

A primer on the use of probability generating functions in infectious disease modeling



Joel C. Miller

Institute for Disease Modeling, Bellevue, WA, USA

ARTICLE INFO

Article history:

Received 21 August 2018

Accepted 30 August 2018

Available online 25 September 2018

Handling Editor: J. Wu

ABSTRACT

We explore the application of probability generating functions (PGFs) to invasive processes, focusing on infectious disease introduced into large populations. Our goal is to acquaint the reader with applications of PGFs, more so than to derive new results. PGFs help predict a number of properties about early outbreak behavior while the population is still effectively infinite, including the probability of an epidemic, the size distribution after some number of generations, and the cumulative size distribution of non-epidemic outbreaks. We show how PGFs can be used in both discrete-time and continuous-time settings, and discuss how to use these results to infer disease parameters from observed outbreaks. In the large population limit for susceptible–infected–recovered (SIR) epidemics PGFs lead to survival-function based models that are equivalent to the usual mass-action SIR models but with fewer ODEs. We use these to explore properties such as the final size of epidemics or even the dynamics once stochastic effects are negligible. We target this primer at biologists and public health researchers with mathematical modeling experience who want to learn how to apply PGFs to invasive diseases, but it could also be used in an applications-based mathematics course on PGFs. We include many exercises to help demonstrate concepts and to give practice applying the results. We summarize our main results in a few tables. Additionally we provide a small python package which performs many of the relevant calculations.

© 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The spread of infectious diseases remains a public health challenge. Increased interaction between humans and wild animals leads to increased zoonotic introductions, and modern travel networks allows these diseases to spread quickly. Many mathematical approaches have been developed to give us insight into the early behavior of disease outbreaks. An important tool for understanding the stochastic behavior of an outbreak soon after introduction is the probability generating function (PGF) (Allen, 2010; Wilf, 2005; Yan, 2008).

Specifically, PGFs frequently give insight about the statistical behavior of outbreaks before they are large enough to be affected by the finite-size of the population. In these cases, both susceptible–infected–recovered (SIR) disease (for which nodes recover with immunity) and susceptible–infected–susceptible (SIS) disease (for which nodes recover and can be reinfected immediately) are equivalent. In the case of SIR disease PGFs can also be used to study the dynamics of disease once an epidemic is established in a large population.

We can investigate properties such as the early growth rate of the disease, the probability the disease becomes established, or the distribution of final sizes of outbreaks that fail to become established. Similar questions also arise in other settings

E-mail address: joel.c.miller.research@gmail.com.

Peer review under responsibility of KeAi Communications Co., Ltd.

<https://doi.org/10.1016/j.idm.2018.08.001>

2468-0427/© 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

where some introduced agent can reproduce or die, such as invasive species in ecological settings (Lewis, Petrovskii, & Potts, 2016), early within-host pathogen dynamics (Conway & Coombs, 2011), and the accumulation of mutations in precancerous and cancerous cells (Antal & Krapivsky, 2011; Durrett, 2015) or in pathogen evolution (Volz, Romero-Severson, & Leitner, 2017). These are all examples of *branching processes*, and PGFs are a central tool for the analysis of branching processes (Bartlett, 1949; Kendall, 1949; Kimmel & Axelrod, 2002). Except for Section 4 where we develop deterministic equations for later-time SIR epidemics, based on (Miller, 2011; Miller, Slim, & Volz, 2012; Volz, 2008), the approaches we describe here have direct application in these other branching processes as well.

Before proceeding, we define what a PGF is. Let r_i denote the probability of drawing the value i from a given distribution of non-negative integers. Then $f(x) = \sum_i r_i x^i$ is the PGF of this distribution. We should address a potential confusion caused by the name. A “generating function” is a function which is defined from (or “generated by”) a sequence of numbers a_i and takes the form $\sum_i a_i x^i$. So a “probability generating function” is a generating function defined from a probability distribution on integers. It is *not* a function that generates probabilities when values are plugged in for x . There are other generating functions, including the “moment generating function”, defined to be $\sum_m \langle i^m \rangle x^m$ where $\langle i^m \rangle = \sum_i r_i i^m$ (the moment and probability generating functions turn out to be closely related).

PGFs have a number of useful properties which we derive in Appendix A. We have structured this paper so that a reader can skip ahead now and read Appendix A in its entirety to get a self-contained introduction to PGFs, or wait until a particular property is referenced in the main text and then read that part of the appendix.

As we demonstrate in Table 1, for many important distributions the PGF takes a simple form. We derive this for the Poisson distribution.

Example 1.1. Consider the Poisson distribution with mean λ

$$r_i = \frac{e^{-\lambda} \lambda^i}{i!}.$$

For this we find

$$\begin{aligned} f(x) &= \sum_i \frac{e^{-\lambda} \lambda^i}{i!} x^i = e^{-\lambda} \sum_i \frac{(\lambda x)^i}{i!} = e^{-\lambda} e^{\lambda x} \\ &= e^{\lambda(x-1)}. \end{aligned}$$

In this primer, we explore the application of PGFs to the study of disease spread. We will use PGFs to answer questions about the early-time behavior of an outbreak (neglecting depletion of susceptibles):

- What is the probability an outbreak goes extinct within g generations (or by time t) in an arbitrarily large population?
- What is the probability an index case causes an epidemic?
- What is the final size distribution of small outbreaks?
- What is the size distribution of outbreaks at generation g (or time t)?
- How fast is the initial growth for those outbreaks that do not go extinct?

Although we present these early-time results in the context of SIR outbreaks they also apply to SIS outbreaks and many other invasive processes.

Table 1
A few common probability distributions and their PGFs.

Distribution	PGF $f(x) = \sum_i r_i x^i$
Poisson, mean λ : $r_i = \frac{e^{-\lambda} \lambda^i}{i!}$	$e^{\lambda(x-1)}$
Uniform: $r_\lambda = 1$	x^λ
Binomial: n trials, with success probability p : $r_i = \binom{n}{i} p^i q^{n-i}$ for $q = 1 - p$	$[q + px]^n$
Geometric ^a : $r_i = q^i p$ for $q = 1 - p$ and $i = 0, 1, \dots$	$p / (1 - qx)$
Negative binomial ^b : $r_i = \binom{i + \hat{r} - 1}{i} q^{\hat{r}} p^i$ for $q = 1 - p$	$\left(\frac{q}{1 - px}\right)^{\hat{r}}$

^a Another definition of the geometric distribution with different indexing, $r_i = q^{i-1} p$ for $i = 1, 2, \dots$, gives a different PGF.

^b Typically the negative binomial is expressed in terms of a parameter r which is the number of failures at which the experiment stops, assuming each with success probability p . For us r_i plays an important role, so to help distinguish these, we use \hat{r} rather than r . Then r_i is the probability of i successes.

We can also use PGFs for some questions about the full behavior accounting for depletion of susceptibles. Specifically:

- In a continuous-time Markovian SIR or SIS outbreak spreading in a finite population, what is the distribution of possible system states at time t ?
- In the large-population limit of an SIR epidemic, what fraction of the population is eventually infected?
- In the large-population limit of an SIR epidemic, what fraction of the population is infected or recovered at time t ?

We will consider both discrete-time and Markovian continuous-time models of disease. In the discrete-time case each infected individual transmits to some number of “offspring” before recovering. In the continuous-time case each infected individual transmits with a rate β and recovers with a rate γ .

In Section 2 we begin our study investigating properties of epidemic emergence in a discrete-time, generation-based framework, focusing on the probability of extinction and the sizes of outbreaks assuming that the disease is invading a sufficiently large population with enough mixing that we can treat the infections caused by any one infected individual as independent of the others. We also briefly discuss how we might use our observations to infer disease parameters from observed small outbreaks. In Section 3, we repeat this analysis for a continuous-time case treating transmission and recovery as Poisson processes, and then adapt the analysis to a population with finite size N . Next in Section 4 we use PGFs to derive simple models of the large-time dynamics of SIR disease spread, once the infection has reached enough individuals that we can treat the dynamics as deterministic. Finally, in Section 5 we explore multitype populations in which there are different types of infected individuals, which may produce different distributions of infections. We provide three appendices. In Appendix A, we derive the relevant properties of PGFs, in Appendix B we provide elementary (*i.e.*, not requiring Calculus) derivations of two important theorems, and in Appendix C we provide details of a Python package `Invasion_PGF` available at https://github.com/joelmiller/Invasion_PGF that implements most of the results described in this primer. Python code that uses this package to implement the figures of Section 2 is provided in the supplement.

Our primary goal here is to provide modelers with a useful PGF-based toolkit, with derivations that focus on developing intuition and insight into the application rather than on providing fully rigorous proofs. Throughout, there are exercises designed to increase understanding and help prepare the reader for applications. This primer (and Appendix A in particular) could serve as a resource for a mathematics course on PGFs. For readers wanting to take a deep dive into the underlying theory, there are resources that provide a more technical look into PGFs in general (Wilf, 2005) or specifically using PGFs for infectious disease (Yan, 2008).

1.1. Summary

Before presenting the analysis, we provide a collection of tables that summarize our main results. Table 2 summarizes our notation. Tables 3 and 4 summarize our main results for the discrete-time and continuous-time models. Table 5 shows

Table 2

Common function and variable names. When we use a PGF for the number of susceptible individuals, active infections, and/or completed infections x and s correspond to susceptible individuals, y and i to active infections, and z and r to completed infections.

Function/variable name	Interpretation
$f(x) = \sum_i p_i x^i$ $g(x) = \sum_i q_i x^i$	Arbitrary PGFs.
$\mu(y) = \sum_i p_i y^i$ $\hat{\mu}(y) = (\beta y^2 + \gamma) / (\beta + \gamma)$ $\hat{\mu}(y, z) = (\beta y^2 + \gamma z) / (\beta + \gamma)$	Without hats: The PGF for the offspring distribution in discrete time. With hats: The PGF for the outcome of an unknown event in a continuous-time Markovian outbreak: y accounts for active infections and z accounts for completed infections.
$\alpha, \alpha_g, \alpha(t)$	Probability of either eventual extinction, extinction by generation g , or by time t in an infinite population.
$\Phi_g(y) = \sum_i \varphi_i(g) y^i$ $\Phi(y, t) = \sum_i \varphi_i(t) y^i$	PGF for the number of active infections in generation g or at time t in an infinite population.
$\Omega_\infty(z) = \sum_{r < \infty} \omega_r z^r + \omega_\infty z^\infty$ $\Omega_g(z) = \sum_r \omega_r(g) z^r$ $\Omega(z, t) = \sum_r \omega_r(t) z^r$	The PGF for the distribution of completed infections at the end of a small outbreak, in generation g , or at time t in an infinite population. If $\mathcal{R}_0 > 1$, then one of the terms in the expansion of $\Omega_\infty(z)$ is $\omega_\infty z^\infty$ where ω_∞ is the probability of an epidemic.
$\Pi_g(y, z) = \sum_{i,r} \pi_{i,r}(g) y^i z^r$ $\Pi(y, z, t) = \sum_{i,r} \pi_{i,r}(t) y^i z^r$	The PGF for the joint distribution of current infections and completed infections either at generation g or time t in an infinite population.
$\Xi(x, y, t) = \sum_{s,i} \xi_{s,i}(t) x^s y^i$	The PGF for the joint distribution of susceptibles and current infections at time t in a finite population of size N (used for continuous time only). In the SIR case we can infer the number recovered from this and the total population size.
$\chi(x) = \sum_i p_i x^i$	PGF for the “ancestor distribution”, analogous to the offspring distribution.
$\psi(x) = \sum_k P(k) x^k$	PGF for the distribution of susceptibility for the continuous time model where rate of receiving transmission is proportional to κ .
β, γ	The individual transmission and recovery rates for the Markovian continuous time model.

Table 3

A summary of our results for application of PGFs to discrete-time SIS and SIR disease processes in the infinite population limit. The function $\mu(\mathbf{x})$ is the PGF for the offspring distribution. The notation $[g]$ in the exponent denotes function composition g times. For example, $\mu^{[2]}(\mathbf{y}) = \mu(\mu(\mathbf{y}))$.

Question	Section	Solution
Basic Reproductive Number \mathcal{R}_0 [the average number of transmissions an infected individual causes early in an outbreak].	Intro to 2	$\mathcal{R}_0 = \mu'(1)$.
Probability of extinction, α , given a single introduced infection.	2.1	$\alpha = \lim_{g \rightarrow \infty} \mu^{[g]}(0)$ or, equivalently, the smallest x in $[0, 1]$ for which $x = \mu(x)$.
Probability of extinction within g generations	2.1.2	$\alpha_g = \mu^{[g]}(0)$.
PGF of the distribution of the number of infected individuals in the g -th generation.	2.2	$\Phi_g(y)$ where Φ_g solves $\Phi_g(y) = \mu^{[g]}(y)$.
Average number of active infections in generation g and average number if the outbreak has not yet gone extinct.	2.2	\mathcal{R}_0^g , and $\frac{\mathcal{R}_0^g}{1 - \alpha_g}$.
PGF of the number of completed cases at generation g in an infinite population.	2.3.1	$\Omega_g(z)$ where Ω_g solves $\Omega_g(z) = z\mu(\Omega_{g-1}(z))$ with $\Omega_0(z) = 1$.
PGF of the joint distribution of the number of current and completed cases at generation g in an infinite population.	2.3.2	$\Pi_g(y, z)$ where Π_g solves $\Pi_g(y, z) = z\mu(\Pi_{g-1}(y, z))$ with $\Pi_0(y, z) = y$.
PGF of the final size distribution.	2.4	$\Omega_\infty(z)$ where Ω_∞ solves $\Omega_\infty(z) = \lim_{g \rightarrow \infty} \Omega_g(z)$. It also solves $\Omega_\infty(z) = z\mu(\Omega_\infty(z))$. This has a discontinuity at $ z = 1$ if epidemics are possible.
Probability an outbreak infects exactly j individuals	2.4	$\frac{p_{j-1}^{(j)}}{j}$ where $p_i^{(j)}$ is the coefficient of y^i in the expansion of $[\mu(y)]^j$.
Probability a disease has a particular set of parameters Θ given a set of observed independent outbreak sizes $X = (j_1, \dots, j_k)$ and a prior belief $P(\Theta)$.	2.4.1	$P(\Theta X) = \frac{P(j_1 \Theta) \cdots P(j_k \Theta)P(\Theta)}{\sum_{\Theta'} P(j_1 \Theta') \cdots P(j_k \Theta')P(\Theta')}$, which can be solved numerically using our prior knowledge $P(\Theta)$ and our knowledge of the probability of each j_i given Θ .

Table 4

A summary of our results for application of PGFs to the continuous-time disease process. We assume individuals transmit with rate β and recover with rate γ . The functions $\hat{\mu}(\mathbf{y}) = (\beta\mathbf{y}^2 + \gamma)/(\beta + \gamma)$ and $\hat{\mu}(\mathbf{y}, \mathbf{z}) = (\beta\mathbf{y}^2 + \gamma\mathbf{z})/(\beta + \gamma)$ are given in System (14).

Question	Section	Solution
Probability of eventual extinction α given a single introduced infection.	3.1	$\alpha = \min(1, \gamma/\beta)$
Probability of extinction by time t , $\alpha(t)$.	3.1.1	$\alpha(t)$ where $\dot{\alpha} = (\beta + \gamma)[\hat{\mu}(\alpha) - \alpha]$ and $\alpha(0) = 0$.
PGF of the distribution of number of infected individuals at time t (assuming one infection at time 0).	3.2	$\Phi(y, t)$ where $\Phi(y, 0) = y$ and Φ solves either $\frac{\partial}{\partial t} \Phi = (\beta + \gamma)[\hat{\mu}(y) - y] \frac{\partial}{\partial y} \Phi$ or $\frac{\partial}{\partial t} \Phi = (\beta + \gamma)[\hat{\mu}(\Phi) - \Phi]$.
PGF of the number of completed cases at time t .	3.4	$\Omega(z, t)$ where $\Omega(z, 0) = 1$ and Ω solves $\frac{\partial}{\partial t} \Omega = (\beta + \gamma)[\hat{\mu}(\Omega, z) - \Omega]$
PGF of the joint distribution of the number of current and completed cases at time t (assuming one infection at time 0).	3.3	$\Pi(y, z, t)$ where $\Pi(y, z, 0) = y$ and Π solves either $\frac{\partial}{\partial t} \Pi = (\beta + \gamma)[\hat{\mu}(y, z) - y] \frac{\partial}{\partial y} \Pi$ or $\frac{\partial}{\partial t} \Pi = (\beta + \gamma)[\hat{\mu}(\Pi, z) - \Pi]$.
PGF of the final size distribution.	3.4	$\Omega_\infty(z) = \lim_{t \rightarrow \infty} \Omega(z, t)$. This also solves $\Omega_\infty(z) = \hat{\mu}(\Omega_\infty(z), z)$. If epidemics are possible this has a discontinuity at $ z = 1$.
Probability an outbreak infects exactly j individuals	3.4	$\frac{1}{j} \frac{\beta^{j-1} \gamma^j}{(\beta + \gamma)^{2j-1}} \binom{2j-2}{j-1}$.
PGF for the joint distribution of the number susceptible and infected at time t for SIS dynamics in a population of size N .	3.5.1	$\Xi(x, y, t)$ where Ξ solves $\frac{\partial}{\partial t} \Xi = \frac{\beta}{N}(y^2 - xy) \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi + \gamma(x - y) \frac{\partial}{\partial y} \Xi$
PGF for the joint distribution of the number susceptible and infected at time t for SIR dynamics in a population of size N .	3.5.2	$\Xi(x, y, t)$ where Ξ solves $\frac{\partial}{\partial t} \Xi = \frac{\beta}{N}(y^2 - xy) \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi + \gamma(1 - y) \frac{\partial}{\partial y} \Xi$

applications of PGFs to the continuous-time dynamics of SIR epidemics once the disease has infected a non-negligible proportion of a large population, effectively showing how PGFs can be used to replace most common mass-action models. Finally, Table 6 provides the probability of each finite final outbreak size assuming a sufficiently large population that susceptible depletion never plays a role.

1.2. Exercises

We end each section with a collection of exercises. We have designed these exercises to give the reader more experience applying PGFs and to help clarify some of the more subtle points.

Table 5

A summary of our results for application of PGFs to the final size and large-time dynamics of SIR disease. The PGFs χ and ψ encode the heterogeneity in susceptibility. The PGF χ is the PGF of the ancestor distribution (an ancestor of u is any individual who, if infected, would infect u). The PGF $\psi(\mathbf{x}) = \sum_{\mathbf{k}} \mathbf{p}(\mathbf{k}) \mathbf{x}^{\mathbf{k}}$ encodes the distribution of the contact rates.

Question	Section	Solution
Final size relation for an SIR epidemic assuming a vanishingly small fraction ρ randomly infected initially with $\rho N \gg 1$.	4.2	$r(\infty) = 1 - \chi(1 - r(\infty))$. [For standard assumptions, including the usual continuous-time assumptions, $\chi(x) = e^{-\beta_0(1-x)}$.]
Discrete-time number susceptible, infected, or recovered in a population with homogeneous susceptibility and given \mathcal{R}_0 , assuming an initial fraction ρ is randomly infected with $\rho N \gg 1$.	4.3	For $g > 0$: $S(g) = N(1 - \rho)e^{-\beta_0(1-S(g-1)/N)}$ $I(g) = N - S(g) - R(g)$ $R(g) = R(g-1) + I(g-1)$ with the initial condition $S(0) = (1 - \rho)N$, $I(0) = \rho N$, and $R(0) = 0$.
Discrete-time number susceptible, infected, or recovered in a population with heterogeneous susceptibility for SIR disease after g generations with an initial fraction ρ randomly infected where $\rho N \gg 1$.	4.3	For $g > 0$: $S(g) = N(1 - \rho)\chi(S(g-1)/N)$ $I(g) = N - S(g) - R(g)$ $R(g) = R(g-1) + I(g-1)$ with the initial condition $S(0) = (1 - \rho)N$, $I(0) = \rho N$, and $R(0) = 0$.
Continuous time number susceptible, infected, or recovered for SIR disease as a function of time with an initial fraction ρ randomly infected where $\rho N \gg 1$. Assumes u receives infection at rate $\beta I_{\kappa_u}/N(K)$	4.4	$S(t) = (1 - \rho)N\psi(\theta(t))$ $I(t) = N - S(t) - R(t)$ For $t > 0$: $R(t) = \frac{\gamma N(K)}{\beta} \ln \theta(t)$ with the initial condition $\dot{\theta}(t) = -\frac{\beta}{N(K)} I\theta(t)$ $\theta(0) = 1$.

Table 6

The probability of j total infections in an infinite population for different offspring distributions, derived using Theorem 2.7 and the corresponding log-likelihoods. For any one of these, if we sum the probability of j over (finite) j , we get the probability that the outbreak remains finite in an infinite population. This is particularly useful when inferring disease parameters from observed outbreak sizes (Section 2.4.1). The parameters' interpretations are given in Table 1.

Distribution	PGF	Probability of j infections	Log-Likelihood of Parameters given j
Poisson	$e^{\lambda(y-1)}$	$\frac{(j\lambda)^{j-1}}{j!} e^{-j\lambda}$	$-j\lambda + (j-1)\log(j\lambda) - \log(j!)$
Uniform	y^λ	$\begin{cases} 1 & j = 1, \lambda = 0 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 0 & j = 1, \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$
Binomial	$(q + py)^n$	$\frac{1}{j} \binom{nj}{j-1} p^{j-1} q^{nj-j+1}$	$\log((nj)!) - \log((nj-j+1)!) - \log(j!) + (j-1)\log p + (nj-j+1)\log q$
Geometric	$p/(1-xy)$	$\frac{1}{j} \binom{2j-2}{j-1} p^j q^{j-1}$	$\log((2j-2)!) - \log((j-1)!) - \log(j!) + j\log p + (j-1)\log q$
Negative Binomial	$\left(\frac{q}{1-py}\right)^{\hat{r}}$	$\frac{1}{j} \binom{\hat{r}j+j-2}{j-1} q^{\hat{r}j} p^{j-1}$	$\log((\hat{r}j+j-1)!) - \log((\hat{r}j-1)!) - \log(j!) + \hat{r}j\log q + (j-1)\log p$

Exercise 1.1. Except for the Poisson distribution handled in Example 1.1, derive the PGFs shown in Table 1 directly from the definition $f(x) = \sum_i r_i x^i$.

For the negative binomial, it may be useful to use the binomial series:

$$(1 + \delta)^\eta = 1 + \eta\delta + \frac{\eta(\eta-1)}{2!}\delta^2 + \dots + \frac{\eta(\eta-1)\dots(\eta-i+1)}{i!}\delta^i + \dots$$

using $\eta = -\hat{r}$ and $\delta = -px$.

Exercise 1.2. Consider the binomial distribution with n trials, each having success probability $p = \lambda/n$. Using Table 1, show that the PGF for the binomial distribution converges to the PGF for the Poisson distribution in the limit $n \rightarrow \infty$, if λ is fixed.

2. Discrete-time spread of a simple disease: early time

We begin with a simple model of disease transmission using a discrete-time setting. In the time step after becoming infected, an infected individual causes some number of additional cases and then recovers. We let p_i denote the probability of causing exactly i infections (referred to as “offspring”) before recovering. It will be useful to define the PGF for the offspring distribution

$$\mu(y) = \sum_{i=0}^{\infty} p_i y^i. \tag{1}$$

For results related to early extinction or early-time dynamics, we will assume that the population is large enough and sufficiently well-mixed that the transmissions in successive generations are all independent events and unaffected by depletion of susceptible individuals. Before deriving our results for the early-time behavior of our discrete-time model, we offer a summary in Table 3.

Often in disease spread we are interested in the expected number of infections caused by an infected individual early in an outbreak, which we define to be \mathcal{R}_0 .

$$\mathcal{R}_0 = \sum_i ip_i = \mu'(1) \tag{2}$$

where $\mu'(x) = \frac{d}{dx}\mu(x)$. The value of \mathcal{R}_0 is related to disease dynamics, but it is not the only important property of μ .

Example 2.1. We demonstrate a few sample outbreaks in Fig. 1. Here we take a bimodal case with $\mathcal{R}_0 = 0.9$ such that a proportion 0.3 of the population cause 3 infections and the remaining 0.7 cause none. Most of the outbreaks die out immediately, but some persist, surviving multiple generations before extinction.

Example 2.2. Throughout Section 2 we compare simulated SIR outbreaks with the theoretical predictions which we calculate using the Python package *Invasion_PGF* described in Appendix C. We assume that all individuals are equally likely to be infected by any transmission, and we focus on $\mathcal{R}_0 = 0.75$ and $\mathcal{R}_0 = 2$. For each \mathcal{R}_0 , we consider two distributions for the number of new infections an infected individual causes:

- a Poisson-distributed number of infections with mean \mathcal{R}_0 , or
- a bimodal distribution with either 0 or 3 infections, with the proportion chosen to give a mean of \mathcal{R}_0 . The probabilities are $p_0 = 1 - \mathcal{R}_0/3$ and $p_3 = \mathcal{R}_0/3$ ($\mathcal{R}_0 > 3$ is impossible).

The bimodal distribution is similar to that of Fig. 1, but with different probabilities of 0 or 3. After an individual chooses the number of infections to cause, the recipients are selected uniformly at random (with replacement) from the population. If they are susceptible, an infection occurs at the next time step, otherwise nothing happens. We use 5×10^5 simulations for $N = 100$ and $N = 1000$.

Fig. 2 looks at the final size distribution. The distribution of the number infected in small outbreaks (insets) is not significantly affected by the total population size. This is because they do not grow large enough to “see” the system size. They would die out even in an infinite population. Large outbreaks, or epidemics, on the other hand would grow without bound in an infinite population, and their growth is limited by the finiteness of the population. We will see that (assuming homogeneous susceptibility and the large population limit), the proportion infected in an SIR epidemic depends only on \mathcal{R}_0 .

2.1. Early extinction probability

A common misconception is that if $\mathcal{R}_0 > 1$ an epidemic is inevitable. In fact, if we are lucky an outbreak can die out stochastically before the number infected is large. Conversely, if we are not lucky it may initially grow faster than our deterministic models predict.

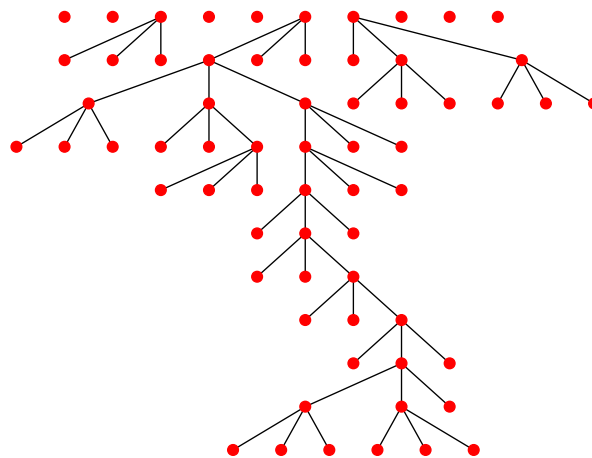


Fig. 1. A sample of 10 outbreaks starting with a bimodal distribution having $\mathcal{R}_0 = 0.9$ in which 3/10 of the population causes 3 infections and the rest cause none. The top row denotes the initial states, showing each of the 10 initial infections. An edge from one row to the next denotes an infection from the higher node to the lower node. Most outbreaks die out immediately.

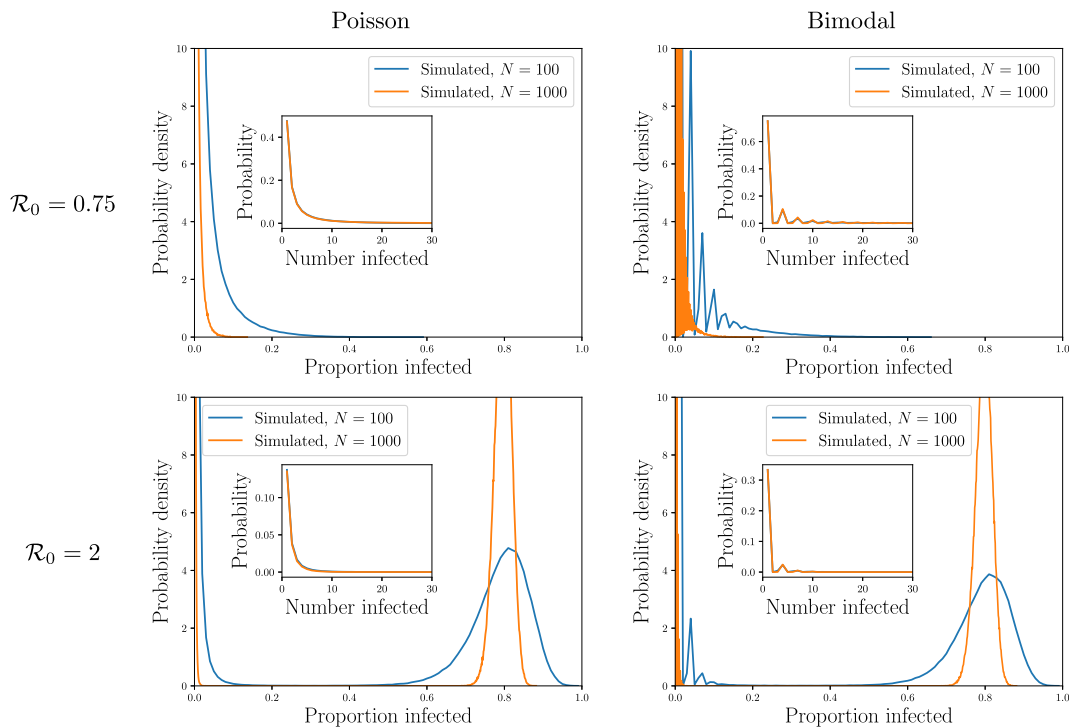


Fig. 2. Simulated outcomes of SIR outbreaks in populations as described in Example 2.2. Outbreaks tend to be either small or large. The typical number infected in small outbreaks (insets) is affected by the details of the offspring distribution, but not the population size. The typical proportion infected in large outbreaks (epidemics) appears to depend on the average number of transmissions an individual causes, but not the population size or the offspring distribution. These observations will be explained later. These simulations are reused throughout this section to show how PGFs capture different properties of the distributions.

In any finite population a disease will eventually go extinct because the disease interferes with its own spread. Our observations show that the typical final outcomes of an outbreak are either an “epidemic” which grows until the number infected is limited by the finiteness of the population or a small outbreak which dies out before it can see the system size. One of our first questions about a possible disease emergence is “what is the probability that an outbreak will grow into an epidemic?”. We focus on the equivalent question, “what is the probability the outbreak goes extinct before causing an epidemic?”. We aim to calculate the probability that the disease would go extinct if it never interferes with its own spread, or in other words, if it were spreading through an unlimited population. Throughout we assume that disease is introduced with a single randomly chosen index case.

The theory for the extinction probability in an unbounded population has been developed extensively in the context of Galton–Watson processes (Watson & Galton, 1875). It has been applied to infectious disease many times, e.g. (Easley & Kleinberg, 2010, section 21.8), and (Getz & Lloyd-Smith, 2006; Lloyd-Smith, Schreiber, Kopp, & Getz, 2005).

2.1.1. Derivation as a fixed point equation

We present two derivations of the extinction probability. Our first is quicker, but gives less insight. We start with the *a priori* observation that the extinction probability takes some value between 0 and 1 inclusive. Our goal is to filter out the vast majority of these options by finding a property of the extinction probability that most values between 0 and 1 do not have.

Let α be the probability of extinction if the spread starts from a single infected individual. Then from Property A.1 of Appendix A we have $\alpha = \sum_i p_i \hat{\alpha}^i = \mu(\hat{\alpha})$ where $\hat{\alpha}$ is the probability that, in isolation, an offspring of the initial infected individual would not cause an epidemic. Because we assume that the offspring distribution of later cases is the same as for the index case, we must have $\hat{\alpha} = \alpha$ and so the extinction probability solves $\alpha = \mu(\alpha)$.

We have established:

Theorem 2.1. Assuming that each infected individual produces an independent number of offspring i chosen from a distribution having PGF $\mu(y)$, then α , the probability an outbreak starting from a single infected individual goes extinct, satisfies

$$\alpha = \mu(\alpha). \tag{3}$$

Not all solutions to $x = \mu(x)$ must give the extinction probability.

There can be more than one x solving $x = \mu(x)$. In fact $1 = \mu(1)$ is always a solution, and from Property A.9 it follows that there is another solution if and only if $\mathcal{R}_0 = \mu'(1) > 1$. In this case, our derivation of Theorem 2.1 does not tell us which of the solutions is correct. However, Section 2.1.2 shows that the correct solution is the smaller solution when it exists. More specifically the extinction probability is $\alpha = \lim_{g \rightarrow \infty} \alpha_g$ where $\alpha_g = \mu(\alpha_{g-1})$ starting with $\alpha_0 = 0$. This gives a condition for a nonzero epidemic probability. Namely $\mathcal{R}_0 = \mu'(1) = \sum_i i p_i > 1$.

Example 2.3. We now consider the Poisson and bimodal offspring distributions described in Example 2.2. We saw that typically an outbreak either affects a small proportion of the population (a vanishing fraction in the infinite population limit) or a large number (a nonzero fraction in the infinite population limit).

By plotting the cumulative density function (cdf) of proportion infected in Fig. 3, we extend our earlier observations. The cdf is steep near zero (becoming vertical in the infinite population limit). Then it is effectively flat for a while. Finally if $\mathcal{R}_0 > 1$ it again grows steeply at some proportion infected well above 0 (the size of epidemic outbreaks).

The plateau's height is the probability that an outbreak dies out while small. Fig. 3 shows that this is well-predicted by choosing the smaller of the solutions to $x = \mu(x)$.

For a fixed $\mathcal{R}_0 > 1$, the the plateau's height (i.e., the early extinction probability) depends on the details of the offspring distribution and not simply \mathcal{R}_0 . However, the critical value at which the cdf increases for the second time depends only on \mathcal{R}_0 . This suggests that even though the probability of an epidemic depends on the details of the offspring distribution, the proportion infected in an SIR epidemic depends only on \mathcal{R}_0 , the reproductive number. We explore this in more detail in Section 4.2.

2.1.2. Derivation from an iterative process

In our second derivation, we calculate the probability that the outbreak dies out within g “generations”. Then the probability the outbreak would die out after a finite number of steps in an infinite population is simply the limit of this as $g \rightarrow \infty$. In our counting of “generations”, we consider the index case to be generation 0. An individual's generation is equal to the number of transmissions occurring in the chain from the index case to that individual.

We define α_g to be the probability that the longest chain an index case will initiate has fewer than g transmissions. So because there are always at least 0 transmissions, $\alpha_0 = 0$. The probability that there is no transmission is by definition α_1 . Recalling that the probability the index case causes zero infections is p_0 , we have

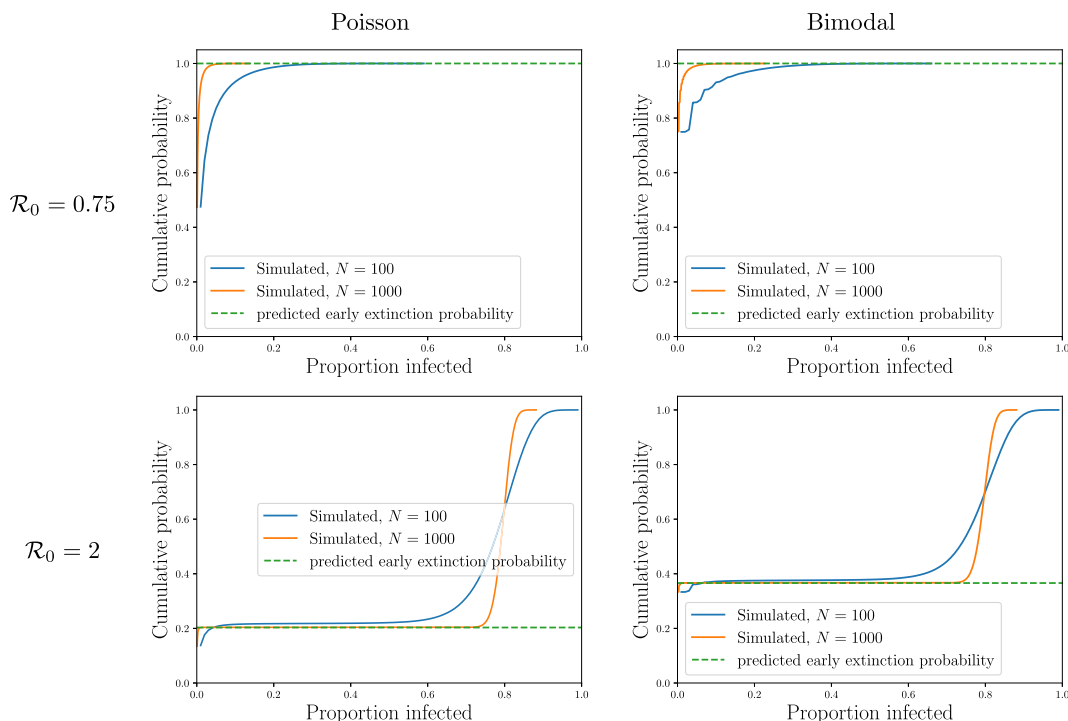


Fig. 3. Illustration of Theorem 2.1. The cumulative density function (cdf) for the total proportion ever infected (effectively the integral of Fig. 2). For small \mathcal{R}_0 , all outbreaks die out without affecting a sizable portion of the population. For larger \mathcal{R}_0 , there are many small outbreaks and many large outbreaks, but very few outbreaks in between, so the cdf is flat in this range. The height of this plateau is the probability the outbreak dies out while small. This is approximately the predicted extinction probability for an infinite population (dashed). The probability of a small outbreak is different for the different offspring distributions, but the proportion infected corresponding to epidemics is the same (for given \mathcal{R}_0).

$$\alpha_1 = p_0 = \mu(0) = \mu(\alpha_0)$$

is the probability that the index case does not cause a chain of 1 or more transmissions. The probability that all chains die out after at most 1 transmission (that is, there are no second generation cases) is the probability that the index case causes i infections, p_i , times the probability none of those i individuals causes further infections, α_1^i , summed over all i . We introduce the notation $\mu^{[g]}(x)$ to be the result of iterative applications of μ to x g times, so $\mu^{[1]}(x) = \mu(x)$ and for $g > 1$, $\mu^{[g]}(x) = \mu(\mu^{[g-1]}(x))$. Then following Property A.1 we have

$$\alpha_2 = p_0 + p_1\alpha_1 + p_2\alpha_1^2 + \dots = \mu(\alpha_1) = \mu^{[2]}(0)$$

We generalize this by stating that the probability an initial infection fails to initiate any length g chains is equal to the probability that all of its i offspring fail to initiate a chain of length $g - 1$.

$$\alpha_g = \sum_i p_i \alpha_{g-1}^i = \mu(\alpha_{g-1}) = \mu^{[g]}(0).$$

So the probability of not starting a chain of length at least g is found by iteratively applying the function μ g times to $x = 0$. Taking $g \rightarrow \infty$ gives the extinction probability (Getz & Lloyd-Smith, 2006):

$$\alpha = \lim_{g \rightarrow \infty} \mu^{[g]}(0). \quad (4)$$

The fact that there is a biological interpretation of α_g starting with $\alpha_0 = 0$ is important. It effectively guarantees that the iterative process converges and that the speed of convergence reflects the typical speed of extinction. Iteration appears to be an efficient way to solve $x = \mu(x)$ numerically and because of the biological interpretation, we can avoid questions that might arise about whether there are multiple solutions of $x = \mu(x)$ and, if so, which of them corresponds to the biological problem. Instead we simply iterate starting from 0 and the result must converge to the probability that in an infinite population the outbreak would go extinct in finite time, regardless of what other solutions $x = \mu(x)$ might have.

Exercise 2.1 shows that if $\mu(0) \neq 0$ then the limit of the sequence α_g is 1 if $\mathcal{R}_0 \leq 1$ and some $\alpha < 1$ satisfying $\alpha = \mu(\alpha)$ if $\mathcal{R}_0 > 1$. This proves:

Theorem 2.2. Assume that each infected individual produces an independent number of offspring i chosen from a distribution having PGF $\mu(y)$. Then

- The probability an outbreak goes extinct within g generations is

$$\alpha_g = \mu^{[g]}(0). \quad (5)$$

- The probability of extinction in an infinite population is

$$\alpha = \lim_{g \rightarrow \infty} \alpha_g.$$

- If $\mathcal{R}_0 = \mu'(1) \leq 1$ and $\mu(0) \neq 0$ then $\alpha = 1$. If $\mathcal{R}_0 > 1$ extinction occurs with probability $\alpha < 1$.

Example 2.4. We now consider the Poisson and bimodal offspring distributions described in Example 2.2

Fig. 4 shows that starting with $\alpha_0 = 0$ and defining $\alpha_g = \mu(\alpha_{g-1})$, the values of α_g emerging from the iterative process correspond to the observed probability that outbreaks have gone extinct by generation g for early values of g .

In the infinite population limit, this provides a match for all g . So this gives the probability the outbreak goes extinct by generation g assuming it has not grown large enough to see the finite-size of the population (i.e., assuming it has not become an epidemic). For SIR epidemics in the finite populations we use for simulations, the plateaus eventually give way to extinction because eventually there are not enough remaining susceptibles.

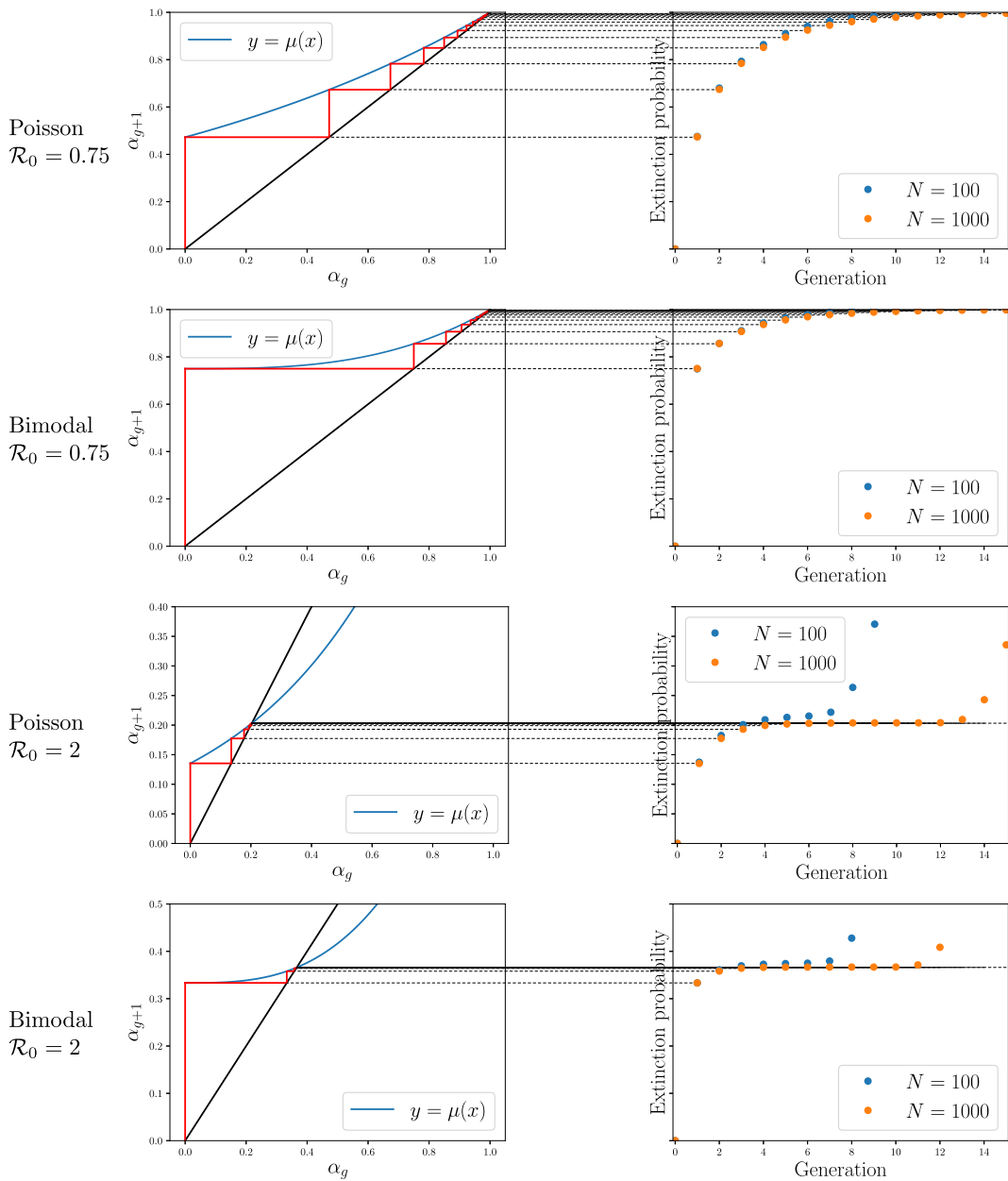


Fig. 4. Illustration of Theorem 2.2. Left: Cobweb diagrams showing convergence of iterations to the predicted outbreak extinction probability (see Fig. A.10). **Right:** Observed probabilities of no infections remaining after each generation for simulations of Fig. 2 showing the probability of extinction by generation g . Thin lines show the relation between the cobweb diagram and the extinction probabilities. The simulated probability initially rises quickly, representing outbreaks that die out early on, then it remains steady at a level representing the probability of outbreaks dying out while small. For $\mathcal{R}_0 > 1$ it increases again because the epidemics burn through the finite population (and so the infinite population theory breaks down). The values match the corresponding iteration of the cobweb diagrams.

2.2. Early-time outbreak dynamics

We now explore the number of active infections present in generation g . We will need to use $i = \sqrt{-1}$ in this subsection, so to avoid confusion we use ℓ as our indexing variable rather than i . Setting $\varphi_\ell(g)$ to be the probability ℓ active infections exist at generation g , we define the PGF $\Phi_g(y) = \sum_\ell \varphi_\ell(g) y^\ell$. Assuming at generation 0 there is a single infection ($\varphi_1(0) = 1$) then the initial condition is $\Phi_0(y) = y$. From inductive application of Property A.8 for composition of PGFs (exercise 2.7) it is straightforward to conclude that for $g > 0$, $\Phi_g(y) = \mu^{|g|}(y)$ where $\mu(y)$ is the PGF for the offspring distribution.

Theorem 2.3. Assuming that each infected individual produces an independent number of offspring chosen from a distribution with PGF $\mu(y)$, the number infected in the g -th generation has PGF

$$\Phi_g(y) = \sum_{\ell} \varphi_{\ell}(g)y^{\ell} = \mu^{[g]}(y) \tag{6}$$

where $\varphi_{\ell}(g)$ is the probability there are ℓ active infections in generation g . This does not provide information about the cumulative number infected.

It is worth highlighting that for general distributions, the calculation of coefficients of $\Phi_g(y)$ may seem quite challenging. Luckily, it is not so difficult. Property A.3 states (taking $i = \sqrt{-1}$)

$$\varphi_{\ell}(g) \approx \frac{1}{M} \sum_{m=1}^M \frac{\Phi_g(\text{Re}^{2\pi im/M})}{R^{\ell} e^{2\ell\pi im/M}}$$

for large M and any $R \leq 1$. For each $y_m = \text{Re}^{2\pi im/M}$ we can calculate $\Phi_g(y_m) = \mu^{[g]}(y_m)$ by numerically iterating μ g times. Then for large enough M , this gives a remarkably accurate and efficient approximation to the individual coefficients.

Example 2.5. We demonstrate Theorem 2.3 in Fig. 5, using the simulations from Example 2.2. Simulations and predictions are in excellent agreement.

There is a mismatch noticeable for the bimodal distribution with $\mathcal{R}_0 = 2$ particularly with $N = 100$, which is a consequence of the fact that the population is finite. In stochastic simulations, occasionally an individual receives multiple transmissions even early in the outbreak, but in the PGF theory this does not happen. We are often interested in the expected number of active infections in generation g , $\sum_{\ell} \ell \varphi_{\ell}(g)$ (however, as seen below this is not the most relevant measure to use if $\mathcal{R}_0 > 1$). Property A.5 shows that this is given by $\frac{\partial}{\partial y} \Phi_g(y)|_{y=1}$. To calculate this we use $\Phi_g(1) = 1$ for all g (Property A.4) and $\mu'(1) = \mathcal{R}_0$. Then through induction and the chain rule we show that $\frac{\partial}{\partial y} \Phi_g(y)|_{y=1} = \mathcal{R}_0^g$:

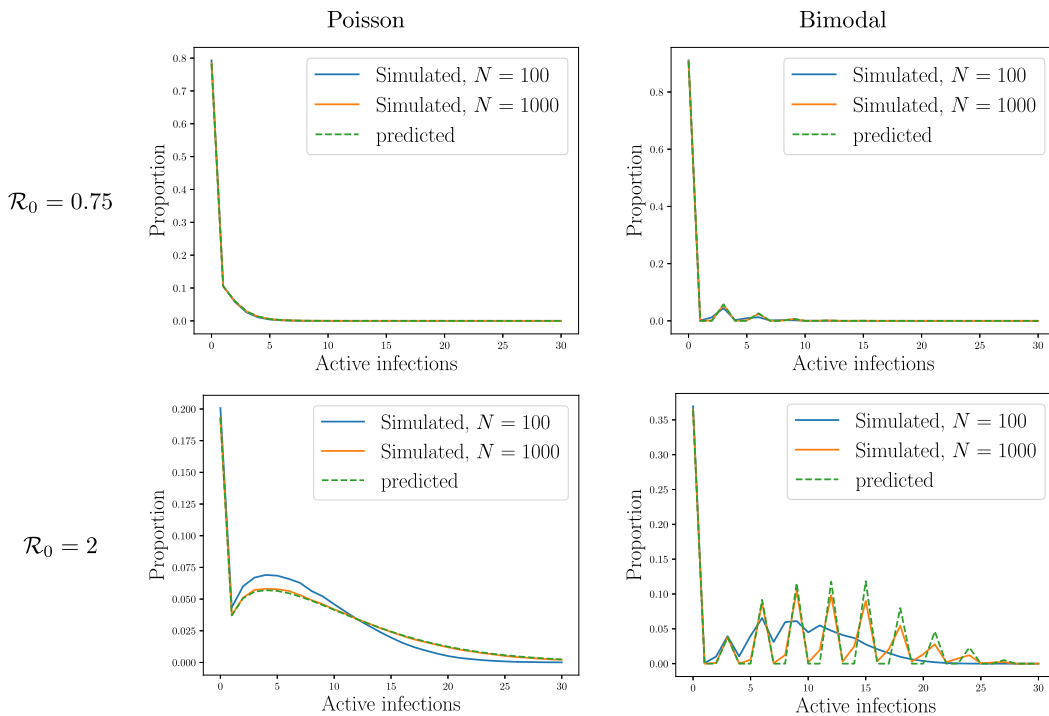


Fig. 5. Illustration of Theorem 2.3. Comparison of predictions and the simulations from Fig. 2 for the number of active infections in the third generation. The bimodal case with $N = 100$ shows a clear impact of population size as a sizable number of transmissions fail because the population is finite. The predictions were made numerically using the summation in Property A.3.

$$\begin{aligned} \frac{\partial}{\partial y} \Phi_g(y)|_{y=1} &= \frac{\partial}{\partial y} \mu(\Phi_{g-1}(y))|_{y=1} \\ &= \left(\mu'(\Phi_{g-1}(y)) \times \frac{\partial}{\partial y} \Phi_{g-1}(y) \right) |_{y=1} \\ &= \mu'(1) \times \mathcal{R}_0^{g-1} \\ &= \mathcal{R}_0^g. \end{aligned}$$

we initialized the induction with the case $g = 1$ which is the definition of \mathcal{R}_0 . If $\mathcal{R}_0 < 1$, this shows that we expect decay.

If $\mathcal{R}_0 > 1$, there is a more relevant measure. On average we see growth, but a sizable fraction of outbreaks may go extinct, and these zeros are included in the average, which alters our prediction. This is closely related to the “push of the past” effect observed in phylodynamics (Nee, Holmes, May, & Harvey, 1994). For policy purposes, we are more interested in the expected size if the outbreak is not yet extinct because a response that is scaled to deal with the average size (where the average includes those that are extinct) is either too big (if the disease has gone extinct) or too small (if the disease has become established) (Miller, Davoudi, Meza, Slim, & Pourbohloul, 2010). It is very unlikely to be just right. The expected number infected in generation g conditional on the outbreaks not dying out by generation g is $\mathcal{R}_0^g / (1 - \alpha_g)$. This has an important consequence. We can have different extinction probabilities for different offspring distributions with the same \mathcal{R}_0 . The disease with a higher extinction probability tends to have considerably more infections in those outbreaks that do not go extinct.

We have

Corollary 2.1. *In the infinite population limit, the expected number infected in generation g starting from a single infection is*

$$[I]_g = \mathcal{R}_0^g \tag{7}$$

and the expected number starting from a single infection conditional on the disease persisting to generation g is

$$\langle I \rangle_g = \frac{\mathcal{R}_0^g}{1 - \alpha_g} \tag{8}$$

We can explore higher moments of the distribution of the number infected by taking more derivatives of $\Phi_g(y)$ and evaluating at $y = 1$.

2.3. Cumulative size distribution

We now look at the total number infected while the outbreak is small. There are multiple ways to calculate how the cumulative size of small outbreaks is distributed. We look at two of these. The first focuses just on the number of completed infections by generation g . The second calculates the joint distribution of the number of completed infections and the number of active infections at generation g . Later we address the distribution of final sizes of small outbreaks.

2.3.1. Focused approach to find the cumulative size distribution

We begin by calculating just the number of completed infections at generation g . We define $\omega_j(g)$ to be the probability that there are j completed infections at generation g (by “completed” we only include individuals who are no longer infectious in generation g). We will use PGFs of the variable z when focusing on completed infections.

We define

$$\Omega_g(z) = \sum_j \omega_j(g) z^j$$

to be the PGF for the number of completed infections j at generation g . Although we use j to represent recoveries, this model is still appropriate for SIS disease because we are interested in small outbreak sizes in a well-mixed infinite population for which we can assume no previously infected individuals have been reexposed. If the outbreak begins with a single infection, then

$$\Omega_0(z) = 1 \quad \text{and} \quad \Omega_1(z) = z$$

showing that the first individual (infectious during generation 0) completes his infection at the start of generation 1. For generation 2 we have the initial individual and his direct offspring, so $\Omega_2(z) = z\mu(z)$.

More generally, to calculate for $g > 1$, the completed infections consist of

- the initial infection

- the active infections in generation 1.
- any descendants of those active infections in generation 1 that will have recovered by generation g .

The distribution of the number of descendants of a generation 1 individual (including that individual) who have recovered by generation g is given by $\Omega_{g-1}(z)$. That is each generation 1 individual and its descendants for the following $g - 1$ infections have the same distribution as an initial infection and its descendants after $g - 1$ generations.

From Property A.8 the number of descendants by generation g (not counting the initial infection) that have recovered is distributed like $\mu(\Omega_{g-1}(z))$. Accounting for the initial individual requires that we increment the count by 1 which requires increasing the exponent of z by 1. So we multiply by z . This yields

$$\Omega_g(z) = z\mu(\Omega_{g-1}(z))$$

To sustain an outbreak up to generation g there must be at least one infection in each generation from 0 to $g - 1$. So any outbreak with fewer than g completed infections at generation g must be extinct. So the coefficient of z^j does not change once $g > j$. Thus we have shown

Theorem 2.4. Assuming a single initial infection in an infinite population, the PGF $\Omega_g(z) = \sum_j \omega_j(g)z^j$ for the distribution of the number of completed infections at generation $g > 1$ is given by

$$\Omega_g(z) = z\mu(\Omega_{g-1}(z)) \tag{9}$$

with $\Omega_1(z) = z$. Once $g > j$, the coefficient $\omega_j(g)$ is constant.

Example 2.6. We test Theorem 2.4 in Fig. 6, using the simulations from Example 2.2. Simulations and predictions are in excellent agreement.

Example 2.7. Expected cumulative size It is instructive to calculate the expected number of completed infections at generation g . Note that $\Omega_g(z)|_{z=1} = 1$, $\mu(1) = 1$, and $\mu'(1) = \mathcal{R}_0$. We use induction to show that for $g \geq 1$ the expected number of completed infections is $\sum_{j=0}^{g-1} \mathcal{R}_0^j$:

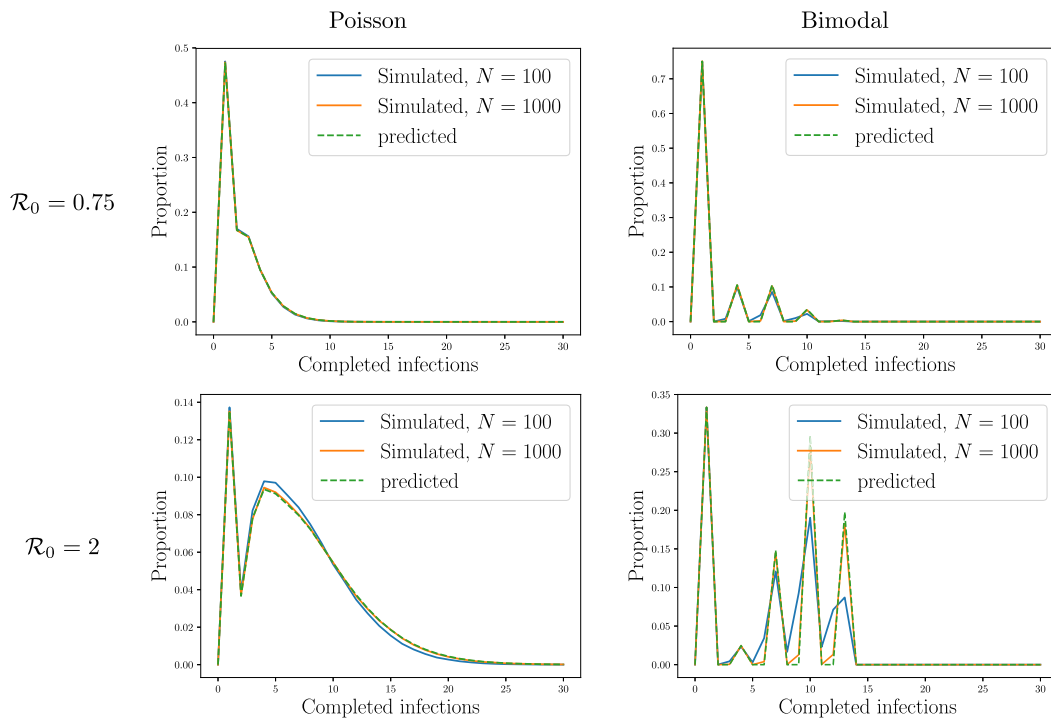


Fig. 6. Illustration of Theorem 2.4 Comparison of predictions with the simulations from Fig. 2 for the number of completed infections at the start of the third generation. The predictions were calculated using Property A.3.

$$\begin{aligned} \frac{\partial}{\partial z} \Omega_g(z)|_{z=1} &= \frac{\partial}{\partial z} z\mu(\Omega_{g-1}(z))|_{z=1} \\ &= \mu(\Omega_{g-1}(z)) + z\mu'(\Omega_{g-1}(z)) \frac{\partial}{\partial z} \Omega_{g-1}(z)|_{z=1} \\ &= \mu(1) + \mu'(1) \sum_{j=0}^{g-2} \mathcal{R}_0^j \\ &= 1 + \mathcal{R}_0 \sum_{j=0}^{g-2} \mathcal{R}_0^j \\ &= \sum_{j=0}^{g-1} \mathcal{R}_0^j \end{aligned}$$

This is in agreement with our earlier result that the expected number that are infected in generation j is \mathcal{R}_0^j . This is

$$\frac{\partial}{\partial z} \Omega'_g(z)|_{z=1} = \begin{cases} \frac{1 - \mathcal{R}_0^g}{1 - \mathcal{R}_0} & \mathcal{R}_0 \neq 1 \\ g & \mathcal{R}_0 = 1 \end{cases}$$

As with our previous results, the sum shows a threshold behavior at $\mathcal{R}_0 = 1$. If $\mathcal{R}_0 < 1$, then in the limit $g \rightarrow \infty$, the expected cumulative outbreak size converges to the finite value $1/(1 - \mathcal{R}_0)$. If $\mathcal{R}_0 \geq 1$, it diverges.

This example shows

Corollary 2.2. In the infinite population limit the expected number of completed infections at the start of generation g assuming a single randomly chosen initial infection is

$$\frac{\partial}{\partial z} \Omega'_g(z)|_{z=1} = \begin{cases} \frac{1 - \mathcal{R}_0^g}{1 - \mathcal{R}_0} & \mathcal{R}_0 \neq 1 \\ g & \mathcal{R}_0 = 1 \end{cases} \tag{10a}$$

For $\mathcal{R}_0 \geq 1$ this diverges as $g \rightarrow \infty$. Otherwise it converges to $1/(1 - \mathcal{R}_0)$.

2.3.2. Broader approach

An alternate approach calculates both the current and cumulative size at generation g . We let $\pi_{i,r}(g)$ be the probability that there are i actively infected individuals and r completed infections in generation g . We define $\Pi_g(y, z) = \sum_{i,r} \pi_{i,r}(g) y^i z^r$, so y represents the active infections and z the completed infections.

Assume we know the values i_{g-1} and r_{g-1} for generation $g - 1$. Then r_g is simply $i_{g-1} + r_{g-1}$ and i_g is distributed according to $\mu(y)^{i_{g-1}}$. So given those known i_{g-1} and r_{g-1} , the distribution for the next generation would be $[z\mu(y)]^{i_{g-1}} z^{r_{g-1}}$. Summing over all possible i_{g-1} and r_{g-1} yields

$$\begin{aligned} \Pi_g(y, z) &= \sum_{i,r} \pi_{i,r}(g-1) [z\mu(y)]^i z^r \\ &= \Pi_{g-1}(z\mu(y), z) \end{aligned}$$

with the initial condition

$$\Pi_0(y, z) = y$$

The first few iterations are

$$\begin{aligned} \Pi_1(y, z) &= z\mu(y) \\ \Pi_2(y, z) &= z\mu(z\mu(y)) \end{aligned}$$

and we can use induction on this to show that in general

$$\Pi_g(y, z) = z\mu(\Pi_{g-1}(y, z))$$

Theorem 2.5. Given a single initial infection in an infinite population, the PGF $\Pi_g(y, z) = \sum_{i,r} \pi_{i,r}(g) y^i z^r$ for the joint distribution of the number of active i and completed infections r in generation g is given by

$$\Pi_g(y, z) = z\mu(\Pi_{g-1}(y, z)) \tag{11}$$

with $\Pi_0(y, z) = y$.

Example 2.8. We demonstrate Theorem 2.5 in Fig. 7, using the same simulations as in Example 2.2. Simulations and predictions are in excellent agreement.

2.4. Small outbreak final size distribution

There are many diseases for which there have been multiple small outbreaks in recent years but no large-scale epidemics (such as Nipah, H5N1 avian influenza, Pneumonic Plague, Monkey pox, and — prior to 2013 — Ebola). A natural question emerges: what can we infer about the epidemic potential of these diseases? The size distribution may help us to infer properties of the disease and in particular to estimate the probability that $\mathcal{R}_0 > 1$ (Blumberg & Lloyd-Smith, 2013; Kucharski & Edmunds, 2015; Nishiura, Yan, Sleeman, & Mode, 2012).

We have found that $\Omega_g(z)$ gives the PGF for the number of completed infections by generation g . We noted earlier that for a given r , once $g > r$, the coefficient of z^r in $\Omega_g(z)$ is fixed and equal to the probability that the outbreak goes extinct after exactly r infections. Motivated by this, we look for the limit as $g \rightarrow \infty$. We define

$$\Omega_\infty(z) = \lim_{g \rightarrow \infty} \Omega_g(z)$$

We expect this to be the PGF for the final size of the outbreaks.

We can express the pointwise limit¹ as

$$\Omega_\infty(z) = \sum_r \omega_r z^r + \omega_\infty z^\infty$$

where for $r < \infty$ the coefficient ω_r is the probability an outbreak causes exactly r infections in an infinite population. We use ω_∞ to denote the probability that the outbreak is infinite in an infinite population (i.e., that it is an epidemic), and we interpret z^∞ as 1 when $z = 1$ and 0 for $0 \leq z < 1$. So if epidemics are possible, $\Omega_\infty(z)$ has a discontinuity at $z = 1$, and the limit as $z \rightarrow 1$ from below gives $\sum_{r < \infty} \omega_r = 1 - \omega_\infty$ which is the extinction probability α .

We now look for a recurrence relation for $\Omega_\infty(z)$ in the infinite population limit. Each offspring of the initial infection independently causes a set of infections. The distribution of these new infections (including the original offspring) also has PGF $\Omega(z)$. So the distribution of the number of descendants of the initial infection (but not including the initial infection) has PGF $\mu(\Omega_\infty(z))$. To include the initial infection, we must increase the exponent of z by one, which we do by multiplying by z . We conclude that $\Omega_\infty(z) = z\mu(\Omega_\infty(z))$. Although we have shown that $\Omega_\infty(z)$ solves $f(z) = z\mu(f(z))$, we have not shown that there is only one function that solves this.

We may be interested in the outbreak size distribution conditional on the outbreak going extinct. For this we are looking at $\Omega_\infty(z)/\alpha$ for any $z < 1$, and at $z = 1$, this is simply 1. Note that if $\mathcal{R}_0 < 1$ then $\alpha = 1$.

Summarizing this we have

Theorem 2.6. Given a single initial infection in an infinite population, consider $\Omega_\infty(z)$, the PGF for the final size distribution: $\Omega_\infty(z) = (\sum_{r < \infty} \omega_r z^r) + \omega_\infty z^\infty$ where $z^\infty = 0$ if $|z| < 1$ and 1 if $|z| = 1$.

- Then

¹ Although this converges for any given z in $[0, 1]$, it does not do so “uniformly” if $\mathcal{R}_0 > 1$. That is, for $\mathcal{R}_0 > 1$ no matter how large g is, there are always some values of $z < 1$, but sufficiently close to 1, which are far from converged.

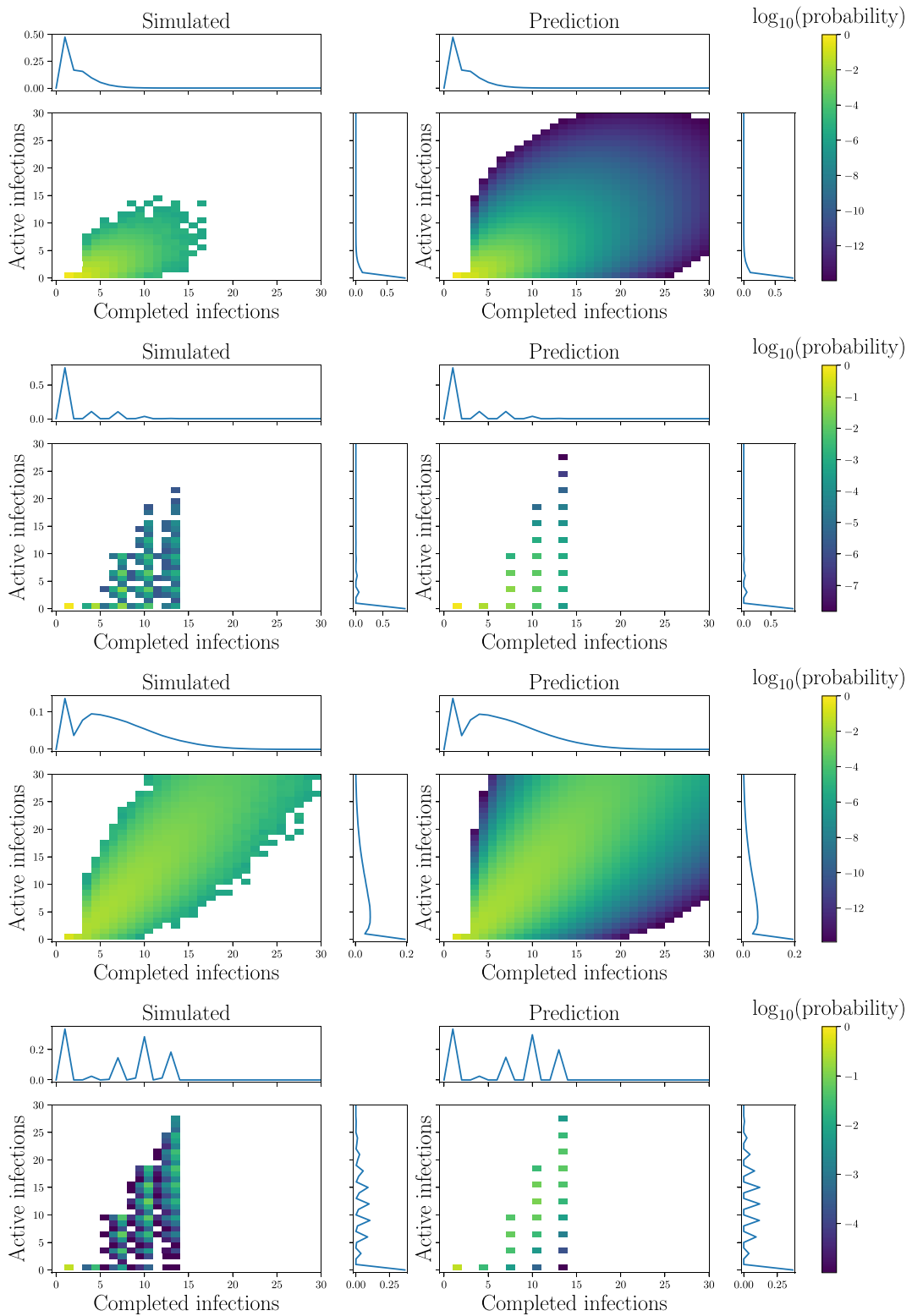


Fig. 7. Illustration of Theorem 2.5. Comparison of predictions and simulations for the joint distribution of the number of current and completed infections at generation $g = 3$. The predictions were calculated using Property A.3. **Left:** simulations from Fig. 2 for $N = 1000$ and **Right:** predictions (note vertical scales on left and right are the same). **Top to Bottom:** Poisson $\mathcal{R}_0 = 0.75$, Bimodal $\mathcal{R}_0 = 0.75$, Poisson $\mathcal{R}_0 = 2$, and Bimodal $\mathcal{R}_0 = 2$. The predictions match our observations, with some difference for two reasons: 1) because 5×10^5 simulations cannot resolve events with probabilities as small as 10^{-12} , but the PGF approach can, and 2) due to finite-size effects as occasionally an individual receives multiple transmissions even early on. The plots also show the marginal distributions, matching Figs. 5 and 6.

$$\Omega_\infty(z) = \begin{cases} z\mu(\Omega_\infty(z)) & z \neq 1 \\ 1 & z = 1 \end{cases} \tag{12}$$

- We have $\lim_{z \rightarrow 1^-} \Omega_\infty(z) = \alpha = 1 - \omega_\infty$. If $\mathcal{R}_0 > 1$ then $\Omega_\infty(z)$ is discontinuous at $z = 1$, with a jump discontinuity of ω_∞ , the probability of an epidemic.
- The PGF for outbreak size distribution conditional on the outbreak being finite is

$$\begin{cases} \Omega_\infty(z)/\alpha & 0 < z < 1 \\ 1 & z = 1 \end{cases}$$

Perhaps surprisingly we can often find the coefficients of $\Omega_\infty(z)$ analytically if $\mu(y)$ is known. We use a remarkable result showing that the probability of infecting exactly n individuals is equal to the coefficient of z^{n-1} in $[\mu(z)]^n$ (Blumberg & Lloyd-Smith, 2013; Dwass, 1969; van der Hofstad & Keane, 2008; Wendel, 1975). The theorem is

Theorem 2.7. Given an offspring distribution with PGF $\mu(y)$, for $j < \infty$ the coefficient of z^j in $\Omega_\infty(z)$ is $\frac{1}{j} p_{j-1}^{(j)}$ where $p_i^{(j)}$ is defined by $[\mu(y)]^j = \sum_i p_i^{(j)} y^i$.

That is, for $j < \infty$ the probability of having exactly j infections in an outbreak starting from a single infection is $\frac{1}{j}$ times the coefficient of y^{j-1} in $[\mu(y)]^j$.

We prove this theorem in Appendix B. The proof is based on observing that if we draw a sequence of j numbers from the offspring distribution, the probability they sum to $j - 1$ (corresponding to $j - 1$ transmissions and hence j infected individuals including the index case) is the coefficient of z^{j-1} in $[\mu(z)]^j$. A fraction $1/j$ of these satisfy additional constraints needed to correspond to a valid transmission tree² and thus the probability of a valid transmission tree with exactly $j - 1$ transmissions is $1/j$ times $p_{j-1}^{(j)}$.

Because the coefficient of y^{j-1} in $[\mu(y)]^j$ is $\frac{1}{(j-1)!} \left(\frac{d}{dy}\right)^{j-1} [\mu(y)]^j|_{y=0}$ (by Property A.2), we have that the probability of an outbreak of size j is

$$\frac{1}{j!} \left(\frac{d}{dy}\right)^{j-1} [\mu(y)]^j|_{y=0}$$

It is enticing to think there may be a similar theorem for coefficients of $\Pi(y, z)$, but we are not aware of one. The theorem has been generalized to models having multiple types of individuals (Kucharski & Edmunds, 2015).

Example 2.9. We demonstrate Theorems 2.6 and 2.7 in Fig. 8, using the same simulations as in Example 2.2.

Example 2.10. The PGF for the negative binomial distribution with parameters p and \hat{r} (with $q = 1 - p$) is

$$\mu(y) = \left(\frac{q}{1 - py}\right)^{\hat{r}}$$

We can rewrite this as

$$\mu(y) = q^{\hat{r}}(1 - py)^{-\hat{r}}$$

We will use this to find the final size distribution. We expand $[\mu(y)]^j = q^{\hat{r}j}(1 - py)^{-\hat{r}j}$ using the binomial series

$$(1 + \delta)^\eta = 1 + \eta\delta + \frac{\eta(\eta - 1)}{2!}\delta^2 + \dots + \frac{\eta(\eta - 1)\dots(\eta - i + 1)}{i!}\delta^i + \dots$$

which holds for integer or non-integer η . Then with $-py$, $-\hat{r}j$, and $j - 1$ playing the role of δ , η , and i :

² If the index case causes 0 infections and its first offspring causes 1 infection, we have a sequence of two numbers that sum to 1, but it is biologically meaningless because it does not make sense to talk about the first offspring of an individual who causes no infections.

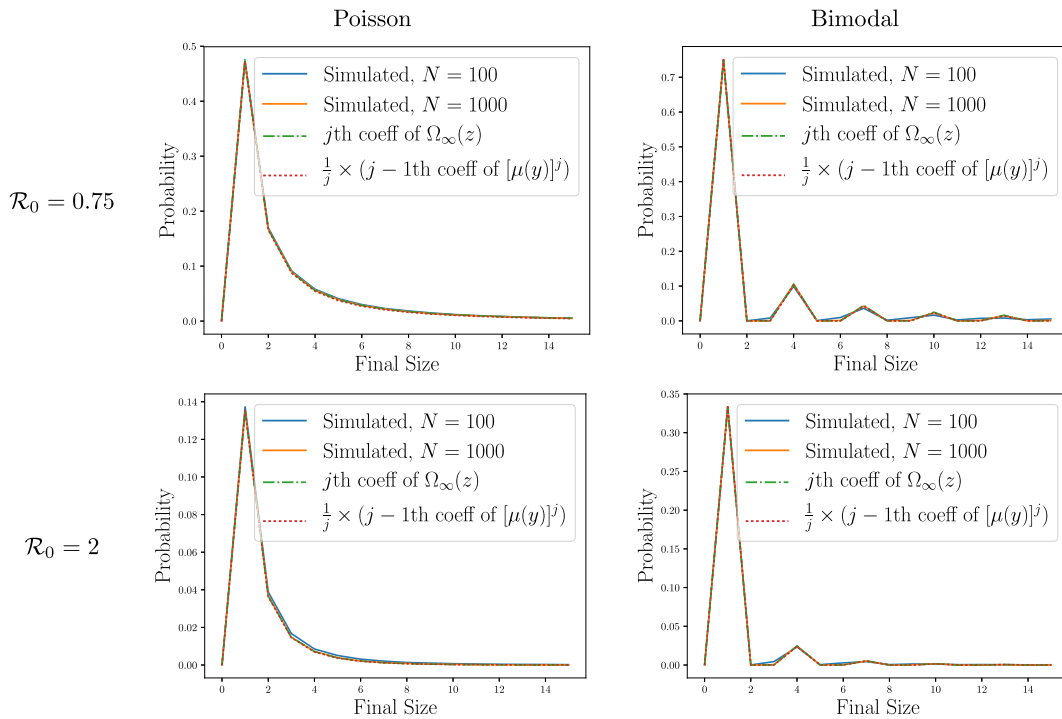


Fig. 8. Illustration of Theorems 2.6 and 2.7. The final size of small outbreaks predicted by Theorem 2.6 and by Theorem 2.7 as calculated using Property A.3 matches observations from the simulations in Fig. 2 (see also insets of Fig. 2).

$$\begin{aligned}
 [\mu(y)]^j &= q^{\hat{r}j} (1 - py)^{-\hat{r}j} \\
 &= q^{\hat{r}j} \left(1 + \hat{r}jpy + \frac{\hat{r}j(\hat{r}j + 1)}{2!} p^2 y^2 + \dots + \frac{\hat{r}j(\hat{r}j + 1) \cdots (\hat{r}j + j - 2)}{(j - 1)!} p^{j-1} y^{j-1} + \dots \right)
 \end{aligned}$$

[the negatives all cancel]. So the coefficient of y^{j-1} is $q^{\hat{r}j} p^{j-1} \frac{(\hat{r}j + j - 2)!}{(\hat{r}j - 1)!(j - 1)!} = \binom{\hat{r}j + j - 2}{j - 1} q^{\hat{r}j} p^{j-1}$ (assuming \hat{r} is an integer). Looking at $1/j$ times this, we conclude that the probability an outbreak infects exactly j individuals is

$$\frac{1}{j} \binom{\hat{r}j + j - 2}{j - 1} q^{\hat{r}j} p^{j-1}$$

A variation of this result for non-integer \hat{r} is commonly used in work estimating disease parameters (Blumberg & Lloyd-Smith, 2013; Nishiura et al., 2012). Exercise 2.12 generalizes the formula for this.

Applying Theorem 2.7 to several different families of distributions yields Table 6 for the probability of a final size j .

2.4.1. Inference based on outbreak sizes

A major challenge in infectious disease modeling is inferring the parameters of an infectious disease. In Section 2.4 we alluded to the use of PGFs to infer disease properties from observations of the size distribution of small outbreaks. In this section we describe how to do this using a Bayesian approach, using the probabilities given in Table 6. A number of researchers have used this approach to estimate disease parameters (Blumberg & Lloyd-Smith, 2013; Kucharski & Edmunds, 2015; Nishiura et al., 2012).

We assume that we know what type of distribution the offspring distribution, but that there are some unknown parameters (often it is assumed to be a negative binomial distribution). We also assume that we have some prior belief about the probability of various parameters. For practical purposes, we will assume that we have some finite number of possible parameter values, each with a probability.

We use Bayes' Theorem (Hoff, 2009):

$$P(\Theta|X) = \frac{P(\Theta, X)}{P(X)} = \frac{P(X|\Theta)P(\Theta)}{P(X)} \quad (13)$$

Here we think of Θ as the specific parameter values and X as the observed data (typically the observed size of an outbreak or sizes of multiple independent outbreaks, in which case $P(X|\Theta)$ comes from [Theorem 2.7](#) or [Table 6](#)). In our calculations we can simply use the fact that $P(\Theta|X) \propto P(X|\Theta)P(\Theta)$ with a normalization constant which can be dealt with at the end.

The *prior* for Θ is the probability distribution we assume for the parameter values before observing the data, given by $P(\Theta)$. We often simply assume that all parameter values are equally probable initially.

The *likelihood* of the parameters Θ is defined to be $P(X|\Theta)$, the probability that we would observe X for the given parameter values. If we are choosing between two sets of parameter values Θ_1 and Θ_2 and the observations have consistently higher likelihood for Θ_2 , then we intuitively expect that Θ_2 is the more probable parameter value.

In practice the likelihood may be very small which can lead to numerical error. It is often useful to instead look at *log-likelihood*,³ $\log P(X|\Theta)$. For example, if we have many observed outbreak sizes, the likelihood $P(X|\Theta)$ under independence is the product of the probabilities of each individual outbreak size. The likelihood is thus quite small (perhaps less than machine precision), while the log-likelihood is simply the sum of the log-likelihoods of each individual observation.

We know that

$$\log P(\Theta|X) - C = \log P(X|\Theta) + \log P(\Theta)$$

where C is the logarithm of the proportionality constant $1/P(X)$ in Equation (13). If we have a prior and the likelihood, the right hand side can be calculated. It is often possible (and advisable) to calculate the log likelihood $\log P(X|\Theta)$ directly rather than calculating $P(X|\Theta)$ and then taking the logarithm.

Exponentiating the right hand side and then finding the appropriate normalization constant will yield $P(\Theta|X)$. Numerically the numbers may be very small when we exponentiate, so to prior to exponentiating it is advisable to add a constant value to all of the expressions. This constant is corrected for in the final normalization step.

We now provide the steps for a numerical calculation of $P(\Theta|X)$ given the prior $P(\Theta)$, the observations X , and the log likelihood $\log P(X|\Theta)$.

1. For each Θ , calculate $f(\Theta) = \log P(X|\Theta) + \log P(\Theta)$.
2. Find the maximum X_{\max} over all Θ and subtract it to yield $\hat{f}(\Theta) = \log P(X|\Theta) + \log P(\Theta) - X_{\max}$. Note that $X_{\max} \leq 0$, and this brings all of our numbers closer to zero.
3. Calculate $g(\Theta) = e^{\hat{f}(\Theta)}$. This will be proportional to $P(\Theta|X)$. Note that by using $e^{\hat{f}(\Theta)}$ rather than $e^{f(\Theta)}$ we have reduced the impact of roundoff error.
4. Find the normalization constant $\sum_{\Theta} g(\Theta')$. Then

$$P(\Theta|X) = \frac{g(\Theta)}{\sum_{\Theta} g(\Theta')}$$

Note that if Θ comes from a continuous distribution rather than a discrete distribution, then the same approach works, except that P is a probability density and the summation in the final step becomes an integral.

Example 2.11. A frequent assumption is that the offspring distribution is negative binomial. Let us make this assumption with unknown p and \hat{r} .

To artificially simplify the problem, we assume that we know that there are only two possible pairs of $\Theta = (p, \hat{r})$, namely $\Theta_1 = (p_1, \hat{r}_1) = (0.02, 40)$ or $\Theta_2 = (p_2, \hat{r}_2) = (0.03, 20)$, and that our a priori belief is that they are equally probable.

After observing 2 independent outbreaks, with total sizes $j_1 = 8$ and $j_2 = 7$, we want to use our observations to update $P(\Theta)$. From [Table 6](#), the likelihood of a given Θ given the two independent observations is

³ Throughout this section, we assume that log is taken with base e .

$$\begin{aligned}
 f(\Theta) &= \left(\log \prod_{j=7,8} \frac{1}{j} \binom{\hat{r}j + j - 2}{j - 1} q^{\hat{r}j} p^{j-1} \right) + \log 0.5 \\
 &= \left(\sum_{j=7,8} \log \frac{1}{j} \binom{\hat{r}j + j - 2}{j - 1} q^{\hat{r}j} p^{j-1} \right) + \log 0.5 \\
 &= \left(\sum_{j=7,8} \log((\hat{r}j + j - 2)!) - \log(j!) - \log((\hat{r}j - 1)!) + \hat{r}j \log q + (j - 1) \log p \right) + \log 0.5
 \end{aligned}$$

In problems like this, we will often encounter logarithms of factorials. Many programming languages provide this, typically using Stirling’s approximation. For example, Python, R, and C++ all have a special function `lgamma` which calculates the natural log of the absolute value of the gamma function.⁴ We find

$$\begin{aligned}
 f(\Theta_1) &\approx -8.495 \\
 f(\Theta_2) &\approx -9.135
 \end{aligned}$$

So $\hat{f}(\Theta_1) = 0$ and $\hat{f}(\Theta_2) \approx -0.640$. Exponentiating, we have

$$\begin{aligned}
 g(\Theta_1) &= 1 \\
 g(\Theta_2) &\approx 0.5277
 \end{aligned}$$

So now

$$\begin{aligned}
 p(\Theta_1|X) &\approx \frac{1}{1.5277} \approx 0.6546 \\
 p(\Theta_2|X) &\approx \frac{0.5277}{1.5277} \approx 0.3454
 \end{aligned}$$

So rather than the two parameter sets being equally probable, Θ_2 is now about half as likely as Θ_1 given the observed data.

2.5. Generality of discrete-time results

Thus far we have measured time in generations. However, many models measure time differently and different generations may overlap. For both SIS and SIR disease, our results above about final size distribution or extinction probability still apply. To see this, we note first that our results have been derived assuming that the population is infinite and well-mixed so no individuals receive multiple transmissions. Regardless of the clock time associated with transmission and recovery, there is still a clear definition of the length of the transmission chain to an infected individual. Once we group individuals by length of the transmission chain, we get the generation-based model used above. This equivalence is studied more in (Ludwig, 1975; Yan, 2008).

2.6. Exercises

Exercise 2.1. Monotonicity of α_g

- By considering the biological interpretation of α_g , explain why the sequence of inequalities $0 = \alpha_0 \leq \alpha_1 \leq \dots \leq 1$ should hold. That is, explain why $\alpha_0 = 0$, why the α_i form a monotonically increasing sequence, and why all of them are at most 1.
- Show that α_g therefore converges to some non-negative limit α that is at most 1 and that $\alpha = \mu(\alpha)$.
- Use Property A.9 to show that if $\mu(0) \neq 0$ there exists a unique $\alpha < 1$ solving $\alpha = \mu(\alpha)$ if and only if $\mathcal{R}_0 = \mu'(1) > 1$.
- Assuming $\mu(0) \neq 0$, use Property A.9 to show that if $\mathcal{R}_0 > 1$ then α_g converges to the unique $\alpha < 1$ solving $\alpha = \mu(\alpha)$, and otherwise α_g converges to 1.

Exercise 2.2. Use Theorem 2.2 to prove Theorem 2.1.

⁴ The Gamma function is an analytic function that satisfies $\Gamma(n) = (n - 1)!$ for positive integer values so to calculate $\log(n!)$ we use `lgamma(n + 1)`.

Exercise 2.3. Show that if $\mu(0) = 0$, then $\lim_{g \rightarrow \infty} \alpha_g = 0$. By referring to the biological interpretation of $\mu(0) = 0$, explain this result.

Exercise 2.4. Find all PGFs $\mu(y)$ with $\mathcal{R}_0 \leq 1$ and $\mu(0) = 0$. Why were these excluded from [Theorem 2.2](#)?

Exercise 2.5. Larger initial conditions

Assume that disease is introduced with m infections rather than just 1, or that it is not observed by surveillance until m infections are present. Assume that the offspring distribution PGF is $\mu(y)$.

- If m is known, find the extinction probability.
- If m is unknown but its distribution has PGF $h(y)$, find the extinction probability.

Exercise 2.6. Extinction probability

Consider a disease in which $p_0 = 0.1$, $p_1 = 0.2$, $p_2 = 0.65$, and $p_3 = 0.05$ with a single introduced infection.

- Numerically approximate the probability of extinction within 0, 1, 2, 3, 4, or 5 generations up to five significant digits (assuming an infinite population).
- Numerically approximate the probability of eventual extinction up to five significant digits (assuming an infinite population).
- A surveillance program is being introduced, and detection will lead to a response. But it will not be soon enough to affect the transmissions from generations 0 and 1. From then on $p_0 = 0.3$, $p_1 = 0.4$, $p_2 = 0.3$, and $p_3 = 0$. Numerically approximate the new probability of eventual extinction after an introduction in an unbounded population [be careful that you do the function composition in the right order – review [Properties A.1](#) and [A.8](#)].

Exercise 2.7. We look at two inductive derivations of $\Phi_g(y) = \mu^{[g]}(y)$. They are similar, but when adapted to the continuous-time dynamics we study later, they lead to two different models. We take as given that $\Phi_{g-1}(y)$ gives the distribution of the number of infections caused after $g - 1$ generations starting from a single case. One argument is based on discussing the results of outcomes attributable to the infectious individuals of generation $g - 1$ in the next generation. The other is based on the outcomes indirectly attributable to the infectious individuals of generation 1 through their descendants after another $g - 1$ generations.

- Explain why [Property A.8](#) shows that $\Phi_g(y) = \Phi_{g-1}(\mu(y))$.
- (without reference to **a**) Explain why [Property A.8](#) shows that $\Phi_g(y) = \mu(\Phi_{g-1}(y))$.

Exercise 2.8. Use [Theorem 2.3](#) to prove the first part of [Theorem 2.2](#).

Exercise 2.9. How does [Corollary 2.1](#) change if we start with k infections?

Exercise 2.10. Assume the PGF of the offspring size distribution is $\mu(y) = (1 + y + y^2)/3$.

- What offspring size distribution yields this PGF?
- Find the PGF $\Omega_g(z)$ for the number of completed infections at 0, 1, 2, 3, and 4 generations [it may be helpful to use a symbolic math program once $g > 2$].
- Check that for these cases, once $g > r$, the coefficient of z^r does not change.

Exercise 2.11. By setting $y = 1$, use [Theorem 2.5](#) to prove [Theorem 2.4](#).

Exercise 2.12. Redo [Example 2.10](#) if \hat{r} is a real number, rather than an integer. It may be useful to use the Γ -function, which satisfies $\Gamma(x + 1) = x\Gamma(x)$ for any x and $\Gamma(n + 1) = n!$ for integer n .

Exercise 2.13. Except for the negative binomial case done in [Example 2.10](#), derive the probabilities in [Table 6](#).

- For the Poisson distribution, use [Property A.2](#).
- For the Uniform distribution, use [Property A.2](#).
- For the Binomial distribution, use the binomial theorem: $(a + b)^c = \sum_{i=0}^c \binom{c}{i} a^i b^{c-i}$.
- For the Geometric distribution, follow [Example 2.10](#) (noting that p and q interchange roles).

Exercise 2.14. To help model continuous-time epidemics, [Section 3](#) will use a modified version of μ , which in some contexts will be written as $\hat{\mu}(y, z)$. To help motivate the use of two variables, we reconsider the discrete case. We think of a recovery as an infected individual disappearing and giving birth to a recovered individual and a collection of infected individuals. Look back at the discrete-time calculation of Ω_g and Π_g . Define a two-variable version of μ as $\mu(y, z) = z \sum_i r_i y^i = z\mu(y)$.

- a. What is the biological interpretation of $\mu(y, z) = z\mu(y)$?
- b. Rewrite the recursive relations for Ω_g using $\mu(y, z)$ rather than $\mu(y)$.
- c. Rewrite the recursive relations for Π_g using $\mu(y, z)$ rather than $\mu(y)$.

The choice to use $\mu(y, z)$ versus $\mu(y)$ is purely a matter of convenience.

Exercise 2.15. Consider Example 2.11. Assume that a third outbreak is observed with 4 infections. Calculate the probability of Θ_1 and Θ_2 given the data starting

- a. with the assumption that $P(\Theta_1) = P(\Theta_2) = 0.5$ and X consists of the three observations $j = 7, j = 8,$ and $j = 4$.
- b. with the assumption that $P(\Theta_1) = 0.6546$ and $P(\Theta_2) = 0.3454$ and X consists only of the single observation $j = 4$.
- c. Compare the results and explain why they should have the relation they do.

Exercise 2.16. Assume that we know a priori that the offspring distribution for a disease has a negative binomial distribution with $p = 0.02$. Assume that our a priori knowledge of \hat{r} is that it is an integer uniformly distributed between 1 and 80 inclusive. Given observed outbreaks of sizes 1, 4, 5, 6, and 10:

- a. For each \hat{r} , calculate $P(\hat{r}|X)$ where X is the observed outbreak sizes. Plot the result.
- b. Find the probability that $\mathcal{R}_0 = \mu'(1)$ is greater than 1.

3. Continuous-time spread of a simple disease

We now develop PGF-based approaches adapting the results above to continuous-time processes. In the continuous-time framework, generations will overlap, so we need a new approach if we want to answer questions about the probability of being in a particular state at time t rather than at generation g . Questions about the final state of the population can be answered using the same techniques as for the discrete case, but the techniques introduced here also apply and yield the same predictions. Unlike Section 2, we do not do a detailed comparison with simulation.

In the continuous-time model, infected individuals have a constant rate of recovery γ and a constant rate of transmission β . Then $\gamma/(\beta + \gamma)$ is the probability that the first event is a recovery, while $\beta/(\beta + \gamma)$ is the probability it is a transmission. If the event is a recovery, then the individual is removed from the infectious population. If the event is a transmission, then the individual is still available to transmit again, with the same rate. If the recipient of a transmission is susceptible, it becomes infectious.

Unlike the discrete-time case, we do not focus on the offspring distribution. Rather, we focus on the resulting number of infected individuals after an event. Early on we treat the process as if as if each infected individual were removed and replaced by either 2 or 0 new infections. Although this is not the true process (she either recovers or she creates one additional infection and remains present), it is equivalent as far as the number of infections at any early time is concerned. We focus on a PGF for the outcome of the next event.

We define $\hat{\mu}(y) = \sum_i \hat{p}_i y^i$ and so

$$\hat{\mu}(y) = \frac{\beta}{\beta + \gamma} y^2 + \frac{\gamma}{\beta + \gamma} \tag{14a}$$

When we are calculating the number of completed cases, it will be useful to have a two-variable version of $\hat{\mu}$:

$$\hat{\mu}(y, z) = \frac{\beta}{\beta + \gamma} y^2 z + \frac{\gamma}{\beta + \gamma} z. \tag{14b}$$

Most of the results in this section are the continuous-time analog of the discrete-time results above for the infinite population limit. In the discrete-time approach we did not attempt to address outbreaks in finite populations. However, we end this section by deriving the equations for $\Xi(x, y, t)$, the PGF for the joint distribution of the number of susceptibles and active infections in a population of finite size N .

3.1. Extinction probability

For the extinction probability, we can apply the same methods derived in the discrete case to $\hat{\mu}(y)$. Thus we can find the extinction probability iteratively starting from the initial guess $\alpha_0 = 0$ and setting $\alpha_g = \hat{\mu}(\alpha_{g-1})$.

Exercises 3.1 and 3.2 each show that

Theorem 3.1. For the continuous-time Markovian model of disease spread in an infinite population, the probability of extinction given a single initial infection is

$$\alpha = \min(1, \gamma/\beta) \tag{15}$$

3.1.1. Extinction probability as a function of time

In the discrete-time case, we were interested in the probability of extinction after some number of generations. When we are using a continuous-time model, we are generally interested in “what is the probability of extinction by time t ?”

To answer this, we set $\alpha(t)$ to be the probability of extinction within time t . We will calculate the derivative of α at time t by using some mathematical sleight of hand to find $\alpha(t + \Delta t) - \alpha(t)$. Then dividing this by Δt and taking $\Delta t \rightarrow 0$ will give the result. Our approach is closely related to *backward Kolmogorov equations* (described later below).

We choose the time step Δt to be small enough that we can assume that at most one event happens between time 0 and Δt . The probabilities of having 0, 1, or 2 infections are $P(I(\Delta t) = 0) = \gamma\Delta t + \mathcal{O}(\Delta t)$, $P(I(\Delta t) = 1) = 1 - (\beta + \gamma)\Delta t + \mathcal{O}(\Delta t)$ and $P(I(\Delta t) = 2) = \beta\Delta t + \mathcal{O}(\Delta t)$ where the \mathcal{O} notation means that the error goes to zero fast enough that $\mathcal{O}(\Delta t)/\Delta t \rightarrow 0$ as $\Delta t \rightarrow 0$. The probability of having 3 or more infections (that is, multiple transmission events in the interval) is $\mathcal{O}(\Delta t)$ as well.

If there are two infected individuals at time Δt , then the probability of extinction by time $t + \Delta t$ is $\alpha(t)^2$. Similarly, if there is one infected at time Δt , the probability of extinction by time $t + \Delta t$ is $\alpha(t)$; and if there are no infections at time Δt , then the probability of extinction by time $t + \Delta t$ is $1 - \alpha(t)^0$. So up to $\mathcal{O}(\Delta t)$ we have

$$\begin{aligned} \alpha(t + \Delta t) &= \sum_{i=0}^{\infty} P(I(\Delta t) = i) \alpha(t)^i \\ &= [\gamma\Delta t]\alpha(t)^0 + [1 - (\beta + \gamma)\Delta t]\alpha(t) + [\beta\Delta t]\alpha(t)^2 + \mathcal{O}(\Delta t) \\ &= \alpha(t) + \Delta t(\beta + \gamma)[\hat{\mu}(\alpha(t)) - \alpha(t)] + \mathcal{O}(\Delta t) \end{aligned} \tag{16}$$

Thus

$$\dot{\alpha} = \lim_{\Delta t \rightarrow 0} [\alpha(t + \Delta t) - \alpha(t)]/\Delta t = (\beta + \gamma)[\hat{\mu}(\alpha) - \alpha]$$

and so

Theorem 3.2. *Given an infinite population with constant transmission rate β and recovery rate γ , then $\alpha(t)$, the probability of extinction by time t assuming a single initial infection at time 0 solves*

$$\dot{\alpha} = (\beta + \gamma)[\hat{\mu}(\alpha) - \alpha] \tag{17}$$

with $\hat{\mu}(y) = (\beta y^2 + \gamma)/(\beta + \gamma)$ and the initial condition $\alpha(0) = 0$.

We could solve this analytically (Exercise 3.4), but most results are easier to derive directly from the ODE formulation.

3.2. Early-time outbreak dynamics

We now explore the number of infections at time t . We define the PGF

$$\Phi(y, t) = \sum_i \varphi_i(t) y^i$$

where $\varphi_i(t)$ is the probability of i actively infected individuals at time t . We will derive equations for the evolution of $\Phi(y, t)$. We assume that $\Phi(y, 0) = y$ so a single infected individual exists at time 0.

Our goal is to derive equations telling us how Φ changes in time. We will use two approaches which were hinted at in Exercise 2.7, yielding two different partial differential equations. Although their appearance is different, for the appropriate initial condition, their solutions are the same. These equations are called the *forward* and *backward Kolmogorov equations*.

We briefly describe the analogy between the forward and backward Kolmogorov equations and Exercise 2.7:

- Our first approach finds the *forward Kolmogorov equations*. This is akin to Exercise 2.7 where we found $\Phi_g(y)$ by knowing the PGF $\Phi_{g-1}(y)$ for the number infected in generation $g - 1$ and recognizing that since the PGF for the number of infections each of them causes is $\mu(y)$, we must have $\Phi_g(y) = \Phi_{g-1}(\mu(y))$.
- Our second approach finds the *backward Kolmogorov equations* which are more subtle and can be derived similarly to how we derived the ODE for extinction probability in Theorem 3.2. This is akin to Exercise 2.7 where we found $\Phi_g(y)$ by knowing that the PGF for the number infected in generation 1 is $\mu(y)$, and recognizing that after another $g - 1$ generations each of those creates a number of infections whose PGF is $\Phi_{g-1}(y)$ and so $\Phi_g(y) = \mu(\Phi_{g-1}(y))$.

For both approaches, we make use of the observation that for $\Delta t \ll 1$, we can write the PGF for the number of infections resulting from a single infected individual at time $t = 0$ to be

$$\Phi(y, \Delta t) = y + (y^2 - y)\beta\Delta t + (1 - y)\gamma\Delta t + \mathcal{O}(\Delta t).$$

This says that with probability approximately $\beta\Delta t$ a transmission happens and we replace y by y^2 , and with probability approximately $\gamma\Delta t$ a recovery happens and we replace y by 1. With probability $\mathcal{O}(\Delta t)$ multiple events happen. We can rewrite this as

$$\Phi(y, \Delta t) = y + (\beta + \gamma)[\widehat{\mu}(y) - y]\Delta t + \mathcal{O}(\Delta t).$$

Note that $\Phi(y, 0) = y$ and $\frac{\partial}{\partial t}\Phi(y, 0) = (\beta + \gamma)[\widehat{\mu}(y) - y]$.

Both of our approaches rely on the observation that $\Phi(y, t_1 + t_2) = \Phi(\Phi(y, t_2), t_1)$ by Property A.8. This states that if we take the PGF at time t_1 , and then substitute for each y the PGF for the number of descendants of a single individual after t_2 units of time, the result is the PGF for the total number at time $t_1 + t_2$.

Forward equations. For this we use $\Phi(y, t_1 + t_2) = \Phi(\Phi(y, t_2), t_1)$ with t_2 playing the role of Δt and t_1 playing the role of t .

So $\Phi(y, t + \Delta t) = \Phi(\Phi(y, \Delta t), t)$. For small Δt (and taking $\Phi_y(\Phi(y, 0), t)$ to be the partial derivative of Φ with respect to its first argument), we have

$$\begin{aligned} \Phi(y, t + \Delta t) &= \Phi(\Phi(y, \Delta t), t) \\ &= \Phi(\Phi(y, 0), t) + (\Delta t)\Phi_y(\Phi(y, 0), t)\frac{\partial}{\partial t}\Phi(y, 0) + \mathcal{O}(\Delta t) \\ &= \Phi(y, t) + (\Delta t)(\beta + \gamma)[\widehat{\mu}(y) - y]\frac{\partial}{\partial y}\Phi(y, t) + \mathcal{O}(\Delta t). \end{aligned}$$

Then

$$\begin{aligned} \dot{\Phi}(y, t) &= \lim_{\Delta t \rightarrow 0} \frac{\Phi(y, t + \Delta t) - \Phi(y, t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Phi(y, t) + (\Delta t)(\beta + \gamma)[\widehat{\mu}(y) - y]\frac{\partial}{\partial y}\Phi(y, t) + \mathcal{O}(\Delta t) - \Phi(y, t)}{\Delta t} \\ &= (\beta + \gamma)[\widehat{\mu}(y) - y]\frac{\partial}{\partial y}\Phi(y, t). \end{aligned}$$

More generally, we can directly apply Property A.10 to get this result. [Exercise 3.6](#) provides an alternate direct derivation of these equations.

Backward equations. In the backward direction we have $\Phi(y, t_1 + t_2) = \Phi(\Phi(y, t_2), t_1)$ with t_2 playing the role of t and t_1 playing the role of Δt .

So $\Phi(y, t + \Delta t) = \Phi(y, \Delta t + t) = \Phi(\Phi(y, t), \Delta t)$. Note that because $\Phi(y, 0) = y$, we have $\Phi(\Phi(y, t), 0) = \Phi(y, t)$. Thus for small Δt , we expand Φ as a Taylor Series in its second argument t

$$\begin{aligned} \Phi(y, t + \Delta t) &= \Phi(\Phi(y, t), \Delta t) \\ &= \Phi(\Phi(y, t), 0) + (\Delta t)\Phi_t(\Phi(y, t), 0) + \mathcal{O}(\Delta t) \\ &= \Phi(y, t) + (\Delta t)\Phi_t(\Phi(y, t), 0) + \mathcal{O}(\Delta t) \\ &= \Phi(y, t) + (\Delta t)(\beta + \gamma)[\widehat{\mu}(\Phi(y, t)) - \Phi(y, t)] + \mathcal{O}(\Delta t). \end{aligned}$$

To avoid ambiguity, we use Φ_t above to denote the partial derivative of Φ with respect to its second argument t . So

$$\begin{aligned} \dot{\Phi}(y, t) &= \lim_{\Delta t \rightarrow 0} \frac{\Phi(y, t + \Delta t) - \Phi(y, t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Phi(y, t) + (\Delta t)(\beta + \gamma)[\widehat{\mu}(\Phi(y, t)) - \Phi(y, t)] + \mathcal{O}(\Delta t) - \Phi(y, t)}{\Delta t} \\ &= (\beta + \gamma)[\widehat{\mu}(\Phi(y, t)) - \Phi(y, t)]. \end{aligned}$$

This result also follows directly from Property A.12.

So we have

Theorem 3.3. The PGF $\Phi(y, t)$ for the distribution of the number of current infections at time t assuming a single introduced infection at time 0 solves

$$\frac{\partial}{\partial t} \Phi(y, t) = (\beta + \gamma)[\hat{\mu}(y) - y] \frac{\partial}{\partial y} \Phi(y, t) \quad (18)$$

as well as

$$\frac{\partial}{\partial t} \Phi(y, t) = (\beta + \gamma)[\hat{\mu}(\Phi(y, t)) - \Phi(y, t)]. \quad (19)$$

both with the initial condition $\Phi(y, 0) = y$.

It is perhaps remarkable that such seemingly different equations yield the same solution for the given initial condition.

Example 3.1. The expected number of infections in the infinite population limit is given by $[I] = \sum_i i p_i(t) = \frac{\partial}{\partial y} \Phi(1, t)$. From this we have

$$\begin{aligned} \frac{d}{dt} [I] &= \frac{\partial}{\partial t} \frac{\partial}{\partial y} \Phi(y, t) \Big|_{y=1} \\ &= \frac{\partial}{\partial y} \left[(\beta + \gamma)[\hat{\mu}(y) - y] \frac{\partial}{\partial y} \Phi(y, t) \right] \Big|_{y=1} \\ &= (\beta + \gamma) \left[\hat{\mu}'(y) - 1 \right] \frac{\partial}{\partial y} \Phi(y, t) + (\beta + \gamma) \left[\hat{\mu}(y) - y \right] \frac{\partial^2}{\partial y^2} \Phi(y, t) \Big|_{y=1} \\ &= (\beta + \gamma)[\hat{\mu}'(1) - 1][I] + (\beta + \gamma)[\hat{\mu}(1) - 1] \left[\frac{\partial^2}{\partial y^2} \Phi(y, t) \right] \Big|_{y=1} \\ &= (\beta + \gamma)[(2\beta)/(\beta + \gamma) - 1][I] \\ &= (\beta - \gamma)[I] \end{aligned}$$

We used $\hat{\mu}(1) = 1$ to eliminate the $\frac{\partial^2}{\partial y^2} \Phi(y, t)$ term and replaced $\hat{\mu}'(1)$ with $2\beta/(\beta + \gamma)$. Using this and $[I](0) = 1$, we have

$$[I] = e^{(\beta - \gamma)t}.$$

This example proves

Corollary 3.1. In the infinite population limit, if a disease starts with a single infection, then the expected number of active infections at time t solves

$$[I] = e^{(\beta - \gamma)t} \quad (20)$$

3.3. Cumulative and current outbreak size distribution

Let $\pi_{i,r}(t)$ be the probability of having i currently infected individuals and r completed infections at time t . We define $\Pi(y, z, t) = \sum_{i,r} \pi_{i,r}(t) y^i z^r$ to be the PGF at time t . We have $\Pi(y, z, 0) = y$. As before we assume the population is large enough that the spread of the disease is not limited by the size of the population.

We give an abbreviated derivation of the Kolmogorov equations for Π . A full derivation is requested as an exercise.

Forward Kolmogorov formulation. To derive the forward Kolmogorov equations for the PGF $\Pi(y, z, t)$, we use Property A.11, noting that all transition rates are proportional to i . The rate of transmission is βi and the rate of recovery is γi . There are no interactions to consider. So

$$\begin{aligned} \frac{\partial}{\partial t} \Pi(y, z, t) &= (\beta + \gamma) \left(\frac{\beta y^2}{\beta + \gamma} + \frac{\gamma z}{\beta + \gamma} - y \right) \frac{\partial}{\partial y} \Pi(y, z, t) \\ &= (\beta + \gamma)[\hat{\mu}(y, z) - y] \frac{\partial}{\partial y} \Pi(y, z, t) \end{aligned}$$

Backward Kolmogorov formulation. To derive the backward Kolmogorov equations for the PGF Π , we use a modified version of Property A.12 to account for two types of individuals (Exercise A.14, with events proportional only to the infected individuals). We find

$$\dot{\Pi}(y, z, t) = (\beta + \gamma)[\widehat{\mu}(\Pi(y, z, t), z) - \Pi(y, z, t)].$$

Combining our backward and forward Kolmogorov equation results, we get

Theorem 3.4. *Assuming a single initial infection in an infinite population, the PGF $\Pi(y, z, t)$ for the joint distribution of the number of current and completed infections at time t solves*

$$\frac{\partial}{\partial t} \Pi(y, z, t) = (\beta + \gamma)[\widehat{\mu}(y, z) - y] \frac{\partial}{\partial y} \Pi(y, z, t) \tag{21}$$

as well as

$$\frac{\partial}{\partial t} \Pi(y, z, t) = (\beta + \gamma)[\widehat{\mu}(\Pi(y, z, t), z) - \Pi(y, z, t)] \tag{22}$$

both with the initial condition $\Pi(y, z, 0) = y$.

It is again remarkable that these seemingly very different equations have the same solution.

Example 3.2. *The expected number of completed infections at time t is*

$$[R] = \sum_{j,k} k p_{jk} = \frac{\partial}{\partial z} \Pi(y, z, t)|_{y=z=1}$$

(although we use R , this approach is equally relevant for counting completed infections in the SIS model because of the infinite population assumption). Its evolution is given by

$$\begin{aligned} \frac{d}{dt} [R] &= \frac{\partial}{\partial t} \frac{\partial}{\partial z} \Pi(y, z, t)|_{y=z=1} \\ &= \frac{\partial}{\partial z} \left[(\beta + \gamma)[\widehat{\mu}(y, z) - y] \frac{\partial}{\partial y} \Pi(y, z, t) \right] |_{y=z=1} \\ &= (\beta + \gamma) \left[\frac{\partial}{\partial z} \widehat{\mu}(y, z) \frac{\partial}{\partial y} \Pi(y, z, t) + [\widehat{\mu}(y, z) - y] \frac{\partial}{\partial z} \frac{\partial}{\partial y} \Pi(y, z, t) \right] |_{y=z=1} \\ &= (\beta + \gamma) \left[\frac{\gamma}{\beta + \gamma} \frac{\partial}{\partial y} \Pi(y, z, t) + 0 \frac{\partial}{\partial z} \frac{\partial}{\partial y} \Pi(y, z, t) \right] |_{y=z=1} \\ &= \gamma [I] \end{aligned}$$

where we use the fact that $\widehat{\mu}(1, 1) = 1$, $\frac{\partial}{\partial z} \widehat{\mu}(y, z) = \gamma/(\beta + \gamma)$, and $[I] = \frac{\partial}{\partial y} \Pi(y, z, t)|_{y=z=1}$. Our result says that the rate of change of the expected number of completed infections is γ times the expected number of current infections.

This example proves

Corollary 3.2. *In the infinite population limit the expected number of recovered individuals as a function of time solves*

$$\frac{d}{dt} [R] = \gamma [I] \tag{23}$$

We will see that this holds even in finite populations.

3.4. Small outbreak final size distribution

We define

$$\Omega_\infty(z) = \left(\sum_{j < \infty} \omega_j z^j \right) + \omega_\infty z^\infty$$

to be the PGF of the distribution of outbreak final sizes in an infinite population, with $\omega_\infty z^\infty$ representing epidemics and for $j < \infty$ ω_j representing the probability that an outbreak infects exactly j individuals. We use the convention that $z^\infty = 0$ for $z < 1$ and 1 for $z = 1$. To calculate Ω_∞ , we make observations that the outbreak size coming from a single infected individual is 1 if the first thing that individual does is a recovery or it is the sum of the outbreak sizes of two infected individuals if the first thing the individual does is to transmit (yielding herself and her offspring).

Thus we have

$$\begin{aligned} \Omega_\infty(z) &= \frac{\beta}{\beta + \gamma} [\Omega_\infty(z)]^2 + \frac{\gamma}{\beta + \gamma} z \\ &= \hat{\mu}(\Omega_\infty(z), z) \end{aligned}$$

As for the discrete-time case we may solve this iteratively, starting with the guess $\Omega_\infty(z) = z$. Once n iterations have occurred, the first n coefficients of $\Omega_\infty(z)$ remain constant. Note that unlike the discrete case, here $\Omega_\infty(z) \neq z\hat{\mu}(\Omega_\infty(z))$. This yields

Theorem 3.5. *The PGF $\Omega_\infty(z) = \sum_j \omega_j z^j + \omega_\infty z^\infty$ for the final size distribution assuming a single initial infection in an infinite population solves*

$$\Omega_\infty(z) = \hat{\mu}(\Omega_\infty(z), z) \tag{24}$$

with $\Omega_\infty(1) = 1$. This function is discontinuous at $z = 1$. For the final size distribution conditional on the outbreak being finite, the PGF is continuous and equals

$$\begin{cases} \Omega_\infty(z)/\alpha & 0 \leq z < 1 \\ 1 & z = 1 \end{cases}$$

As in the discrete-time case, we can find the coefficients of $\Omega_\infty(z)$ analytically.

Theorem 3.6. *Consider continuous-time outbreaks with transmission rate β and recovery rate γ in an infinite population with a single initial infection. The probability the outbreak causes exactly j infections for $j < \infty$ [that is, the coefficient of z^j in $\Omega_\infty(z)$] is*

$$\omega_j = \frac{1}{j} \frac{\beta^{j-1} \gamma^j}{(\beta + \gamma)^{2j-1}} \binom{2j-2}{j-1}$$

We prove this theorem in [Appendix B](#). The proof is based on observing that if there are j total infected individuals, this requires $j - 1$ transmissions and j recoveries. Of the sequences of $2j - 1$ events that have the right number of recoveries and transmissions, a fraction $1/(2j - 1)$ of these satisfy additional constraints required to be a valid sequence leading to j infections (the sequence cannot lead to 0 infections prior to the last step). Alternately, we can note that the offspring distribution is geometric and use [Table 6](#).

3.5. Full dynamics in finite populations

We now derive the PGFs for continuous time SIS and SIR outbreaks in a finite population.

PGF-based techniques are easiest when we can treat events as independent. In the continuous-time model, when we look at the system in a given state, each event is independent of the others. Once the next event happens the possible events change, but conditional on the new state, they are still independent. Thus we can use the forward Kolmogorov approach.

The backward Kolmogorov approach will not work because in a finite population descendants of any individual are not independent. We do not look at the discrete-time version because in a single time step, multiple events can occur, some of which affect one another. So we would lose independence as we go from one time step to another. For these reasons we focus on the forward Kolmogorov formulations for the continuous-time models. Much of our approach here was derived previously in ([Bailey, 1953](#); [Bartlett, 1949](#)). See also ([Allen, 2008](#)).

For a given population size N , we let s , i , and r be the number of susceptible, infected and immune (removed) individuals. For the SIS model $r = 0$ and we have $s + i = N$ while for the SIR model we have $s + i + r = N$.

3.5.1. SIS

We start with the SIS model. We set $\xi_{s,i}(t)$ to be the probability of s susceptible and i actively infected individuals at time t . We define the PGF for the joint distribution of susceptible and infected individuals

$$\Xi(x, y, t) = \sum_i \xi_{s,i}(t) x^s y^i$$

At rate $\frac{\beta}{N}si$, successful transmissions occur, moving the system from the state (s, i) to $(s - 1, i + 1)$, which is equivalent to removing one susceptible individual and one infected individual, and replacing them with two infected individuals. Following Property A.11, this is represented by

$$\frac{\beta}{N}(y^2 - xy) \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi.$$

At rate γi , recoveries occur, moving the system from the state (s, i) to $(s + 1, i - 1)$, which is equivalent to removing one infected individual and replacing it with a susceptible individual. This is represented by

$$\gamma(x - y) \frac{\partial}{\partial y} \Xi.$$

So the PGF solves

$$\dot{\Xi} = \frac{\beta}{N}(y^2 - xy) \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi + \gamma(x - y) \frac{\partial}{\partial y} \Xi$$

It is sometimes useful to rewrite this as

$$\dot{\Xi} = (y - x) \left[\frac{\beta}{N} y \frac{\partial}{\partial x} - \gamma \right] \frac{\partial}{\partial y} \Xi$$

We have

Theorem 3.7. For SIS dynamics in a finite population we have

$$\frac{\partial}{\partial t} \Xi = \frac{\beta}{N}(y^2 - xy) \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi + \gamma(x - y) \frac{\partial}{\partial y} \Xi \tag{25}$$

We can use this to derive equations for the expected number of susceptible and infected individuals.

Example 3.3. We use $[S]$ and $[I]$ to denote the expected number of susceptible and infected individuals at time t . We have

$$[S] = \sum_{s,i} s \xi_{si}(t) = \sum_{s,i} s \xi_{si} 1^{s-1} 1^i = \frac{\partial}{\partial x} \Xi(1, 1, t)$$

$$[I] = \sum_{s,i} i \xi_{si}(t) = \sum_{s,i} i \xi_{si} 1^s 1^{i-1} = \frac{\partial}{\partial y} \Xi(1, 1, t)$$

We also define the expected value of the product si ,

$$[SI] = \sum_{s,i} si \xi_{si}(t) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi(1, 1, t).$$

Then we have

$$\begin{aligned} [\dot{S}] &= \frac{\partial}{\partial t} \frac{\partial}{\partial x} \Xi(1, 1, t) \\ &= \frac{\partial}{\partial x} \frac{\partial}{\partial t} \Xi(x, y, t) \Big|_{x=y=1} \\ &= \frac{\partial}{\partial x} \left((y - x) \left[\frac{\beta}{N} y \frac{\partial}{\partial x} - \gamma \right] \frac{\partial}{\partial y} \Xi \right) \Big|_{x=y=1} \\ &= (y - x) \frac{\partial}{\partial x} \left[\left(\frac{\beta}{N} y \frac{\partial}{\partial x} - \gamma \right) \frac{\partial}{\partial y} \Xi \right] - \left[\frac{\beta}{N} y \frac{\partial}{\partial x} - \gamma \right] \frac{\partial}{\partial y} \Xi \Big|_{x=y=1} \\ &= -\frac{\beta}{N} [SI] + \gamma [I] \end{aligned}$$

In the final line, we eliminated the first term because $y - x$ is zero at $x = y = 1$. Similar steps show that

$$[I] = \frac{\beta}{N}[SI] - \gamma[I]$$

but the derivation is faster if we simply note $[S] + [I] = N$ is constant. This proves

Corollary 3.3. For SIS disease, the expected number infected and susceptible solves

$$\frac{d}{dt}[S] = -\frac{\beta}{N}[SI] + \gamma[I] \quad (26)$$

$$\frac{d}{dt}[I] = \frac{\beta}{N}[SI] - \gamma[I] \quad (27)$$

where $[SI]$ is the expected value of the product si .

3.5.2. SIR

Now we consider the SIR model. A review of various techniques (including PGF-based methods) to find the final size distribution of outbreaks in finite-size populations can be found in (House, Ross, & Sirl, 2013). Here we focus on the application of PGFs to find the full dynamics. To reduce the number of variables we track, we focus just on s and i and use $r = N - s - i$ to find the number recovered. So $\Xi(x, y, t)$ does not have any dependence on z . For a given s and i , infection occurs at rate $\beta si/N$. It appears as a departure from the state (s, i) and entry into $(s - 1, i + 1)$. Following Property A.11, this is captured by

$$\frac{\beta}{N}(y^2 - xy) \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi.$$

Recovery is captured by

$$\gamma(1 - y) \frac{\partial}{\partial y} \Xi$$

[note the difference from the SIS case in the recovery term]. So we have

Theorem 3.8. For SIR dynamics in a finite population we have

$$\frac{\partial}{\partial t} \Xi = \frac{\beta(y^2 - xy)}{N} \frac{\partial}{\partial x} \frac{\partial}{\partial y} \Xi + \gamma(1 - y) \frac{\partial}{\partial y} \Xi \quad (28)$$

We follow similar steps to Example 3.3 to derive equations for $[S]$ and $[I]$ in Exercise 3.16. The result of this exercise should show

Corollary 3.4. For SIR disease, the expected number of susceptible, infected, and recovered individuals solves

$$\frac{d}{dt}[S] = -\frac{\beta}{N}[SI] \quad (29)$$

$$\frac{d}{dt}[I] = \frac{\beta}{N}[SI] - \gamma[I] \quad (30)$$

$$\frac{d}{dt}[R] = \gamma[I] \quad (31)$$

where $[SI]$ is the expected value of the product si .

3.6. Exercises

Exercise 3.1. Extinction Probability. Let β and γ be given with $\hat{\mu}(y) = (\beta y^2 + \gamma)/(\beta + \gamma)$.

- Analytically find solutions to $y = \hat{\mu}(y)$.
- Assume $\beta < \gamma$. Find all solutions in $[0, 1]$.
- Assume $\beta > \gamma$. Find all solutions in $[0, 1]$.

Exercise 3.2. Consistency with discrete-time formulation. Although we have argued that a transmission in the continuous-time disease transmission case can be treated as if a single infected individual has two infected offspring and then disappears, this is not what actually happens. In this exercise we look at the true offspring distribution of an infected individual before recovery, and we show that the ultimate predictions of the two versions are equivalent. Consider a disease in which individuals transmit at rate β and recover at rate γ . Let p_i be the probability an infected individual will cause exactly i new infections before recovering.

- Explain why $p_0 = \gamma/(\beta + \gamma)$.
- Explain why $p_i = \beta^i \gamma / (\beta + \gamma)^{i+1}$. So p_i form a geometric distribution.
- Show that $\mu(y) = \sum_i p_i y^i$ can be expressed as $\mu(y) = \gamma/(\beta + \gamma - \beta y)$. [This definition of μ without the hat corresponds to the discrete-time definition]
- Show that the solutions to $y = \mu(y)$ are the same as the solutions to $y = \hat{\mu}(y) = (\beta y^2 + \gamma)/(\beta + \gamma)$. So the extinction probability can be calculated either way. (You do not have to find the solutions to do this, you can simply show that the two equations are equivalent).

Exercise 3.3. Relation with \mathcal{R}_0 . Take $\mu(y) = \gamma/(\beta + \gamma - \beta y)$ as given in Exercise 3.2 and $\hat{\mu} = (\beta y^2 + \gamma)/(\beta + \gamma)$.

- Show that $\mu'(1) \neq \hat{\mu}'(1)$ in general.
- Show that when $\mathcal{R}_0 = \mu'(1) = 1$, then $\mu'(1) = \hat{\mu}'(1) = 1$. So both are still threshold parameters.

Exercise 3.4. Revisiting eventual extinction probability. We revisit the results of Exercise 3.1 using Eq. (17) (without solving it).

- By substituting for $\hat{\mu}(\alpha)$, show that $\dot{\alpha} = (1 - \alpha)(\gamma - \beta\alpha)$.

We have $\alpha(0) = 0$. Taking this initial condition and expression for $\dot{\alpha}$, show that

- $\alpha \rightarrow 1$ as $t \rightarrow \infty$ if $\beta < \gamma$ (i.e., $\mathcal{R}_0 < 1$) and
- $\alpha \rightarrow \gamma/\beta$ as $t \rightarrow \infty$ if $\beta > \gamma$ (i.e., $\mathcal{R}_0 > 1$).
- Set up (but do not solve) a partial fraction integration that would give $\alpha(t)$ analytically.

Exercise 3.5. Understanding the backward Kolmogorov equations. Let $\varphi_i(t)$ denote the probability of having i active infections at time t given that at time 0 there was a single infection [$\varphi_1(0) = 1$]. We have $\varphi_0(t) = \alpha(t)$. We extend the derivation of Eq. (16) to φ_1 . Assume $\varphi_0(t_0)$ and $\varphi_1(t_0)$ are known.

- Following the derivation of Eq. (16), approximate $\varphi_0(\Delta t)$, $\varphi_1(\Delta t)$, and $\varphi_2(\Delta t)$ for small Δt .
- From biological grounds explain why if there are 0 infections at time Δt then there are also 0 infections at time $t_0 + \Delta t$.
- If there is 1 infection at time Δt , what is the probability of 1 infection at time $t_0 + \Delta t$?
- If there are 2 infections at time Δt , what is the probability of 1 infection at time $t_0 + \Delta t$?
- Write $\varphi_1(t_0 + \Delta t)$ in terms of $\varphi_0(t_0)$, $\varphi_1(t_0)$, $\varphi_1(\Delta t)$, and $\varphi_2(\Delta t)$.
- Using the definition of the derivative, find an expression for $\dot{\varphi}_1$ in terms of $\varphi_1(t)$ and $\varphi_2(t)$.

Exercise 3.6. Derivation of the forward Kolmogorov equations. In this exercise we derive the PGF version of the forward Kolmogorov equations by directly calculating the rate of change of the probabilities of the states. Define $\varphi_j(t)$ to be the probability that there are j active infections at time t .

We have the forward Kolmogorov equations:

$$\dot{\varphi}_j = \beta(j - 1)\varphi_{j-1} + \gamma(j + 1)\varphi_{j+1} - (\beta + \gamma)j\varphi_j.$$

- Explain each term on the right hand side of the equation for $\dot{\varphi}_j$.
- By expanding $\dot{\Phi}(y, t) = \frac{\partial}{\partial t} \sum_j \varphi_j y^j$, arrive at Equation (18).

Exercise 3.7. Derivation of the backward Kolmogorov equations. In this exercise we follow (Allen, 2017; Bailey, 1964) and derive the PGF version of the backward Kolmogorov equations by directly calculating the rate of change of the probabilities of the states. Define $\varphi_{ki}(t)$ to be the probability of i infections at time t given that there were k infections at time 0. Although we assume that at time 0 there is a single infection, we will need to derive the equations for arbitrary k .

a. Explain why

$$\varphi_{ki}(t + \Delta t) = \varphi_{ki}(t) - k(\beta + \gamma)\varphi_{ki}(t)\Delta t + k\left(\beta\varphi_{(k+1)i}(t) + \gamma\varphi_{(k-1)i}(t)\right) + \mathcal{O}(\Delta t)$$

for small Δt .

b. By using the definition of the derivative $\dot{\varphi}_{ki} = \lim_{\Delta t \rightarrow 0} \frac{\varphi_{ki}(t+\Delta t) - \varphi_{ki}(t)}{\Delta t}$, find $\dot{\varphi}_{ki}$

Define $\Phi(y, t|k) = \sum_i \varphi_{ki} y^i$ to be the PGF for the number of active infections assuming that there are k initial infections.

c. Show that

$$\dot{\Phi}(y, t|1) = -(\beta + \gamma)\Phi(y, t|1) + \beta\Phi(y, t|2) + \gamma\Phi(y, t|0)$$

d. Explain why $\Phi(y, t|k) = \Phi(y, t|1)^k$.

e. Complete the derivation of Equation (19).

Exercise 3.8. Define $\Phi(y, t|k)$ to be the PGF for the probability of having i infections at time t given k infections at time 0.

a. Explain why $\Phi(y, t|k) = [\Phi(y, t|1)]^k$.

b. Show that if we substitute $\Phi(y, t|k) = [\Phi(y, t|1)]^k$ in place of $\Phi(y, t)$ in Eq. (18) the equation remains true with the initial condition y^k .

c. Show that if we substitute $\Phi(y, t|k) = [\Phi(y, t|1)]^k$ in place of $\Phi(y, t)$ in equation (19) we do not get a true equation.

So Eq. (18) applies regardless of the initial condition, but Eq. (19) is only true for the specific initial condition of one infection.

Exercise 3.9. Let $\Phi(y, t|k)$ be the PGF for the number of infections assuming there are initially k infections. Derive the backward Kolmogorov equation for $\Phi(y, t|k)$. Note that some of the Φ s in the derivation above would correspond to $\Phi(y, t|1)$ and some of them to $\Phi(y, t|k)$.

Exercise 3.10. Comparison of the formulations.

a. Using Eq. (18) derive an equation for $\dot{\alpha}$ where $\alpha(t) = \Phi(0, t)$. What, if any, additional information would you need to solve this numerically?

b. Using Eq. (19), derive Equation (17) for $\dot{\alpha}$ where $\alpha(t) = \Phi(0, t)$. What, if any, additional information would you need to solve this numerically?

Exercise 3.11. Full solution.

a. Show that Eq. (19) can be written

$$\frac{\partial}{\partial t} \Phi(y, t) = (\gamma - \beta\Phi(y, t))(1 - \Phi(y, t))$$

b. Using partial fractions, set up an integral which you could use to solve for $\Phi(y, t)$ analytically (you do not need to do all the algebra to solve it).

Exercise 3.12. Argue from their definitions that $\Phi(y, t) = \Pi(y, z, t)|_{z=1}$.

Exercise 3.13. Derive Theorem 3.3 from Theorem 3.4.

Exercise 3.14. Derive Theorem 3.5 from Theorem 3.4.

Exercise 3.15. Equivalence of continuous and discrete final size distributions.

Show by direct substitution that if $\Omega_\infty(z) = \hat{\mu}(\Omega_\infty(z), z)$ then $\Omega_\infty(z) = z\mu(\Omega_\infty(z))$ where $\mu(y) = \gamma/(\beta + \gamma - \beta y)$ is the PGF for the offspring distribution found in Exercise 3.2.

Exercise 3.16. We revisit the derivations of the usual mass action SIR ODEs. Following Example 3.3,

- a. Derive $[\dot{S}]$ in terms of $[SI]$.
- b. Derive $[\dot{I}]$ in terms of $[SI]$ and $[I]$.
- c. Using $[S] + [I] + [R] = N$, derive $[\dot{R}]$.

4. Large-time dynamics

We now look at how PGFs can be used to develop simple models of SIR disease spread in the large population limit when the disease infects a nonzero fraction of the population. In this limit, the early-time approaches derived before break down because depletion of the susceptible population is important. The later-time models of Section 3.5 are impractical because of the $N \rightarrow \infty$ limit and are more restricted due to the continuous-time assumption.

4.1. SIR disease and directed graphs

In Section 2.5 we argued that for early times the continuous-time predictions can be framed in terms of the discrete-time predictions because we can classify infections by the length of the transmission chain to them from the index case. For SIR disease this argument extends beyond early times.

To see this, we assume that prior to the disease introduction, we know for each individual what would happen if he ever becomes infected as in Fig. 9. In particular, we know how long his infection would last, to whom he would transmit, and how long the delays from his infection to onwards transmission would be. The process of choosing these in advance, selecting the initial infection(s), and tracing infection from there is equivalent to choosing the initial infection(s) and then choosing the transmissions while the infection process is traced out.

By assigning who transmits to whom (and how long the delays are), we have defined a weighted directed graph whose edges represent the potential transmissions and weights represent the delays (Kenah & Miller, 2011; Kiss, Miller, & Simon). A node v will become infected if and only if there is at least one directed path from an initially infected node u to v . The time of v 's infection is given by the least sum of all paths from initially infected nodes to v . We note that the transmission process could be quite complex: the duration of a node's infection and the delays from time of infection to time of onwards transmissions can have effectively arbitrary distributions, and we could still build a similar directed graph.

This directed graph is a useful structure to study because it encodes the outbreak in a single static object, as opposed to a dynamic process. There is significant study of the structure of such directed graphs (Broder et al., 2000; Dorogovtsev, Mendes, & Samukhin, 2001). Much of it focuses on the size of out-components of a node (that is, for a given node, what fraction of the population can be reached following the edges forwards) or the in-components (that is, from what fraction of the population is it possible to reach a given node by following edges forwards).

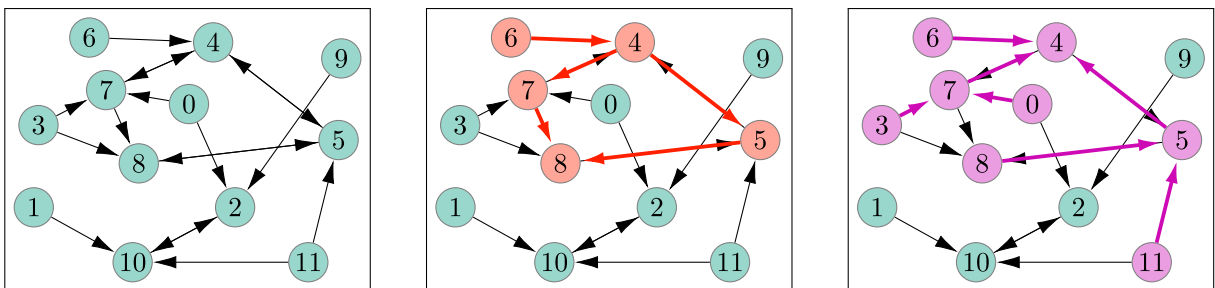


Fig. 9. (Left) A twelve-individual population, after the *a priori* assignment of who would transmit to whom if ever infected by the SIR disease (the delay until transmission is not shown). Half of the nodes have zero potential infectors and half have 3. Half of the nodes have 1 potential offspring and half have 2. So the offspring distribution has PGF $(x + x^2)/2$ while the ancestor distribution has PGF $\chi(x) = (1 + x^3)/2$. (Middle) If node 6 is initially infected, the infection will reach node 4 who will transmit to 5 and 7, and eventually infection will also reach 8 and 2 before further transmissions fail because nodes are already infected. If however, it were to start at 9, then it would reach 2, from which it would spread only to 10. (Right) By tracing backwards from an individual, we can determine which initial infections would lead to infection of that individual. For example individual 4 will become infected if and only if it is initially infected or any of 0, 3, 5, 6, 7, 8, or 11 is an initial infection.

4.2. Final size relations for SIR epidemics

We now derive final size relations for SIR epidemics in the large population limit. We begin with the assumption that a single node is initially infected and that an epidemic happens.

We use the mapping of the SIR epidemic to a directed graph G . Assume that a single node u is chosen to be infected. Consider a node v . The probability v is infected is the probability that u is in her in-component, and so it equals the proportion of G that is in the in-component of v . In the limit as G becomes infinite, there are a few possibilities. We are interested in what happens when an epidemic occurs, so we can assume that u has a large out-component (in the sense that the out-component takes up a non-zero fraction of G in the $N \rightarrow \infty$ limit) (Broder et al., 2000):

- If v has a small in-component, then almost surely u is not in the in-component and so almost-surely v is not infected.
- If v has a large in-component, then almost surely it contains a node w that lies in the out-component of u . The existence of w then implies the existence of a path from u to w to v , so v is in u 's out-component and v becomes infected.

Thus, if u causes an epidemic in the large N limit, then the probability that v becomes infected equals the probability that v has a large in-component. So the size of an epidemic (if it happens) is simply the probability a random individual has a large in-component.

We approach the question of whether v has a large in-component in the same way we approached the question of whether u causes a large chain of infections (i.e., whether u has a large out-component). We define the PGF of the ancestor distribution to be the function $\chi(x)$ defined by

$$\chi(x) = \sum_i p_i x^i$$

where p_i is the probability that a random node in the directed graph has in-degree i . That is, there are exactly i nodes that would directly transmit to the randomly chosen node if they were ever infected. So the probability an individual is not infected $S(\infty)/N$ solves $x = \chi(x)$, choosing the smaller solution when two solutions exist. Since the proportion infected is $r(\infty) = R(\infty)/N = 1 - S(\infty)/N$, we can conclude

Theorem 4.1. Assume that an outbreak begins with a single infected individual and an epidemic results. In the large N limit, the expected cumulative proportion infected $r(\infty) = R(\infty)/N$ solves

$$r(\infty) = 1 - \chi(1 - r(\infty))$$

where $\chi(x)$ is the PGF of the ancestor distribution. If there are multiple solutions we choose the larger solution for $r(\infty)$ in $[0, 1]$.

Under common assumptions, the population is large, the average number of transmissions an individual causes is \mathcal{R}_0 , and the recipient is selected uniformly at random. Under these assumptions the ancestor distribution is Poisson with mean \mathcal{R}_0 . So $\chi(x) = e^{-\mathcal{R}_0(1-x)}$. Then

$$r(\infty) = 1 - e^{-\mathcal{R}_0 r(\infty)}. \quad (32)$$

Deriving this result does not depend on the duration of infections, or even on the distribution of factors affecting infectiousness. The assumptions required are that an epidemic starts from a single infected individual, that each transmission reaches a randomly chosen member of the population, that all individuals have equal susceptibility, and the average individual will transmit to \mathcal{R}_0 others. This result is general across a wide range of assumptions about the infectious process.

Restating this we have:

Corollary 4.1. Assume that an SIR disease is spreading in a well-mixed population with homogeneous susceptibility. Assuming that the initial fraction infected is infinitesimal and an epidemic occurs, the final size satisfies

$$r(\infty) = 1 - e^{-\mathcal{R}_0 r(\infty)} \quad (33)$$

where \mathcal{R}_0 is the reproductive number of the disease.

This explains many of the results of (Ma & Earn, 2006; Miller, 2012), and our observation in Example 2.3 that the epidemic size depended on \mathcal{R}_0 and not on any other property of the offspring distribution. A closely-related derivation is provided by (Diekmann & Heesterbeek, 2000, Section 1.3).

4.3. Discrete-time SIR dynamics

We now take a discrete-time approach, similar to (Miller et al., 2012; Valdez, Macri, & Braunstein, 2012) and (Kiss, Miller, & Simon, chapter 6). We will assume that at generation $g = 0$ the disease is introduced by infecting a proportion p uniformly at

random leaving the remainder susceptible. We assume that the population is very large and that the number of infections is large enough that the dynamics can be treated as deterministic. Our results can be adapted to other initial conditions (for example, to account for nonzero R in the initial condition).

We assume that $\chi(x)$ is known and that there is no correlation between how susceptible an individual is and how infectious that individual is. Thus at generation g , the expected number of transmissions occurring is $\mathcal{R}_0 I(g)$, and how the recipients are chosen depends on χ .

Let v be a randomly chosen member of the population. The probability that v 's randomly chosen ancestor has not yet been infected by generation $g - 1$ is $S(g - 1)/N$. The probability v is susceptible at generation g is the probability v was initially susceptible, $1 - \rho$, times the probability v has not received any transmissions, $\chi(S(g - 1)/N)$ (see [Exercise 4.2](#)).

So for $g > 0$ we arrive at

$$\begin{aligned} S(g) &= (1 - \rho)N\chi(S(g - 1)/N) \\ I(g) &= N - R(g) - S(g) \\ R(g) &= R(g - 1) + I(g - 1) \end{aligned}$$

with

$$S(0) = 1 - \rho, \quad I(0) = \rho, \quad R(0) = 0.$$

So we have

Theorem 4.2. *Assume that $\chi(x)$ is the PGF of the ancestor distribution and assume there is no correlation between infectiousness and susceptibility of a given individual. Further assume that at generation 0 a fraction ρ is randomly infected in the generation-based discrete-time model. Then in the large population limit*

$$S(g) = (1 - \rho)N\chi(S(g - 1)/N) \tag{34a}$$

$$I(g) = N - R(g) - S(g) \tag{34b}$$

$$R(g) = R(g - 1) + I(g - 1). \tag{34c}$$

With initial conditions

$$S(0) = (1 - \rho)N, \quad I(0) = \rho N, \quad R(0) = 0. \tag{34d}$$

We can interpret this in the context of survival functions. The function $(1 - \rho)\chi(S(g - 1)/N)$ gives the probability that a node has lasted g generations without being infected.

4.4. Continuous-time SIR epidemic dynamics

We now move to continuous-time SIR epidemics. We allow for heterogeneity, assuming that each susceptible individual u receives transmissions at some rate $\kappa_u \beta I(t)/N \langle K \rangle$, and that the PGF of κ is $\psi(x) = \sum_{\kappa} P(\kappa) x^{\kappa}$. We assume κ takes only non-negative integer values.

For an initially susceptible individual u with a given κ_u , the probability of not yet receiving a transmission by time t solves $\dot{s}_u = -\kappa_u \beta I(t) s_u / N \langle K \rangle$, which has solution

$$s_u = e^{-\kappa_u \beta \int_0^t \frac{I(\tau) d\tau}{N \langle K \rangle}}.$$

So we can write

$$s_u = \theta^{\kappa_u}$$

where $\theta = e^{-\beta \int_0^t \frac{I(\tau) d\tau}{N \langle K \rangle}}$ and

$$\dot{\theta} = -\beta \theta I / N \langle K \rangle.$$

Considering a random individual of unknown κ , the probability she was initially susceptible is $1 - \rho$ and the probability she has not received any transmissions is $\psi(\theta)$. So

$$S(t) = (1 - \rho)N\psi(\theta)$$

Taking $\dot{R} = \gamma I$, we have

$$\begin{aligned} \dot{R} &= \gamma I \\ &= -\frac{\gamma N \langle K \rangle}{\beta} \frac{\dot{\theta}}{\theta}. \end{aligned}$$

Integrating both sides, taking $\theta(0) = 1$ and $R(0) = 0$, we have

$$R = -\frac{\gamma N \langle K \rangle}{\beta} \ln \theta$$

Taking $I = N - S - R$ we get

$$I = N \left(1 - (1 - \rho)\psi(\theta) + \frac{\gamma \langle K \rangle}{\beta} \ln \theta \right)$$

and so $\dot{\theta}$ becomes

$$\dot{\theta} = -\beta \theta \left(1 - (1 - \rho)\psi(\theta) + \frac{\gamma \langle K \rangle}{\beta} \ln \theta \right) / \langle K \rangle$$

Theorem 4.3. Assuming that at time $t = 0$ a fraction ρ of the population is randomly infected and that the susceptible individuals each have a κ such that they become infected as a Poisson process with rate $\kappa \beta I / N \langle K \rangle$, in the large population limit we have

$$S = N(1 - \rho)\psi(\theta) \tag{35a}$$

$$I = N \left(1 - (1 - \rho)\psi(\theta) + \frac{\gamma \langle K \rangle}{\beta} \ln \theta \right) \tag{35b}$$

$$R = -\frac{\gamma N \langle K \rangle}{\beta} \ln \theta \tag{35c}$$

where $\psi(x) = \sum_k P(k)x^k$ and the system is governed by a single ODE

$$\dot{\theta} = \frac{-\beta \theta \left(1 - (1 - \rho)\psi(\theta) + \frac{\gamma \langle K \rangle}{\beta} \ln \theta \right)}{\langle K \rangle} \tag{35d}$$

with initial condition

$$\theta(0) = 1. \tag{35e}$$

As in the discrete-time case, this can be interpreted as a survival function formulation of the SIR model. Most, if not all, mass-action formulations of the SIR model can be re-expressed in a survival function formulation. Some examples are shown in the Exercises.

Some very similar systems of equations are developed in (Kiss, Miller, & Simon, chapter 6) and (Miller, 2011; Miller et al., 2012; Valdez et al., 2012; Volz, 2008) where the focus is on networks for which the value of κ not only affects the probability of becoming infected, but also of transmitting further. These references focus on the assumption that an individual's infector remains a contact after transmission, but they contain techniques for studying partnerships with varying duration.

4.5. Exercises

Exercise 4.1. Ancestor distribution for homogeneous well-mixed population.

Consider an SIR disease in a well-mixed population having N individuals and a given \mathcal{R}_0 . Let v be a randomly chosen individual from the directed graph created by placing edges from each node to all those nodes they would transmit to if infected.

a Show that if the average number of offspring is \mathcal{R}_0 , then so is the average number of infectors.

- b If there are exactly $\mathcal{R}_0 N$ edges in the directed graph and each recipient is chosen uniformly at random from the population (independent of any previous choice), argue that the number of transmissions v receives has a binomial distribution with $\mathcal{R}_0 N$ trials and probability \mathcal{R}_0/N . (technically we must allow edges from v to v)
- c Argue that if \mathcal{R}_0 remains fixed as $N \rightarrow \infty$, then the number of transmissions v receives is Poisson distributed with mean \mathcal{R}_0 .

Exercise 4.2. Explain why for large N the probability v is still susceptible at generation g if she was initially susceptible is $\chi(S(g-1)/N)$.

Exercise 4.3. Use Theorem 4.2 to derive a result like Theorem 4.1, but with nonzero ρ .

Exercise 4.4. Final size relations

Consider the continuous time SIR dynamics as given in System (35)

- a Assume $\kappa = 1$ for all individuals, and write down the corresponding equations for S, I, R , and θ .
- b At large time $I \rightarrow 0$, so $S(\infty) = N - R(\infty)$. But also $S(\infty) = S(0)\psi(\theta(\infty))$. By writing $\theta(\infty)$ in terms of $R(\infty)$, derive a recurrence relation for $r(\infty) = R(\infty)/N$ in terms of $r(\infty)$ and $\mathcal{R}_0 = \beta/\gamma$.
- c Comment on the relation between your result and Theorem 4.1

Exercise 4.5. Other relations

- a Using the equations from Exercise 4.4, derive the peak prevalence relation, an expression for the maximum value of I . [at the maximum $\dot{I} = 0$, so we start by finding θ so that $\dot{S} + \dot{R} = 0$.]
- b Similarly, find the peak incidence relation, an expression for the maximum rate at which infections occur, $-\dot{S}$.

Exercise 4.6. Alternate derivation of S_u .

If the rate of transmissions to u is $\beta I_{\kappa u}/N\langle K \rangle$, then the expected number of transmissions u has received is $\beta \kappa_u \int_0^t I(\tau) d\tau/N\langle K \rangle$ and this is Poisson distributed.

- a Let $f_u(x)$ be the PGF for the number of transmissions u has received. Find an expression for $f_u(x)$ in terms of the integral $\int_0^t I(\tau) d\tau$.
- b Explain why $f_u(0)$ is the probability u is still susceptible.
- c Find $f_u(0)$.

Exercise 4.7. Alternate derivation of Theorem 4.3 in the homogeneous case.

The usual homogeneous SIR equations are

$$\begin{aligned} \dot{S} &= -\beta IS/N \\ \dot{I} &= \beta IS/N - \gamma I \\ \dot{R} &= \gamma I \end{aligned}$$

We will derive system (35) for fixed $\kappa = 1$ from this system through the use of an integrating factor. Set $\theta = e^{-\beta \int_0^t I(\tau) d\tau/N}$.

- a Show that $\dot{\theta} = -\beta I\theta/N$ and so $\dot{\theta}/\theta = -\beta \dot{R}/N\gamma$.
- b Using the equation for \dot{S} add $\beta IS/N$ to both sides and then divide by θ (the factor $1/\theta$ is an integrating factor). Show that the expression on the left hand side is $\frac{d}{dt} S/\theta$ and so

$$\frac{d}{dt} S/\theta = 0.$$

- c Solve for R in terms of θ .
- d Solve for S in terms of θ .

e Solve for I in terms of θ using $S + I + R = N$.

This equivalence was found in (Miller, 2012) and (Harko, Lobo, & Mak, 2014).

Exercise 4.8. Alternate derivation of Theorem 4.3.

Consider now a population having many subgroups of susceptibles denoted by κ with the group κ receiving transmissions at rate $\beta\kappa I/N$ per individual. Once infected, each individual transmits with rate $\beta\langle K \rangle$ and recovers with rate γ . These assumptions lead to

$$\begin{aligned} \dot{S}_\kappa &= -\beta\kappa \frac{I}{N\langle K \rangle} S_\kappa \\ \dot{I} &= -\gamma I + \beta \frac{I}{N\langle K \rangle} \sum_\kappa \kappa S_\kappa \\ \dot{R} &= \gamma I \end{aligned}$$

Following Exercise 4.7, set $\theta = e^{-\beta \int_0^t I(\tau) d\tau / N}$ and derive system (35) from these equations by use of an integrating factor.

5. Multitype populations

We now briefly discuss how PGFs can be applied to multitype populations. This section is intended primarily as a pointer to the reader to show that it is possible to apply these methods to such populations. We do not perform a detailed analysis.

Many populations can be divided into subgroups. These may be patches in a metapopulation model, genders in a heterosexual sexually transmitted infection model, age groups in an age-structured population, or any of a number of other groupings. Applications of PGFs to such models have been studied in multiple contexts (Kucharski & Edmunds, 2015; Reluga, Meza, Walton, & Galvani, 2007).

5.1. Discrete-time epidemic probability

We begin by considering the probability of an epidemic in a discrete-time model. To set the stage, assume there are M groups and let $p_{i_1, i_2, \dots, i_M | k}$ be the probability that an individual of group k will cause i_ℓ infections in group ℓ . Define $\alpha_{g|k}$ to be the probability that a chain of infections starting from an individual of group k becomes extinct within g generations.

It is straightforward to show that if we define

$$\psi_k(x_1, x_2, \dots, x_M) = \sum_{i_1, i_2, \dots, i_M} p_{i_1, i_2, \dots, i_M | k} x_1^{i_1} x_2^{i_2} \dots x_M^{i_M}$$

then

$$\begin{aligned} \alpha_{g|k} &= \sum_{i_1, i_2, \dots, i_M} p_{i_1, i_2, \dots, i_M | k} \alpha_{g-1|1}^{i_1} \alpha_{g-1|2}^{i_2} \dots \alpha_{g-1|M}^{i_M} \\ &= \psi_k(\alpha_{g-1|1}, \alpha_{g-1|2}, \dots, \alpha_{g-1|M}) \end{aligned}$$

After converting this into vectors we get $\vec{\alpha}_1 = \vec{\psi}(\vec{0})$. Iterating g times we have

$$\vec{\alpha}_g = \vec{\psi}^{[g]}(\vec{0}) \tag{36}$$

Setting $\vec{\alpha}$ to be the limit as g goes to infinity, we find the extinction probabilities. Specifically, the k -th component of $\vec{\alpha}$ is the probability of extinction given that the first individual is of type k . Thus we have:

Theorem 5.1. *Let*

- $\vec{\alpha}_g = (\alpha_{g|0}, \alpha_{g|1}, \dots, \alpha_{g|M})$ where $\alpha_{g|k}$ is the probability a chain of infections starting with a type k individual will end within g generations
- and $\vec{\psi} = (\psi_1, \psi_2, \dots, \psi_M)$ where $\psi_k(\vec{x}) = \sum_{i_1, i_2, \dots, i_M} p_{i_1, i_2, \dots, i_M | k} x_1^{i_1} x_2^{i_2} \dots x_M^{i_M}$.

Then $\vec{\alpha}_g = \vec{\psi}^{[g]}(\vec{0})$.

The vector of eventual extinction probabilities in the infinite population limit is given by $\vec{\alpha}_\infty = \lim_{g \rightarrow \infty} \vec{\alpha}_g$ and is a solution to

$$\vec{\alpha}_\infty = \vec{\psi}(\vec{\alpha}_\infty).$$

We could have derived this directly by showing that the extinction probabilities solve $\vec{\alpha} = \vec{\psi}(\vec{\alpha})$. In this case it might not be obvious how to solve this multidimensional system of nonlinear equations or how to be certain that the solution found is the appropriate one. However, by interpreting the iteration in Eqn. (36) in terms of the extinction probability after g generations, it is clear that simply iterating starting from $\vec{\alpha}_0 = \vec{0}$ will converge to the appropriate values. Additionally the values calculated in each iteration have a meaningful interpretation.

Example 5.1. Consider a population made up of many large communities. We assume an unfamiliar disease is spreading through the population. When the disease begins to spread in a community, the community learns to recognize the disease symptoms and infectiousness declines. We assume that we can divide the population into 3 types: primary cases T_0 , secondary cases T_1 , and tertiary cases T_2 . The infectiousness of primary cases is higher than that of secondary cases which is higher than that of tertiary cases. Within a community a primary case can cause secondary cases, while secondary and tertiary cases can cause tertiary cases. All cases can cause new primary cases in other communities. We ignore multiple introductions to the same community.

We define n_{ij} to be the number of infections of type T_i caused by a type T_j individual, and we assume that we know the joint distribution $p_{n_{00},n_{10}}, p_{n_{01},n_{21}}$, and $p_{n_{02},n_{22}}$. We define

$$\begin{aligned} \psi_1(x, y, z) &= \sum_{n_{00}, n_{10}} p_{n_{00}, n_{10}} x^{n_{00}} y^{n_{10}} \\ \psi_2(x, y, z) &= \sum_{n_{01}, n_{21}} p_{n_{01}, n_{21}} x^{n_{01}} z^{n_{21}} \\ \psi_3(x, y, z) &= \sum_{n_{02}, n_{22}} p_{n_{02}, n_{22}} x^{n_{02}} z^{n_{22}} \end{aligned}$$

Note that ψ_1 does not depend on z while ψ_2 and ψ_3 do not depend on y .

We define $\vec{\alpha}_0 = (0, 0, 0)$ and set $\vec{\alpha}_g = (\psi_1(\vec{\alpha}_{g-1}), \psi_2(\vec{\alpha}_{g-1}), \psi_3(\vec{\alpha}_{g-1}))$. Then taking $\vec{\alpha}$ to be the limit as $g \rightarrow \infty$, the first entry of $\vec{\alpha}$ is the probability that the disease goes extinct starting from a single primary case.

5.2. Continuous-time SIR dynamics

Now we consider a continuous-time version of SIR dynamics in a heterogeneous population.

Assume again that there are M groups and let β_{ij} be the rate at which an individual in group j causes transmissions that go to group i . Let ξ_i be the expected number of transmissions that an individual in group i has received since time 0. Finally assume that individuals in group i recover at rate γ_i . Then the expected number of transmissions an individual in group i has received by time t is Poisson distributed with mean ξ_i . The PGF for the number of transmissions received is thus $e^{-\xi_i(1-x)}$. Setting $x = 0$, the probability of having received zero transmissions is $e^{-\xi_i(t)}$. Thus $S_i = S_i(0)e^{-\xi_i(t)}$. We have $I_i = N_i - S_i - R_i$ and $\dot{R}_i = \gamma_i I_i$. To find ξ_i , we simply note that the total rate that group i is receiving infection is $\sum_j I_j \beta_{ij}$, and so

$$\dot{\xi}_i = \frac{\sum_j I_j \beta_{ij}}{N_i}.$$

Thus:

Theorem 5.2. If the rate of transmission from an infected individual in group j to group i is β_{ij} , then

$$S_i = S_i(0)e^{-\xi_i(t)} \tag{37a}$$

$$I_i = N_i - S_i - R_i \tag{37b}$$

$$\dot{R}_i = \gamma_i I_i \tag{37c}$$

$$\dot{\xi}_i = \frac{\sum_j I_j \beta_{ij}}{N_i} \tag{37d}$$

with $\xi(0) = 0$.

5.3. Exercises

Exercise 5.1. Consider a vector-borne disease for which each infected individual infects a Poisson-distributed number of vectors, with mean λ . Each infected vector causes i infections with probability $p_i = \pi^i(1 - \pi)$ for some $\pi \in [0, 1]$. This scenario corresponds to human infection lasting for a fixed time with some constant transmission rate to vectors, and each vector having probability π of living to bite again after each bite and transmitting with probability 1 if biting.

- Let α_{g1} and α_{g2} be the probability that an outbreak would go extinct in g generations starting with an infected human or vector respectively. Find the vector-valued function $\vec{\psi}(\vec{x}) = (\psi_1(\vec{x}), \psi_2(\vec{x}))$. That is, what are the PGFs $\psi_1(x_1, x_2)$ and $\psi_2(x_1, x_2)$?
- Set $\lambda = 3$ and $\pi = 0.5$. Find the probability of an epidemic if one infected human is introduced or if one infected vector is introduced.
- For the same values, find the probability of an epidemic if one infected vector is introduced.
- Find $\psi_2(\psi_1(0, x), 0)$. How should we interpret the terms of its Taylor Series expansion?

Exercise 5.2. Starting from the equations

$$\begin{aligned} \dot{S}_i &= -\frac{S_i}{N_i} \sum_j \beta_{ij} I_j \\ \dot{I}_i &= -\gamma_i I_i + \frac{S_i}{N_i} \sum_j \beta_{ij} I_j \\ \dot{R}_i &= \gamma_i I_i \end{aligned}$$

use integrating factors to derive System (37).

Exercise 5.3. Assume the population is grouped into subgroups of size N_i with $N = \sum_i N_i$ and the i -th subgroup has a parameter κ_i representing their rate of contact with others. Take

$$\beta_{ji} = \kappa_j \frac{\kappa_i N_i}{\sum_k N_k \kappa_k} \beta$$

to be the transmission rate from type i individuals to a single type j individual, and assume all infected individuals recover with the same rate γ .

Define $\theta = e^{-\beta(\sum_j \kappa_j \int_0^t I_j(\tau) d\tau) / \sum_j \kappa_j N_j}$ and define the PGF $\psi(x) = \sum_i \frac{N_i}{N} x^i$. Let $S = \sum_i S_i$, $I = \sum_i I_i$, and $R = \sum R_i$.

$$\begin{aligned} S &= N\psi(\theta) \\ I &= N - S - R \\ \dot{R} &= \gamma I \\ \dot{\theta} &= -\beta\theta \frac{\sum_j \kappa_j I_j}{\sum_j \kappa_j N_j} \end{aligned}$$

with $\theta(0) = 1$.

$$\dot{\theta} = -\beta\theta + \beta \frac{\theta^2 \psi'(\theta)}{\psi'(1)} - \theta\gamma \ln \theta$$

- Explain what assumptions this model makes about interactions between individuals in group i and j .
- Show that $\sum_j \frac{\kappa_j I_j}{\sum_j \kappa_j N_j} = 1 - \frac{\sum_j \kappa_j S_j}{\sum_j \kappa_j N_j} - \frac{\sum_j \kappa_j R_j}{\sum_j \kappa_j N_j}$.
- Show that $\frac{\sum_j \kappa_j S_j}{\sum_j \kappa_j N_j} = \frac{\theta \psi'(\theta)}{\psi'(1)}$.
- Show that $\frac{d}{dt} \frac{\sum_j \kappa_j R_j}{\sum_j \kappa_j N_j} = -(\gamma/\beta) \frac{\dot{\theta}}{\theta}$ and solve for $\frac{\sum_j \kappa_j R_j}{\sum_j \kappa_j N_j}$ in terms of θ assuming $R_j = 0$ for all j .
- Thus conclude that

6. Discussion

There are many contexts where we are interested in how a newly introduced infectious disease would spread. We encounter situations like this in the spread of zoonotic infections such as Monkey Pox or Ebola as well as the importation of novel diseases such as the Zika in the Americas or the reintroduction of locally eliminated diseases such as Malaria.

PGFs are an important tool for the analysis of epidemics, particularly at early stages. They allow us to relate the individual-level transmission process to the distribution of outcomes. This allows us to take data about the transmission process and make predictions about the possible outcomes, but it also allows us to take observed outbreaks and use them to infer the individual-level transmission properties.

For SIR disease PGFs also provide a useful alternative formulation to the usual mass-action equations. This formulation leads to a simple derivation of final-size relations and helps explain why previous studies have shown that a wide range of disease assumptions give the same final size relation.

Our goal with this primer has been to introduce researchers to the many applications of PGFs to disease spread. We have used the appendices to derive some of the more technical properties of PGFs. Additionally we have developed a Python package `Invasion_PGF` which allows for quick calculation of the results in the first three sections of this primer. A detailed description of the package is in [Appendix C](#). The software can be downloaded at https://github.com/joelmiller/Invasion_PGF. Documentation is available within the repository, starting with the file `docs/_build/html/index.html`. The supplementary information includes code that uses `Invasion_PGF` to generate the figures of [Section 2](#).

Acknowledgments

This work was funded by Global Good.

I thank Linda Allen for useful discussion about the Kolmogorov equations. Hao Hu played an important role in inspiring this work and testing the methods. Hil Lyons and Monique Ambrose provided valuable feedback on the discussion of inference. Amelia Bertozzi-Villa and Monique Ambrose read over drafts and recommended a number of changes that have significantly improved the presentation.

The python code and output in [Appendix C](#) was incorporated using Pythontex ([Poore, 2015](#)). I relied heavily on <https://tex.stackexchange.com/a/355343/70067> by “touhami” in setting up the solutions to the exercises.

Appendix A. Important properties of PGFs

In this appendix, we give some theoretical background behind the important properties of PGFs which we use in the main part of the primer. We attempt to make each subsection self-contained so that the reader has a choice of reading through the appendix in its entirety, or waiting until a property is used before reading that section. Because we expect the appendix is more likely to be read piecemeal, the exercises are interspersed through the text where the relevant material appears.

A PGF has been described as “a clothesline on which we hang up a sequence of numbers for display” ([Wilf, 2005](#)). Similarly ([Pólya, 1990](#)) says “A generating function is a device somewhat similar to a bag. Instead of carrying many little objects detachedly, which could be embarrassing, we put them all in a bag, and then we have only one object to carry, the bag.” Indeed for many purposes mathematicians use PGFs primarily because once we have the distribution put into this “bag”, many more mathematical tools are available, allowing us to derive interesting and sometimes surprising identities ([Wilf, 2005](#)).

However, for our purposes there is a meaningful direct interpretation of a PGF. Assume that we are interested in the probability that an event does not happen given some unknown number i of independent identical Bernoulli trials with probability α the event does not happen in any one trial. Let r_i represent the probability that there are i trials. Then the probability that the event does not occur in any trial is

$$\sum_i r_i \alpha^i = f(\alpha),$$

and so PGFs emerge naturally in this context.

In infectious disease, this context occurs frequently and many results in this primer can be expressed in this framework. For reference, we make this property more formal:

Property A.1. Assume we have a process consisting of a random number i independent identical Bernoulli trials. Let r_i denote the probability of a given i and $f(x) = \sum_i r_i x^i$ be its PGF. If α is the probability that a random trial fails, then $f(\alpha)$ is the probability all trials fail.

Appendix A.1. Properties related to individual coefficients

We start by investigating how to find the coefficients of a PGF if we can calculate the numeric value of the PGF at any point.

This section makes use of the imaginary number $i = \sqrt{-1}$, and so in this section we avoid using i as an index in the sum of $f(x)$.

Property A.2. Given a PGF $f(x) = \sum_n r_n x^n$, the coefficient of x^n in its expansion for a particular n can be calculated by taking n derivatives, evaluating the result at $x = 0$, and dividing by $(n!)$. That is

$$r_n = \frac{1}{n!} \left(\frac{d}{dx} \right)^n f(x)|_{x=0}$$

This result holds for any function with a Taylor Series (it does not use any special properties of PGFs).

Exercise A.1. Prove Property A.2 [write out the sum and show that the derivatives eliminate any r_m for $m < n$, the leading coefficient of the result is $n!r_n$, and the later terms are all zero].

There are many contexts in which we can only calculate a function numerically. In this case the calculation of these derivatives is likely to be difficult and inaccurate. An improved way to calculate it is given by a Cauchy integral (Moore & Newman, 2000). This is a standard result of Complex Analysis, and initially we simply take it as given.

$$r_n = \frac{1}{2\pi i} \oint \frac{f(z)}{z^{n+1}} dz$$

This integral can be done on a closed circle around the origin $z = Re^{i\theta}$, in which case $dz = izd\theta$. Then r_n can be rewritten as

$$r_n = \frac{1}{2\pi} \int_0^{2\pi} \frac{f(Re^{i\theta})}{(Re^{i\theta})^{n+1}} d\theta$$

Using another substitution, $\theta = 2\pi u$, we find $d\theta = 2\pi du$ with u varying from 0 to 1. This integral becomes

$$r_n = \int_0^1 \frac{f(Re^{2\pi i u})}{R^{n+1} e^{2n\pi i u}} du$$

The integral on the right hand side can be approximated by a simple summation and we find

$$r_n \approx \frac{1}{M} \sum_{m=1}^M \frac{f(Re^{2\pi i m/M})}{R^{n+1} e^{2n\pi i m/M}}$$

for large M .

A few technical steps show that the PGF $f(z)$ converges for any z with $|z| \leq 1$ (any PGF is analytic within the unit circle $R = 1$ and that the PGF converges everywhere on the unit circle [the coefficients are all positive or zero and the sum converges for $z = 1$, so it converges absolutely on the unit circle]). Thus this integral can be performed for any positive $R \leq 1$ (and in many cases it can be done for larger R). We have found that the unit circle ($R = 1$) yields remarkably good accuracy, so we recommend using it unless there is a good reason not to. Some discussion of identifying the optimal radius appears in (Bornemann, 2011).

Thus we have

Property A.3. Given a PGF $f(x)$, the coefficient of x^n in its expansion can be calculated by the integral

$$r_n = \int_0^1 \frac{f(Re^{2\pi i u})}{R^{n+1} e^{2n\pi i u}} du \tag{A.1}$$

This is well-approximated by the summation

$$r_n \approx \frac{1}{M} \sum_{m=1}^M \frac{f(Re^{2\pi i m/M})}{R^{n+1} e^{2n\pi i m/M}} \tag{A.2}$$

with $R = 1$ and $M \gg 1$.

It turns out that this approach is closely related to the approach to get a particular coefficient of a Fourier Series. Once the variable is changed from z to θ , our function is effectively a (complex) Fourier Series in θ , and the integral corresponds to the standard approach to finding the n th coefficient of a Fourier Series.

Exercise A.2. Verification of Equation (A.1):

In this exercise we show that the formula in Equation (A.1) yields r_n . Assume that the integral is performed on a circle of radius $R \leq 1$ about the origin.

a. Write $f(z) = \sum_m r_m z^m$ and rewrite $\int_0^1 \frac{f(Re^{2\pi i u})}{R^n e^{2n\pi i u}} du$ as a sum

$$\int_0^1 \frac{f(Re^{2\pi i u})}{R^n e^{2n\pi i u}} du = \sum_m r_m \int_0^1 R^{m-n} e^{2(m-n)\pi i u} du$$

- b. Show that for $m = n$ the integral in the summation on the right hand side is 1.
- c. Show that for $m \neq n$, the integral in the summation on the right hand side is 0.
- d. Thus conclude that the integral on the left hand side must yield r_n .

Exercise A.3. Let $f(z) = e^z = 1 + z + z^2/2 + z^3/6 + z^4/24 + z^5/120 + \dots$. Write a program that estimates r_0, r_1, \dots, r_5 using Equation (A.2) with $R = 1$. Report the values to four significant figures for

- a. $M = 2$
- b. $M = 4$
- c. $M = 5$
- d. $M = 10$
- e. $M = 20$.
- f. How fast is convergence for different r_n ?

Appendix A.2. Properties related to distribution moments

We next look at two straightforward properties about the moments of the distribution r_i having PGF $f(x)$. We return to using $i = 0, 1, \dots$ as an indexing variable, so i is no longer $\sqrt{-1}$. We have

$$\begin{aligned} f(1) &= \sum_i r_i 1^i \\ &= \sum_i r_i \\ &= 1 \end{aligned}$$

where the final equality is because the r_i determine a probability distribution. With mildly more effort, we have

$$\begin{aligned} f'(1) &= \sum_i r_i i 1^{i-1} \\ &= \sum_i i r_i \\ &= \mathbb{E}(i) \end{aligned}$$

where $\mathbb{E}(i)$ denotes the expected value of i . These arguments show

Property A.4. Any PGF $f(x)$ must satisfy $f(1) = 1$.

Property A.5. The expected value of a random variable i whose distribution has PGF $f(x)$ is given by $\mathbb{E}(i) = f'(1)$.

It is straightforward to derive relationships for $\mathbb{E}(i^2)$ and higher order moments by repeated differentiation of f and evaluating the result at 1.

Appendix A.3. Properties related to function composition

To motivate function composition, we start with an example.

Example A.1. Consider a weighted coin which comes up ‘Success’ with probability p and ‘Failure’ with probability $1 - p$. We play a game in which we stop at the first failure, and otherwise flip it again. Define $f(x) = px + 1 - p$

Let α_g be the probability of failure within the first g flips. Then $\alpha_0 = 0$ and $\alpha_1 = 1 - p = f(0)$ are easily calculated.

More generally the probability of starting the game and failing immediately is $\alpha_0 = 1 - p = f(0)$ while the probability of having a success and flipping again is p , at which point the probability of failure within $g - 1$ flips is α_{g-1} . So we have $\alpha_g = (1 - p) + p\alpha_{g-1} = f(\alpha_{g-1})$. So using induction we can show that the probability of failure within g generations is $f^{[g]}(0)$.

Exercise A.4. The derivation in Example A.1 was based on looking at what happened after a single flip and then looking $g - 1$ flips into the future in the inductive step. Derive $\alpha_g = f(\alpha_{g-1})$ by instead looking $g - 1$ flips into the future and then considering one additional step. [the distinction between this argument and the previous one becomes useful in the continuous-time case where we use the ‘backward’ or ‘forward’ Kolmogorov equations.]

Exercise A.5. Consider a fair six-sided die with numbers 0, 1, ..., 5, rather than the usual 1, ..., 6. We roll the die once. Then we look at the result, and roll that many copies (if zero, we stop), then we look at the sum of the result and repeat.

$$f(x) = \frac{1 + x + \dots + x^5}{6} = \begin{cases} \frac{x^6 - 1}{6(x - 1)} & x \neq 1 \\ 1 & x = 1 \end{cases}$$

Define α_g to be the probability the process stops after g iterations (with $\alpha_0 = 0$ and $\alpha_1 = 1/6$).

- a. Find an expression for α_g , the probability that by the g 'th iteration the process has stopped, in terms of $f(x)$.
- b. Rephrase this question in terms of the extinction probability for an infectious disease.

Processes like that in Exercise A.1 can be thought of as “birth-death” processes where each event generates a discrete number of new events. Our examples above show that function composition arises naturally in calculating the probability of extinction in a birth-death process. We show below that it also arises naturally when we want to know the distribution of population sizes after some number of generations rather than just the probability of 0.

Specifically, we often assume an initially infected individual causes some random number of new infections i from some distribution. Then we assume that each of those new infections independently causes an additional random number of infections from the same distribution. We will be interested in how to get from the one-generation PGF to the PGF for the distribution after g generations.

We derive this in a few stages.

- We first show that if we take two numbers from different distributions with PGFs $f(x)$ and $h(x)$, then their sum has distribution $f(x)h(x)$ [Property A.6]. Then inductively applying this we conclude that the distribution of the sum of n numbers from a distribution with PGF $f(x)$ has PGF $[f(x)]^n$.
- We also show that if the probability we take a number from the distribution with PGF $f(x)$ is π_1 and the probability we take it from the distribution with PGF $h(x)$ is π_2 , then the PGF of the resulting distribution is $\pi_1 f(x) + \pi_2 h(x)$ [Property A.7].
- Putting these two properties together, we can show that if we choose i from a distribution with PGF $f(x)$ and then choose i different values from a distribution with PGF $h(x)$, then the sum of the i values has PGF $f(h(x))$ [Property A.8].

Our main use of Properties A.6 and A.7 is as stepping stones towards Property A.8.

Consider two probability distributions, let r_i be the probability of i for the first distribution and q_j be the probability of j for the second distribution. Assume they have PGFs $f(x) = \sum_i r_i x^i$ and $h(x) = \sum_j q_j x^j$ respectively.

We are first interested in the process of choosing i from the first distribution, j from the second, and adding them. In the disease context this arises where the two distributions give the probability that one individual infects i and another infects j and we want to know the probability of a particular sum.

The probability of obtaining a particular sum k is

$$\sum_i r_i q_{k-i}$$

So the PGF of the sum is $\sum_k \sum_{i=0}^k r_i q_{k-i} x^k$. By inspection, this is equal to the product $f(x)h(x)$. This means that the PGF of the process where we choose i from the first and j from the second and look at the sum is the product $f(x)h(x)$.

We have shown

Property A.6. Consider two probability distributions, r_0, r_1, \dots and q_0, q_1, \dots with PGFs $f(x) = \sum_i r_i x^i$ and $h(x) = \sum_j q_j x^j$. Then if we choose i from the distribution r_i and j from the distribution q_j , the PGF of their sum is $f(x)h(x)$.

Usually we want the special case where we choose two numbers from the same distribution having PGF $f(x)$. The PGF for the sum is $[f(x)]^2$. The PGF for the sum of three numbers from the same distribution can be thought of as the result of $[f(x)]^2$ and $f(x)$, yielding $[f(x)]^3$. By induction, it follows that the PGF for the sum of i numbers sum is $[f(x)]^i$.

Now we want to know what happens if we are not sure what the current system state is. For example, we might not know if we have 1 or 2 infected individuals, and the outcome at the next generation is different based on which it is.

We use the distributions r_i and q_j . We assume that with probability π_1 we choose a random number k from the r_i distribution, while with probability $\pi_2 = 1 - \pi_1$ it is chosen from the q_j distribution. Then the probability of a particular value k occurring is $\pi_1 r_k + \pi_2 q_k$, and the resulting PGF is $\sum_k (\pi_1 r_k + \pi_2 q_k) x^k = \pi_1 f(x) + \pi_2 h(x)$. This becomes:

Property A.7. Consider two probability distributions, r_0, r_1, \dots and q_0, q_1, \dots with PGFs $f(x) = \sum_i r_i x^i$ and $h(x) = \sum_j q_j x^j$. We consider a new process where with probability π_1 we choose k from the r_i distribution and with probability $\pi_2 = 1 - \pi_1$ we choose k from the q_j distribution. Then the PGF of the resulting distribution is $\pi_1 f(x) + \pi_2 h(x)$.

We finally consider a process in which we have two distributions with PGFs $f(x) = \sum_i r_i x^i$ and $h(x) = \sum_j q_j x^j$. We choose the number i from the distribution r_i and then take the sum of i values chosen from the q_j distribution, $\sum_{\ell=1}^i j_\ell$. Both the number of terms in the sum and their values are random variables. Using the results above, the PGF of the resulting sum is $\sum_i r_i [h(x)]^i = f(h(x))$. Thus we have

Property A.8. Consider two probability distributions, r_0, r_1, \dots and q_0, q_1, \dots with PGFs $f(x) = \sum_i r_i x^i$ and $h(x) = \sum_j q_j x^j$. Then if we choose i from the distribution r_i and then take the sum of i values chosen from the distribution q_j , the PGF of the sum of those i values is $f(h(x))$.

This property is closely related to the spread of infectious disease. An individual may infect i others, and then each of them causes additional infections. The number of these second generation cases is the sum of i random numbers $\sum_{\ell=1}^i j_\ell$ where j_ℓ is the number of additional infections caused by the ℓ -th infection caused by the initial individual. So if $f(x)$ is the PGF for the distribution of the number of infections caused by the first infection and $h(x)$ is the PGF for the distribution of the number of infections caused by the offspring, then $f(h(x))$ is the PGF for the number infected in the second generation [and if the two distributions are the same this is $f^{[2]}(x)$]. Repeated iteration gives us the distribution after g generations.

Exercise A.6. Note that if we interchange p and q in the PGF of the negative binomial distribution in Table 1, it is simply the PGF of the geometric distribution raised to the power \hat{r} . A number chosen from the negative binomial can be defined as the number of successful trials (each with success probability p) before the \hat{r} th failure.

Using this and Property A.8, derive the PGF of the negative binomial.

Exercise A.7. Sicherman dice (Gallian & Rusin, 1979; Gardner, 1978).

To motivate this exercise consider two tetrahedral dice, numbered 1, 2, 3, 4. When we roll them we get sums from 2 to 8, each with its own probability, which we can infer from this table:

	2	3	4	5
	3	4	5	6
	4	5	6	7
	5	6	7	8

However another pair of tetrahedral dice, labelled 1, 2, 2, 3 and 1, 3, 3, 5 yields the same sums with the same probabilities:

	2	3	3	4
	4	5	5	6
	4	5	5	6
	6	7	7	8

We now try to find a similar pair for 6-sided dice. First consider a pair of standard 6-sided dice.

- Show that the PGF of each die is $f(x) = (x + x^2 + x^3 + x^4 + x^5 + x^6)/6$.
- Fill in the tables showing the possible sums from rolling two dice (fill in each square with the sum of the two entries) and multiplication for two polynomials (fill in each square with the product of the two entries):

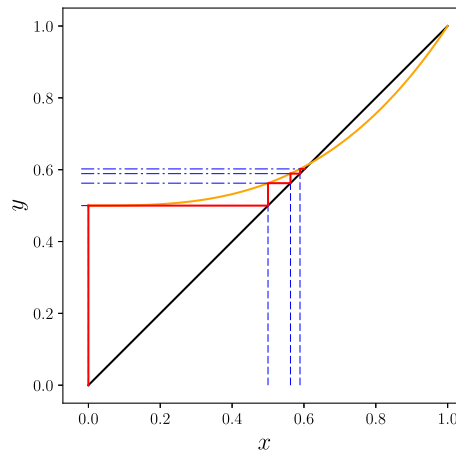


Fig. A.10. Cobweb diagrams: We take the function $f(x) = (1 + x^2)/2$. A cobweb diagram is built by alternately drawing vertical lines from the diagonal to $f(x)$ and then horizontal lines from $f(x)$ to the diagonal. The dashed lines show $\alpha_g = f(\alpha_{g-1})$ starting with $\alpha_0 = 0$ and highlight the relation to the iterative process.

Exercise A.8. Early-time outbreak dynamics

- a Consider normal dice. The PGF is $f(x) = (x + x^2 + x^3 + x^4 + x^5 + x^6)/6$. Consider the process where we roll a die, take the result i , and then roll i other dice and look at their sum. What is the PGF of the resulting sum in terms of f ?
- b If an infected individual causes anywhere from 1 to 6 infections, all with equal probability, find the PGF for the number of infections in generation 2 if there is one infection in generation 0. [you can express the result in terms of f]
- c And in generation g (assuming depletion of susceptibles is unimportant)?

Appendix A.4. Properties related to iteration of PGFs

There are various contexts in which we might iterate to calculate $f^{[n]}(x)$ (the result of applying f n times to x).

In the disease context, this occurs most frequently in calculating the probability of outbreak extinction. If we think of α as the probability that the outbreak goes extinct from a single individual, then from Property A.1 we would expect that $\alpha = f(\hat{\alpha})$ where $\hat{\alpha}$ is the probability that an offspring of the individual fails to produce an epidemic. However, under common assumptions, the number of infections from the offspring should be from the same distribution as from the parent. In this case we would conclude $\alpha = \hat{\alpha}$ and so $\alpha = f(\alpha)$.

It turns out that a good way to solve for α is iteration, starting with the guess $\alpha_0 = 0$. We will show that this converges to the correct value [$x = f(x)$ can have multiple solutions, only one of which is the correct α].

Fig. A.10 demonstrates how the iterative process can be represented by a “cobweb diagram” (May 1976; Peitgen, Jürgens, & Saupe, 2006) To use a cobweb diagram to study the behavior of $f^{[g]}(x_0)$, we draw the line $y = x$ and the curve $y = f(x)$. Then at x_0 we draw a vertical line to the curve $y = f(x)$. We draw a horizontal line to the line $y = x$ [which will be at the point (x_1, x_1)]. We then repeat these steps, drawing a vertical line to $y = f(x)$ and a horizontal line to $y = x$. Cobweb diagrams are particularly useful in studying behavior near fixed or periodic points.

Exercise A.9. Understanding cobweb diagrams

From Fig. A.10 the origin of the term “cobweb” may be unclear. Because of properties of PGFs, the more interesting behavior does not occur for our applications. Here we investigate cobweb diagrams in more detail for non-PGF functions. Since we use $f(x)$ to denote a PGF, in this exercise we use $z(x)$ for an arbitrary function.

- a. Consider the line $z(x) = 2(1 - x)/3$. Starting with $x_0 = 0$, show how the first few iterations of $x_i = z(x_{i-1})$ can be found using a cobweb diagram (do not explicitly calculate the values).
- b. Now consider the line $z(x) = 2(1 - x)$. The solution to $z(x) = x$ is $x = 2/3$. Starting from an initial x_0 close to (but not quite equal to) $2/3$, do several iterations of the cobweb diagram graphically.
- c. Repeat this with the lines $z(x) = 1/4 + x/2$ starting at $x_0 = 0$ and $z(x) = -1 + 3x$ starting close to where $x = z(x)$.
- d. What is different when the slope is positive or negative?
- e. Can you predict what condition on the slope’s magnitude leads to convergence to or divergence from the solution to $x = z(x)$ when z is a line?

So far we have considered lines $z(x)$. Now assume $z(x)$ is nonlinear and consider the behavior of cobweb diagrams close to a point where $x = z(x)$.

f. Use Taylor Series to argue that (except for degenerate cases where z' is 1 at the intercept) it is only the slope at the intercept that determines the behavior sufficiently close to the intercept.

Exercise A.10. Structure of fixed points of $f(x)$.

Consider a PGF $f(x) = \sum_i r_i x^i$, and assume $r_0 > 0$.

- Show that $f(1) = 1$ and $f(0) > 0$.
- Show that $f(x)$ is convex (that is $f''(x) \geq 0$) for $x > 0$. [hint $r_i \geq 0$ for all i]
- Thus argue that if $f'(1) \leq 1$, then $x = f(x)$ has only one solution to $x = f(x)$ in $[0, 1]$, namely $f(1) = 1$. It may help to draw pictures of $f(x)$ and the function $y = x$ for x in $[0, 1]$.
- Explain why if there is a point $x_0 \neq 1$ where $f(x_0) = x_0$ and $f(x) > x$ for x in some region (x_0, x_1) then $0 < f'(x_0) < 1$.
- Thus show that if $f'(1) > 1$ then there are exactly two solutions to $x = f(x)$ in $[0, 1]$, one of which is $x = 1$.

These results suggest:

Property A.9. Assume $f(x) = \sum_i r_i x^i$ is a PGF, and $f(0) > 0$.

- If $f'(1) \leq 1$ then the only intercept of $x = f(x)$ in $[0, 1]$ is at $x = 1$.
- Otherwise, there is another intercept x^* , $0 < x^* < 1$, and if $x < x^*$ then $x < f(x) < x^*$ while if $x > x^*$ then $x > f(x) > x^*$ and for $0 \leq x_0 < 1$, $f^{[g]}(x_0)$ converges monotonically to x^* .

The assumption $r_0 > 0$ was used to rule out $f(x) = x$. Excluding this degenerate case, these results hold even if $r_0 = 0$, in which case we can show $f'(1) > 1$ and $x^* = 0$.

To sketch the proof of this property, we note that clearly $f(1) = 1$, so if $f(0) > 0$ then either $f(x)$ crosses $y = x$ at some intermediate $0 < x^* < 1$ or it does not cross until $x = 1$. Then using the fact that for $x > 0$ the slope of f is positive and increasing, we can inspect the cobweb diagram to see these results.

Appendix A.5. Finding the Kolmogorov Equations

To study continuous-time dynamics, we will want to have partial differential equations (PDEs) where we write the time derivative of a PGF $f(x, t)$ or $f(x, y, t)$ in terms of f and its spatial derivatives.

We will use two approaches to find the derivative. Both start with the assumption that we know $f(x, t)$, and calculate the derivative by finding $f(x, t + \Delta t)$ and use the definition of the derivative:

$$\frac{\partial}{\partial t} f(x, t) = \lim_{\Delta t \rightarrow 0} \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t}$$

The methods differ in how they find $f(x, t + \Delta t)$. The distinction is closely related to the observation in [Exercise 2.7](#) that $\mu^{[g]}(x)$ can be written as either $\mu^{[g-1]}(\mu(x))$ or $\mu(\mu^{[g-1]}(x))$.

- The first involves assuming we know $f(x, t)$ and then looking through all of the possible transitions to find how the system changes going from t to $t + \Delta t$. This will yield the forward Kolmogorov equations.
- The second involves starting from the initial condition $f(x, 0)$ and finding $f(x, \Delta t)$ by investigating all of the possible transitions. Then taking $f(x, \Delta t)$ and $f(x, t)$ we are able to find $f(x, t + \Delta t)$. This will yield the backward Kolmogorov equations.

Appendix A.5.1. Forward Kolmogorov Equations

We start with the forward Kolmogorov equations. We let $r_i(t)$ denote the probability that at time t there are i individuals, and define the PGF

$$f(x, t) = \sum_i r_i(t) x^i$$

We begin by looking at events that can be treated as if they remove one individual and replace it with m individuals. Thus i is replaced by $i + m - 1$:

$$i \mapsto i + m - 1 .$$

For example early in an epidemic, we may assume that an infected individual causes new infections at rate β . The outcome of an infection event is equivalent to the removal of the infected individual and replacement by two infected individuals. Similarly, a recovery event occurs with rate γ and is equivalent to removal with no replacement. So $\lambda_2 = \beta$, $\lambda_0 = \gamma$, and all other λ_m are 0.

Our events happen at a per-individual rate λ_m , so the total rate an event occurs across the population of i individuals is $\lambda_m i$. Events that can be modeled like this include decay of a radioactive particle, recovery of an infected individual, or division of a cell. We assume that different events may be possible, each having a different m . If multiple events have the same effect on m (for example emigration or death), we can combine their rates into a single λ_m .

It will be useful to define

$$\Lambda = \sum_m \lambda_m$$

to be the combined per-capita rate of all possible events and

$$h(x) = \sum_m \lambda_m x^m / \Lambda$$

We can think of $h(x)$ as the PGF for the number of new individuals given that a random event happens (since λ_m / Λ is the probability that the random event introduces m individuals).

We start with one derivation of the equation for $\dot{f}(x, t)$ based on directly calculating $f(x, t + \Delta t)$ and using the definition of the derivative. An alternate way is shown in Exercise A.11. For small Δt the probability that multiple events occur in the same time interval is $\mathcal{O}(\Delta t)$, and we will see that this is negligible. Let us assume the system has i individuals at time t , which occurs with probability $r_i(t)$. For a given m , the probability that the event occurs in the time interval given i is $\lambda_m i \Delta t + \mathcal{O}(\Delta t)$, and $(1 - \sum_m \lambda_m i \Delta t) + \mathcal{O}(\Delta t)$ measures the probability that none of the events occur in the time interval and the system remains in state i . If the event occurs, the system leaves the state corresponding to x^i and enters the state corresponding to x^{i+m-1} . Summing over m and i , we have

$$f(x, t + \Delta t) = \sum_i \left(r_i(t) \left[\sum_m (\lambda_m i \Delta t) x^{i+m-1} + \left(1 - \sum_m \lambda_m i \Delta t \right) x^i \right] \right) + \mathcal{O}(\Delta t)$$

The $\mathcal{O}(\Delta t)$ corrects for the possibility of multiple events happening in the time interval. A bit of algebra and separating the i and m summations shows that

$$\begin{aligned} f(x, t + \Delta t) &= \sum_i r_i(t) x^i + \sum_m \lambda_m (\Delta t) (x^m - x) \sum_i r_i(t) i x^{i-1} + \mathcal{O}(\Delta t) \\ &= f(x, t) + \sum_m \lambda_m (x^m - x) \Delta t \sum_i r_i(t) \frac{\partial}{\partial x} x^i + \mathcal{O}(\Delta t) \\ &= f(x, t) + (\Delta t) \left(\sum_m \lambda_m x^m - x \sum_m \lambda_m \right) \frac{\partial}{\partial x} \sum_i r_i(t) x^i + \mathcal{O}(\Delta t) \\ &= f(x, t) + \Lambda (\Delta t) [h(x) - x] \frac{\partial}{\partial x} f(x, t) + \mathcal{O}(\Delta t) \end{aligned}$$

So we now have

$$\begin{aligned} \frac{\partial}{\partial t} f(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{f(x, t) + \Lambda \Delta t [h(x) - x] \frac{\partial}{\partial x} f(x, t) + \mathcal{O}(\Delta t) - f(x, t)}{\Delta t} \\ &= \Lambda [h(x) - x] \frac{\partial}{\partial x} f(x, t) \end{aligned}$$

We finally have

Property A.10. Let $f(x, t) = \sum_i r_i(t) x^i$ be the PGF for the probability of having i individuals at time t . Assume several events indexed by m can occur, each with rate $\lambda_m i$, that remove one individual and replace it with m . Let $\Lambda = \sum_m \lambda_m$ be the total per-capita rate and $h(x) = \sum_m \lambda_m x^m / \Lambda$ be the PGF of the outcome of a random event. Then

$$\frac{\partial}{\partial t} f(x, t) = \Lambda[h(x) - x] \frac{\partial}{\partial x} f(x, t) \tag{A.3}$$

We look at a heuristic way to interpret this. We can rewrite Equation (A.3) as

$$\dot{f}(x, t) = \left[\sum_m (\lambda_m x^m - \lambda_m x) \right] \frac{\partial}{\partial x} f(x, t)$$

Then if we expand f on the right hand side, we have

$$\sum_m \sum_i \lambda_m (x^m - x) i r_i x^{i-1}$$

The derivative serves the purpose of getting the factor i into the coefficient of each term which addresses the fact that the rate events happen is proportional to the total count. The derivative has the additional effect of reducing the exponent by 1, corresponding to the removal of one individual. The λ_m in the remaining factor gives the per-capita rate of changing state. The $x^m - x$ captures the fact that when moving to that new state m individuals are added but the system is leaving the current state (which has an exponent of x^i) with the same rate.

Exercise A.11. Alternate derivation of Equation (A.3)

An alternate way to derive Equation (A.3) is through directly calculating \dot{r}_i .

- a. Explain why $\dot{r}_i = -\sum_m \lambda_m i r_i + \sum_m \lambda_m (i - m + 1) r_{i-m+1}$.
- b. Taking $\dot{f}(x, t) = \sum_i \dot{r}_i x^i$, derive Equation (A.3).

We can generalize this to the case where there are multiple types of individuals. For the forward Kolmogorov equations, it is relatively straightforward to allow for interactions between individuals. We may be interested in this generalization when considering predator-prey interactions or interactions between infected and susceptible individuals if we are interested in depletion of susceptibles. We assume that there are two types of individuals A and B with counts i and j respectively, and we let $r_{ij}(t)$ denote the probability of a given pair i and j . We define the PGF

$$f(x, y, t) = \sum_{ij} r_{ij}(t) x^i y^j$$

We assume that interactions between an A and a B individual occur with some rate proportional to the product ij . We assume that the interaction removes both individuals and replaces them by m of type A and n of type B . We denote the rate as $\mu_{m,n} ij$, and the sum

$$\mathfrak{M} = \sum_{m,n} \mu_{m,n}$$

We also assume that individuals of type A spontaneously undergo changes as they did above, but they can be replaced by type A and/or type B individuals. So one individual of type A is removed and replaced by m individuals of type A and n of type B with rate $\lambda_{m,n}$, and the combined rate for one specific transition over the entire set of individuals is $\lambda_{m,n} i$. We define

$$\Lambda = \sum_{m,n} \lambda_{m,n}$$

We will ignore spontaneous changes by nodes of type B , but the generalization to include these can be found by following the same method.

Finally, let

$$h(x, y) = \sum_{m,n} \lambda_{m,n} x^m y^n / \Lambda$$

and

$$g(x, y) = \sum_{m,n} \mu_{m,n} x^m y^n / \mathfrak{M}$$

be the PGFs for the outcomes of the two types of events.

Then

$$\begin{aligned}
 f(x, y, t + \Delta t) &= \sum_{i,j} r_{i,j}(t) \left[\sum_{m,n} \left[(\lambda_{m,n} i \Delta t) x^{i+m-1} y^j + (\mu_{m,n} i j \Delta t) x^{i+m-1} y^{j+n-1} \right. \right. \\
 &\quad \left. \left. + \left(1 - \sum_{m,n} \left[\lambda_{m,n} i \Delta t + \mu_{m,n} i j \Delta t \right] \right) x^i y^j \right] + \mathcal{O}(\Delta t) \right] \\
 &= \sum_{i,j} r_{i,j}(t) x^i y^j + \sum_{m,n} \lambda_{m,n} (x^m y^n - x) \Delta t \sum_{i,j} r_{i,j}(t) i x^{i-1} y^j \\
 &\quad + \sum_{m,n} \mu_{m,n} (x^m y^n - xy) \Delta t \sum_{i,j} r_{i,j}(t) i j x^{i-1} y^{j-1} + \mathcal{O}(\Delta t) \\
 &= f(x, y, t) + \sum_{m,n} \lambda_{m,n} (x^m y^n - x) \Delta t \frac{\partial}{\partial x} f(x, y, t) \\
 &\quad + \sum_{m,n} \mu_{m,n} (x^m y^n - xy) \Delta t \frac{\partial}{\partial x} \frac{\partial}{\partial y} f(x, y, t) + \mathcal{O}(\Delta t) \\
 &= f(x, y, t) + (\Delta t) \left(\Lambda [h(x, y) - x] \frac{\partial}{\partial x} f(x, y, t) + \mathfrak{W} [g(x, y) - xy] \frac{\partial}{\partial x} \frac{\partial}{\partial y} f(x, y, t) \right) + \mathcal{O}(\Delta t)
 \end{aligned}$$

So

$$\begin{aligned}
 \frac{\partial}{\partial t} f(x, y, t) &= \lim_{\Delta t \rightarrow 0} \frac{f(x, y, t + \Delta t) - f(x, y, t)}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{(\Delta t) \Lambda [h(x, y) - x] \frac{\partial}{\partial x} f(x, y, t) + (\Delta t) \mathfrak{W} [g(x, y) - xy] \frac{\partial}{\partial x} \frac{\partial}{\partial y} f(x, y, t) + \mathcal{O}(\Delta t)}{\Delta t} \\
 &= \Lambda [h(x, y) - 1] x \frac{\partial}{\partial x} f(x, y, t) + \mathfrak{W} [g(x, y) - 1] xy \frac{\partial}{\partial x} \frac{\partial}{\partial y} f(x, y, t)
 \end{aligned}$$

We have shown:

Property A.11. Let $f(x, y, t) = \sum_{i,j} r_{i,j}(t) x^i y^j$ be the PGF for the probability of having i type A and j type B individuals. Assume that events occur with rate $\lambda_{m,n} i$ or $\mu_{m,n} i j$ to replace a single type A individual or one of each type with m type A and n type B individuals. Let $\Lambda = \sum_{m,n} \lambda_{m,n}$ and $\mathfrak{W} = \sum_{m,n} \mu_{m,n}$. Then

$$\frac{\partial}{\partial t} f(x, y, t) = \Lambda [h(x, y) - x] \frac{\partial}{\partial x} f(x, y, t) + \mathfrak{W} [g(x, y) - xy] \frac{\partial}{\partial x} \frac{\partial}{\partial y} f(x, y, t) \tag{A.4}$$

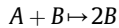
where $h(x, y) = \sum_{m,n} \lambda_{m,n} x^m y^n / \Lambda$ is the PGF for the outcome of a random event whose rate is proportional to i and $g(x, y) = \sum_{m,n} \mu_{m,n} x^m y^n / \mathfrak{W}$ is the PGF for the outcome of a random event whose rate is proportional to ij .

This can be generalized further if there are events whose rates are proportional only to j or if there are more than two types. The exercise below shows how to generalize this if the rate of events depend on i in a more complicated manner.

Exercise A.12. In many cases interactions between two individuals of the same type are important. These may occur with rate $i(i - 1)$ or i^2 depending on the specific details. Assume we have only a single type of individual with PGF $f(x, t) = \sum_i r_i(t) x^i$.

- a. If a collection of events to replace two individuals with m individuals occur with rate $\beta_m i(i - 1)$, find how write a PDE for f . Your final result should contain $\frac{\partial^2}{\partial x^2} f(x, t)$. Use $\mathfrak{B} = \sum_m \beta_m$ and $g(x) = \sum_m \beta_m x^m / \mathfrak{B}$. Follow the derivation of Equation (A.3).
- b. If instead the events replace two individuals with m individuals and occur with rate $\beta_m i^2$, find how to incorporate them into a PDE for f . Your final result should contain $\frac{\partial}{\partial x} \left(x \frac{\partial}{\partial x} f(x, t) \right)$ or equivalently $\frac{\partial}{\partial x} f(x, t) + x \frac{\partial^2}{\partial x^2} f(x, t)$.

Exercise A.13. Consider a chemical system that begins with some initial amount of chemical A. Let i denote the number of molecules of species A. A molecule of A spontaneously degrades into a molecule of B, with rate ξ per molecule. Let j denote the number of molecules of species B. Species B reacts with A at rate η_{ij} to produce new molecules of species B. The reactions are denoted



Let $r_{ij}(t)$ denote the probability of i molecules of A and j molecules of B at time t . Let $f(x, y, t) = \sum_i r_i(t) x^i y^j$ be the PGF. Find the forward Kolmogorov equation for $f(x, y, t)$.

Appendix A.5.2. Backward Kolmogorov Equations

We now look for another derivation of $\frac{\partial}{\partial t} f(x, t)$, and as before we find it by first finding $f(x, t + \Delta t)$ for small Δt and then using the definition of the derivative. We will assume that each individual acts independently, and at rate λ_m an individual may be removed and replaced by m new individuals. So if there are i total individuals, at rate $\lambda_m i$ the count i is replaced by $i - 1 + m$.

Property A.8 plays an important role in our derivation. We define $f_1(x, t) = \sum_i r_i(t) x^i$ where we assume that $r_1(0) = 1$, that is we start with exactly one individual at time 0. Then Property A.8 shows that $f_1(x, t_1 + t_2) = f_1(f_1(x, t_2), t_1)$. Then from our initial condition $f_1(x, 0) = x$, and

$$f_1(x, \Delta t + t) = f_1(f_1(x, t), \Delta t) \tag{A.5}$$

We need to find $f_1(x, \Delta t)$. We have

$$\begin{aligned} f_1(x, \Delta t) &= \sum_i r_i(0) x^i \left(1 - \sum_m i \lambda_m (\Delta t) + \sum_m i \lambda_m (\Delta t) x^{m-1} \right) + \mathcal{O}(\Delta t) \\ &= x \left(1 - \sum_m \lambda_m (\Delta t) + \sum_m \lambda_m (\Delta t) x^{m-1} \right) + \mathcal{O}(\Delta t) \\ &= x - x(\Delta t) \sum_m \lambda_m + \sum_m \lambda_m x^m + \mathcal{O}(\Delta t) \\ &= x + (\Delta t) \Lambda [h(x) - x] + \mathcal{O}(\Delta t) \end{aligned}$$

where, as in the forward Kolmogorov case, $\Lambda = \sum_m \lambda_m$ and $h(x) = \sum_m \lambda_m x^m / \Lambda$ is the PGF of the number of new individuals created given that an event occurs. In the first step we used the fact that for $f_1(x, t)$, $r_i(0) = 1$ if $i = 1$ and otherwise it is 0. Thus Equation (A.5) implies

$$f_1(x, t + \Delta t) = f_1(x, t) + (\Delta t) \Lambda [h(f_1(x, t)) - f_1(x, t)] + \mathcal{O}(\Delta t) .$$

Now taking the definition of the derivative, we have

$$\begin{aligned} \frac{\partial}{\partial t} f_1(x, t) &= \lim_{\Delta t \rightarrow 0} \frac{f_1(x, t + \Delta t) - f_1(x, t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{f_1(x, t) + (\Delta t) \Lambda [h(f_1(x, t)) - f_1(x, t)] + \mathcal{O}(\Delta t) - f_1(x, t)}{\Delta t} \\ &= \Lambda [h(f_1(x, t)) - f_1(x, t)] \end{aligned}$$

Thus we have an ODE for $f_1(x, t)$.

In general, our initial condition may not be a single individual, but some other number (or perhaps a value chosen from a distribution). Let the initial condition have PGF $f(x, 0)$. Then it follows from Property A.8 that

$$f(x, t) = f(f_1(x, t), 0)$$

So we have

Property A.12. Consider a process in which the number of individuals change in time such that when an event occurs one individual is destroyed and replaced with m new individuals. The associated rate associated with an event that changes the population size by m is $\lambda_m i$ where i is the number of individuals. Let $f_1(x, t)$ be the PGF for this process beginning from a single individual and $\Lambda = \sum_m \lambda_m$. Then

$$\dot{f}_1(x, t) = \Lambda[h(f_1(x, t)) - f_1(x, t)] \tag{A.6}$$

where $h(x)$ is the PGF for the number of new individuals created in a random event. If the initial number of individuals is not 1, let $f(x, 0)$ denote the PGF for the initial condition. Then

$$f(x, t) = f(f_1(x, t), 0) \tag{A.7}$$

is the PGF at arbitrary positive time.

This is fairly straightforward to generalize to multiple types as long as none of the events involve interactions.

Exercise A.14. In this exercise we generalize Property A.12 for the case where there are two types of individuals A and B with counts i and j .

Assume events occur spontaneously with rate $\lambda_{m,n,i}$ to remove an individual of type A and replace it with m of type A and n of type B, or they occur spontaneously with rate $\zeta_{m,n,j}$ to remove an individual of type B and replace it with m of type A and n of type B.

Set $\Lambda = \sum_{m,n} \lambda_{m,n}$ and $\mathfrak{Z} = \sum_{m,n} \zeta_{m,n}$. Let $f_{1,0}(x, y, t)$ denote the outcome beginning with one individual of type A and $f_{0,1}(x, y, t)$ denote the outcome beginning with one individual of type B.

- Write $f_{1,0}(x, y, \Delta t)$ and $f_{0,1}(x, y, \Delta t)$ in terms of $h(x, y) = \sum \lambda_{m,n} x^m y^n / \Lambda$ and $g(x, y) = \sum \zeta_{m,n} x^m y^n / \mathfrak{Z}$.
- Use Property A.8, write $f_{1,0}(x, \Delta t + t)$ and $f_{0,1}(x, \Delta t + t)$ in terms of $f_{1,0}$ and $f_{0,1}$ evaluated at t and Δt . The answer should resemble Equation (A.5).
- Derive expressions for $\frac{\partial}{\partial t} f_{1,0}(x, y, t)$ and $\frac{\partial}{\partial t} f_{0,1}(x, y, t)$.
- Use this to derive Equation (22).

Appendix B. Proof of Theorems 2.7 and 3.6

We now prove Theorems 2.7 and 3.6.

Appendix B.1. Theorem 2.7

We take as given a probability distribution so that p_i is the probability of i offspring.

We will first show a way to represent a (finite) transmission tree as a sequence of integers representing the number of offspring of each node. Additionally we show that the possible sequences coming from a tree can be characterized by a few specific properties. Then the probability of such a sequence corresponds to the probability of the corresponding tree.

Given a finite transmission tree \mathcal{T} , we first order the offspring of any individual (randomly) from “left” to “right”. We then construct a sequence \mathcal{S} by performing a depth-first traversal of the tree and recording the number of offspring as we visit the nodes of the tree, as shown in Fig. B.11. A sequence constructed in this way is called a Łukasiewicz word (Stanley, 2001).

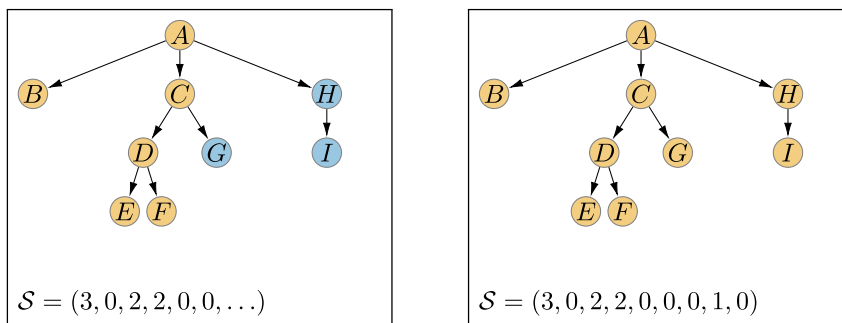


Fig. B11. Demonstration of the steps mapping the tree \mathcal{T} to the sequence \mathcal{S} . The nodes are traced in a depth-first traversal and their number of offspring is recorded. For the labeling given, a depth-first traversal traces the nodes in alphabetical order. At an intermediate stage (left) the traversal has not finished the sequence. The final sequence (right) is uniquely determined once the order of a node’s offspring is (randomly) chosen.

It is straightforward to see that if we are given a Łukasiewicz word \mathcal{S} , we can uniquely reconstruct the (ordered) tree \mathcal{T} from which it came. We now demonstrate the relation between the probability of a given tree \mathcal{T} and the probability of observing a given sequence \mathcal{S} .

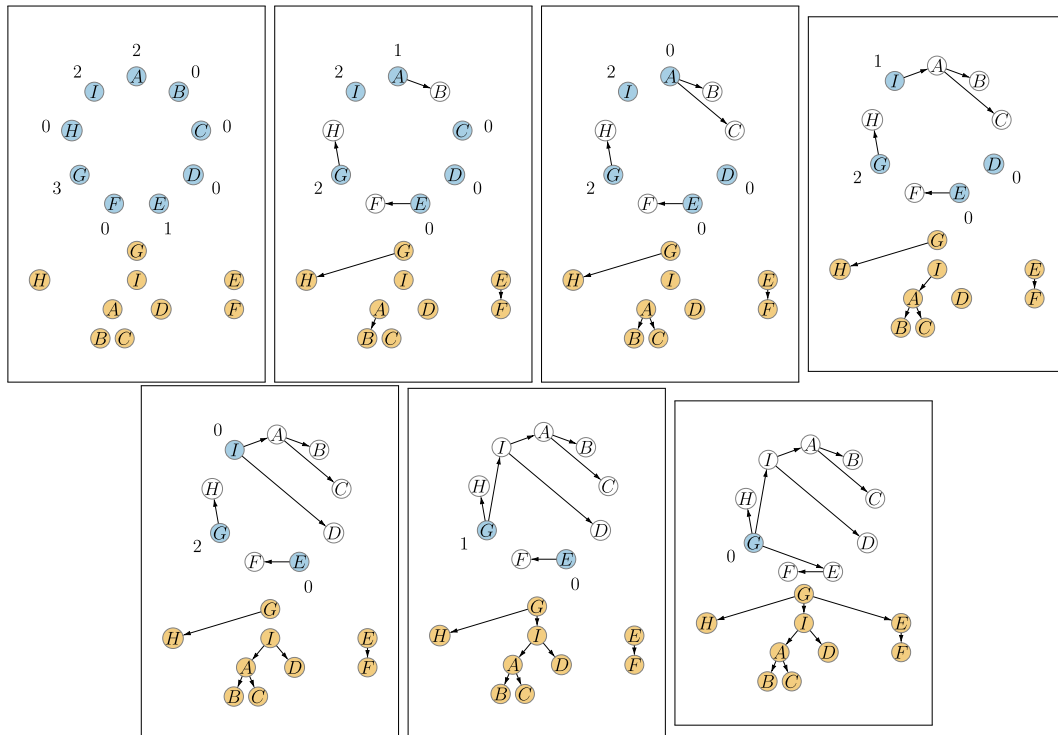


Fig. B.12. The steps of the construction of a tree with $\mathcal{S} = (2, 0, 0, 0, 1, 0, 3, 0, 2)$ [a cyclic permutation of the previous \mathcal{S}]. Each frame shows next step in building a tree on a ring. The resulting tree is not rooted at the top. The names of the nodes in the tree are a cyclic permutation of the original.

We first note that the probability of observing a given length- j sequence \mathcal{S} by choosing j numbers from the offspring distribution is simply $\pi_{\mathcal{S}} = \prod_{s_i \in \mathcal{S}} p_{s_i}$.

Similarly, as infection spreads, each infected individual infects some number s_i with probability p_{s_i} . The probability of a given tree is thus $\prod_{i=1}^j p_{s_i}$. Note that the different trees may have different sizes j . However, for a given tree with a given j and corresponding Łukasiewicz word \mathcal{S} , the probability of observing this sequence by choosing j numbers from the offspring distribution is $\prod_{i=1}^j p_{s_i}$, equal to the *a priori* probability of observing the tree.

Now we look for the probability that a random length- j sequence created by choosing numbers from the offspring distribution actually corresponds to a tree, that is it is actually a Łukasiewicz word.⁵ To be a Łukasiewicz word, the sequence clearly must satisfy that $\sum_{s_i \in \mathcal{S}} s_i = j - 1$ because the sum is the total number of transmissions occurring which is one less than

the total number of infections. Momentarily we will show that given a length- j sequence which sums to $j - 1$, exactly one of its j cyclic permutations is a Łukasiewicz word. Since each cyclic permutation has the same probability of being observed as a sequence, we will conclude that the probability of observing a tree of length j is exactly $1/j$ times the probability that the entries in a length- j sequence of values chosen from the offspring distribution sum to $j - 1$.

We now prove the final detail. Given a length j sequence \mathcal{S} of non-negative integers that sum to $j - 1$, we place j nodes on a ring starting at the top and ordered clockwise. We label each i th node with s_i . If a node v is labelled with 0 and the adjacent position in the counter-clockwise direction has node u with a positive label, we place an edge from u to v (with v to the right of any previous edge from u to another node) and remove v . We decrease u 's label by one.

This process reduces both the number of nodes and the sum by one. So the sum remains one less than the number of nodes. This guarantees at least one zero and at least one nonzero value until only one node remains. Thus we can always find an appropriate pair u and v until only a single node remains. The process constructs a tree (the final outcome has j nodes and $j - 1$ edges and is connected). Fig. B.12 demonstrates the steps.

If the tree is rooted at the node that began at the top of the ring, then \mathcal{S} corresponds to a depth-first traversal of that tree. If the tree is not rooted at the top, then a rotation of the sequence so that the root is at the top will result in a sequence which corresponds to a tree. All cyclic rotations of any length- j sequence summing to $j - 1$ are equally probable as sequences, but

⁵ If the sequence is not a Łukasiewicz word, then either it is the start of a sequence corresponding to a larger (possibly infinite) tree, or some initial subsequence corresponds to a completed tree.

only one corresponds to a tree. Thus the probability that a random length- j sequence summing to $j - 1$ corresponds to a tree is $1/j$.

So we finally conclude that the probability of a tree of j nodes is equal to $1/j$ times the probability that j randomly-chosen values from the offspring distribution sum to $j - 1$. By repeated application of Property A.6, this is $1/j$ times the coefficient of y^{j-1} in $[\mu(y)]^j$ as [Theorem 2.7](#) claims.

Appendix B.2. Theorem 3.6

We can prove [Theorem 3.6](#) as a special case of [Theorem 2.7](#) by calculating the offspring distribution ([Exercise B.1](#)). However, a more illuminating proof is by noting that if we treat a transmission event as a node disappearing and being replaced by two infected nodes and a recovery event as a node disappearing with no offspring, then we have a tree where each node has 2 or 0 offspring. The total number of actual individuals infected in the outbreak is equal to the number of nodes with 0 offspring in the tree.

Following the arguments above, we are looking for sequences of length $2j - 1$ in which 2 appears $j - 1$ times and 0 appears j times. There are $\binom{2j-1}{j-1}$ such sequences. The probability of each is $\beta^{j-1}\gamma^j/(\beta + \gamma)^{2j-1}$ and a fraction $1/(2j - 1)$ of these correspond to trees. Thus, the probability a length $2j - 1$ sequence is a Łukasiewicz word is

$$\frac{1}{2j-1} \frac{\beta^{j-1}\gamma^j}{(\beta + \gamma)^{2j-1}} \binom{2j-1}{j-1} = \frac{1}{j} \frac{\beta^{j-1}\gamma^j}{(\beta + \gamma)^{2j-1}} \binom{2j-2}{j-1}$$

Using the same approach as before, we conclude that this is the probability of exactly j infections.

Exercise B.1. *If we do not think of an infected individual as disappearing and being replaced by two infected individuals when a transmission happens, but rather, we count up all of the transmissions the individual causes, we get a geometric distribution with $q = \beta/(\beta + \gamma)$. The details are in [Exercise 3.2](#). Use this along with [Theorem 2.7](#) and [Table 6](#) (which was derived in [Exercise 2.13](#)) to give a different proof of [Theorem 3.6](#).*

Appendix C. Software

We have produced a python package, `Invasion_PGF` which can be used to solve the equations of [Section 2](#) or [Section 3](#) once the PGF of the offspring distribution or β and γ are determined. Because the numerical method involves solving differential equations in the complex plane, it requires an integration routine that can handle complex values. For this we use `odeintw` ([Weckesser](#)).

[Table C.7](#) briefly summarizes the commands available in `Invasion_PGF`.

We now demonstrate a sample session with these commands.

```
>>> import Invasion_PGF as pgf
>>> def mu(x):
...     return (1 + x + x**2 + x**3)/4.
...
>>> pgf.R0(mu)
1.5000001241105565
>>> #extinction probabilities up to generation 5
>>> pgf.extinction_prob(mu, 5, intermediate_values = True)
array([ 0.          ,  0.25         ,  0.33203125,  0.36972018,  0.38923784,
        0.39992896])
>>> #following commands look at possible states in generation 3
>>> #probability of 0..9 active infections in generation 3
>>> pgf.active_infections(mu, 3, 10)
array([ 0.36972018,  0.05259718,  0.07178445,  0.09609134,  0.07393309,
        0.07334027,  0.06617007,  0.05007252,  0.04119097,  0.03182213])
>>> #probability of 0..9 completed infections in generation 3
>>> pgf.completed_infections(mu, 3, 10)
```

Table C.7

Commands of Invasion_PGF. Many of these have an optional boolean argument `intermediate_values` which, if `True`, will result in returning values from generation 0 to generation `gen` in the discrete-time case or at some intermediate times in the continuous-time case. For the discrete-time results, the input μ is the offspring distribution PGF. For the continuous-time version, β and γ are the transmission and recovery rates respectively.

Command	Output
<code>R0(μ)</code>	Approximation of \mathcal{R}_0 .
<code>extinction_prob(μ, <code>gen</code>)</code>	Probability α_{gen} of extinction by generation <code>gen</code> given offspring PGF μ .
<code>cts_time_extinction_prob(β, γ, <code>T</code>)</code>	Probability $\alpha(T)$ of extinction by time <code>T</code> given transmission and recovery rates β and γ .
<code>active_infections(μ, <code>gen</code>, <code>M</code>)</code>	Array containing probabilities $\varphi_0, \dots, \varphi_j, \dots, \varphi_{M-1}$ of having j active infections in generation <code>gen</code> given offspring PGF μ .
<code>cts_time_active_infections(β, γ, <code>T</code>)</code>	Array containing probabilities $\varphi_0, \dots, \varphi_j, \dots, \varphi_{M-1}$ of having j active infections at time <code>T</code> given transmission and recovery rates β and γ .
<code>completed_infections(μ, <code>gen</code>, <code>M</code>)</code>	Array containing probabilities $\omega_0, \dots, \omega_j, \dots, \omega_{M-1}$ of having j completed infections in generation <code>gen</code> given offspring PGF μ .
<code>cts_time_completed_infections(β, γ, <code>T</code>)</code>	Array containing probabilities $\omega_0, \dots, \omega_j, \dots, \omega_{M-1}$ of having j completed infections at time <code>T</code> given transmission and recovery rates β and γ .
<code>active_and_completed(μ, <code>gen</code>, <code>M1</code>, <code>M2</code>)</code>	$M_1 \times M_2$ array containing probabilities $\pi_{i,r}$ of i active infections and r completed infections in generation <code>gen</code> given offspring PGF μ .
<code>cts_time_active_and_completed(β, γ, <code>T</code>)</code>	$M_1 \times M_2$ array containing probabilities $\pi_{i,r}$ of i active infections and r completed infections at time <code>T</code> given transmission and recovery rates β and γ .
<code>final_sizes(μ, <code>M</code>)</code>	Array containing probabilities $\omega_0, \dots, \omega_j, \dots, \omega_{M-1}$ of having j total infections in an outbreak given offspring PGF μ .
<code>cts_time_final_sizes(β, γ, <code>T</code>)</code>	Array containing probabilities $\omega_0, \dots, \omega_j, \dots, \omega_{M-1}$ of having j total infections in an outbreak given transmission and recovery rates β and γ .

```

array([-2.04281037e-17,  2.50000000e-01,  6.25000000e-02,
        7.81250000e-02,  9.76562500e-02,  1.21093750e-01,
        8.59375000e-02,  8.59375000e-02,  7.81250000e-02,
        6.25000000e-02])
>>> #joint probabilities of 0..4 active infections and 0..4 completed
>>> #infections in generation 3
>>> pgf.active_and_completed(mu, 3, 5, 5)
array([[ 0.          ,  0.25         ,  0.0625        ,  0.03125       ,  0.015625      ],
       [ 0.          ,  0.          ,  0.            ,  0.015625     ,  0.015625     ],
       [ 0.          ,  0.          ,  0.            ,  0.015625     ,  0.01953125   ],
       [ 0.          ,  0.          ,  0.            ,  0.015625     ,  0.0234375    ],
       [ 0.          ,  0.          ,  0.            ,  0.            ,  0.01171875   ]])
>>> #check that marginals match, increase sizes considered to improve match
>>> act_and_complete = pgf.active_and_completed(mu, 3, 20, 20)
>>> act_and_complete.sum(axis=1) #Active infections
array([ 3.69720176e-01,  5.25971800e-02,  7.17844516e-02,
        9.60913450e-02,  7.39330947e-02,  7.33402669e-02,
        6.61700666e-02,  5.00725210e-02,  4.11909670e-02,
        3.18221301e-02,  2.31783241e-02,  1.72899812e-02,
        1.21286511e-02,  8.08435678e-03,  5.25146723e-03,
        3.23349237e-03,  1.90655887e-03,  1.08598173e-03,
        5.83335757e-04,  2.96160579e-04])
>>> act_and_complete.sum(axis=0) #Completed infections
array([ 0.          ,  0.25         ,  0.0625        ,  0.078125     ,  0.09765625    ,
        0.12109375   ,  0.0859375   ,  0.0859375   ,  0.078125     ,  0.0625         ,
        0.0390625    ,  0.02342606   ,  0.01163167   ,  0.00376529   ,  0.            ,
        0.            ,  0.            ,  0.            ,  0.            ,  0.            ])
>>> #yes, these match previous calculations, with a small mismatch because
>>> #e.g., there may be 21 cumulative cases and 8 active cases. To accurately
>>> #calculate the probability of 8 active cases we would need to
>>> #increase the sizes to include this.

>>> #
>>> #Now look at the final sizes
>>> pgf.final_sizes(mu, 20)

array([ 0.00000000e+00,  2.50000000e-01,  5.93750000e-02,
        2.82031250e-02,  1.67456055e-02,  1.03404114e-02,
        6.80080902e-03,  4.66611063e-03,  3.29263648e-03,
        2.37637247e-03,  1.74605802e-03,  1.30159459e-03,
        9.81970183e-04,  7.48352208e-04,  5.75249662e-04,
        4.45491477e-04,  3.47250478e-04,  2.72225362e-04,
        2.14494366e-04,  1.69773210e-04])

>>> #
>>> #Now consider the continuous-time model
>>> beta = 2
>>> gamma = 1
>>> #In next command, first returned array is the times and second
>>> #is the extinction probabilities at those times
>>> pgf.cts_time_extinction_prob(beta, gamma, 5, intermediate_values =
... True, numvals = 6)
(array([ 0.          ,  1.          ,  2.          ,  3.          ,  4.          ,  5.]),
array([[ 0.          ,  0.38730017,  0.46371057,  0.48723549,  0.49537878,
        0.49830983]]))
>>> #following commands look at possible states at time 3
>>> pgf.cts_time_active_infections(beta, gamma, 3, 10)
array([ 0.48723548,  0.01309038,  0.0127562 ,  0.01243055,  0.01211321,
        0.01180397,  0.01150263,  0.01120897,  0.01092282,  0.01064397])
>>> pgf.cts_time_completed_infections(beta, gamma, 3, 10)
array([ 0.00037014,  0.33477236,  0.07721546,  0.03805734,  0.02535527,
        0.02008499,  0.01755497,  0.0161637 ,  0.01527602,  0.0146211 ])
>>> #check that the joint distribution has the same marginals
>>> cts_time_act_and_complete = pgf.cts_time_active_and_completed(beta, gamma, 3, 20, 20)
>>> cts_time_act_and_complete.sum(axis=1) #Active infections
array([ 0.48717492,  0.01298644,  0.01257732,  0.0121518 ,  0.01170803,
        0.0112452 ,  0.01076355,  0.01026434,  0.00974976,  0.00922277,
        0.00868697,  0.00814639,  0.00760529,  0.00706802,  0.0065388 ,
        0.00602164,  0.00552019,  0.00503763,  0.00457669,  0.00413954])
>>> cts_time_act_and_complete.sum(axis=0) #Completed infections
array([ 0.00036997,  0.334771 ,  0.07720946,  0.03803859,  0.02530855,
        0.01998599,  0.0173696 ,  0.01584881,  0.0147816 ,  0.01389348,
        0.01306482,  0.01224544,  0.01141776,  0.01057991,  0.0097376 ,
        0.00889996,  0.00807718,  0.0072792 ,  0.00651487,  0.00579152])
>>> #yes, these match previous calculations
>>> #Now look at the final sizes at time infinity
>>> pgf.cts_time_final_sizes(beta, gamma, 20)
array([ 0.00000000e+00,  3.33333333e-01,  7.40740741e-02,
        3.29218107e-02,  1.82898948e-02,  1.13803790e-02,
        7.58691934e-03,  5.29880081e-03,  3.82691169e-03,
        2.83474940e-03,  2.14181066e-03,  1.64421828e-03,
        1.27883644e-03,  1.00558079e-03,  7.98079995e-04,
        6.38463996e-04,  5.14318219e-04,  4.16833066e-04,
        3.39641758e-04,  2.78069275e-04])

```

Appendix D. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.idm.2018.08.001>.

References

- Allen, L. J. S. (2008). An introduction to stochastic epidemic models. In *Mathematical epidemiology* (pp. 81–130). Springer.
- Allen, L. J. S. (2010). *An introduction to stochastic processes with applications to biology*. CRC Press.
- Allen, L. J. S. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2), 128–142.
- Antal, T., & Kravivsky, P. L. (2011). Exact solution of a two-type branching process: Models of tumor progression. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(08), P08018.
- Bailey, N. T. J. (1953). The total size of a general stochastic epidemic. *Biometrika*, 177–185.
- Bailey, N. T. J. (1964). *The elements of stochastic processes with applications to the natural sciences*. John Wiley & Sons.
- Bartlett, M. S. (1949). Some evolutionary stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2), 211–229.
- Blumberg, S., & Lloyd-Smith, J. O. (2013). Inference of \mathcal{R}_0 and transmission heterogeneity from the size distribution of stuttering chains. *PLoS Computational Biology*, 9(5), e1002993.
- Bornemann, F. (2011). Accuracy and stability of computing high-order derivatives of analytic functions by Cauchy integrals. *Foundations of Computational Mathematics*, 11(1), 1–63.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., et al. (2000). Graph structure in the web. *Computer Networks*, 33, 309–320.
- Conway, J. M., & Coombs, D. (2011). A stochastic model of latently infected cell reactivation and viral blip generation in treated HIV patients. *PLoS Computational Biology*, 7(4), e1002033.
- Diekmann, O., & Heesterbeek, J. A. P. (2000). *Mathematical epidemiology of infectious diseases*. Wiley Chichester.
- Dorogovtsev, S. N., Mendes, J. F. F., & Samukhin, A. N. (Jul 2001). Giant strongly connected component of directed networks. *Physical Review E*, 64(2), 025101.
- Durrett, R. (2015). Branching process models of cancer. In *Branching process models of cancer* (pp. 1–63). Springer.
- Dwass, M. (1969). The total progeny in a branching process and a related random walk. *Journal of Applied Probability*, 6(3), 682–686.
- Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.
- Gallian, J. A., & Rusin, D. J. (1979). Cyclotomic polynomials and nonstandard dice. *Discrete Mathematics*, 27(3), 245–259.
- Gardner, M. (1978). Mathematical games. *Scientific American*, 238, 19–32.
- Getz, W. M., & Lloyd-Smith, J. O. (2006). Basic methods for modeling the invasion and spread of contagious diseases. In *Disease evolution: Models, concepts, and data analyses* (pp. 87–112).
- Harko, T., Lobo, F. S. N., & Mak, M. K. (2014). Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*, 236, 184–194.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer Science & Business Media.
- van der Hofstad, R., & Keane, M. (2008). An elementary proof of the hitting time theorem. *The American Mathematical Monthly*, 115(8), 753–756.
- House, T., Ross, J. V., & Sirl, D. (2013). How big is an outbreak likely to be? methods for epidemic final-size calculation. *Proceedings of the Royal Society A*, 469(2150), 20120436.
- Kenah, E., & Miller, J. C. (2011). Epidemic percolation networks, epidemic outcomes, and interventions. *Interdisciplinary Perspectives on Infectious Diseases*, 2011.
- Kendall, D. G. (1949). Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2), 230–282.
- Kimmel, M., & Axelrod, D. E. (2002). Branching processes in biology. *Interdisciplinary Applied Mathematics*, 19.
- Kiss, I. Z., Miller, J. C., & Simon, P. L. (2017). *Mathematics of epidemics on networks: From exact to approximate models*. Springer.
- Kucharski, A. J., & Edmunds, W. J. (2015). Characterizing the transmission potential of zoonotic infections from minor outbreaks. *PLoS Computational Biology*, 11(4), e1004154.
- Lewis, M. A., Petrovskii, S. V., & Potts, J. R. (2016). *The mathematics behind biological invasions* (Vol. 44). Springer.
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066), 355.
- Ludwig, D. (1975). Final size distributions for epidemics. *Mathematical Biosciences*, 23, 33–46.
- Ma, J. J., & Earn, D. J. D. (2006). Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*, 68(3), 679–702.
- May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261(5560), 459–467.
- Miller, J. C. (2011). A note on a paper by Erik Volz: SIR dynamics in random networks. *Journal of Mathematical Biology*, 62(3), 349–358.
- Miller, J. C. (2012). A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology*, 74(9), 2125–2141.
- Miller, J. C., Davoudi, B., Meza, R., Slim, A. C., & Pourbohloul, B. (2010). Epidemics with general generation interval distributions. *Journal of Theoretical Biology*, 262(1), 107–115.
- Miller, J. C., Slim, A. C., & Volz, E. M. (2012). Edge-based compartmental modelling for infectious disease spread. *Journal of The Royal Society Interface*, 9(70), 890–906.
- Moore, C., & Newman, M. E. J. (2000). Exact solution of site and bond percolation on small-world networks. *Physical Review E*, 62(5), 7059.
- Nee, S., Holmes, E. C., May, R. M., & Harvey, P. H. (1994). Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society London B*, 344(1307), 77–82.
- Nishiura, H., Yan, P., Sleeman, C. K., & Mode, C. J. (2012). Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *Journal of Theoretical Biology*, 294, 48–55.
- Peitgen, H.-O., Jürgens, H., & Saupe, D. (2006). *Chaos and fractals: New frontiers of science*. Springer Science & Business Media.
- Pólya, G. (1990). *Mathematics and plausible reasoning: Induction and analogy in mathematics* (Vol. 1). Princeton University Press.
- Poore, G. M. (2015). Pythontex: Reproducible documents with LaTeX, Python, and more. *Computational Science & Discovery*, 8(1), 014010.
- Reluga, T., Meza, R., Walton, D. B., & Galvani, A. P. (2007). Reservoir interactions and disease emergence. *Theoretical Population Biology*, 72(3), 400–408.
- Stanley, R. P. (2001). *Enumerative combinatorics* (Vol. II). Cambridge University Press.
- Valdez, L. D., Macri, P. A., & Braunstein, L. A. (2012). Temporal percolation of the susceptible network in an epidemic spreading. *PLoS One*, 7(9), e44188.
- Volz, E. M. (2008). SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology*, 56(3), 293–310.
- Volz, E. M., Romero-Severson, E., & Leitner, T. (2017). Phylogenetic inference across epidemic scales. *Molecular Biology and Evolution*, 34(5), 1276–1288.
- Watson, H. W., & Galton, F. (1875). On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4, 138–144.
- W. Weckesser odeintw. "https://github.com/WarrenWeckesser/odeintw".
- Wendel, J. G. (1975). Left-continuous random walk and the Lagrange expansion. *American Mathematical Monthly*, 494–499.
- Wilf, H. S. (2005). *generatingfunctionology* (3rd ed.). A K Peters, Ltd.
- Yan, P. (2008). Distribution theory, stochastic processes and infectious disease modelling. *Mathematical Epidemiology*, 229–293.