Given a disease with $R0$ of 2, the standard SIR model predicts an outbreak to infect 79.681% of the population before running its course. When we simulate such an outbreak (homogeneous, with $R0$ of 2) in a population of 1,000, we see outbreaks of about this size (shown in Figure 1.), but we also see some number of simulations in which there's no large outbreak at all.

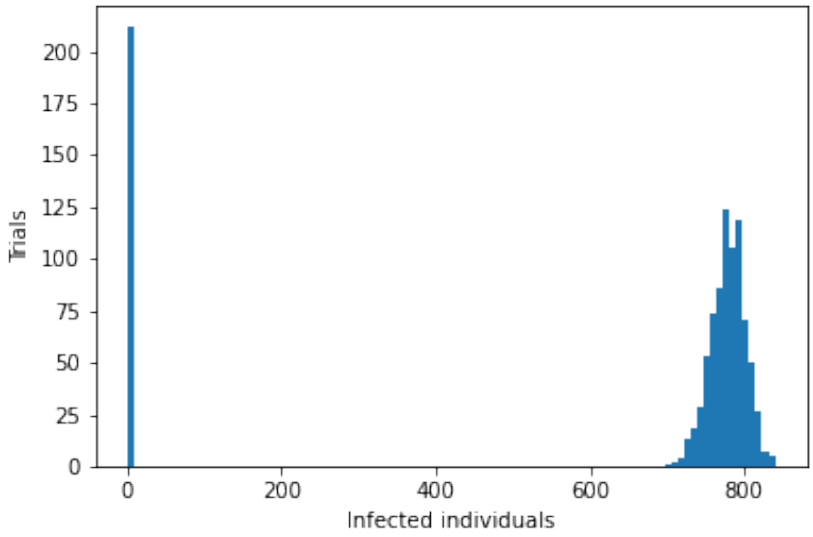

Figure 1: Figure 1. figure showing histogram of outbreak sizes

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[1]. These predictions are shown against the simulated results in Figure 2.

How does this change in the problem place model?

To investigate, we fix R0 and vary the contribution of "problem place" spread

[1]In the homogeneous simulation, an infected individual has probability $p(n)$ of infecting n individuals before recovering, where $p(n)$ is nearly the probability mass function of the binomial distribution with parameters β (infectiousness) and N (total population); except that N is not correct since not every individual is susceptible and things are further complicated by the possibility of multiple infected invididuals at once. In the branching process, $p(n)$ is exactly the probability mass function of the binomial distribution with parameters β and N. So to find the probability of extinction γ, we follow the normal formula:

$$\gamma = p(0) + p(1)\gamma + p(2)\gamma^2 + ... + p(N)\gamma^N$$
$$\gamma = G_{Binomial(\beta,N)}(\gamma)$$
$$\gamma = (1 - \beta + \beta\gamma)^N$$

This can be easily computed and is how we produce the approximation in Figure 2.

to its value from zero (all community spread) to one (all problem place spread) and observe the chance of disease extinction before an outbreak. This is shown in Figure 3 for each of the riskyness distributions discussed in (Figure in the Introduction).

Figure 3. Figure showing R0=1.5, 2.0, 3.0, 4.0

We find that higher problem place spread means a higher chance of disease extinction for the same level of R0, an effect that is more pronounced in riskyness distributions with lower means, but which suprisingly does not vary between different distributions with the same means.

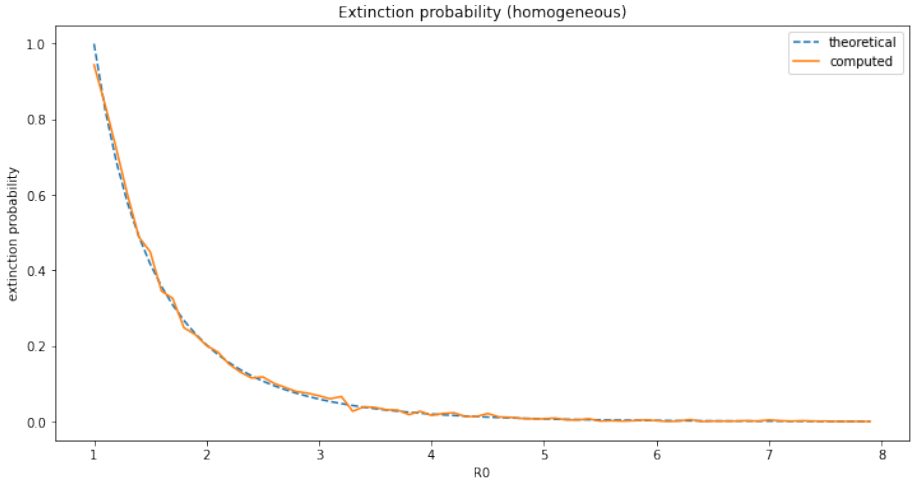

Figure 2: Figure 2. Branching process extinction probability vs likelihood of outbreak above 50 individuals. The star indicates the trial shown in figure 1.

To explain this effect, we again approximate with branching process.

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 $j = 1, \ldots, N$ 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:

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

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

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

For small $\beta_c + \bar{\rho}\beta_r$, the overlap is very small so:

$$
\begin{aligned}
P(X=x) &\approx (1-\bar{\rho})B(\beta_c, N, x) + \bar{\rho}(B(\beta_c, N, x) + B(\bar{\rho}\beta_r, N, x)) \\
&= \bar{\rho}B(\bar{\rho}\beta_r, N, x) + B(\beta_c, N, x)
\end{aligned}
$$

And

$$G_X(s) \approx \bar{\rho}[(1-\bar{\rho}\alpha_r) + \bar{\rho}\alpha_r s]^N + [(1-\alpha_c)+\alpha_c s]^N$$