Epidemics and branching processes

Daniel Ahlberg, Stockholm University *May 20, 2021*

Understanding the dynamics of infectious disease is an important problem in order for authorities to be able to take preventive measures and to plan for future epidemics. In particular, we may want to understand under what circumstances an infectious disease may result in a major outbreak, the size of that outbreak and how the number of infectious individuals develop over time. It may also be important to understand how different preventive measures, such as vaccination and social distancing, may affect the course of an epidemic.

An often mentioned quantity in epidemiological discussions is the *basic reproduction number* R_0, loosely defined as the expected number of newly infected that one infected individual will be responsible for (in an otherwise susceptible population). We shall below put up a probabilistic model for the spread of an infectious disease, which will help us understand why an epidemic may result in a large outbreak if and only if $R_0 > 1$. We shall also see how we can estimate the probability of large outbreak, and the final size of an epidemic. Along the way we will come in contact with the theory of branching processes, that we will get a glimpse of.

The questions that we primarily set out to address here concern the early phase of an epidemic. A probabilistic model is therefore suitable as it accounts for different random factors that affect whether an epidemic will take off or not. At later stage of the epidemic also a deterministic model will give a resonable description of the evolution of the epidemic. This is due to the fact that when a large number of individuals already have become infected, then the random effects associated with the spread of the infection cancel out, much like in the law of large numbers, and the epidemic will follow its average behaviour. We shall below mainly focus on the probabilistic model.

A probabilistic model for an epidemic

The most basic model for an epidemic is a so-called SIR-model, in which the individuals of a population can be in either of three states:

(S) susceptible,
(I) infectious,
(R) recovered.

Individual can jump from state S to I, and from state I to R, and once in state R they remain in that state forever. Individuals in state R are thought of as immune to further infection (or possibly deceased), and do no longer participate in the spread of the infection. This is a reasonable assumption when the duration of an epidemic is much shorter than the time an individual remains immune to further infection.

Apart from the different states an individual may be in, we need an assumption on how individuals move between the different states. As infection

is transmitted when different individuals interact, we need to describe this interaction. The most basic assumption ignores social structures, and simply assumes that any two individuals are equally likely to interact at any given time point. This is referred to as *homogeneous mixing*. This is a reasonable assumption for the spread of an infectious disease in a metropolitan area, but gives a less accurate description in a population consisting of both more and less densely populated areas.

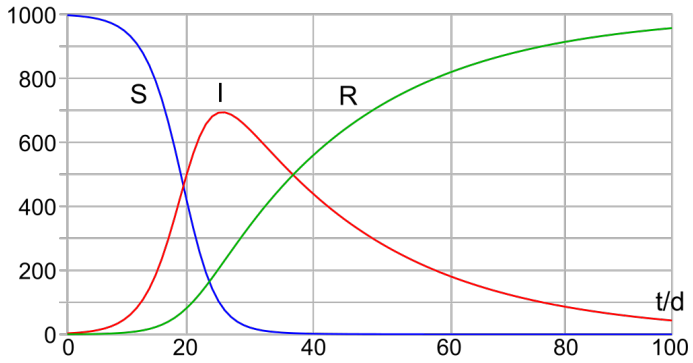


Figure 1: The typical shape of an epidemic. Simulation corresponds to a basic reproduction number $R_0 = 10$. (Figure from [3].)

Consider a population consisting of N individuals. Let $S(t)$, $I(t)$ and $R(t)$ denote the number of individuals in state S, I and R, respectively, at time $t \geq 0$. Let $S(0) = N - 1$, $I(0) = 1$ and $R(0) = 0$. Given $\lambda, \gamma > 0$, the probabilistic SIR-model under the assumption of homogeneous mixing takes the following form:

(i) For each individual in state I, infectious contacts occur according to a Poisson process at rate λ.
(ii) At each infectious contact an individual is chosen uniformly from the population, and if in state S, then the individual transitions to state I.
(iii) An individual remains infectious for a random time T with mean $1/\gamma$.

The parameters λ and γ govern how individuals go from state S to I, and from I to R. The parameter λ corresponds to the *rate*, i.e. the average number of occurrences per unit time, at which infectious contacts occur. The parameter λ corresponds to the rate at which an individual recovers from infection. Note that the uniformity assumption in (ii) corresponds to homogeneous mixing. Above we have not specified the distribution of T. It is common to assume that T is exponentially distributed or simply constant.

Let Y denote the number of infectious contacts of an arbitrary infectious individual. Under the assumption that all other individuals are susceptible, then each infectious contact will result in a new infected individual. Hence, we can identify the basic reproduction number of the SIR-model as $R_0 = \mathbb{E}[Y]$. By assumption, the number of infectious contacts an individual has during a time interval of length t is Poisson distributed with parameter λt. Consequently,

the conditional distribution of Y, given that the time T the individual remains infectious equals t, is Poisson with parameter λt. Hence, $E[Y|T=t] = \lambda t$, so

$$R_0 = \mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|T]] = \mathbb{E}[\lambda T] = \lambda/\gamma. \tag{1}$$

In the early phase of an epidemic it is unlikely that an infectious contact does not lead to a new infectious individual. That is because the number of infectious or recovered individuals is very small in comparison to the number of susceptibles. Counting the number of infectious contacts will therefore give an accurate estimate of the number of individuals infected in the early phase of the epidemic. This estimate will remain accurate, with high probability, as long as the number of infectious contacts is much lower than $\sqrt{N}$. The number of infectious contacts can be described by what in known as a branching process.

Branching processes

A branching process is a stochastic process describing the growth of a population where every individual gives rise to a random number of offspring, independently of one another. The branching process that we will consider is known as a Galton-Watson process.

Definition 1. *For $n \geq 0$, let $Y_1^{(n)}, Y_2^{(n)}, \ldots$ be independent non-negative and integer-valued random variables with common distribution F. Assume further that these variables are independent for different n. Set $X_0 = 1$ and for each $n \geq 0$ let*

$$X_{n+1} := Y_1^{(n)} + \ldots + Y_{X_n}^{(n)}.$$

The sequence $(X_n)_{n\geq 0}$ is referred to as a Galton-Watson process.

If we interpret $Y_k^{(n)}$ as the number of offspring of individual k in generation n, then X_{n+1} takes the significance of the number of individuals in generation $n+1$, in a population starting with a single ancestor.

We want to study the behaviour of a Galton-Watson process over time. In particular, we want to determine what affects survival or eventual extinction of a population described by a Galton-Watson process. A Galton-Watson process $(X_n)_{n\geq 0}$ is said to *survive* if $X_n \geq 1$ for all $n \geq 1$, and to eventually go *extinct* if $X_n = 0$ for some $n \geq 1$. We shall by η denote the probability that the Galton-Watson process eventually goes extinct. (η will depend on the distribution F.)

Lemma 2. *The probability η of eventual extinction satisfies*

$$\eta = \lim_{n\to\infty} \mathbb{P}(X_n = 0).$$

Proof. Note that by definition

$$\eta = \mathbb{P}\big(X_n = 0 \text{ for some } n \geq 1\big) = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} \{X_n = 0\}\Big).$$

Note also that if $X_n = 0$, then $X_{n+1} = 0$ too. So $\{X_1 = 0\} \subseteq \{X_2 = 0\} \subseteq \ldots$, and the events $A_n := \{X_n = 0\} \setminus \{X_{n-1} = 0\}$ are mutually disjoint. By countable additivity (one of the probability axioms) it follows that

$$\eta = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} \{X_n = 0\}\Big) = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{P}(A_n).$$

Again by countable additivity, and since $\{X_{n-1} = 0\} \subseteq \{X_n = 0\}$, we have

$$\sum_{n=1}^{N} \mathbb{P}(A_n) = \mathbb{P}\Big(\bigcup_{n=1}^{N} A_n\Big) = \mathbb{P}\Big(\bigcup_{n=1}^{N} \{X_n = 0\}\Big) = \mathbb{P}(X_N = 0),$$

from which the lemma follows. □

The next proposition gives a first indication for when a population is bound to go extinct and when it has a chance of survival.

Proposition 3. *Let $(X_n)_{n\geq 0}$ be a Galton-Watson process with an offspring distribution F with mean $m < \infty$. Then*

$$\mathbb{E}[X_n] = m^n \quad \textit{for all } n \geq 1.$$

Proof. From the definition of a Galton-Watson process we have that X_n is a random sum of X_{n-1} independent offspring variables. Since all offspring variables are independent of one another, and since X_{n-1} is a function of the offspring variables Y_k^{ℓ} for $\ell \leq n-2$, it follows that X_{n-1} is independent of $Y_1^{(n-1)}, Y_2^{(n-1)}, \ldots$. We can therefore compute the mean of X_n as

$$\mathbb{E}[X_n] = \mathbb{E}\left[\sum_{k=1}^{X_{n-1}} Y_k^{(n-1)}\right] = \mathbb{E}\left[Y_1^{(n-1)}\right] \mathbb{E}[X_{n-1}] = m\,\mathbb{E}[X_{n-1}],$$

given that $\mathbb{E}[X_{n-1}]$ is finite. Repeating the argument, we obtain

$$\mathbb{E}[X_n] = m\,\mathbb{E}[X_{n-1}] = m^2\,\mathbb{E}[X_{n-2}] = \ldots = m^n\,\mathbb{E}[X_0] = m^n,$$

given that $\mathbb{E}[X_\ell]$ is finite for each $\ell = 0, 1, \ldots, n-1$. However, since $\mathbb{E}[X_0] = 1$ we conclude that $\mathbb{E}[X_1] = \mathbb{E}[X_0]m$ is finite, that $\mathbb{E}[X_2] = \mathbb{E}[X_1]m$ is finite, and so on. Consequently, $\mathbb{E}[X_n]$ is finite for all n, and given by m^n. □

The above proposition shows that when $m > 1$ the expected size of each generation grows exponentially fast, which *indicates* that the population has a chance to survive. On the other hand, when $m < 1$, Lemma 2 and Proposition 3 together show that extinction is inevitable:

$$\eta = \lim_{n\to\infty} [1 - \mathbb{P}(X_n \geq 1)] \geq \lim_{n\to\infty} (1 - \mathbb{E}[X_n]) = \lim_{n\to\infty} (1 - m^n) = 1,$$

where the inequality $\mathbb{P}(X_n \geq 1) \leq \mathbb{E}[X_n]$ follows from Markov's inequality.

To obtain a full characterization of the destiny of a population, we shall study the probability generating function of a Galton-Watson process. This will result in the following theorem.

Theorem 4. *Let $(X_n)_{n\geq 0}$ be a Galton-Watson process with offspring distribution F with mean $m < \infty$ and $0 < F(0) < 1$. Write g_1 for the probability generating function of F. Then the following holds:*

(a) η satisfies the equation $\eta = g_1(\eta)$.
(b) η is the smallest nonnegative root to the equation $t = g_1(t)$.
(c) $\eta = 1$ if $m \geq 1$ and $\eta < 1$ if $m > 1$.

Note that if $F(0) = 0$, then each individual will have at least one offspring, and the population will live on forever. If $F(0) = 1$ then the population will go extinct already in the first generation. The assumption that $0 < F(0) < 1$ is therefore no restriction.

To prove the theorem we shall need to analyze the probability generating function of X_n, which we shall denote by g_n. Since $X_1 = Y_1^{(0)}$, this is consistent with the definition of g_1 in the theorem. We begin with a lemma.

Lemma 5. *Let $(X_n)_{n\geq 0}$ be a Galton-Watson process and let g_n denote the probability generating function of X_n. Then, for all $k = 1, 2, \ldots, n$,*

$$g_n(t) = g_{n-k} \circ g_k(t) = g_{n-k}(g_k(t)).$$

Proof. From the definition of a Galton-Watson process we have that X_n is a random sum of X_{n-1} independent offspring variables. Since all offspring variables are independent of one another, and since X_{n-1} is a function of the offspring variables Y_k^ℓ for $\ell \leq n-2$, it follows that X_{n-1} is independent of $Y_1^{(n-1)}, Y_2^{(n-1)}, \ldots$. The probability generating function of X_n can thus be computed as

$$g_n(t) = g_{n-1}(g_1(t)) = g_{n-1} \circ g_1(t).$$

Continuing the above argument yields that

$$g_n(t) = g_1 \circ g_1 \circ \ldots \circ g_1(t).$$

By grouping the first $n-k$ into g_{n-k} and the last k into g_k ends the proof. □

Proof of Theorem 4. For part *(a)*, we condition on the number of individuals in the first generation. The law of total probability gives that

$$\eta = \sum_{k=0}^{\infty} \mathbb{P}\big(\text{eventual extinction}\big|X_1 = k\big)\mathbb{P}(X_1 = k).$$

Given that $X_1 = k$, eventual extinction of the population is equivalent to each of the k individuals of the first generation having a finite number of descendants. Since the subpopulation of descendants of each of these k individuals are independent, and equal in distribution to the (unconditional) population as a whole, it follows that

$$\eta = \sum_{k=0}^{\infty} \mathbb{P}\big(\text{eventual extinction}\big)^k \mathbb{P}(X_1 = k) = \sum_{k=0}^{\infty} \eta^k \mathbb{P}(X_1 = k) = g_1(\eta).$$

This proves the first part of the theorem.

For part *(b)*, let $\eta_n := \mathbb{P}(X_n = 0)$. Since $\{X_n = 0\} \subseteq \{X_{n+1} = 0\}$ it follows that $\eta_n \leq \eta_{n+1}$ so that $(\eta_n)_{n\geq 1}$ is an increasing sequence. By Lemma 2 it converges to η.

Let $\eta^* \geq 0$ be any root to the equation $t = g_1(t)$. ($t = 1$ is always a solution, so at least one root exists.) Since $g_n(0) = \mathbb{P}(X_n = 0)$, we have by Lemma 5 that

$$\eta_n = g_n(0) = g_1 \circ g_1 \circ \ldots \circ g_1(0).$$

Since $g_1(t) = \sum_{k=0}^{\infty} t^k \, \mathbb{P}(X_1 = k)$ is a power series with nonnegative coefficients, its derivative g_1' is nonnegative on the interval $[0, 1]$, and hence g_1 is nondecreasing on $[0, 1]$. Consequently, since $\eta^* \geq 0$,

$$\eta_n = g_1 \circ g_1 \circ \ldots \circ g_1(0) \leq g_1 \circ g_1 \circ \ldots \circ g_1(\eta^*) = \eta^*,$$

and hence $\eta = \lim_{n\to\infty} \eta_n \leq \eta^*$. Since η is a root to the equation $t = g_1(t)$ (proved in the first step), it must be the smallest nonnegative root.

For part *(c)*, recall that g_1 is a power series with positive coefficients. Then its derivative is a power series with positive coefficients too. It follows that that the second derivative g_1'' is nonnegative, and g_1 therefore convex, on $[0, 1]$. Moreover, since $g_1(0) = F(0) < 1$, we have that $g_1'(t) > 0$ for all $t \in (0, 1)$. That is, g_1 is both strictly increasing and convex on $[0, 1]$. Moreover, properties regarding the probability generating function shows that

$$\lim_{t\to 1} g_1'(t) = \lim_{t\to 1} \sum_{k=1}^{\infty} k t^{k-1} \, \mathbb{P}(X_1 = k) = \sum_{k=1}^{\infty} k \, \mathbb{P}(X_1 = k) = \mathbb{E}[X_1] = m.$$

Since $g_1(0) = F(0) > 0$ and $g_1(1) = 1$, it follows that g_1 must behave as in Figure 2. More precisely, if $m < 1$ then $g_1'(t) < 1$ for all $t \in (0, 1)$ so $g_1(t)$

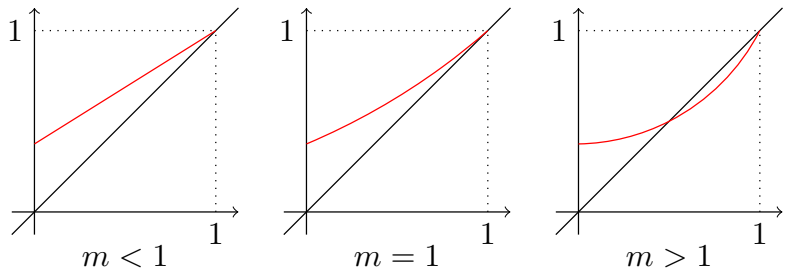


Figure 2: Illustration of g_1 on the interval $[0, 1]$ in the three cases $m < 1$, $m = 1$ and $m > 1$. Note that $g_1(0) = F(0)$ and $g_1' = m$.

cannot intersect the line $y = t$ before $t = 1$. Hence, when $m < 1$ the least nonnegative root to the equation $t = g_1(t)$ is $t = 1$. On the other hand, if $m > 1$ then $g_1'(t) > 1$ for t close to 1, and $g_1(t)$ must intersect the line $y = t$ for some $t \in (0, 1)$. Finally, if $m = 1$ and $F(0) > 0$, then we cannot have $F(1) = 1$, and some mass must fall on integers larger than or equal to 2. This implies that g_1 has degree at least two. (That is, for some $k \geq 2$ the

coefficient is positive.) Then the derivative g_1' is strictly increasing on $(0,1)$, and $g_1'(t) < m = 1$ for $t \in (0,1)$. Again this shows that $g_1(t)$ cannot intersect the line $y = t$ in $(0,1)$ when $m = 1$. □

Back to epidemics

We argued above that in the early phase of an epidemic, the number of infectious contacts give an accurate estimate on the number of infected individuals so far in the epidemic. The number of infectious contacts a given infectious individual give rise to is in the model independent from, and identically distributed to, the number of infectious contacts other infectious individuals give rise to. Hence, the number of infectious contacts will, in the early phase of the epidemic, evolve as a Galton-Watson process. This will accurately describe the evolution of the epidemic until an infectious individual for the first time has an infectious contact with a non-susceptible individual.

Consider the SIR-model in a population consisting of N individuals, of which one is initially infected and all others susceptible.

Lemma 6. *The probability that the first n infectious contacts all lead to a new infectious individuals is close to 1 as long as $n \ll \sqrt{N}$ (and N is large).*

Proof. Note that the only reason for an infectious contact *not* to lead to a new infectious individual is that the infectious contact involves an individual that is already infectious or recovered. Whenever an infectious contact occurs, and there are k susceptible individuals in the population, the probability that the infectious contact leads to a new infectious individual is $k/(N-1)$. (There are k susceptibles among $N-1$ individuals, since the infectious individual responsible for the infectious contact is removed.) Consequently, the probability that the first n infectious contacts all lead a new infectious individual equals

$$\frac{N-1}{N-1}\frac{N-2}{N-1}\cdots\frac{N-n}{N-1} = \frac{(N-1)!}{(N-n-1)!}(N-1)^{-n}$$

Stirling's formula says that $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$, where $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$. An application of Stirling's formula shows that the above expression is asymptotically equal to (as $N \to \infty$)

$$\frac{\sqrt{N-1}(N-1)^{N-1}e^{N-n-1}}{\sqrt{N-n-1}(N-n-1)^{N-n-1}e^{N-1}}(N-1)^{-n} = e^{-n}\left(1+\frac{n}{N-n-1}\right)^{N-n-1/2}.$$

Recall that $(1 + a/m) \to e^a$ as $m \to \infty$. For fixed n the above expression therefore tends to 1 as $N \to \infty$. A more careful analysis (using that $(1 + a/m)^m \le e^a \le (1 + a/m)^{m+a}$ for all $m \ge 1$ and $a > 0$) it is possible to verify that the same holds as long as $n \ll \sqrt{N}$. □

Recall that Y denotes the number of infectious contacts of a typical infected individual, and write g_Y for its probability generating function. The connection to branching processes allows us to draw the following conclusions:

- For large populations, the probability of a large outbreak is close to zero for $R_0 \leq 1$ and positive for $R_0 > 1$.
- For large populations, the probability of a large outbreak is close to $1-\eta$, where η is the least nonnegative root of the equation $t = g_Y(t)$.

Note that R_0 only depends of the mean of Y, whereas η is a consequence of the probability generating function of Y, which depends more intricate manner on the distribution of Y.

In the above model we have started out from a single infectious individual, and estimated the probability of a large outbreak. If the number of initially infectious people is k, then the probability of no large outbreak will roughly equal η^k. So if $R_0 > 1$, then it is essentially inevitable for a large outbreak to occur when infection in brought into a population at many independent occasions. This explains (due to the absence of quarantine guidelines and other restrictions) why it was just a matter of time before the covid-19 pandemic would take off in Sweden during the spring of 2020.

Epilogue: A deterministic model for an epidemic

The probabilistic model described above is suitable to describe the early phase of an epidemic. By further analysis of the probabilistic model it is also possible to determine the later course of the epidemic. Once a large number of individuals have become infected, however, the random effects of the model will cancel out. As a consequence, the main phase of an epidemic is rather predictable, and can be reasonably well described by a deterministic model. Let us end this discussion by a brief overview of a deterministic model, which in the main phase is roughly analogous to the probabilistic model analyzed above.

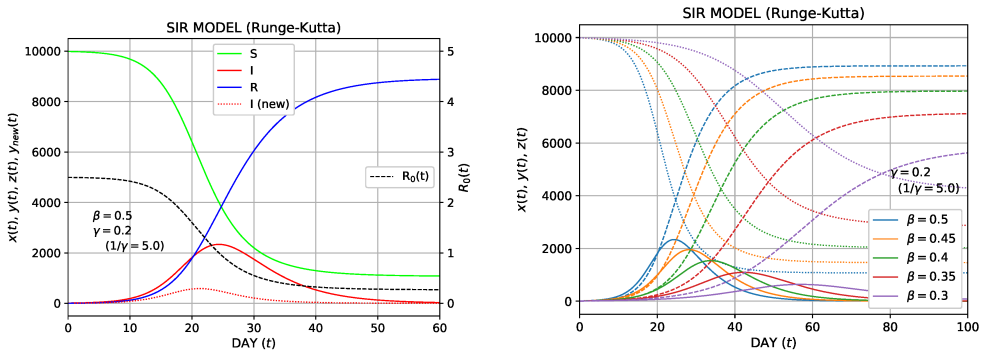
Let $s(t)$, $i(t)$ and $r(t)$ denote the proportion of the population in state S, I and R, respectively, at time $t \geq 0$. Suppose that $s(0) = 1 - \varepsilon$, $i(0) = \varepsilon$ and $r(0) = 0$ for some $\varepsilon > 0$. Since infection spreads as a result of interaction between individuals in state S and I, the rate of change of $s(t)$ ought to be proportional to the product $s(t)i(t)$. Recall that λ and γ are the parameters that correspond to the rate of transmission and recovery. The deterministic model corresponding to the probabilistic model described above will thus be described by the set of equations

$$\begin{aligned} s'(t) &= -\lambda s(t)i(t), \\ i'(t) &= \lambda s(t)i(t) - \gamma i(t), \\ r'(t) &= \gamma i(t), \end{aligned} \tag{2}$$

From the second equation we find that

$$i'(t) = \big[\lambda s(t) - \gamma\big] i(t),$$

so that the proportion of infected individuals will continue to rise as long as $s(t) > \gamma/\lambda$. That is, for an epidemic to get started we must have $\gamma/\lambda < 1$, or



(a) (b)

Figure 3: Numerical solution of the deterministic SIR-model (where β is used in place of λ). In the left figure $R_0 = 2.5$. The right figure illustrates how the spread of an epidemic is affected by lowering the rate of infectious contacts, in this case from $R_0 = 2.5$ down to 1.5. This has an effect on both the maximal number of infectious individuals and the final size of the epidemic, but also on the duration of the epidemic. (Simulation from [2].)

equivalently $\lambda/\gamma > 1$, which corresponds to our previous conclusions from (1) about the reproductive number $R_0 = \lambda/\gamma$.

If we divide the first equation (2) with the third we obtain

$$\frac{ds(t)}{dr(t)} = -\frac{\lambda}{\gamma}s(t).$$

This *suggests* (but does not prove) that $s(t)$ and $r(t)$ have the relation

$$s(t) = s(0)e^{-R_0 r(t)}.$$

Since every infected individual eventually recovers from infection (or dies), the quantity $r(\infty)$ corresponds to the finial size of the epidemic, i.e. the total number of individuals ever infected. Since we at the end of the epidemic have $s(\infty) = 1 - r(\infty)$, we *should* (but this needs to be motivated) be able to compute the final size of the epidemic from the expression

$$1 - r(\infty) = e^{-R_0 r(\infty)}.$$

It is possible to verify that the probabilistic model analyzed above indeed will behave as suggested by the deterministic equations. However, to verify that so is the case will require some work that takes us deeper into the world of probability theory. For a survey on further aspects of the epidemics modelling I direct the reader to [1].

References

[1] Tom Britton. Stochastic epidemic models: A survey. *Mathematical Biosciences*, 225(1):24-35, 2010.

[2] Yutaka Okabe and Akira Shudo. A Mathematical Model of Epidemics–A Tutorial for Students. *Mathematics*, 8(7):1174, 2020.

[3] Wikipedia contributors. Compartmental models in epidemiology. *Wikipedia, The Free Encyclopedia*, accessed April 26, 2021. `https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology`