8 Toward Realism I: Stochastic Models

This week, we'll take an important step toward realism by asking: what if transmission is, indeed, stochastic? Recall that in our SIR model formulation, we let c be the rate of contact between people, and let p be the probability of transmission given contact between an S and an I. *In expectation*, the number of new infections will be $cpSI$, but what about in practice? What if, in fact, the probabilities really are probabilities? And when does it matter?

8.1 Expected Epidemic Final Sizes

Imagine choosing a person from the population uniformly at random.[1] Let's call this person u. What is the probability that u gets infected during an epidemic?

There are a few ways to work with such a calculation. A typical approach starts like this:

$$\begin{aligned} P(u \text{ gets infected}) &= P(u \text{ gets infected by one or more people}) \\ &= 1 - P(u \text{ doesn't get infected by anybody}) \\ &= 1 - \prod_v P(u \text{ doesn't get infected by } v) \\ &= 1 - \prod_v \left[1 - P(u \text{ gets infected by } v)\right] . \end{aligned} \tag{1}$$

This feels straightforward, and it is. But what assumptions have we made in getting to the last row? Critically, we have assumed independence between these events. Whether an individual v transmits to u depends on whether v becomes infected, and *that* could come about from a chain of infections that passed through u to get to v in the first place. In other words, the probability that $v \to u$ depends on whether u has been infected, i.e., not independent events!

To sidestep the question of independence, we're going to add an assumption: u cannot transmit, even when infected. In other words, u is a dead-end for the epidemic. This makes it OK to make independence assumptions. Now what?

We're going to define a **consistency relation** for π, the probability that u gets infected. It's called a consistency relation because the right-hand side will be a function of π, and the left-hand side will *be* π, and for it all to work the notions of the LHS's π and the RHS's π will have to be consistent with each other, i.e. LHS=RHS.

Before continuing, let's revisit this idea of "the probability that u gets infected during an epidemic." Remember that u was chosen from the population uniformly at random. This means that if the probability that

[1] Notice that we say "*uniformly* at random" instead of just "randomly." This precision is meant to be clear that our process gives equal probability of N^{-1} to choosing each person.

u gets infected is, say, $\frac{2}{3}$, then in the typical epidemic any person chosen uniformly at random has a probability $\frac{2}{3}$ of having been infected. In other words, the epidemic's final size, as a proportion of the population, is $\frac{2}{3}$. We therefore now have a new way to interpret π. In one sense, it's the probability that a randomly chosen person was infected. In another, and equivalent sense, it's the epidemic's final size.

The following two subsections follow the derivation of Joel Miller closely, and his original paper, while moving quickly, provides a fantastic set of references if this type of mathematics fascinates you.

8.1.1 Homogeneous susceptibility

Let's imagine that everyone in the population is equally susceptible, and that we begin with an initially susceptible population proportion of s_0. Our test individual u is a dead-end, and can be infected but can infect no one else. The expected number of transmissions from a randomly chosen individual, when everyone else is susceptible, is R_0. Transmissions to susceptibles take place, and would-be transmissions to non-susceptibles have no effect.

First, the probability that u begins as susceptible is s_0. Given that u is susceptible, the probability that a uniformly randomly chosen infected individual transmits to u is $\frac{R_0}{N}$, so the probability that the person does *not* transmit to u is $1 - \frac{R_0}{N}$. Because a fraction π become infected eventually, the total number of people who attempt to transmit to u is πN. Putting this together using Eq. (1), we get

$$\pi = 1 - s_0 \left(1 - \frac{R_0}{N}\right)^{\pi N} ,$$

which becomes in the limit of large N,

$$\pi = 1 - s_0 e^{-R_0 \pi} . \tag{2}$$

Note that if we let $s_0 \to 1$ and write π as r_∞, we get the familiar

$$r_\infty = 1 - e^{R_0 r_\infty} \tag{3}$$

which we derived using a completely different approach (differential equations) from the SIR model in past lectures! This is one of the joys of mathematical modeling.

What assumptions did we make about the structure of the model? In fact, *we assumed nothing about the structure of our model*, be it SIR, SEIR, or even exotic models with various E and I stages, varying levels of infectiousness, etc. The model could even be one in which we don't have differential equations or typical compartments. In fact, we could have a scenario where some individuals are barely infectious while others are extremely highly infectious—a so-called superspreader scenario—but this, too, reduces to the Eq. (2). In fact, our key assumptions here were that (1) everyone is equally susceptible, and (2) that the population is well mixed. Then, provided that the expected number of transmissions is R_0, the equation above holds. This is one of the exciting and powerful strengths of the probabilistic approach.

8.1.2 Heterogeneous susceptibility

What if people vary in their susceptibility? Previously, when an individual infection led to R_0 infections in expectation, each person (and our test individual u in particular) was on the receiving end of that attempted transmission with probability $\frac{1}{N}$, making the total probability that each person is on the receiving end of an attempted transmission $\frac{R_0}{N}$.

We now simply modify this to say that each individual has some susceptibility level x, such that $\frac{x}{N}$ is the probability that a random infected individual will transmit to an individual of level x. To characterize the distribution of susceptibility in the population, let $p(x)$ be the PDF of susceptibilities, and let $s_0(x)$ be the probability that an individual of susceptibility x begins susceptible.

Let $\pi(x)$ be the probability that an individual of susceptibility level x is ultimately infected. This means that the mean probability of infection over the whole population is given by $\bar{\pi} = \int_0^\infty \pi(x)p(x)\,dx$.[2] Just as before, this also means that $\bar{\pi}$ is the proportion of the population that is infected!

In parallel to our previous derivation, we now need to compute the expected number of transmissions to our test individual from a single infection. Previously, this was simply $R_0\pi$, but now, if our test individual is of susceptibility level x, it is now $x\bar{\pi}$. The probability that u remains susceptible is

$$1 - \pi(x) = s_0(x)e^{-x\bar{\pi}}$$

which, taking the complement of both sides and integrating against $p(x)\,dx$ from 0 to ∞, yields

$$\bar{\pi} = 1 - \int_0^\infty s_0(x)e^{-x\bar{\pi}}p(x)\,dx\ . \tag{4}$$

and if the population begins with everyone susceptible $s_0(x) \to 1$ and we get

$$\bar{\pi} = 1 - \int_0^\infty e^{-x\bar{\pi}}p(x)\,dx\ . \tag{5}$$

This equation, too, is exciting. We derived it without consideration of whether natural history of infection was complicated or simple. In fact, our key assumptions here were that the population was well mixed, and that there was no association between effects like superspreading and susceptibility[3]

[2] This note is just a reminder that this is just the standard way to compute an expected value of a function of a continuous random variable. It takes the values that the function can achieve $\pi(x)$, and then weights them by their probability $p(x)$. Instead of a weighted average (like a sum) which we would get for a discrete random variable with discrete outcomes, we get an integral for a continuous random variable. Ask in office hours if you need a refresher!

[3] You could imagine a scenario in which the people who are highly susceptible are also the people who go on to infect a great many others. This type of scenario would violate the assumptions of this derivation.

8.1.3 Immunity from prior infection, vaccination, and/or both

What if the population contains some folks who are naïve, and various others with different levels of partial immunity to infection? In this scenario, Equation (5) applies right out of the box, with little additional modification, provided that we can assume that anyone who does become infected is equally infectious.[4]

In a simple case, suppose that a vaccinated proportion of the population v has partial protection VE, and that everyone else is fully susceptible. For the vaccinated, $x = (1 - VE)R_0$, and for the unvaccinated, $x = R_0$. This means that

$$p\big(R_0\big) = 1 - v \qquad \text{and} \qquad p\big((1 - VE)R_0\big) = v \; .$$

Applying Eq. (5), we get

$$\bar{\pi} = 1 - \left[(1 - v)e^{-R_0\bar{\pi}} + ve^{-(1-VE)R_0\bar{\pi}}\right] \; . \tag{6}$$

As usual we can always reduce a more complicated scenario to a simpler one. For example, taking Eq. (6) in which no one is vaccinated $v = 0$ leads back to our homogeneous equation, Eq. (2). Similarly, if $VE = 1$ for perfect immunity, we get

$$\begin{aligned}
\bar{\pi} &= 1 - \left[(1 - v)e^{-R_0\bar{\pi}} + ve^{-(1-VE)R_0\bar{\pi}}\right] \\
&= 1 - \left[(1 - v)e^{-R_0\bar{\pi}} + v\right] \\
&= 1 - \left[(1 - v)e^{-R_0\bar{\pi}} + v\right] \\
&= (1 - v) - (1 - v)e^{-R_0\bar{\pi}} \\
&= (1 - v)\left[1 - e^{-R_0\bar{\pi}}\right] \; .
\end{aligned}$$

Note that $\bar{\pi}$ is the proportion of the total population that gets infected. Defining simply π as the proportion of the unvaccinated population that gets infected, we have $\pi(1 - v) = \bar{\pi}$, and therefore

$$\begin{aligned}
\pi &= 1 - e^{-R_0(1-v)\pi} \\
&= 1 - e^{-R_e\pi} \; ,
\end{aligned} \tag{7}$$

where $R_e = (1 - v)R_0$ is the effective reproductive number. This equation tells us that, when a proportion of the population v is perfectly vaccinated, the epidemic takes places solely in the unvaccinated population but with a reproductive number of R_e. This gives a slightly different interpretation to the herd immunity threshold: vaccination reduces the final epidemic size in the unvaccinated population, eventually eliminated the epidemic all together.

[4] This would include the leaky vaccine model, but not a multi-factor model in which the vaccine reduces infectiousness.

8.1.4 Reflections

Joel Miller's paper, A Note on the Derivation of Epidemic Final Sizes contains a number of interesting cases, including population contact structures. Joel repeatedly uses the same approach of a test individual u, and the same mathematical sleight of hand: the probability that u is infected is the same as the final epidemic size.

The lovely part about this approach is that it does not rely on a particular model structure, only model assumptions. The natural history of the infection could be appropriate for an SIR model, or it could be complicated with varying levels of infectiousness over time, or varying levels of infectiousness from one individual to another.

One important assumption of the mathematics in our lecture notes is that the population is well mixed, meaning that there is not an explicit and heterogeneous contact network, or other contact structure. However, these assumptions can be relaxed, for which I refer you to Miller's paper and the references therein.

8.2 Branching Processes and Epidemic Sizes

The previous sections considered the expected epidemic size in a *large* population in which the epidemic successfully takes off. In other words, at no point did we even consider the fact that the epidemic might not grow. Now, we'll ask a simple question: what are the chances that, even if $R_0 > 1$, an epidemic never takes off?

8.2.1 A Poisson Branching Process

A branching process is a *mathematical* process, meaning it is a mathematical description or model of stochastic events. One way to think of a branching process is as a description of a population of individuals and how they reproduce from one generation to the next.

Suppose that we start with a single infection in an otherwise susceptible (and, let's assume, effectively infinitely large) population. Suppose that, on average, each infection will produce two additional infections. Suppose further that the *number* of additional infections has a known distribution: a Poisson distribution with mean $\lambda = 2$. The probability mass function of $X \sim Pois(2)$ is shown in Fig. 1.

Look at the Poisson PMF and think about that single infected individual. Yes, in expectation, there will be 2 infections in the next generation. However, it is also clear that in the majority of cases, there will not be exactly 2 infections, and in fact, there is a good chance that there will be zero infections.

What is the probability that the transmission chain dies after just one generation? This is equivalent to calculating $P(X = 0)$. Recalling that the PMF for a Poisson with rate λ is

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} ,$$

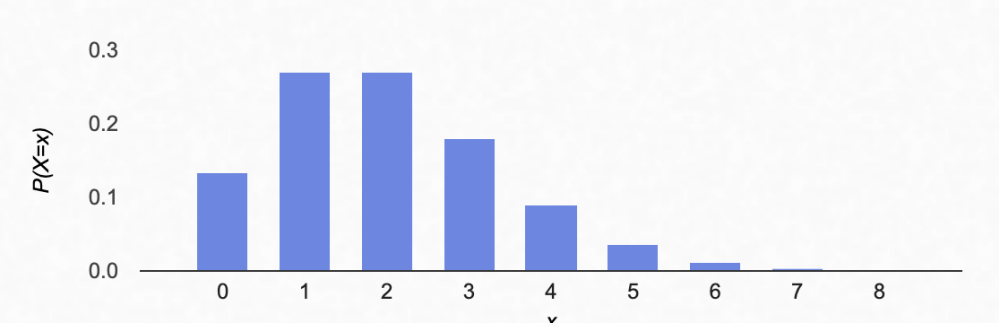

Figure 1: The probability mass function for a Poisson distribution with rate $\lambda = 2$.

we can see that

$$P(X = 0) = e^{-\lambda} ,$$

which is e^{-2} for $\lambda = 2$ or approximately 0.135.

What is the probability that the transmission chain has ceased after two generations? This complicates our question. Now we must consider a variety of possible ways in which the chain could cease. It could either cease in the first generation, *or* it could create one new infection and then cease, *or* it could create two new infections and then both of those could cease, *or* it could create three new infections and then both of those could cease, and so on.

This is a form of the Law of Total Probability. The outcome we are interested in is that the transmission has ceased by the end of the second generation, but there are many possible intermediate states. We'll introduce some notation to make things easier. Let $Z_0 = 1$ be the number of infections at time 0 and Z_1 be the number of infections at time 1. We can then write

$$P(\text{end by } t = 2) = \sum_{Z_1=0}^{\infty} P(Z_1)\ P(\text{all } Z_1 \text{ infections create zero new infections}) \tag{8}$$

We will assume that the population is sufficiently large that the transmission chains are independent, i.e. that they don't crash into each other. This allows us to write the probability that all Z_1 chains cease as $P(\text{one chain ceases})^{Z_1}$, which is equal to $(e^{-\lambda})_1^Z = e^{-\lambda Z_1}$. The first term is also addressable, because $P(Z_1)$ asks what the probability is that a Poisson random variable with rate λ is equal to Z_1. As a result, we get

$$P(\text{end by } t = 2) = \sum_{Z_1=0}^{\infty} \frac{\lambda^{Z_1}}{Z_1!} e^{-\lambda}\ e^{-\lambda Z_1}$$

and letting $\lambda = R_0$, we get

$$P(\text{end by } t = 2) = \sum_{Z_1=0}^{\infty} \frac{R_0^{Z_1}}{Z_1!} e^{-R_0}\, e^{-R_0 Z_1} \tag{9}$$

This formula is not particularly interesting because it doesn't address the big question: what's the probability that, in spite of $R_0 > 1$, that eventually, just by chance, the transmission chain ceases *without* requiring the depletion of susceptibles (and thus an epidemic) to do so?

To calculate the probability that transmission ceases without an epidemic, we will again introduce a consistency relation. Let q be the probability that the transmission chain from a single infection eventually dies. This is helpful because it is the quantity we hope to calculate, *and* because it can help us in its own calculation. In particular, imagine that we get $Z_1 = 5$ infections in generation 1. The probability that all 5 infections are dead ends is q^5, since we require each of those new transmission chains to eventually cease. The consistency relation is therefore

$$q = P(Z_1 = 0) + P(Z_1 = 1) \cdot q + P(Z_1 = 2) \cdot q^2 + P(Z_1 = 3) \cdot q^3 + \dots \tag{10}$$

Substituting in our Poisson probabilities, and explicitly writing $1 = q^0$ and $q = q^1$ to guide the eye, we have

$$q = \frac{R_0^0}{0!} e^{-R_0} \cdot q^0 + \frac{R_0^1}{1!} e^{-R_0} \cdot q^1 + \frac{R_0^2}{2!} e^{-R_0} \cdot q^2 + \frac{R_0^3}{3!} e^{-R_0} \cdot q^3 + \dots \tag{11}$$

Writing this out in summation notation, we can see that

$$\begin{aligned} q &= \sum_{k=0}^{\infty} \frac{R_0^k}{k!} e^{-R_0} \cdot q^k \\ &= e^{-R_0} \sum_{k=0}^{\infty} \frac{q R_0^k}{k!} \\ &= e^{-R_0} e^{q R_0} \\ &= e^{-R_0(1-q)} \end{aligned}$$

where we used the Taylor series $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. In short, we can now define the probability q that we get no epidemic *in spite of there being* $R_0 > 1$, and p the complementary probability that the transmission chain never ceases, as

$$q = e^{-R_0(1-q)} \qquad \text{and} \qquad p = 1 - e^{-R_0 p}\,. \tag{12}$$

Hold the phone! This equation for p is identical to Eq. (3), where we were calculating the epidemic final size under deterministic conditions *or* calculating the probability that a randomly chosen person in the population becomes infected during the epidemic. The fact that these calculations coincide with the probability that a Poisson outbreak is large is stimulating. Why might this be the case?

8.2.2 An SIR branching process

What if, instead of a Poisson-type branching process, we instead consider a stochastic version of the typical SIR model? Here, we get a slightly different result for the probabilities p (that a large outbreak occurs) and q (that all transmission chains die out). Once more, our approach will take the form of a consistency relation.

When we introduced the SIR model, we phrased our parameters β and γ in terms of rates, but in our thinking about the micro-scale processes that lead to these expected rates, we can think about stochastic effects. [5] Imagine a single infected individual. The rate at which they create a new infection is β, and the rate at which they recover is γ. In terms of the branching process, we have a race between these two events: will this person create a second infection *or* recover first? Thus

$$1 \to 0 \text{ (recovery) w.p. } \frac{\gamma}{\gamma+\beta} \qquad \text{and} \qquad 1 \to 2 \text{ (transmission) w.p. } \frac{\beta}{\gamma+\beta s_0} \, .$$

Thus,

$$q = \frac{\gamma}{\gamma+\beta} \cdot 1 + \frac{\beta s_0}{\gamma+\beta} \cdot q^2 \tag{13}$$

which can be solved to find

$$q = \frac{1}{R_0} \qquad \text{and thus} \qquad p = 1 - \frac{1}{R_0} \, . \tag{14}$$

A key conceptual difference in this approach compared to the Poisson branching process is that the Poisson branching process was assumed to take place in discrete time, generation after generation. Each infection creates some number of new infections in the next generation, and we assume that the original infection recovers. Here, we consider a branching process in continuous time. Each infection is in a race against the clock to create a secondary infection or recover. Because the SIR process considers an infection to last for an exponential amount of time, this model has the property of being **memoryless** which means that the probability that a person recovers in the next hour, day, or two days is not influenced by whether they have already been infected for a minute or a week. Thus, when the process branches to create two infections, the first infection is just as "fresh" as the second, mathematically speaking.[6]

An important mathematical difference is that here we argue that the probability of an outbreak is $p = 1 - \frac{1}{R_0}$ while under a Poisson branching process we argue that the probability of an outbreak is $p = 1 - e^{-R_0 p}$. For example, with $R_0 = 3$, $p_{\text{SIR}} = 0.667$ while $p_{\text{Poisson}} = 0.940$. With $R_0 = 6$, $p_{\text{SIR}} = 0.833$ while $p_{\text{Poisson}} = 0.998$. Those are *big* differences! In fact, they generalize: the more variable the number of

[5] Recall that, during past homework problems, you have explicitly thought about γ in terms of such probabilistic recovery.

[6] Memorylessness is a fascinating property, and often counterintuitive. You might think about it as the ultimate version of living in the present, because the past states in a memoryless system don't matter: the only thing that determines the future state is the exact present. If the present state doesn't change, the distribution of future states doesn't change—even if a considerable amount of time has elapsed.

secondary infections, the higher the probability of extinction. Here, the SIR model produces a broader distribution of secondary infections, while the Poisson model produces a more narrow distribution.[7]

8.2.3 Multiple introductions

What if a transmission chain is seeded not with just one infection, but with m initial infections? Here, rather than rebooting our derivation from the top, we can simply compute q, the probability that one chain is finite, and then compute q^m, the probability that all chains are finite. The probability of an epidemic is the probability that at least one of the chains is *not* finite, which is the complement of the probability that all chains are finite. In other words,

$$p(m) = 1 - q^m = 1 - [1 - p(1)]^m$$

is the probability that we get an infinite transmission chain due to m introductions. Note that here, we use the consistency relation to compute q (or $p(1)$), and then develop $p(m)$. Importantly, this relationship holds regardless of the type of branching process that we consider.

8.2.4 Reflections

In our derivations, we took a different view of transmission chains. A transmission chain either ceases, eventually, without depletion of susceptibles, or it does not. Put in the language of mathematics, a branching process either ceases in finite time, or continues indefinitely.

From the view of modeling infectious diseases, we can think about these two outcomes as different modes of transmission. We either have a small outbreak[8] that dies out on its own, and is thus finite in time and size, or we have a large outbreak[9] that never ceases and is thus infinite in time and size. In the former case, we may have $R_0 > 1$ and yet "get lucky" while in the latter, we are unlucky. In fact, during an epidemic, it may be the case that there are multiple introductions, and most chains cease, but one does not.

Conceptually, the results presented in these lecture notes should be stimulating. So far, we have relied on SIR-family models, and yet the SIR equations played no role in any of our analyses. In fact, we used solely information about R_0, susceptibility, and the distribution of the number of infections subsequent to each existing infection. This should be stimulating to the reader, because it is always valuable as a modeler to be able to derive results with the minimal set of assumptions about model structure. The less we assume, the more powerful the model and its results!

[7] Indeed, if the number of secondary infections isn't variable at all, and each infection deterministically produces R_0 secondary infections, then $q \to 0$ and $p \to 1$. Similarly, if the distribution of secondary cases is maximally broad, with 99% of infections causing 0 secondary infections, and 1% of infections causing $100 \times R_0$ secondary infections, it is clear that q is at least 0.99 and p is at most 0.01.

[8] Asymptotically irrelevant compared to the total population size.

[9] Asymptotically comparable to the total population size.