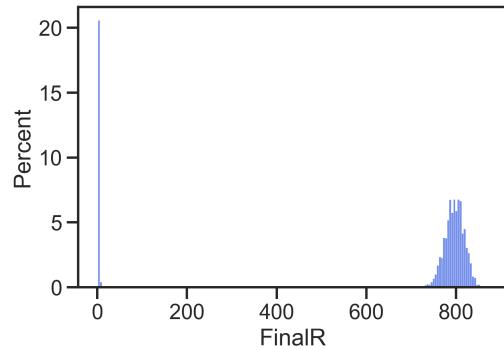


Figure 2: Histogram for homogeneous cases

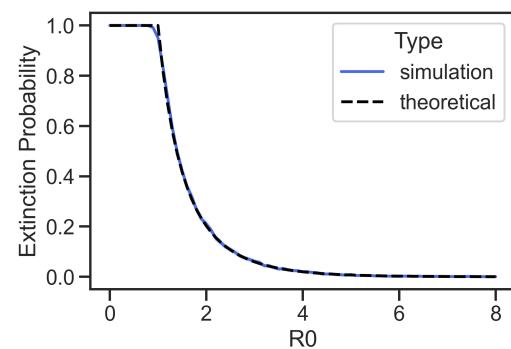


Figure 3: Figure that compares predicted extinction probability to actual, for homogeneous case

such an outbreak in a population of 1,000, we see outbreaks of about this size, but we also see some number of simulations in which there's no large outbreak at all.\

We can predict the probability of a large vs small outbreak with reasonable accuracy by replacing the outbreak scenario with a similarly parameterized *branching process*, and then asking how likely this branching process is to go *extinct*. This probability τ is a function of $p(n)$ – the probability of one individual infecting n individuals before recovering. We make the simplifying assumptions that: every infected individual can simultaneously go on to infect N more individuals; essentially:

1. every infected agent can go on to infect up to N more individuals (rather than $N - (I + R)$), and
2. secondary infections from multiple agents are independent (rather than considering how they are all infecting a shared population).

Then under these assumptions, the extinction probability τ obeys:

$$\tau = p(0) + p(1)\tau + p(2)\tau^2 + \dots$$

In the approximation, $p(n)$ is exactly the probability mass function of a Binomial distribution with parameters $p = 1 - (1 - \beta_c)^\gamma$ and N , so the RHS of the equation above is the generating function of the binomial distribution, which gives:

$$\tau = (\tau(1 - (1 - \beta_c)^\gamma) + (1 - \beta_c)^\gamma)^N$$

This allows us to quite accurately predict the probability of a large outbreak for any value of \mathcal{R}_0 and compare to the simulated results. We show this in [the figure above].

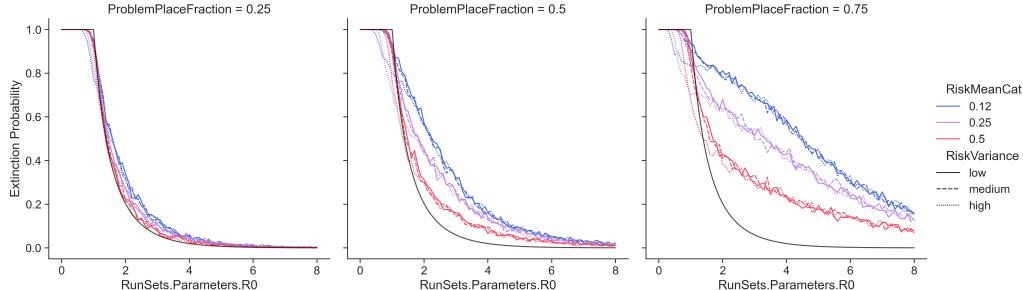


Figure 4: Simulated extinction probabilities Panels from left to right show 25%, 50%, and 75% of spread due to hot spot dynamics. Each curve is extinction probability as R_0 ranges from 0 to 8. Black lines show extinction probability of the homogeneous (hot spot weight = 0%) case for comparison.

Hotspot case Now we consider the extinction probabilities observed in the agent-based simulation, shown in [the figure above]. These deviate noticeably from the homogeneous case even when controlling for R_0 . Interestingly, the mean of the risk taking distribution

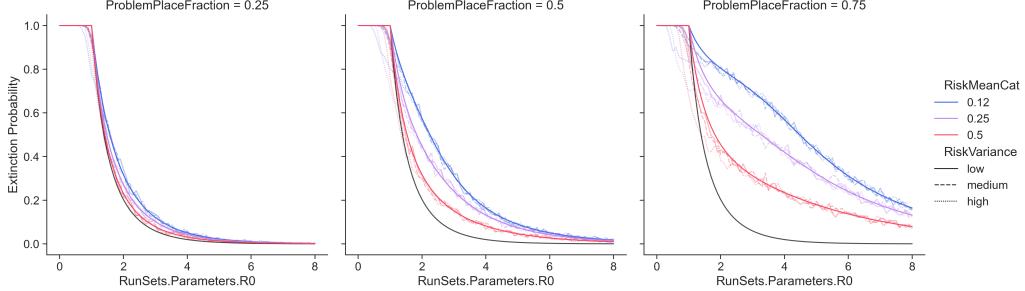


Figure 5: Simulated vs theoretical extinction probabilities

$\bar{\rho}$ (line color) appears to play a large factor; while the skew of the risk taking distribution (line dash style) plays apparently little or no role.

We find that a similar branching process approximation helps explain these findings. To the assumptions (1.) and (2.) above, we now add the following:

3. We now want $E[\tau]$ – the *expected* extinction probability taking into account that the risk taking ρ of all individuals are drawn i.i.d. from the same distribution.

Over the course of their infection, infected individual i with riskiness ρ_i infects susceptible individual j (riskiness ρ_j) through community spread with probability β_c , or through hot spot spread with probability $\rho_i \rho_j \beta_r$ (they must both go to the hot spot). Combined, this is:

$$\beta_c + (1 - \beta_c)(\rho_i \rho_j \beta_r)$$

In expectation, since the ρ are drawn iid, this is:

$$\beta_c + (1 - \beta_c)(\bar{\rho}^2 \beta_r) = \beta_{\text{effective}}$$

And as before, $p(n)$ is a Binomial probability mass; here with $\beta_{\text{effective}}$ and N .

Finally, we make one more simplifying assumption:

4. Individuals infected after the first have expected riskiness $\bar{\rho}$.

This neglects the fact that individuals infected after the first should tend to have a higher riskiness – but this is necessary to use a branching process model.

With these assumptions, following a similar approach (see appendix) to above we find that:

$$E[\tau] = \bar{\rho}[(1 - \bar{\rho}\beta_r)(1 - \beta_c) + (1 - (1 - \bar{\rho}\beta_r)(1 - \beta_c)\tau)^N + (1 - \bar{\rho})[(1 - \beta_c) + \beta_c\tau]^N]$$

These predictions are shown in [figure above] compared to the simulated results from [fig:extinction_results], showing great agreement. Both predict that nearly across the

board disease extinction is more probable than in the homogeneous case. In keeping with [superspreaders papers], spread is concentrated in some individuals more than others. When these individuals are among the first infected, the outbreak may be more explosive; but conversely there the initial infection may not find a superspreaders and even for very high \mathcal{R}_0 the disease peters out. Conversely, we see some possibility of an outbreak for $\mathcal{R}_0 < 1$ – a probabilistic impossibility in the homogeneous scenario. In this case, what we’re seeing is a small outbreak localized among the high risk-taking subpopulation – if we looked at this subpopulation alone there is $\mathcal{R}_0 > 1$.

Additionally, the fact of this model agreement alone has interesting implications. It suggests that the factors we’re leaving out – (a) the differences in the shapes of distributions beyond their mean; and (b) the accelerative effect as later generations of infected agents have generally higher riskiness – only become significant later on, at a point in which outbreak extinction has already become vanishingly unlikely.

3.2 Max I, Total R

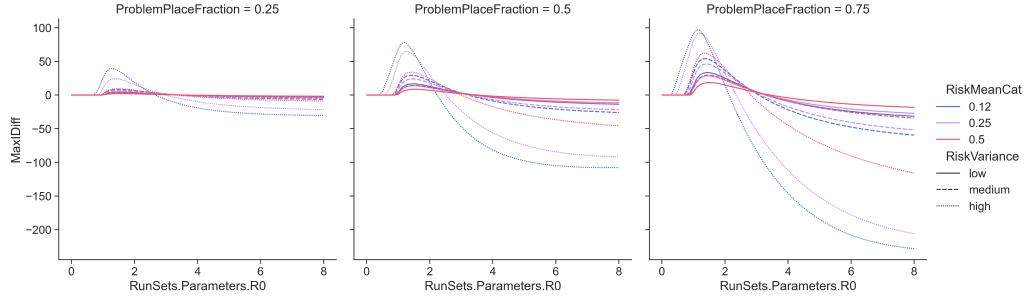


Figure 6: Max I

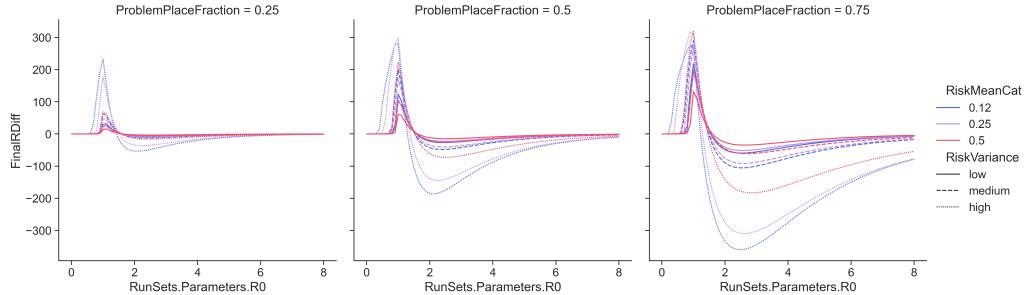


Figure 7: Final R

Next we look at only those scenarios in which an outbreak occurs and ask how problem place dynamics affect the size and severity of the outbreak. We consider Max I – the peak number of infected individuals – and Final R – the total number individuals infected (and recovered) over the entire course of the outbreak. Similar to [Extinction Probability Section] we fix a value of \mathcal{R}_0 , then for the risk distributions discussed in [Intro Section]

we vary the contribution of problem place spread and observe the effects. This is shown in [fig:outbreak_results].

- For very low \mathcal{R}_0 (below about 1.5), we see significantly larger outbreaks. (Here, as above, an outbreak is occurring mostly just among the subpopulation with very large riskyness.)
- For moderate \mathcal{R}_0 (1.5 to somewhere between 2.0 and 3.0), we see a mixed result: outbreaks have a higher number of infections at their peak, but overall affect fewer individuals.
- And finally, for high \mathcal{R}_0 (somewhere above 2.0 to 3.0) outbreaks are less severe both in the total and the peak number of infections.

3.3 Disease Evolution

Finally, to explain the findings of [3.2] we analyze how the disease plays out over time during an outbreak. Here the differential equation model is useful.

In the analysis section we show that the basic reproduction number $R(t)$ is given by:

TODO: Remake this figure. can convey much more information if this is nicely done.

In [dynamics figure] we plot the dynamics of several outbreaks over time. We see all three scenarios: a much larger wave of infections for low R_0 , one with a narrower but slightly higher peak for moderate R_0 , and a lower and smaller wave entirely for large R_0 . The peak of the outbreak occurs significantly earlier in all three cases.

Comparing $R(t)$ helps explain these differences. In the homogeneous case, Rt decreases monotonically. In the hot spot model, Rt increases initially, peaks sharply, then falls off; causing the entire outbreak to accelerate.

This is evident in [figure above], and can be shown universally. First, recall the equation for effective Rt :

$$R(t) = \frac{\bar{S}}{\gamma} (\beta_c + \beta_r \bar{\rho}_S \bar{\rho}_I)$$

Rt has similar form to the homogeneous $Rt = \frac{S}{\gamma}\beta$; except β is replaced by $(\beta_c + \beta_r \bar{\rho}_S \bar{\rho}_I)$ - which we could call “effective beta” and which is governed by $\bar{\rho}_I$ and $\bar{\rho}_S$.

Claim 1: $\bar{\rho}_S$ decays monotonically in proportion to β_r , the total infected population size, $\bar{\rho}_I$ and the variance of ρ_S .

Claim 2: $\bar{\rho}_I$ moves towards a fixed value between $\bar{\rho}_S$ and $\bar{\rho}_S + \frac{\text{Var}(\rho_S)}{\bar{\rho}_S}$.

These claims follow from the following results, which are derived in the appendix:

$$\frac{d}{dt} \bar{\rho}_S = -\beta_r \bar{I} \bar{\rho}_I \text{Var}(\rho_s)$$

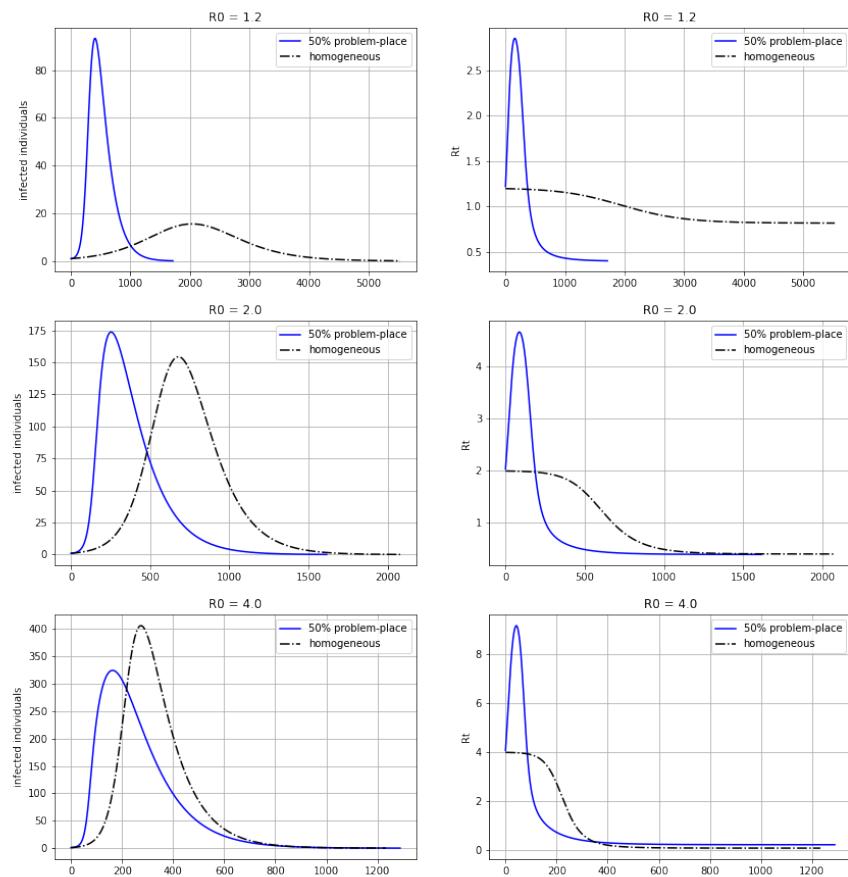


Figure 8: Dynamics Comparison

$$\frac{d}{dt}\bar{\rho}_I = \bar{S} \left[\beta_c(\bar{\rho}_S - \bar{\rho}_I) + \beta_r \bar{\rho}_S \bar{\rho}_I ((\bar{\rho}_S + \frac{\text{Var}(\rho_S)}{\bar{\rho}_S}) - \bar{\rho}_I) \right]$$

(The latter equation contains the sum of two restorative forces: one pushes $\bar{\rho}_I$ towards $\bar{\rho}_I$, and the other towards $\bar{\rho}_S + \frac{\text{Var}(\rho_S)}{\bar{\rho}_S}$. So there must be some value in between towards which $\bar{\rho}_I$ is forced, as claimed.)

4. Discussion

I wanted to get some feedback on this outline for the discussion section before trying too hard to write everything out.

(note to myself: hourglass back out)

Outline:

- review caveats and model assumptions (though I suppose that's covered in appendices A.5 and A.6)
- discuss these “gravity law of movement” papers I've been reading; and use them to inform which distributions make the most sense
- discuss laurant's papers about superspreading events
- discuss dobromir's & mark's paper about the more general case of risk structure
- further work:
 - could look into the implications of this on intervention timing (could cite my other paper if it gets published in time??) (I did look into this briefly but it wasn't super interesting at first pass)

A. Appendix

- A.1 branching process derivations
- A.2 differential equations - moments
- A.3 basic reproduction number
- A.4 differential equations - mean riskyness results
- A.5 choice of γ and N
- A.6 model comparisons

A.1 Disease Extinction

Derivation of the branching process approximation:

$$E[\tau] = \bar{\rho}[(1 - \bar{\rho}\beta_r)(1 - \beta_c) + (1 - (1 - \bar{\rho}\beta_r)(1 - \beta_c)\tau)^N + (1 - \bar{\rho})[(1 - \beta_c) + \beta_c\tau]^N]$$

Here we show that we can apply the same branching process approximation to the problem-place model to predict the probability of a large outbreak under any set of parameters.

Let X be a random variable that represents the number of infections caused by a single infected individual with riskyness ρ_i before recovering, and let $j = 1, \dots, N$ index the susceptible population so that ρ_j is the riskyness of individual j .

Then

$$P(X = 0) = (1 - \rho_i)[(1 - \beta_c)^N] + \rho_i[\prod_j(1 - (\beta_c + \beta_r\rho_j))]$$

By assumption riskyness for each individual is drawn independently from one distribution, so in expectation (over riskyness values) this is:

$$\begin{aligned} E[P(X = 0)] &= E[(1 - \rho_i)[(1 - \beta_c)^N] + \rho_i[\prod_j(1 - (\beta_c + \beta_r\rho_j))]] \\ &= (1 - E[\rho_i])[{(1 - \beta_c)^N}] + E[\rho_i][\prod_j(1 - (\beta_c + \beta_rE[\rho_j]))]] \\ &= (1 - \bar{\rho})[{(1 - \beta_c)^N}] + \bar{\rho}[{(1 - (\beta_c + \beta_r\bar{\rho}))^N}] \\ &= (1 - \bar{\rho})B(\beta_c, N, 0) + \bar{\rho}B(\beta_c + \bar{\rho}\beta_r, N, 0) \end{aligned}$$

Where $B(a, b, x)$ is the Binomial probability mass at x with parameters a and b .

Similarly, $E[P(X) = x]$ is given by

$$(1 - \bar{\rho})B(\beta_c, N, x) + \bar{\rho}B(\beta_c + \bar{\rho}\beta_r, N, x)$$

So

$$\begin{aligned} G_X(s) &= P(X = 0) + P(X = 1)s + P(X = 2)s^2 + \dots \\ &= [\bar{\rho}B_1(0) + (1 - \bar{\rho})B_2(0)] + [\bar{\rho}B_1(1) + (1 - \bar{\rho})B_2(1)]s + [\bar{\rho}B_1(2) + (1 - \bar{\rho})B_2(2)]s^2 + \dots \\ &= \bar{\rho}G_{B_1}(s) + (1 - \bar{\rho})G_{B_2}(s) \\ &= \bar{\rho}[(1 - \bar{\rho}\beta_r)(1 - \beta_c) + (1 - (1 - \bar{\rho}\beta_r)(1 - \beta_c)s]^N + (1 - \bar{\rho})[(1 - \beta_c) + \beta_c s]^N \\ \tau &= \bar{\rho}[(1 - \bar{\rho}\beta_r)(1 - \beta_c) + (1 - (1 - \bar{\rho}\beta_r)(1 - \beta_c)\tau]^N + (1 - \bar{\rho})[(1 - \beta_c) + \beta_c \tau]^N \end{aligned}$$

A.2 Moment equations

Initial equations:

$$\begin{aligned} \frac{\partial S(\rho, t)}{\partial t} &= -\beta_c S(\rho, t) \int_0^1 I(u, t) du - \beta_r S(\rho, t) \rho \int_0^1 I(u, t) u du \\ \frac{\partial I(\rho, t)}{\partial t} &= \beta_c S(\rho, t) \int_0^1 I(u, t) du + \beta_r S(\rho, t) \rho \int_0^1 I(u, t) u du - \gamma I(p, t) \end{aligned}$$

For convenience, we introduce the following shorthands for the “moments” of $I(\rho)$ and $S(\rho)$:

$$\begin{aligned}
\bar{S} &:= \int_0^1 S(\rho) d\rho & \bar{I} &:= \int_0^1 I(\rho) d\rho \\
\hat{S} &:= \int_0^1 S(\rho) \rho d\rho & \hat{I} &:= \int_0^1 I(\rho) \rho d\rho \\
\hat{\hat{S}} &:= \int_0^1 S(\rho) \rho^2 d\rho & \hat{\hat{I}} &:= \int_0^1 I(\rho) \rho^2 d\rho \\
&\vdots &&\vdots \\
{}^{(n)}S &:= \int_0^1 S(\rho) \rho^n d\rho & {}^{(n)}I &:= \int_0^1 I(\rho) \rho^n d\rho
\end{aligned}$$

Now we can more concisely write the initial equations:

$$\begin{aligned}
\frac{\partial S(\rho, t)}{\partial t} &= -\beta_c S(\rho, t) \bar{I} - \beta_r S(\rho, t) \rho \hat{I} \\
\frac{\partial I(\rho, t)}{\partial t} &= \beta_c S(\rho, t) \bar{I} + \beta_r S(\rho, t) \rho \hat{I} - \gamma I(p, t)
\end{aligned}$$

Finally, notice that $S(\rho)$ and $I(\rho)$ are defined over $\rho \in [0, 1]$, and that these give the population at a given risk level. We can equivalently think of how risk is distributed over the S and I populations:

We introduce ρ_S and ρ_I : these are random variables drawn from the distributions of risk in the S and I populations, respectively.

The means of these variables are:

$$\begin{aligned}
\bar{\rho}_S &= \frac{\int_0^1 S(\rho) \rho d\rho}{\int_0^1 S(\rho) d\rho} = \frac{\hat{S}}{\bar{S}} \\
\bar{\rho}_I &= \frac{\int_0^1 I(\rho) \rho d\rho}{\int_0^1 I(\rho) d\rho} = \frac{\hat{I}}{\bar{I}}
\end{aligned}$$

(This is just setting up interpretable variables for the terms $\frac{\hat{S}}{\bar{S}}$ and $\frac{\hat{I}}{\bar{I}}$, which start to pop up in a bunch of places.)

First we have a general result relating the “moments” of S and I.

Suppose we want to know how the total susceptible population is changing. We can integrate $\frac{\partial S}{\partial t}$ over ρ :

$$\begin{aligned}
\frac{d\bar{S}}{dt} &= \int_0^1 \frac{\partial S(\rho)}{\partial t} d\rho = \int_0^1 (-\beta_c S(\rho) \bar{I} - \beta_r S(\rho) \rho \hat{I}) d\rho \\
&= -\beta_c \left(\int_0^1 S(\rho) d\rho \right) \bar{I} - \beta_r \left(\int_0^1 S(\rho) \rho d\rho \right) \hat{I} \\
&= -\beta_c \bar{S} \bar{I} - \beta_r \hat{S} \hat{I}
\end{aligned}$$

Similary, we can find that:

$$\frac{d\bar{I}}{dt} = \int_0^1 \frac{\partial I(\rho)}{\partial t} d\rho = \beta_c \bar{S} \bar{I} + \beta_r \hat{S} \hat{I} - \gamma \bar{I}$$

If we want to see how the first moments are changing, we follow a similar method:

$$\begin{aligned} \frac{d\hat{S}}{dt} &= \int_0^1 \frac{\partial S(\rho)}{\partial t} \rho d\rho = \int_0^1 (-\beta_c S(\rho) \bar{I} - \beta_r S(\rho) \rho \hat{I}) \rho d\rho \\ &= -\beta_c \left(\int_0^1 S(\rho) \rho d\rho \right) \bar{I} - \beta_r \left(\int_0^1 S(\rho) \rho^2 d\rho \right) \hat{I} \\ &= -\beta_c \hat{S} \bar{I} - \beta_r \hat{\hat{S}} \hat{I} \end{aligned}$$

And for any general moment:

$$\begin{aligned} \frac{d^{(n)} S}{dt} &= -\beta_c {}^{(n)} S \bar{I} - \beta_r {}^{(n+1)} S \hat{I} \\ \frac{d^{(n)} I}{dt} &= \beta_c {}^{(n)} S \bar{I} + \beta_r {}^{(n+1)} S \hat{I} - \gamma {}^{(n)} I \end{aligned}$$

A.3 Basic Reproduction Number

In the homogeneous SIR model, the basic reproduction number (number of secondary infections per infection) is:

$$\mathcal{R}_t = \frac{S\beta}{\gamma}$$

In this model, we can derive a similar form for the expected number of secondary infections per infection by dividing $\frac{d\bar{S}}{dt}$ by \hat{I} (current number of infected individuals) then multiplying by the mean duration of an infection $\frac{1}{\gamma}$.

$$\begin{aligned} \mathcal{R}_t &= -\frac{d\bar{S}}{dt} \frac{1}{\gamma \bar{I}} \\ &= \frac{1}{\gamma} \beta_c \bar{S} \frac{\bar{I}}{\bar{I}} + \frac{1}{\gamma} \beta_r \hat{S} \frac{\hat{I}}{\bar{I}} \\ &= \frac{1}{\gamma} \bar{S} \left(\beta_c + \beta_r \frac{\hat{S}}{\bar{S}} \frac{\hat{I}}{\bar{I}} \right) \\ &= \frac{1}{\gamma} \bar{S} (\beta_c + \beta_r \bar{\rho}_S \bar{\rho}_I) \end{aligned}$$

This gives a nicely analogous result, where the homogenous β is replaced by what we can think of as an effective β : $(\beta_c + \beta_r \bar{\rho}_I \bar{\rho}_S)$, which is a straightforward function of both β terms and the mean riskiness in both populations.

We then would like to know how $\bar{\rho}_S$ and $\bar{\rho}_I$ are changing. We'd expect $\bar{\rho}_S$ (mean riskiness of the susceptible population) to monotonically decrease, as more risk-taking- susceptible individuals are more likely to be infected. We'd expect $\bar{\rho}_I$ to behave almost like a chemostat as higher-risk- individuals flow in, and all individuals flow out at a rate of γ . It should increase initially, and eventually decrease after the mean of riskiness decreases sufficiently in the susceptible population.

A.4 $\bar{\rho}_S$ and $\bar{\rho}_I$ (Mean Riskiness)

Fortunately, we can find explicit expressions for $\frac{d}{dt}\bar{\rho}_S$ and $\frac{d}{dt}\bar{\rho}_I$.

Recall that $\bar{\rho}_S = \hat{S}/\bar{S}$ and differentiate using the quotient rule, then substitute expressions from the moment differential equations from above for $\frac{d}{dt}(\hat{S})$ and $\frac{d}{dt}(\bar{S})$.

$$\begin{aligned}\frac{d}{dt}\bar{\rho}_S &= \frac{d}{dt}\left(\frac{\hat{S}}{\bar{S}}\right) \\ &= \frac{\frac{d}{dt}(\hat{S})\bar{S} - \hat{S}\frac{d}{dt}(\bar{S})}{\bar{S}^2} \\ &= \frac{1}{\bar{S}}\left(-\beta_c\hat{S}\bar{I} - \beta_r\hat{S}\hat{I}\right) - \frac{\hat{S}}{\bar{S}^2}\left(-\beta_c\bar{S}\bar{I} - \beta_r\hat{S}\hat{I}\right) \\ &= -\beta_c\frac{\hat{S}}{\bar{S}}\bar{I} - \beta_r\frac{\hat{S}}{\bar{S}}\hat{I} + \beta_c\frac{\hat{S}}{\bar{S}}\bar{I} + \beta_r\left(\frac{\hat{S}}{\bar{S}}\right)^2 \\ &= -\beta_r\frac{\hat{S}}{\bar{S}}\hat{I} + \beta_r\left(\frac{\hat{S}}{\bar{S}}\right)^2 \\ &= -\beta_r\hat{I}\left(\frac{\hat{S}}{\bar{S}} - \left(\frac{\hat{S}}{\bar{S}}\right)^2\right)\end{aligned}$$

Here, notice \hat{S} sums ρ^2 over $[0, 1]$ and so $\frac{\hat{S}}{\bar{S}}$ we can write as $E[\rho_S^2]$. And similarly $\frac{\hat{S}}{\bar{S}} = E[\rho_S]$ so

$$\frac{\hat{S}}{\bar{S}} - \left(\frac{\hat{S}}{\bar{S}}\right)^2 = E[\rho_S^2] - E[\rho_S]^2$$

which is the variance of the riskiness of the susceptible population $Var(\rho_S)$. This allows us to rewrite the equation as:

$$\frac{d}{dt}\bar{\rho}_S = -\beta_r\hat{I}Var(\rho_s)$$

or

$$\frac{d}{dt}\bar{\rho}_S = -\beta_r\bar{I}\bar{\rho}_I Var(\rho_s)$$

This confirms that ρ_S decreases monotonically (as long as there is some infected population with nonzero mean riskiness), and further shows that it decreases proportionally to the variance of the distribution of riskiness in the S population.

Mean Infected Riskiness

Similarly, start with $\bar{\rho}_I = \hat{I}/\bar{I}$ and differentiate with the quotient rule and then substitute in from the moment equations for $\frac{d}{dt}(\hat{I})$ and $\frac{d}{dt}(\bar{I})$.

$$\begin{aligned}\frac{d}{dt}\bar{\rho}_I &= \frac{d}{dt}\left(\frac{\hat{I}}{\bar{I}}\right) \\ &= \frac{\frac{d}{dt}(\hat{I})\bar{I} - \hat{I}\frac{d}{dt}(\bar{I})}{\bar{I}^2} \\ &= \left[(\beta_c\hat{S}\bar{I} + \beta_r\hat{S}\hat{I} - \gamma\hat{I})\bar{I} - \hat{I}(\beta_c\bar{S}\bar{I} + \beta_r\hat{S}\hat{I} - \gamma\bar{I})\right]/\bar{I}^2 \\ &= \left[(\beta_c(\hat{S}\bar{I}^2 - \bar{S}\hat{I}\bar{I}) + \beta_r(\hat{S}\hat{I}\bar{I} - \hat{S}\hat{I}^2)) + \gamma(\hat{I}\bar{I} - \hat{I}\bar{I})\right]/\bar{I}^2\end{aligned}$$

Here, remarkably all terms with γ cancel out. We continue by dividing through by \bar{I}^2 while factoring out \bar{S} . Then we rearrange factors in the β_r term.

$$\begin{aligned}&= \bar{S} \left[\beta_c((\hat{S}/\bar{S}) - (\hat{I}/\bar{I})) + \beta_r((\hat{S}/\bar{S})(\hat{I}/\bar{I}) - (\hat{S}/\bar{S})(\hat{I}/\bar{I})^2) \right] \\ &= \bar{S} \left[\beta_c((\hat{S}/\bar{S}) - (\hat{I}/\bar{I})) + \beta_r(\hat{S}/\bar{S})((\hat{S}/\hat{S})(\hat{I}/\bar{I}) - (\hat{I}/\bar{I})^2) \right]\end{aligned}$$

Now we substitute $\hat{I}/\bar{I} = \bar{\rho}_I$ and $\hat{S}/\bar{S} = \bar{\rho}_S$:

$$\begin{aligned}&= \bar{S} \left[\beta_c(\bar{\rho}_S - \bar{\rho}_I) + \beta_r\bar{\rho}_S((\hat{S}/\hat{S})\bar{\rho}_I - \bar{\rho}_I^2) \right] \\ &= \bar{S} \left[\beta_c(\bar{\rho}_S - \bar{\rho}_I) + \beta_r\bar{\rho}_I\bar{\rho}_S((\hat{S}/\hat{S}) - \bar{\rho}_I) \right]\end{aligned}$$

Finally consider:

$$(\hat{S}/\hat{S}) = \frac{\hat{S}/\bar{S}}{\hat{S}/\bar{S}} = \frac{E[\rho_S^2]}{\bar{\rho}_S} = \frac{\text{Var}(\rho_S) + E[\rho_S]^2}{\bar{\rho}_S} = \frac{\text{Var}(\rho_S)}{\bar{\rho}_S} + \frac{\bar{\rho}_S^2}{\bar{\rho}_S} = \frac{\text{Var}(\rho_S)}{\bar{\rho}_S} + \bar{\rho}_S$$

And we have:

$$= \bar{S} \left[\beta_c(\bar{\rho}_S - \bar{\rho}_I) + \beta_r\bar{\rho}_S\bar{\rho}_I((\bar{\rho}_S + \frac{\text{Var}(\rho_S)}{\bar{\rho}_S}) - \bar{\rho}_I) \right]$$

as desired.

A.5 Choice of γ and N parameters

In general, I use $\gamma = 1$ and $N = 1000$ for everything. In the differential equation model; these choices literally do not matter - they can be parameterized out.

In the simulation model and difference equation model, $\gamma = 1$ is nice for runtime: a single disease generation takes one time step. But it causes a little bit of a discrepancy between these two models and the continuous model - there's kind of a blocky-ness to some of the results (in particular the maximum number of infected agents). Setting $\gamma = 1$ also makes the difference between some of the risk distribution shapes less significant.

Is it worth touching on that in the methods section? Or just kind of gloss over it in the main text and discuss it briefly here? Or really dig into it here?

A.6 Differences between the different models

This is related to the previous (A.5) section. But if I'm only presenting the topline results from one or another model per Results section (i.e. only the differential equation model for final R) it could be nice to at least include the same figure from each different model run in an appendix section.

Most of the differences seem to disappear when γ is larger though. And N shouldn't really matter as long as it's not tiny. So I'm not sure how detailed it makes sense to get.

