medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Hardness of Herd Immunity and Success Probability of Quarantine Measures: A Branching Process Approach

Sujit Kumar Nath[1,2*]

[1] *School of Computing, University of Leeds, LS2 9JT, Leeds, UK and*

[2] *Faculty of Biological Sciences, University of Leeds, LS2 9JT, Leeds, UK*

We model the propagation of an infection, in a population, as a simplified age-dependent branching process. We analytically estimate the fraction of population, needed to be infected or immuned, to achieve herd immunity for an infection. We calculate this estimation as a function of the incubation period of the contagion, contact probability among the infected and susceptible population, and the probability of disease transmission from an infected to a susceptible individual. We show how herd immunity is strongly dependent on the incubation period, and it may be extremely difficult to achieve herd immunity in case of large incubation period. We derive the distribution of generation time from basic principles, which, by far, has been assumed in an ad hoc manner in epidemiological studies. We quantify the success probability of quarantine measures before achieving herd immunity, and discuss a novel method for designing effective quarantine measures in the absence of any pharmaceutical interventions. We also compare the effectiveness of an early imposition against a delayed imposition of lockdown, of the same duration, in mitigating infection from a population.

* Email id: s.k.nath@leeds.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Keywords: Age-dependent branching process; extinction probability; herd immunity; generation time distribution; infectious disease.

I. INTRODUCTION

The propagation of an infectious disease in a population can be viewed as a branching process where an infected individual comes in contact with susceptible individuals in the population, and transmits the infection to the susceptible individuals with some probability of transmission. Branching process model of infection propagation has been widely used to study the propagation of HIV, Rabies etc. infections, both in the population, as well as in the cells or tissues of an infected individual [1–3]. There are many reasons for which a stochastic model of an infectious disease is much preferable than its deterministic counterparts [4, 5]. Due to the inherent randomness of the infectious diseases, arising from the incubation period of the contagions [6–9], infectious period of the infected individuals [10–13], contact and transmission probabilities etc. [14, 15], a stochastic model is preferable than a deterministic one[16–19]. However, as all the models have their own advantages and limitations, both of the stochastic [17, 20] and deterministic [21] models, and sometimes a combination of the two [22], are used to model infectious diseases, as per their suitability.

Here we model the propagation of an infectious disease as an age-dependent branching process under some assumptions (described in detail in the later sections). Like every other model, this simplified model also has limitations, and sometimes could be inappropriate to capture the detailed dynamics of some specific disease. However, our goal here is to present some new interesting theoretical results of the disease propagation, and the effects of interventions, under simplified assumptions. This is of particular relevance due to the current situation of SARS-CoV-2 pandemic when researchers, around the globe, are searching for the best exit strategies from the lockdown restrictions [23]. We show that under reasonable assumptions, a simplified age-dependent branching process model of infection propagation can give rise to many important insights about the infection dynamics. We derive a few general theorems on age-dependent branching process, and apply them to study our model of infectious disease, to quantitatively understand the difficulty for a population to achieve herd immunity, depending on the incubation period of the contagion, social interactions, and transmission probability of the contagion from an infected to a susceptible individual.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

A recent study have shown that the herd immunity for SARS-CoV-2 can be achieved at a population wide infection rate at $\sim 43\%$, considering the effect of heterogeneities present into the population [24]. However, we show here that this $\sim 43\%$ infection rate cannot be uniform over all population, since the herd immunity depends on the other factors as well, such as the rate of contacts (which is higher in densely populated countries, than the less populated countries), incubation period of the contagion (which varies with the climate [25, 26]) etc. Also, we quantify the success probabilities of various quarantine measures in mitigating the infection. The distribution of generation time for an infectious disease [27, 28] is a very important quantity to understand the transmission potential, and to calculate the basic reproduction number which is a key parameter to estimate the epidemiological state of an infection in a population [29, 30]. In this paper, a general theorem on generation time is derived which can have large applicability in the study of infectious disease, as well as any kind of age-dependent population dynamics using branching process, such as cell divisions [31, 32], growth of microbial colonies [33, 34] etc. We apply this general theorem to derive the distribution of generation time for our model.

In section II we introduce our age-dependent model of infection propagation. Section III presents the derivation of the functional dependence of herd immunity threshold on the incubation period, contact and transmission probability, and fraction of susceptible populations. We study the success probabilities of various quarantine measures, depending on the state of the infection in the population, and compare early vs. later lockdown in a hypothetical population in sections IV and V, respectively. In section VI we prove a general theorem to derive the distribution of generation time from the generating function of the first generation progenies of any general age-dependent branching process, and apply it to derive the distribution of generation time for our model. We present a concise summary of this work in section VII.

II. MODEL OF INFECTION PROPAGATION USING AN AGE-DEPENDENT BRANCHING PROCESS

To map the infection propagation with a branching process, we identify an infected individual to be the ancestor of the individuals to whom it transmits the infection directly. The individuals who are directly infected by an ancestor, are called the progenies of the first

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

generation. The individuals who are infected directly by a progeny of the first generation, are called the progenies of the second generation, and so on. The set of all progenies or descendants of an individual is called a line. If the number of descendants at the rth generation is zero for an individual, its line is called extinct at the rth generation. Now, let an ancestor gives rise to n number of progenies with probability p_n, where $n \in \mathbb{Z}_+$, the set of all non-negative integers. Then the function

$$G(s) = \sum_{n=0}^{\infty} p_n s^n, \tag{1}$$

is the generating function for the distribution of number of progenies in the first generation. We now state a well-known theorem which establishes the basic reproduction number as a key parameter to determine the fate of the infection in the population, using the generating function of the first generation progenies.

Theorem 1. *Let $G(s)$ be the generating function of a branching process for which the average number of progenies created by an ancestor be μ. If $\mu \leq 1$, the process dies out with probability one. If however, $\mu > 1$ the probability x_r that the process terminates at or before the rth generation tends to the unique root $x < 1$ of the equation $s = G(s)$.*

See [35, 36] for a proof of theorem 1.

In the infectious disease literature, the average number of first generation progenies (newly created infections) created by an ancestor (primary infected individual) is termed as basic reproduction number, symbolically denoted by R_0 [29, 30]. Therefore, it is clear from theorem 1 that the infection will die out after sufficient time when $R_0 \leq 1$, which is called the herd immunity.

The assumptions of our model are as follows.

Assumption 1: *The number of contacts made by an infected individual, with other individuals in the population, is a Poisson process with a rate α.*

Assumption 2: *Contacts made by the individuals in the population are independent of each other.*

Assumption 3: *The rate of contacts is constant over time, unless any intervention is imposed, such as quarantine measures.*

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Assumption 4: *The incubation period and infectious period are used synonymously, because of the uncertainty of the initiation of the infectiousness in an individual carrying the contagion.*

Assumption 5: *Incubation period is memoryless.*

Assumption 6: *The fraction of susceptibles in the population varies very slowly with respect to the incubation/infectious time period.*

The motivation behind assumption 1 is that the Poisson process is the most suitable model for the number of events occurring in a memoryless fashion [37]. Also, assumption 5 says that the number of days, during which an infected individual remains infectious (or incubation period of the contagion, according to our convention) is also a memoryless random variable. This means, the time period over which an infected individual remains infected, starting from a particular time instant, does not depend on the time period over which the individual has already been infected. In other words, the probability that an infected individual remains infected at least for $t_1 + t_2$ days, given that it has already been infected for t_1 days, is the same as the initial probability that it remains infected for at least t_2 days. This assumption leads us to take the incubation time period exponentially distributed with a mean incubation time λ [37]. Therefore, the incubation time period of the contagion in an infected individual has the density function

$$f(t) = \begin{cases} \frac{1}{\lambda} e^{-\frac{t}{\lambda}} & \text{if } t \geq 0, \\ 0 & \text{if } t < 0. \end{cases} \tag{2}$$

We now prove a general theorem to derive the generating function for the number of first generation progenies, created by an ancestor, in an age-dependent branching process; and apply it to calculate the number of contacts made by an infected individual for our model.

Theorem 2. *Let X(t) be a continuous time stochastic process with atomic distribution* $\mathbb{P}(X(t) = n) = p_n(t)$*, and generating function G(s,t), where n is any non-negative integer. Let* $X_1(t)$ *be the process* $X(t)$ *with survival time* T *which is a continuous non-negative random variable with distribution function* $F(t) = \mathbb{P}(T \leq t)$*. Then the generating function* $G_1(s,t)$ *of the process* $X_1(t)$ *is given by*

$$G_1(s,t) = \int_0^t G(s,t')dF(t'). \tag{3}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Proof. The given condition, $\mathbb{P}(X(t) = n) = p_n(t)$, implies the probability that $X(t) = n$, when the process survives at least for time t, is $p_n(t)$. Since the survival time T is distributed as $F(t)$, the probability that $X_1(t) = n$ is

$$\begin{aligned}\mathbb{P}(X_1(t) = n) &= \int_0^t \mathbb{P}(X_1(t') = n \mid T = t')dF(t') \\ &= \int_0^t \mathbb{P}(X(t') = n)dF(t') = \int_0^t p_n(t')dF(t').\end{aligned}$$

Therefore, the generating function for $X_1(t)$ is

$$G_1(s,t) = \sum_{n=0}^{\infty} \mathbb{P}(X_1(t) = n)s^n = \sum_{n=0}^{\infty} \int_0^t s^n p_n(t')dF(t').$$

Now, identifying

$$\sum_{n=0}^{\infty} \mathbb{P}(X_1(t) = n)s^n = \mathbb{E}(s^{X_1(t)}),$$

as an expectation value on a discrete/atomic measure, and applying Fubini's theorem for changing the order of summation and integration, we obtain

$$G_1(s,t) = \int_0^t \left(\sum_{n=0}^{\infty} s^n p_n(t')\right) dF(t') = \int_0^t G(s,t')dF(t').$$

□

An important point to note here is that, in general, the function $G_1(s,t)$ could be the generating function of an improper distribution [37], i.e., $G_1(1,t) \leq 1$. In other words, the total probability that the number of progenies will be $n = 0, 1, 2, 3, ...$ etc., up to time t, could be less than or equal to 1.

Corollary 3. *If a Poisson process with rate α has exponentially distributed survival time with mean λ, then the generating function for the process at time t is given by*

$$C(s,t) = \frac{1 - e^{-t(\alpha - \alpha s + \frac{1}{\lambda})}}{1 + \alpha\lambda(1-s)} \tag{4}$$

Proof. A Poisson process with rate α, has the generating function

$$\begin{aligned}G(s,t) &= \sum_{n=0}^{\infty} \mathbb{P}(X(t) = n)s^n = \sum_{n=0}^{\infty} e^{-\alpha t}\frac{(\alpha t)^n}{n!}s^n \\ &= \sum_{n=0}^{\infty} e^{-\alpha t}\frac{(\alpha t s)^n}{n!} = e^{-\alpha t(1-s)}.\end{aligned} \tag{5}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Using theorem 2 and equation 5, we therefore obtain the generating function for a Poisson process with exponentially distributed survival time having mean λ as

$$C(s,t) = \frac{1}{\lambda}\int_0^t G(s,t')\ e^{-\frac{t'}{\lambda}}\ dt' = \frac{1}{\lambda}\int_0^t e^{-\alpha t'(1-s)}\ e^{-\frac{t'}{\lambda}}\ dt'$$
$$= \frac{1-e^{-t(\alpha-\alpha s+\frac{1}{\lambda})}}{1+\alpha\lambda(1-s)}.$$

□

Remark 4. *Since $|s| < 1$ and $\alpha, \lambda > 0$, we have $(\alpha - \alpha s + \frac{1}{\lambda}) > 0$. Therefore, taking limit $t \to \infty$, in equation (4), we obtain the generating function for the process, described in corollary 3, at the steady state as*

$$C(s) = \frac{1}{1+\alpha\lambda(1-s)}. \tag{6}$$

III. HARDNESS OF HERD IMMUNITY

In our model of infection dynamics, an infected individual makes contacts with other individuals as a Poisson process with rate α, and remains infectious for an exponentially distributed random time period, having mean λ. Therefore, the generating function for the number of infectious contacts (contacts made while the person is still infectious) made by an infected individual up to time t, and during its full lifetime are given by $C(s,t)$ and $C(s)$ in equations (4) and (6) respectively.

We now recall a theorem for deriving the generating functions for compound distributions, which we shall use for some of our calculations.

Theorem 5. *Let N be a non-negative integral valued random variable with generating function $G(s)$ and let $\{X_i\}$ be a sequence of independent and identically distributed (iid) non-negative integral valued random variables with generating function $R(s)$. Then the generating function for the compound random variable $S_N = X_1 + X_2 + ... + X_N$ is $G(R(s))$.*

For a proof of theorem 5 see [35].

With our model assumptions, let us further assume that at time t the fraction of susceptibles present in a population be p_s, and the probability of disease transmission from an infected individual to a susceptible individual, who have been in an infectious contact, be p_c. Therefore, the probability of generating a new infected individual from a random

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

contact (either with susceptible or with already infected/immuned) with an infected individual is $p_s p_c$. This event of creating a new infected individual, from a random contact with an infected individual, can be thought of as a Bernoulli trial X with success probability $\mathbb{P}(X = 1) = p_s p_c$ and $\mathbb{P}(X = 0) = 1 - p_s p_c$. The generating function corresponding to this Bernoulli trial is

$$B(s) = (1 - p_s p_c) + p_s p_c s = 1 - p_s p_c (1 - s). \tag{7}$$

Now, the number of newly infected individuals from an existing infected individual, in time t, is the sum of N number of Bernoulli trials X, where N is a random number representing the number of infectious contacts made by the infected individual. Therefore, using theorem 5, from equation (4) and (7) we obtain the number of newly infected individuals, from an existing infected individual, in time t from the onset of its infection, has generating function

$$\begin{aligned} C(B(s), t) &= \frac{1 - e^{-t(\alpha - \alpha B(s) + \frac{1}{\lambda})}}{1 + \alpha\lambda(1 - B(s))} \\ &= \frac{1 - e^{-t(\alpha p_s p_c (1-s) + \frac{1}{\lambda})}}{1 + \alpha\lambda p_s p_c (1 - s)}. \end{aligned} \tag{8}$$

Taking limit $t \to \infty$ in equation (8) we get the generating function for the same at steady state as

$$C(B(s)) = \frac{1}{1 + \alpha\lambda(1 - B(s))} = \frac{1}{1 + \alpha\lambda p_s p_c (1 - s)}. \tag{9}$$

Equation (9) can also be obtained directly from equation (6), by replacing s by $B(s)$ from equation (7). Note that here we are using assumption 6, i.e. the fraction of susceptibles p_s, in the population, is constant over the infectious time period of an individual. Figure 1 shows the probability of the number of first generation progenies created by an infected individual, for different incubation periods of the contagion (evaluating the Taylor series coefficients of $C(B(s))$ in equation (9)). Equation (9) is the generating function for the number of newly created infected individuals, by an existing infected individual throughout its whole infectious period. Therefore, the average number of new infections generated by an existing infected individual is

$$\left.\frac{d}{ds} C(B(s))\right|_{s=1} = \alpha\lambda p_s p_c. \tag{10}$$

Recalling the definition of basic reproduction number R_0 for an infectious disease, as the average number of new infections generated by an existing infected individual, we identify

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

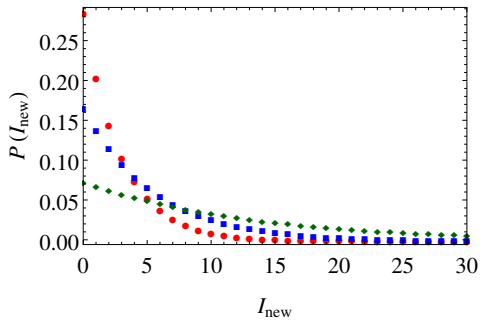

FIG. 1. (Color online) Probability distributions of the number of newly created infections from an existing infected individual. Horizontal axis is represents the number of new infections created, and the vertical axis represents the corresponding probabilities. Red circle, blue square, and green diamond plots are for $\lambda = 10, 20,$ and 50, respectively. $\alpha = 10, p_s = 0.5, p_c = 0.05$ for all plots.

that equation (10) gives the R_0 for our model of infection propagation. It is clear from equation (10) that R_0 is a function of the fraction of susceptibles present in the population (p_s), probability of disease transmission (p_c), the rate of contacts between the individuals in the population (α), and the average incubation period (λ). Now, using theorem 1 we can conclude that the disease will die out over time, with probability one, if $R_0 = \alpha\lambda p_s p_c \leq 1$, i.e. if

$$p_s \leq \frac{1}{\alpha\lambda p_c}. \tag{11}$$

Inequality (11) has many significant implications about the achievement of herd immunity. First of all, it gives an upper bound for the fraction of susceptibles in the population, below which the infection in the population dies out with probability one, which is known as the herd immunity in epidemiological terminology. It is also evident from inequality (11) that when the mean incubation period (λ) is high, the upper bound of p_s is small. This means, to achieve herd immunity in the case of a contagion with high incubation period, $(1 - p_s)$ (which is close to 1) fraction of total population must be either immuned or infected, and hence, very hard to achieve. Proper use of personal protective equipments (PPE), like masks, hand sanitizers, face shields etc. can reduce the probability of disease transmission (p_c), and hence helps to achieve herd immunity easily. Alternatively, imposing quarantine measures can reduce the rate of contacts between infected and susceptible population (α), and helps to achieve herd immunity. Therefore, in absense of any pharmaceutical interventions, like vaccines or other medicines, the alternative way to mitigate infection is by the use of PPE

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

and/or quarantine measures. There is a popular misconception that quarantine measures actually do not help to mitigate infection, rather just slow down (so called "flattening the curve") its propagation to avoid overwhelming the healthcare systems. However, it is clear from inequality (11) that the quarantine measures can also help to achieve the herd immunity and mitigate the infection faster in the population.

IV. SUCCESS PROBABILITY OF QUARANTINE MEASURES

Having obtained the generating function for the number of new infections, generated by an existing infected individual, in equation (9), we can now view our infection propagation model as the following branching process.

1. Every infected individual is a particle which is capable of generating new particles of the same kind.
2. Every particle has probability p_k $(k = 0, 1, 2, ...)$ of generating exactly k new particles during its lifetime.
3. The direct descendants of nth generation form the $(n+1)st$ generation.
4. The particles of each generation acts independent of each other.

Denoting the generating function for the number of first generation progenies created by an ancestor, in the above stated branching process, as $G(s)$, we now recall another theorem for calculating the extinction probability of the line of a single ancestor.

Theorem 6. *For a branching process, as stated in above four points, the extinction probability of a line, i.e. the probability that there will be no particle present at large time, is given by the minimum of the positive root x of the equation $x = G(x)$, and 1.*

See [35, 36] for a proof of theorem 6.

In our model of infection propagation, the generating function for the number of first generation progenies (newly generated infections) from a single infected individual is given by $C(B(s))$ in equation (9). Therefore, the extinction probability of all progenies (of all generations, i.e. the line of an ancestor) of an infected individual is obtained by solving the

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

equation $x = C(B(x))$, which is same as the quadratic equation

$$\alpha\lambda p_s p_c x^2 - (\alpha\lambda p_s p_c + 1)x + 1 = 0. \tag{12}$$

Here we make a strong use of assumption 6, which assumes that p_s varies so slowly that it remains almost constant for a line of an ancestor. This assumption can be supported by realizing that the line of an ancestor is likely to be in the same region or environment in the population. Equation (12) has two roots, 1 and $1/\alpha\lambda p_s p_c$. Therefore, with the help of theorem 6, the extinction probability for all progenies or line of an infected individual is given by

$$p_e = \min\left\{\frac{1}{\alpha\lambda p_s p_c}, 1\right\}. \tag{13}$$

Figure 2 shows the extinction probability p_e as the abscissa of the point where the curves

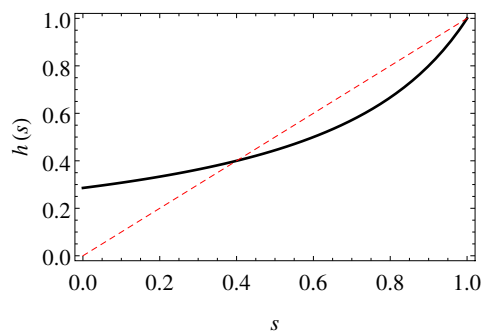

FIG. 2. (Color online) The black solid curve is the plot for $h(s) = C(B(s))$, for $\alpha = 10, \lambda = 10, p_s = 0.5, p_c = 0.05$, and the red dashed curve is $h(s) = s$ straight line. The extinction probability of all progenies or line of an ancestor is the minimum value of $s \in [0, 1]$ at which these two curves intersect, which is 0.4 in this example.

$h(s) = s$ and $h(s) = C(B(s))$ intersect.

According to our model of infection propagation, an infected individual remains infectious for an exponentially distributed time (incubation period of the contagion) with mean incubation period λ. We now intend to calculate how many individuals will remain infected after time $t = T$ if we start observing a fixed N_I number of infected individuals from time $t = 0$.

Theorem 7. *Let the incubation period T_I of a contagion has a distribution function $\mathbb{P}(T_I \leq t) = F(t)$. If we start observing N_I number of infected individuals, each infected from time*

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

$t = 0$, then the generating function for the number of individuals who will still remain infected after time $t = T$ is given by

$$H(s) = [F(T) + (1 - F(T))s]^{N_I} \tag{14}$$

Proof. An infected individual remains infectious after time T, only when $T_I > T$. Since the incubation time T_I has distribution function $F(t)$,

$$\begin{aligned}\mathbb{P}(T_I > T) &= 1 - \mathbb{P}(T_I \leq T)\\ &= 1 - F(T).\end{aligned}$$

Now, we can think of N_I number of Bernoulli trials with success probability $(1 - F(T))$. The total number of successes of these Bernoulli trials results the number of individuals who will still remain infectious after time T. In other words, this number is given by the random variable $S_{N_I} = X_1 + X_2 + ... + X_{N_I}$, where each X_i is a Bernoulli random variable with $\mathbb{P}(X_i = 1) = 1 - F(T)$, and $\mathbb{P}(X_i = 0) = F(T)$. The generating function for each Bernoulli trial X_i is $[F(T) + (1 - F(T))s]$. Therefore, using theorem 5 we obtain the generating function of S_{N_I} as

$$H(s) = [F(T) + (1 - F(T))s]^{N_I}.$$

□

Corollary 8. *Let the incubation period of a contagion be exponentially distributed with mean λ. If we start observing N_I number of infected individuals from time $t = 0$, then the generating function for the number of individuals who will still remain infected after time $t = T$ is given by*

$$H(s) = (1 - e^{-\frac{T}{\lambda}} + e^{-\frac{T}{\lambda}}s)^{N_I} \tag{15}$$

Proof. As the exponential distribution is memoryless, we need not worry about the instants when each individual got infected. Therefore, we can assume that all of the N_I individuals got the infection, simultaneously at $t = 0$, i.e. the time we start our observation. Since the incubation period T_I is exponentially distributed with mean λ, its distribution function is given by

$$F(t) = \mathbb{P}(T_I \leq t) = \int_0^t \frac{1}{\lambda} e^{-\frac{t'}{\lambda}} dt' = 1 - e^{-\frac{t}{\lambda}}.$$

Therefore, using theorem 7, we obtain the desired result. □

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 3 shows the probabilities of the number of remaining infected individuals (among those who are being observed) after the time duration of observation, by plotting the Taylor series coefficients of equation (15).

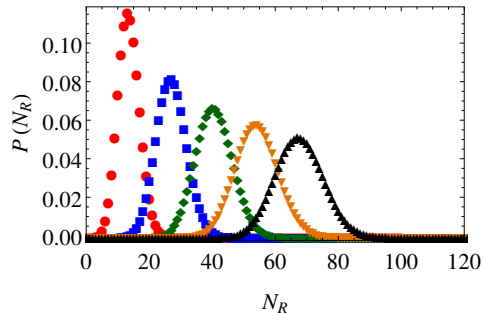

FIG. 3. (Color online) Probability distribution for the number of infected individuals remaining after the withdrawal of quarantine measure. The horizontal axis denotes the number of infected, remaining after $T = 40$ days of quarantine. The average incubation period λ, of the contagion, is taken to be 20 days. The number of infected N_I at time $t = 0$ is taken to be $100, 200, 300, 400,$ and 500, for the red circle, blue square, green diamond, orange inverted triangle, and black triangle plots, respectively.

We now prove a theorem which will have very useful applications in calculating the extinction probability of the infectious disease by quarantine measures.

Theorem 9. *Let N number of iid Bernoulli trials, each having success probability p, are performed, where N is a positive integral valued random variable with generating function $Q(s)$. Then the probability of obtaining all success is $Q(p)$.*

Proof. Let $\mathbb{P}(N = k) = q_k$, for $k = 0, 1, 2, \ldots$. Then by the definition of generating function

$$Q(s) = \sum_{k=0}^{\infty} q_k s^k.$$

By conditioning on the number of trials N, we obtain the probability of all success in random number of Bernoulli trials as

$$\begin{aligned}\mathbb{P}(\text{All success}) &= \sum_{k=0}^{\infty} \mathbb{P}(k \text{ success} \mid N = k)\mathbb{P}(N = k)\\ &= \sum_{k=0}^{\infty} p^k q_k = Q(p).\end{aligned}$$

□

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Theorems 7 and 9 together can be used to estimate the success probability of any quarantine measure in mitigating infection. Here we assume an ideal quarantine measure where no contact between an infected person and a susceptible person occurs. Let at some point of time there be N_I number of infected individuals present in the population. Let an ideal quarantine measure is imposed at that time for next T days, and after T days the quarantine is withdrawn. During the quarantine period T, some infected individuals can be recovered and others will remain infected. The generating function for the number of individuals (N_R), who will still remain infectious after T days of quarantine, is given by $H(s)$ in equation (15). After the quarantine is withdrawn, the remaining infected individuals will again start interacting with the susceptible individuals in the population and create new infections. The probability that the number of all progenies of a single infected individual will be zero, after sufficient time, is given by p_e in equation (13). Now, the event that the total number of progenies of all the remaining N_R number of infected individuals is zero after sufficient time, is equivalent to the event that N_R number of iid Bernoulli trials, each having success probability p_e, results all successes. Since N_R has the generating function same as in equation (15), using theorem 9, we obtain the probability that the infection will be mitigated after sufficient time from the withdrawal of quarantine is

$$H(p_e) = (1 - e^{-\frac{T}{\lambda}} + e^{-\frac{T}{\lambda}} p_e)^{N_I}. \tag{16}$$

As $T \to \infty$ in equation (16), the extinction probability $H(p_e) \to 1$. This implies that as the duration of quarantine increases, the probability of infection mitigation becomes higher. Also, when $p_e = 1$, i.e. $\alpha\lambda p_s p_c \leq 1$ (by equation (13)), and hence herd immunity is achieved, $H(p_e) = 1$, i.e. the infection is mitigated with probability one. Figure 4 shows the probability of zero infection as a function of the duration of quarantine, before achieving herd immunity.

V. EARLY IMPOSITION OF LOCKDOWN CAN SOMETIMES BE LESS EFFECTIVE THAN A DELAYED IMPOSITION

We now study our infection propagation model in a toy population, and see some counter-intuitive results. Let the size of the total population be $N = 10000$, among which N_I number of individuals are infected at time $t = 0$. Therefore, the number of susceptibles at time $t = 0$

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

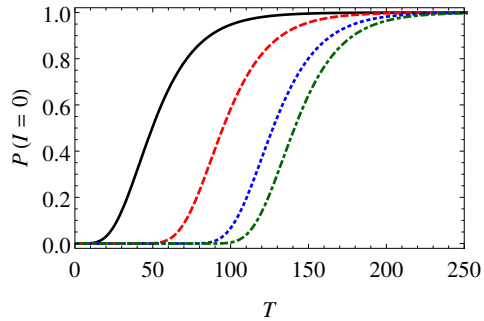

FIG. 4. (Color online) Extinction probabilities of the infection as a function of the duration of quarantine. Horizontal axis represents the duration of quarantine, and vertical axis represents the corresponding extinction probabilities. Black solid, red dashed, blue dotted, and green dot-dashed curves are for $N_I = 10, 100, 500,$ and 1000 respectively. $\alpha = 10, \lambda = 20, p_s = 0.5, p_c = 0.05$ are taken for all plots.

is $N - N_I$. Hence, the fraction of susceptibles in the population is $p_s = (N - N_I)/N$. Let the transmission probability p_c be 0.05. Let us consider that there is no relapse of infection, i.e., if someone has ever been infected, they cannot be susceptible anymore, even after their recovery, or end of their infectious period. Consequently, the fraction of susceptibles cannot be higher than $(N - N_I)/N$ for any $t \geq 0$. If we now calculate the extinction probability as in equation (16), as a function of lockdown/quarantine duration, we obtain Figure 5. We see in Figure 5 that when $N_I = 8500$, the infection dies out with higher probability

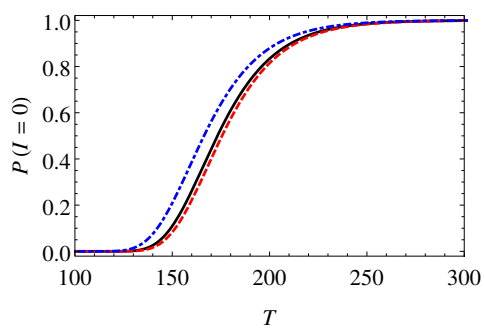

FIG. 5. (Color online) Extinction probabilities of the infection as a function of the duration of quarantine, in a toy population of size $N = 10000$. Horizontal axis represents the duration of quarantine, and vertical axis represents the corresponding extinction probabilities. Black solid, red dashed, and blue dot-dashed curves are for $N_I = 5000, 6000,$ and 8500, respectively. $\alpha = 10, \lambda = 20, p_c = 0.05$ are taken for all plots.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

than the situation when $N_I = 5000$ or 6000, upon the imposition of quarantine measures of same duration. It is to be noted that none of these three cases have achieved herd immunity threshold, as $\alpha\lambda p_s p_c > 1$ for all three cases. The reason for this is, as the size of the infected (or immune, after recovery from infection) population increases, the fraction of susceptibles p_s, in the population, decreases. Therefore, the probability of an infectious contact (contact between an infected and an susceptible individual) decreases. Hence, it is a tug of war between the number of infected N_I, and the fraction of remaining susceptibles $p_s = (N - N_I)/N$, in the population, to maximize the probability of total extinction, $H(p_e)$ in equation (16). This therefore implies that imposing an early quarantine measure may not always be most effective, unless the quarantine is maintained sufficiently long to mitigate infection fully. Otherwise, it is a better strategy to let the infection spread into the population up to some time to ensure the number of remaining susceptibles is relatively low, and then impose the quarantine measure for the same duration. The later strategy will then have a higher probability of success in mitigating infection from the population.

It is to be noted here that after the withdrawal of lockdown the infection still keeps on propagating in the population to some extent, even before eradication, if it at all happens. Therefore, it needs a detailed calculation (recursive and possibly numerical) to comment on the total number of infections occurring before it is eradicated from the population [16, 17]. However, it is also to be understood that when N_I number of people become infected at some point of time, not all of them remains infected at that time instant. So, the extinction probabilities in figure 5 is prone to underestimation. Therefore, to decide on the best strategy for mitigating an infection from a population, using quarantine measures, it needs an extensive analysis, to save both economy and healthcare system in an optimized way.

VI. THE DISTRIBUTION OF GENERATION TIME

The generation time for an infectious disease is defined as the time interval between the onset of infection in an individual to the first generation of another new infected individual by the primary infected individual [27]. We now calculate the distribution of the generation time for our model of infectious disease. To derive the distribution function of the generation time we need to use the generating function for the tail of a discrete random variable. More specifically, let N be a non-negative integral valued random variable having distribution

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

$\mathbb{P}(N = k) = p_k$ (possibly improper distribution), for $k \in \{0, 1, 2, ...\}$. Let M be defined as the tail of N having distribution $\mathbb{P}(M = k) = \mathbb{P}(N > k) = p_{k+1} + p_{k+2} + ... = q_k$, for $k \in \{0, 1, 2, ...\}$. We now prove a theorem which gives the relation between the generating functions of N and M.

Theorem 10. *If the generating function of N is $P(s)$, then the generating function for its tail M is given by*

$$Q(s) = \frac{P(1) - P(s)}{1 - s}. \tag{17}$$

Proof. The proof goes exactly in the same line as in [35], where the generating function for the tail distribution is derived for a proper distribution. By definition,

$$Q(s) = q_0 + q_1 s + q_2 s^2 + q_3 s^3 + ...,$$

where $q_k = p_{k+1} + p_{k+2} + ...$, for $k \in \{0, 1, 2, ...\}$. Therefore, the coefficient of s^n in $(1-s)Q(s)$ equals $q_n - q_{n-1} = -p_n$ when $n \geq 1$, and equals $q_0 = p_1 + p_2 + p_3 + ... = P(1) - p_0$ when $n = 0$. Therefore,

$$(1 - s)Q(s) = P(1) - p_0 - \sum_{n=1}^{\infty} p_n s^n = P(1) - P(s),$$

and hence the desired result. □

We now derive a general theorem to calculate the distribution of generation time, with the help of theorem 10.

Theorem 11 (Generation time distribution)**.** *Let $G(s, t)$ be the generating function for the number of first generation progenies, created by an ancestor of an age-dependent branching process up to time t, with steady state generating function $G(s)$, i.e. $\lim_{t\to\infty} G(s, t) = G(s)$. Let the number of progenies created by the ancestor throughout its lifetime be greater than or equal to m. Then the time τ_m, required to create the first m number of progenies by the ancestor, has the distribution*

$$\mathbb{P}(\tau_m \leq t) = \frac{\frac{\partial^{(m-1)}}{\partial s^{(m-1)}} \left(\frac{G(1,t) - G(s,t)}{1-s} \right) \Big|_{s=0}}{\frac{d^{(m-1)}}{ds^{(m-1)}} \left(\frac{G(1) - G(s)}{1-s} \right) \Big|_{s=0}}, \tag{18}$$

and hence, the generation time τ has the distribution

$$\mathbb{P}(\tau \leq t) = \frac{G(1, t) - G(0, t)}{G(1) - G(0)}. \tag{19}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Proof. The event that $\tau_m \leq t$ is same as the event that the number of progenies $N(t)$, created up to time t, is greater than or equal to m. This implies

$$\begin{aligned}\mathbb{P}(\tau_m \leq t) &= \mathbb{P}(N(t) \geq m \mid N(\infty) \geq m) \\ &= \mathbb{P}(N(t) > m-1 \mid N(\infty) > m-1) \\ &= \frac{\mathbb{P}(N(t) > m-1, N(\infty) > m-1)}{\mathbb{P}(N(\infty) > m-1)} \\ &= \frac{\mathbb{P}(N(t) > m-1)}{\mathbb{P}(N(\infty) > m-1)}. \end{aligned} \tag{20}$$

Identifying $\mathbb{P}(N(t) > m-1)$ as the tail distribution of the branching process under consideration, at time t, using theorem 10 we conclude that its generating function has the form

$$Q(s,t) = \frac{G(1,t) - G(s,t)}{1-s},$$

and $\mathbb{P}(N(t) > m-1)$ is the coefficient of s^{m-1} of the Taylor series expansion of $Q(s,t)$ with respect to s. Similarly, $\mathbb{P}(N(\infty) > m-1)$ is the coefficient of s^{m-1} in the Taylor series expansion of $Q(s)$, where $Q(s) = \lim_{t\to\infty} Q(s,t)$. Hence, substituting these values of $\mathbb{P}(N(t) > m-1)$ and $\mathbb{P}(N(\infty) > m-1)$ in equation (20) we obtain equation (18). Finally, we obtain equation (19) from equation (18), for $m = 1$. □

Having obtained the general procedure to derive the distribution function of generation time, we now apply theorem 11 to derive the distribution of generation time for our model. In our model we have $G(s,t) = C(B(s),t)$. Therefore, using equations (8) and (19), we obtain the distribution function for generation time as

$$\begin{aligned}\mathbb{P}(\tau \leq t) &= \frac{C(B(1),t) - C(B(0),t)}{C(B(1)) - C(B(0))} \\ &= \frac{\alpha\lambda p_s p_c - e^{-\frac{t}{\lambda}}(1+\alpha\lambda p_s p_c) + e^{-t(\alpha p_s p_c + \frac{1}{\lambda})}}{\alpha\lambda p_s p_c}. \end{aligned} \tag{21}$$

Therefore, the density function for the generation time is given by

$$\begin{aligned} g(t) &= \frac{d}{dt}\left[\mathbb{P}(\tau \leq t)\right] \\ &= \frac{(\alpha p_s p_c + \frac{1}{\lambda})}{\alpha\lambda p_s p_c} e^{-\frac{t}{\lambda}}(1 - e^{-t\alpha p_s p_c}) \end{aligned} \tag{22}$$

Figure 6 plots the probability density function $g(t)$, for the generation time, in case of two different incubation periods. As seen from the plot that the mode of the distribution does

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

not change much with the incubation period. Dependence on other parameters can also be checked, and it can be shown that decreasing α, p_s, and p_c will shift the mode towards right, as it is evident that decreasing these parameters will delay the generation of a new infection.

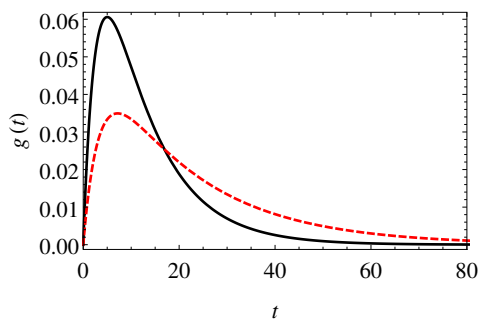

FIG. 6. (Color online) Probability density function for the generation time. Black solid curve is for $\lambda = 10$; and red dashed curve is for $\lambda = 20$. Other parameters for both the curves are taken to be $\alpha = 10, p_s = 0.5, p_c = 0.05$.

VII. CONCLUSION

We model the propagation of an infectious disease, in a population, as an age-dependent branching process, under reasonable assumptions. We obtain that the achievement of herd immunity is highly dependent on the incubation period of the contagion, and is difficult to achieve when the incubation period is high. We show that how the mass use of PPE can help to achieve herd immunity, and to eradicate the infection faster from the population. We also calculate the success probability of various quarantine measures, and show by considering a hypothetical situation that an early imposition of lockdown may not always be a better strategy against a delayed imposition of lockdown of the same duration. Finally, we derive a general theorem to calculate the distribution of generation time, which can be used in the study of any system, modeled as age-dependent branching process. We apply this general theorem to derive the generation time distribution for our model of infection propagation, and obtain a two parameter (effectively) distribution which can be used in epidemiological studies, as a logical replacement of hitherto used ad hoc distributions, such as gamma

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

distribution, log-normal distribution etc.

---

[1] R. Bartoszynski, in *Proceedings of the Fifth Berkeley symposium* (1967) pp. 259–69.

[2] R. Bartoszyński, Mathematical Biosciences **24**, 355 (1975).

[3] S. J. Merrill, Journal of computational and applied mathematics **184**, 242 (2005).

[4] T. Britton, Mathematical biosciences **225**, 24 (2010).

[5] T. Britton and D. Lindenstrand, Mathematical biosciences **222**, 109 (2009).

[6] J. A. Backer, D. Klinkenberg, and J. Wallinga, Eurosurveillance **25**, 2000062 (2020).

[7] P. E. SARTWELL *et al.*, American Journal of Epidemiology **83**, 204 (1966).

[8] C. McAloon, Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O'Grady, and S. J. More, BMJ Open **10** (2020), 10.1136/bmjopen-2020-039652, https://bmjopen.bmj.com/content/10/8/e039652.full.pdf.

[9] J. Lessler, N. G. Reich, R. Brookmeyer, T. M. Perl, K. E. Nelson, and D. A. Cummings, The Lancet infectious diseases **9**, 291 (2009).

[10] M. J. Lydeamore, P. T. Campbell, D. J. Price, Y. Wu, A. J. Marcato, W. Cuningham, J. R. Carapetis, R. M. Andrews, M. I. McDonald, J. McVernon, *et al.*, PLOS Computational Biology **16**, e1007838 (2020).

[11] O. N. Bjørnstad, K. Shea, M. Krzywinski, and N. Altman, Nature methods **17**, 455 (2020).

[12] I. Czumbel, *Systematic review on the incubation and infectiousness/shedding period of communicable diseases in children* (ECDC, 2016).

[13] A. L. Lloyd, Theoretical population biology **60**, 59 (2001).

[14] K. Pandey, S. Basu, R. Pinto, and S. P. Pandey, Bulletin of the American Physical Society (2020).

[15] M. G. Hudgens, I. M. Longini Jr, M. E. Halloran, K. Choopanya, S. Vanichseni, D. Kitayaporn, T. D. Mastro, and P. A. Mock, Journal of the Royal Statistical Society: Series C (Applied Statistics) **50**, 1 (2001).

[16] T. Sellke, Journal of Applied Probability , 390 (1983).

[17] F. Ball, Advances in Applied Probability , 289 (1986).

[18] F. Ball, Advances in applied probability , 1 (1985).

[19] P. Picard and C. Lefevre, Advances in Applied Probability **22**, 269 (1990).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.22.20216481; this version posted October 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

[20] C. L. Addy, I. M. Longini Jr, and M. Haber, Biometrics , 961 (1991).

[21] W. O. Kermack and A. G. McKendrick, Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character **115**, 700 (1927).

[22] Y. Cai, Y. Kang, M. Banerjee, and W. Wang, Journal of Differential Equations **259**, 7463 (2015).

[23] R. N. Thompson, T. D. Hollingsworth, V. Isham, D. Arribas-Bel, B. Ashby, T. Britton, P. Challenor, L. H. Chappell, H. Clapham, N. J. Cunniffe, *et al.*, Proceedings of the Royal Society B **287**, 20201405 (2020).

[24] T. Britton, F. Ball, and P. Trapman, Science **369**, 846 (2020).

[25] J. Xie and Y. Zhu, Science of the Total Environment **724**, 138201 (2020).

[26] H. A. Aboubakr, T. A. Sharafeldin, and S. M. Goyal, Transboundary and Emerging Diseases (2020).

[27] H. Nishiura, Mathematical Biosciences & Engineering **7**, 851 (2010).

[28] Y. Deng, C. You, Y. Liu, J. Qin, and X.-H. Zhou, Biometrics (2020).

[29] K. Dietz, Statistical methods in medical research **2**, 23 (1993).

[30] G. N. Milligan and A. D. Barrett, *Vaccinology: An Essential Guide*.

[31] D. G. Kendall, Biometrika **35**, 316 (1948).

[32] D. G. Hoel and K. S. Crump, Biometrics , 125 (1974).

[33] C. Kelly and O. Rahn, Journal of Bacteriology **23**, 147 (1932).

[34] R. Cowan, Biometrics , 681 (1985).

[35] W. Feller, *An introduction to Probability theory and its applications, Vol. 1* (John Wiley and Sons, Inc., New York, 1950).

[36] T. E. Harris, *The theory of branching process* (Rand Corporation, 1964).

[37] W. Feller, *An introduction to probability theory and its applications, Vol. 2* (John Wiley and Sons, Inc., New York, 1965).