On the spread of epidemics in a closed heterogeneous population

Artem S. Novozhilov *

National Institutes of Health, NCBI, 8600 Rockville Pike, Bldg 38A room 8N811H, Bethesda, MD 20894, USA

ARTICLE INFO

ABSTRACT

Heterogeneity is an important property of any population experiencing a disease. Here we apply general methods of the theory of heterogeneous populations to the simplest mathematical models in epidemiology. In particular, an SIR (susceptible-infective-removed) model is formulated and analyzed when susceptibility to or infectivity of a particular disease is distributed. It is shown that a heterogeneous model can be reduced to a homogeneous model with a nonlinear transmission function, which is given in explicit form. The widely used power transmission function is deduced from the model with distributed susceptibility and infectivity with the initial gamma-distribution of the disease parameters. Therefore, a mechanistic derivation of the phenomenological model, which is believed to mimic reality with high accuracy, is provided. The equation for the final size of an epidemic for an arbitrary initial distribution of susceptibility is found. The implications of population heterogeneity are discussed, in particular, it is pointed out that usual moment-closure methods can lead to erroneous conclusions if applied for the study of the long-term behavior of the models.

Published by Elsevier Inc.

1. Introduction

Mathematical modeling of epidemics arguably began with the pioneer work of Ross [35] and developed considerably ever since (e.g., [4,5]). Especially significant contribution was made by Kermack and McKendrick [26], two students of Ross, who considered the situation of microparasite infection, where contacts between individuals are made according to the law of mass-action, all individuals are identical, the population is closed, and the population size is large enough to apply a deterministic description (for a brief review see [7]). Additionally, if it is assumed that the individuals are infected for an exponentially distributed period of time, then a usual SIR model in the form of ordinary differential equations (ODEs) can be written down for the sizes of susceptible, infective and removed classes.

Since the work of Kermack and McKendrick a great deal of mathematical models were suggested that relax one or more assumptions that led to the Kermack–McKendrick model. A substantial part of this work was devoted to incorporate heterogeneity into the mathematical models, which is also the main subject of the present text. In what follows we retain the assumptions of random mixing, no inflow of susceptible or infected hosts, exponentially distributed infectious period, and validity of deterministic description. Specifically, we will look into heterogeneity in disease parameters (such as susceptibility to a disease); disease parameters are considered as an inherent and invariant property of

individuals, whereas the parameter values can vary between individuals.

The most common way to take into account parametric heterogeneity is to divide population into groups [9,16–18], usually a small number of groups. An important disadvantage of the subgroup approach is that heterogeneity within a group cannot be incorporated if a fixed number of groups is considered. Another approach, which we pursue here, is to consider the population as having a continuous distribution (see, e.g., [5–7,10–12,33,39]), which can be a limiting case of a discrete group model as it was done by May and Anderson [31] (eventually they used a continuous gamma-distribution).

Our approach to formulate mathematical models is close to that applied in, e.g., [12], where known experimental data forced the authors to take into account heterogeneity among hosts in their susceptibility to the virus among other key details, and a simple SIR model was adjusted to account for new information. The major novelty of the present text is to introduce the well developed theory of heterogeneous populations into the epidemiological modeling. Using simple models we are able to obtain known results with less effort, and, more importantly, produce new analytical results. In particular, we show that a heterogeneous SIR model with distributed susceptibility can be reduced to a homogeneous one with a nonlinear transmission function; the exact form of this function is given. It turns out that widely applied nonlinear transmission function in the form of power relationship, $\beta S^p I^q$, can be a consequence of the intrinsic heterogeneity in susceptibility and infectivity parameters. For a heterogeneous SIR model the equation for the final epidemic size is found. The explicit form of the final epidemic

* Tel.: +1 301 402 4146.
  E-mail address: novozhil@ncbi.nlm.nih.gov

size emphasizes and illustrates the fact that the goal to model the evolution of a heterogeneous population for a long time can be accomplished only if the exact initial distribution is available. Any moment-closure methods may lead to erroneous estimates.

Our paper is organized as follows. In Section 2, we formulate the basic models and discuss various assumptions that may lead to them. In Section 3, we review the necessary analytical tools from the theory of heterogeneous populations. In Section 4, a homogeneous model that is equivalent to the heterogeneous one is explicitly constructed, and it is shown that the former has a nonlinear transmission function; moreover, the well known power transmission function is shown to be a consequence of the initial gamma-distribution. In Section 5, the influence of heterogeneity on the disease course is studied for an SIR model with distributed susceptibility, in particular, the final size of an epidemic is found for an arbitrary initial distribution. Section 6 is devoted to discussion and conclusions. Finally, in Appendix we collect the definitions of the probability distributions used throughout the text together with some auxiliary facts from the general theory of heterogeneous models.

2. The basic models with population heterogeneity

2.1. The model with distributed susceptibility

Suppose that each individual of a (sub-)population has its own value of a certain trait (which can be, e.g., susceptibility to a particular disease, social behavior, infectivity, or a hereditary attribute) that describes his or her invariant property and has a marked influence on the spread of the disease at the population level; that is, the key parameters that determine disease evolution in the total population depend on the trait values and we can speak of the trait distribution or the parameter distributions (in general, we speak of a distribution when no ambiguity is expected). The trait value remains unchanged for any given individual during the time period we are interested in, but varies from one individual to another. Any changes of the mean, variance and other characteristics of the trait distribution in the population (or the parameter distributions) are caused only by variation of population structure.

For the moment we assume that the susceptible subpopulation is heterogeneous, and denote $s(t, \omega)$ the density of susceptible individuals at time t having trait value ω (i.e., the size of subpopulation of susceptible hosts having trait values in the range from ω to $\omega + d\omega$ is approximately equal to $s(t, \omega)d\omega$, and the total size of the susceptibles $S(t) = \int_\Omega s(t, \omega) d\omega$, where Ω is the set of trait values). Assuming that the subpopulation of infectives is homogeneous (under the modeling situation), the total size of the population is constant, the disease course keeps within the simplest situation 'susceptible' → 'infective' → 'removed', the contact process is described by the so-called mass-action kinetics (i.e., the contact rate is proportional to the total population size N [5,14,32]), and that the rate of change in the susceptibles is determined by the transmission parameter, which is a function of trait values, we can write down the following equation for the change in the susceptible subpopulation having trait value ω:

$$\frac{\partial}{\partial t} s(t, \omega) = -\beta(\omega) s(t, \omega) I(t), \qquad (1)$$

where $I(t)$ is the size of the subpopulation of infectives, and $\beta(\omega)$ incorporates information on the contact rate and the probability of a successful contact. Hereinafter we assume that the trait that characterizes susceptible individuals is the susceptibility to the disease, although it can be assumed that different individuals have different contact rates (in the latter case it becomes difficult to interpret the equation for the infectives, see below).

The change in the infective class, if the length of being infective is distributed exponentially with the mean time $1/\gamma$, is given by

$$\frac{d}{dt} I(t) = I(t) \int_\Omega \beta(\omega) s(t, \omega) d\omega - \gamma I(t) = \bar{\beta}(t) S(t) I(t) - \gamma I(t), \qquad (2)$$

where we used the following notations:

$$\bar{\beta}(t) = \int_\Omega \beta(\omega) p_s(t, \omega) d\omega, \quad p_s(t, \omega) = \frac{s(t, \omega)}{S(t)}.$$

Hence, $\bar{\beta}(t)$ is the mean value of the function $\beta(\omega)$, and ω has the probability density function (pdf) $p_s(t, \omega)$ for any time moment t. We need to supplement the model (1), (2) with the initial conditions

$$s(0, \omega) = s_0(\omega) = S_0 p_s(0, \omega), \quad I(0) = I_0, \quad R(0) = 0, \qquad (3)$$

and with the third equation for the removed $dR(t)/dt = \gamma I(t)$. Here S_0 and I_0 are the initial sizes of the susceptibles and infectives, respectively, and $p_s(0, \omega)$ is the given initial distribution.

The model (1)–(3) is the basic model we study in this paper. This model was formulated from the first principles, as it was done for conceptually similar models in [12] for the transmission of virus in gypsy moths, in [33] for the effect of antimicrobial agents on microbial populations, in [31] for the spread of HIV in the human population, and in [39] for a class of SIS models (we note, however, that in the last two examples the frequency-dependent transmission was employed [32], and the heterogeneity of the contact rates was modeled).

The model (1)–(3) can be also deduced from the general epidemic equation (see [5])

$$\frac{\partial}{\partial t} s(t, \omega) = s(t, \omega) \int_\Omega \int_0^\infty A(\tau, \omega, \eta) \frac{\partial}{\partial t} s(t - \tau, \eta) d\tau d\eta, \qquad (4)$$

where $A(\tau, \omega, \eta)$ is the expected infectivity of an individual that was infected τ units of time ago while having trait value η towards a susceptible with trait value ω. If we assume that $A(\tau, \omega, \eta) = \beta(\omega) \exp\{-\gamma\tau\}$, and set

$$I(t) = -\int_\Omega \int_0^\infty \exp\{-\gamma\tau\} \frac{\partial}{\partial t} s(t - \tau, \eta) d\tau d\eta,$$

after some algebra we obtain (1)–(3) (see also [7]). As a side remark we note that letting $A(\tau, \omega, \eta) = \beta(\omega) \chi(T - \tau)$, where $\chi(t)$ is the Heaviside function, we obtain the model studied in [12].

2.2. Model with distributed infectivity

It can be also assumed that the population of infectives is heterogeneous. Now let $\beta(\omega)$ be the infectivity of an individual with trait value ω, and $i(t, \omega)$ be the density of the infective hosts with trait value ω at time moment t, $I(t) = \int_\Omega i(t, \omega) d\omega$. For simplicity we assume that the susceptible hosts are homogeneous. The change in the infective subpopulation should incorporate the law that specifies which trait value is assigned to a newly infected individual, and can be described by the following equation:

$$\frac{\partial}{\partial t} i(t, \omega) = S(t) \int_\Omega \psi(\omega, \omega') \beta(\omega') i(t, \omega') d\omega' - \gamma i(t, \omega), \qquad (5)$$

where $\psi(\omega, \omega')$ is the probability that a newly infected individual gets trait value ω if infected by an individual with trait value ω'. The change in the population $S(t)$ is given by

$$\frac{d}{dt} S(t) = -\bar{\beta}(t) S(t) I(t), \qquad (6)$$

where now $\bar{\beta}(t) = \int_\Omega \beta(\omega) p_i(t, \omega) d\omega$, and $p_i(t, \omega) = i(t, \omega)/I(t)$.

The need to specify function $\psi(\omega, \omega')$ precludes the interpretation of the function $\beta(\omega)$ in Section 2.1 as the heterogeneity in the contact rates. It was assumed that the infectives are all identical,

and thus the supposition that an individual that has had the trait value ω turns into another identical infective host is not warranted.

A variety of choices for the function $\psi(\omega, \omega')$ is possible, but we especially interested in the particular case when a newly infected individual gets the same trait value that was possessed by the individual who passed the infection, namely

$$\psi(\omega, \omega') = \delta(\omega' - \omega),$$

where $\delta(\omega)$ is the Dirac delta-function (which can correspond to modeling a situation when several different strains of an infection can be passed on). In this case Eq. (5) simplifies to the equation, which is very similar to (1):

$$\frac{\partial}{\partial t} i(t, \omega) = \beta(\omega) S(t) i(t, \omega) - \gamma i(t, \omega). \tag{7}$$

Model (6), (7) is another example of a simple mathematical model for the spread of an infectious disease in a closed population with heterogeneities. The list of possible models can be easily extended. For instance, it is straightforward to assume that the parameter γ is not constant for the infected individuals, but rather is distributed with a known initial distribution. In this case the equation for the infectives takes the form

$$\frac{\partial}{\partial t} i(t, \omega) = \beta S(t) i(t, \omega) - \gamma(\omega) i(t, \omega),$$

where β is now constant. Another obvious generalization is to assume that several model parameters are distributed (see Section 4.2 for an example).

In general, models (1)–(3) and (6), (7) are infinite-dimensional dynamical systems where the evolutionary operator specifies complex transformations of the initial distributions. Such models are less amenable to qualitative, quantitative, or numerical analysis than their finite-dimensional analogues formulated in terms of ODEs. A usual practice is to formulate an infinite-dimensional system of ODEs for which some approximations methods (e.g., assuming that the initial distribution is close to the normal distribution [33]), moment-closure [10,11,8], or numerical methods [39] can be applied. We show below that in some particular cases, when the analytical form of the heterogeneous models meets certain requirements, the initial model can be reduced to a low-dimensional ODE model, which, in turn, can be effectively analyzed.

3. The necessary facts from the theory of heterogeneous populations

To keep the exposition self-contained and for the sake of convenience of references we briefly survey the necessary results from [21,23]. We present the results in the form suitable for our goal noting that more general cases can be analyzed [23]. For the proofs we refer to [25], where similar models are considered. Some additional facts are given in Appendix.

Let us assume that there are two interacting populations whose dynamics depend on trait values ω_1 and ω_2, respectively. The densities are given by $n_1(t, \omega_1)$ and $n_2(t, \omega_2)$, and the total population sizes $N_1(t) = \int_{\Omega_1} n_1(t, \omega) d\omega_1$ and $N_2(t) = \int_{\Omega_2} n_2(t, \omega) d\omega_2$. Obviously, more than two populations can be considered, or some populations may be supposed to be homogeneous; we choose two not to be drowned in notations. Assume next that the net reproduction rates of the populations have the specific form:

$$\frac{\partial}{\partial t} n_1(t, \omega_1) = n_1(t, \omega_1)[f_1(N_1, N_2, \bar{\varphi}_2(t)) + \varphi_1(\omega_1) g_1(N_1, N_2, \bar{\varphi}_2(t))],$$

$$\frac{\partial}{\partial t} n_2(t, \omega_1) = n_2(t, \omega_2)[f_2(N_1, N_2, \bar{\varphi}_1(t)) + \varphi_2(\omega_2) g_2(N_1, N_2, \bar{\varphi}_1(t))], \tag{8}$$

where $\varphi_i(\omega_i)$ are given functions, $\bar{\varphi}_i(t) = \int_{\Omega_i} \varphi_i(\omega_i) p_i(t, \omega_i) d\omega_i$ are the mean values of $\varphi_i(\omega_i)$, and $p_i(t, \omega_i) = n_i(t, \omega_i)/N_i(t)$ are the corresponding pdfs, $i = 1, 2$. We also assume that $\varphi_i(\omega_i)$, considered as random variables, are independent. The system (8) plus the initial conditions

$$n_i(0, \omega_i) = N_i(0) p_i(0, \omega_i), \quad i = 1, 2, \tag{9}$$

define, in general, a complex transformation of densities $n_i(t, \omega_i)$. For the approach to study such systems based on the analysis of abstract differential equations in Banach spaces we refer to [1–3]. Another approach to analyze models in the form (8) was suggested in [21]. The latter is more attractive because eventually one has to deal with systems of ODEs of low dimensions (examples of model analysis are given in [22,24,25,34]).

Let us denote

$$M_i(t, \lambda) = \int_{\Omega_i} e^{\lambda \varphi_i(\omega_i)} p_i(t, \omega_i) d\omega_i, \quad i = 1, 2,$$

the moment generating functions (mgfs) of the functions $\varphi_i(\omega_i)$, $M_i(0, \lambda)$ are the mgfs of the initial distributions, $i = 1, 2$, which are given.

Let us introduce auxiliary variables $q_i(t)$ as the solutions to the differential equations

$$dq_i(t)/dt = g_i(N_1, N_2, \bar{\varphi}_{i+1}(t)), \quad q_i(0) = 0, \quad i = 1, 2, \tag{10}$$

where indexes that exceed 2 are counted modulo 2.

The following theorem holds

Theorem 1. *Suppose that $t \in [0, T)$, where T is the maximal value of t such that (8), (9) has a unique solution. Then*

(i) *The current means of $\varphi_i(\omega_i), i = 1, 2$, are determined by the formulas*

$$\bar{\varphi}_i(t) = \frac{dM_i(0, \lambda)}{d\lambda} \bigg|_{\lambda = q_i(t)} \frac{1}{M_i(0, q_i(t))}, \tag{11}$$

and satisfy the equations

$$\frac{d}{dt} \bar{\varphi}_i(t) = g_i(N_1, N_2, \bar{\varphi}_{i+1}(t)) \sigma_i^2(t), \tag{12}$$

where $\sigma_i^2(t)$ are the current variances of $\varphi_i(t, \omega_i)$, $i = 1, 2$.

(ii) *The current population sizes $N_1(t)$ and $N_2(t)$ satisfy the system*

$$\frac{d}{dt} N_i(t) = N_i(t)[f_i(N_1, N_2, \bar{\varphi}_{i+1}(t))$$
$$+ \bar{\varphi}_i(t) g_i(N_1, N_2, \bar{\varphi}_{i+1}(t))], \quad i = 1, 2, \tag{13}$$

where indexes that exceed 2 are counted modulo 2.

Theorem 1 gives a method of computation of the main statistical characteristics of $\varphi_i(\omega_i)$; the analysis of model (8), (9) is reduced to analysis of ODE system (10), (11), (13), the only thing we need to know is the mgfs of the initial distributions; the evolution of distributions can also be analyzed [23].

It is worth emphasizing that all the model we consider in this note are deterministic, and the language of probability theory is used because it is very convenient to use probabilistic terms while speaking of evolution (again, deterministic) of distributions. In this respect the discussion of well-posedness of the problem (11), (13) merits some considerations.

It is well known that a mgf, if it exists, defines uniquely the underlying distribution. It is also known that for some distributions the corresponding mgfs do not exist for any $\lambda > 0$ (an example is a log-normal distribution, see Appendix, Eq. (33)). The reason for this is the heavy tails of some distributions. From (11) it can be seen that the mgfs are evaluated at $q_i(t)$, for which the initial conditions $q_i(0) = 0$. Therefore, for some distributions, if $\dot{q}_i > 0$, the

problem (11), (13) cannot be solved at all, and for some distributions, the solution blows up at a finite time (which is the reason to have const T in Theorem 1). From the point of view of mathematical modeling such situations as non-existence of solutions or blow-ups in finite time are the consequences of some unrealistic assumptions. In particular, it is a consequence of the assumption that there are infinitely large parameter values, albeit with extremely small probabilities (more on such situations can be found in [22]). To deal with such situations it can always be assumed that the support of the initial distribution is $[a, b] \in [0, \infty)$, i.e., to consider truncated distributions. In this case the problem (11), (13) is well-posed.

The particular form of the model (8) also specifies exactly what kind of restrictions on the types of heterogeneity can be handled with the considered approach. In the literature on the heterogeneity in epidemiological models the discussions are usually around three aspects. First, this is heterogeneity of hosts when the probability to transmit a disease, given a contact, depends on individuals' susceptibility and infectivity, and this is exactly what we deal with in this paper. Second, this is heterogeneity in contact rates, which cannot be handled with the discussed approach (except in extremely simple cases) because the system (8) yields that each subpopulation should have its *own* parameter, whereas modeling of contact structure would imply that individuals of the whole population are characterized by trait values irrespective of whether they are susceptible or infective. And third, the distributions of latent and infectious periods; for the latter in the considered framework the discussion is restricted to the case when different individuals can have exponential distributions with different means.

In short, using the theory, outlined in this section, we can model communities of populations when each population is characterized by its *own* parameter value, and the relative growth rate depend on this parameter and the *average* characteristics of other distributions.

Concluding this sections we note that, with obvious notation changes, models (1)–(3) and (6), (7) fall into the general framework of the master model (8).

4. Homogeneous models with nonlinear transmission functions

4.1. Reduction to a system of ODEs

We start with Eq. (1), other equations in the system can be quite arbitrary, e.g., the full system can contain the class of exposed or several infective classes. Here we show that a heterogeneous model that contains (1) as a modeling ingredient can be reduced to a homogeneous model with a nonlinear transmission function whose explicit form is determined by the initial distribution of susceptibility. This result is close to the analysis in [39] where it was argued that the model with heterogeneities can be encapsulated in a homogeneous model. Due to the fact that the model considered in [39] is substantially more general than the models we consider here, no explicit formulas were found. In the case of model (1) nonlinear transmission function can be found in the exact form for different initial distributions.

According to Theorem 1 we can rewrite Eq. (1) in the form

$$\frac{d}{dt}S(t) = -\bar{\beta}(t)S(t)I(t), \quad S(0) = S_0,$$

$$\dots$$

$$\frac{d}{dt}q(t) = -I(t), \quad q(0) = 0, \tag{14}$$

$$\bar{\beta}(t) = \frac{dM(0, \lambda)}{d\lambda}\bigg|_{\lambda=q(t)} \frac{1}{M(0, q(t))},$$

where dots denote equations that govern the dynamics of other subpopulations, e.g., those can be usual equations for infected and removed classes. $M(0, \lambda)$ is the given mgf of $p_s(0, \omega)$.

Proposition 1. *Model (14) is equivalent to the following model:*

$$\frac{d}{dt}S(t) = -h(S(t))I(t), \tag{15}$$

$$\dots$$

where dots denote the same as in (14),

$$h(S) = S_0 \left[\frac{dM^{-1}(0, \xi)}{d\xi}\bigg|_{\xi=S/S_0} \right]^{-1}, \tag{16}$$

and $M^{-1}(0, \xi)$ is the inverse function to mgf $M(0, \lambda)$.

Proof. The first equation in (14) can be rewritten in the form

$$\frac{1}{S(t)}\frac{d}{dt}S(t) = \bar{\beta}(t)\frac{d}{dt}q(t).$$

From (11) $\bar{\beta}(t)$ can be represented as $\bar{\beta}(t) = \frac{d\ln M(0,\lambda)}{d\lambda}\big|_{\lambda=q(t)}$, which, together with the previous, gives

$$\frac{d\ln S(t)}{dt} = \frac{d}{dt}\ln M(0, q(t)),$$

or, using the initial conditions $S(0) = S_0, q(0) = 0$,

$$S(t)/S_0 = M(0, q(t)), \tag{17}$$

which is the first integral to system (14). Knowledge of a first integral allows to reduce the order of the system by one. Since $M(0, \lambda)$ is an absolutely monotone function in the case of non-negative $\beta(\omega) \geq 0$, then it follows that

$$q(t) = M^{-1}(0, S(t)/S_0), \tag{18}$$

where $M^{-1}(0, M(0, \lambda)) = \lambda$ for any λ.

Putting (18) into (14) gives

$$\frac{d}{dt}S(t) = \frac{dM(0, \lambda)}{d\lambda}\bigg|_{\lambda=M^{-1}(0,S(t)/S_0)} S_0 I(t),$$

or, by the inverse function theorem, (15) with (16). $\square$

The simple properties of the nonlinear incidence function $h(S)$ are $h(0) = 0$, $h(S_0) = S_0\bar{\beta}(0)$, $h'(S) > 0$.

Let us consider several examples. Definitions for the probability distributions we use can be found in Appendix. In all examples it is assumed that $\beta(\omega) = \omega$, i.e., the transmission coefficient takes the values from the domain of ω with the probabilities corresponding to $p_s(t, \omega)$.

If the initial distribution is a gamma-distribution with parameters k and ν (see Appendix, (30)) we obtain

$$h(S) = \frac{kS}{\nu}\left[\frac{S}{S_0}\right]^{1/k}. \tag{19}$$

If $k = 1$ (i.e., the initial distribution is exponential with mean $1/\nu$), then $h(S) = S^2/(\nu S_0)$. Expression (19) was first obtained in [12] and later used as a nonlinear incidence function in [10].

If the initial distribution is an inverse gaussian (Wald) distribution with parameters μ and ν (the definition is given in Appendix, Eq. (31)), then

$$h(S) = \nu S\left[\frac{\nu}{\mu} - \ln\left(\frac{S}{S_0}\right)\right]^{-1}. \tag{20}$$

In a similar vein other initial distribution can be analyzed, but not all of them have an explicit expression for mgf. Other examples of possible initial distributions are uniform distribution with

parameters a and b, lognormal distribution, beta-distribution, Weibull distribution, Pareto distribution and many others (see Fig. 1 for four different probability distributions).

The theory we briefly presented in Section 3 is equally valid both for continuous and discrete distribution. For instance, if the initial distribution is the Poisson distribution with parameter θ (Appendix, Eq. (35)), then we obtain

$$h(S) = S\left[\theta + \ln\left(\frac{S}{S_0}\right)\right]. \tag{21}$$

Note that in the case of the Poisson distribution there is non-zero probability that a randomly chosen individual has zero transmission coefficient, i.e., the Poisson distribution implies that some individuals are immune.

### 4.2. Derivation of the power transmission function

In standard epidemiological models the incidence rate (the number of new cases in a time unit) was frequently used as a bilinear function of infective and susceptible populations: $\propto SI$. In addition to this it is usually argued that there are a variety of reasons that the standard bilinear form may require modification, including the assumption of heterogeneous mixing [28,36]. We refer to a review paper on the subject [32] for a general account of different models for incidence rates, while noting that one of the most widely used models has the form

$$\beta S^p I^q, \quad p, q > 0, \tag{22}$$

and is of direct interest to our study. The incidence rate in the form (22) was first used in [37], with the restriction that $p, q < 1$, but generally it is only required that $p, q > 0$, see also [15,27,29]. It is interesting to note in the context of our exposition that the exponents p, q in (22) were dubbed as 'heterogeneity parameters,' but the model itself is considered phenomenological and lacking mechanistic derivation [32].

A special case of (22) is $q = 1$, which was considered in [13,38]; the values for parameter p were considered $p > 1$. Comparison of the model (22) with the ODE (15) when the initial distribution of the susceptible subpopulation is a gamma-distribution (see (19)) let us state the following corollary.

Corollary 1. *The power relationship (22) of the incidence rate in an SIR model for the case $q = 1, p = 1 + 1/k > 1$ can be obtained as a consequence of the heterogeneous model (1)–(3) when the initial distribution of the susceptible subpopulation is a gamma-distribution with parameters k and v (see (30)).*
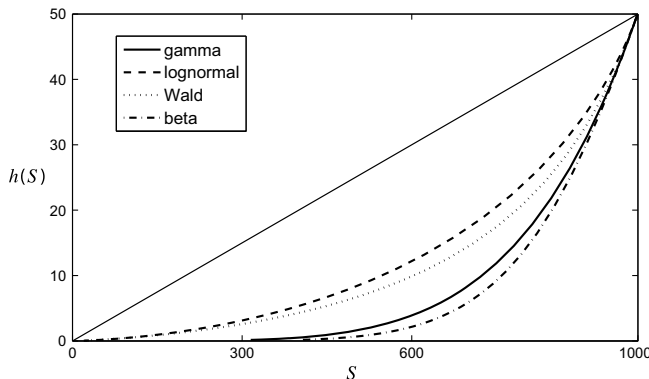
Fig. 1. The functions $h(S)$ given by (16) for four different probability distributions with the same means and variances. The straight line shows the same function for the homogeneous model $h(S) = \beta S$. Definitions of the probability distributions are given in Appendix.

As expected, letting $k \to \infty$ and having k/v fixed we obtain the population where everyone has the same susceptibility, and, consequently, the spread of infection is described with the standard bilinear term.

Let us assume that not only the susceptibles are heterogeneous for some trait that influences the disease evolution, but also the infectives are heterogeneous, and consider the simplest possible SI model. Let $s(t, \omega_1)$ and $i(t, \omega_2)$ be the densities of the susceptibles and infectives, respectively, here we assume that the effects of the two traits on transmission are independent, i.e., $\beta(\omega_1, \omega_2) = \beta_1(\omega_1)\beta_2(\omega_2)$. The last assertion is a serious oversimplification, as it is natural to assume that infectivity might influence the susceptibility of contacted individual, but we need it to apply the general theory from Section 3.

The number of susceptibles with the trait value ω_1 infected by individuals with trait value ω_2 is given by

$$\beta_1(\omega_1)s(t, \omega_1)\beta_2(\omega_2)i(t, \omega_2),$$

and the total change in the infective class with trait value ω_2 is

$$\beta_2(\omega_2)i(t, \omega_2)\int_{\Omega_1}\beta_1(\omega_1)s(t, \omega_1)\,\mathrm{d}\omega_1;$$

an analogous expression applies to the change in the susceptible population. Combining the assumptions we obtain the following model:

$$\begin{aligned}
\frac{\partial}{\partial t}s(t, \omega_1) &= -\beta_1(\omega_1)s(t, \omega_1)\int_{\Omega_2}\beta_2(\omega_2)i(t, \omega_2)\,\mathrm{d}\omega_2 \\
&= -\beta_1(\omega_1)s(t, \omega_1)\bar{\beta}_2(t)I(t) \\
\frac{\partial}{\partial t}i(t, \omega_2) &= \beta_2(\omega_2)i(t, \omega_2)\int_{\Omega_1}\beta_1(\omega_1)s(t, \omega_1)\,\mathrm{d}\omega_1 \\
&= \beta_2(\omega_2)i(t, \omega_2)\bar{\beta}_1(t)S(t).
\end{aligned} \tag{23}$$

Model (23) is supplemented with initial conditions $s(0, \omega_1) = S_0 p_s(0, \omega_1), i(0, \omega_2) = I_0 p_i(0, \omega_2)$. In (23) it is assumed that if an individual having trait value ω_1 was infected by an individual with trait value ω_2 he or she becomes an infective with trait value ω_2 (see Section 2.2). The global dynamics of (23) is simple and is similar to the simplest homogeneous SI model.

According to Theorem 1 the system (23) can be reduced to a four-dimensional system of ODEs. Reasoning exactly as in the proof of Proposition 1 we obtain

Proposition 2. *The model (23) is equivalent to the model*

$$\frac{d}{dt}S(t) = -h_1(S)h_2(I),$$
$$\frac{d}{dt}I(t) = h_1(S)h_2(I),$$

where $h_i(x)$ is given by (16).

Combining Proposition 2 with (19) we get

Corollary 2. *The power relationship (22) of the incidence rate in an SI model for the case $q = 1 + 1/k_2 > 1, p = 1 + 1/k_1 > 1$ can be obtained as a consequence of the heterogeneous model (23) when the initial distributions of the susceptibles and the infectives are gamma-distributions with parameters k_1, v_1 and k_2, v_2, respectively.*

Consequently, it turns out that the power relationship (22), at least for the case $p, q \geqslant 1$, can be explained on the mechanistic basis by the inherent heterogeneities in susceptibility and infectivity of the populations. Its exact form is the consequence of the initial gamma-distributions, but we note that any of the transmission functions given in the previous subsection can be well approximated by (19) (see also Fig. 1).

5. The influence of population heterogeneity on the disease course

Here we mainly restrict our attention to the model (1)–(3) and study its global behavior. First we state almost obvious proposition:

Proposition 3. *Let* $S_1(t) = \int_\Omega s_1(t, \omega) \, d\omega$ *be the solution of* (1)–(3) *with the initial condition* $s_1(0, \omega) = S_0 p_1(0, \omega)$, *and* $S_2(t) = \int_\Omega s_2(t, \omega) \, d\omega$ *be the solution of* (1)–(3) *with the initial condition* $s_2(0, \omega) = S_0 p_2(0, \omega)$, *such that* $\bar{\beta}_1(0) = \bar{\beta}_2(0)$ *and* $\sigma_1^2(0) > \sigma_2^2(0)$, *where* $\bar{\beta}_i(0) = \int_\Omega \beta(\omega) p_i(0, \omega) \, d\omega$ *and*

$$\sigma_i^2(0) = \int_\Omega (\beta(\omega) - \bar{\beta}_i(0))^2 p_i(0, \omega) \, d\omega, \; i = 1, 2.$$

Then there exists $\varepsilon > 0$ *such that* $S_1(t) > S_2(t)$ *for all* $t \in (0, \varepsilon)$.

The gist of this proposition is very simple: the more heterogeneous the susceptible hosts, the less severe the initial disease progression under the model (1)–(3).

Proof. Differentiating the first equation in (14) and using (12) we get

$$S''(t) = I^2 S(\sigma^2(t) + \bar{\beta}(t)) - \bar{\beta}(t) I' S,$$

or, at the initial time moment, $S_1''(0) > S_2''(0)$. Since $S(t)$ is continuous the proposition follows. □

We remark that this proposition also holds for more general model (14). Moreover, we can replace Eq. (1) with equation

$$\frac{\partial}{\partial t} s(t, \omega) = -\beta(\omega) s(t, \omega) I(t) + s(t, \omega)[\dots]$$

where dots denote terms describing demography, migration or the lost of immunity by removed individuals, the only condition is that these terms cannot depend on $\beta(\omega)$. Even in this case Proposition 3 still holds. For the model (5), (6) the opposite proposition is true: the more heterogeneous the infective class in infectivity, the more severe the disease progression, which follows from the fact that $\bar{\beta}'(t) > 0$ in this case, and, consequently, $S_1''(0) < S_2''(0)$.

It is interesting to note that in the case of frequency-dependent transmission knowledge of only the initial variances of the parameter distributions does not allow inference on the short ran behavior [39].

One of the main characteristics of SIR models is the final size of the disease, which is often expressed in the number (or proportion) of susceptibles that are never infected. For the Kermack–McKendrick model

$$dS/dt = -\beta SI, \quad dI/dt = \beta SI - \gamma I, \quad dR/dt = \gamma I,$$

it is well known that the desired number, which we denote as $S(\infty)$, is given by the root of the equation

$$S(\infty) = S_0 \exp\left\{-\frac{\beta}{\gamma}(N - S(\infty))\right\} \tag{24}$$

on the interval $(0, S_0)$. Here N is the constant size of the population. It is easy to show that this root always exists.

Recall that $M(0, \lambda)$ is the mgf of the initial parameter distribution. For the model (1)–(3) the following theorem holds.

Theorem 2. *The size of the susceptible subpopulation that escapes infection within the framework of the model* (1)–(3) *is given by the solution of the equation*

$$S(\infty) = S_0 M\big(0, (S(\infty) - N)\gamma^{-1}\big) \tag{25}$$

satisfying condition $0 < S(\infty) < S_0$.

Proof. First we note that exactly as it was done in Proposition 1, we can reduce the system (1)–(3) to the system

$$dS(t)/dt = -h(S(t))I(t),$$
$$dI(t)/dt = h(S(t))I(t) - \gamma I(t), \tag{26}$$
$$dR(t)/dt = \gamma I(t).$$

Using (16) and dividing the first equation in (26) by the third one we obtain

$$\left.\frac{dM^{-1}(0, \xi)}{d\xi}\right|_{\xi = S/S_0} \frac{dS}{S_0} = -\gamma^{-1} dR.$$

Integrating from 0 to ∞ gives

$$\int_1^{S(\infty)/S_0} dM_0^{-1}(\xi) = -\frac{R(\infty)}{\gamma}.$$

Using the identities $R(\infty) = N - S(\infty)$ (since $I(\infty) = 0$) and $M^{-1}(0, 1) = 0$ we obtain

$$M_0^{-1}(S(\infty)/S_0) = -\frac{N - S(\infty)}{\gamma},$$

from which (25) follows. Due to the fact that $M(0, \lambda)$ is an increasing function, the solution of (25) satisfying $0 < S(\infty) < S_0$ is unique. □

Remark. If we consider undistributed parameter (formally, we can let $\beta(\omega) = \beta = \text{const}$, or, equivalently, $s(0, \omega) = S_0 \delta(\omega - \bar{\omega})$, $\beta(\bar{\omega}) = \text{const}$, where $\delta(\omega)$ is the delta-function), we obtain (24) from (25).

Arguing in the same spirit as it is done in [5] (e.g., p. 183), the problem of the epidemic invasion can be considered. We speak of the epidemic invasion here within the deterministic models: that is, an epidemic can invade the population if its basic reproductive number R_0[6] is more than unity, and cannot if $R_0 < 1$. Assuming that initially the whole population is susceptible (formally, for our model, $S(-\infty) = N, q(-\infty) = 0$), from (25) the equation for the fraction of susceptible population that does not get infected follows:

$$z = M(0, -N(1 - z)/\gamma). \tag{27}$$

Here $z = S(\infty)/N$. Eq. (27) always has the root $z = 1$. If the basic reproductive number, defined here as

$$R_0 = \frac{\bar{\beta}(0)N}{\gamma}, \tag{28}$$

satisfies the condition $R_0 > 1$, then there is another root of (27) in the interval $0 < z < 1$. This root gives the sought fraction. The proof of the existence of this root under the threshold condition $R_0 > 1$ is straightforward and can be conducted similar to the homogeneous case (e.g., [5]). This result can be illustrated by the case when the initial distribution is exponential with parameter v: the equation for z is quadratic: $R_0 z^2 - (1 + R_0)z + 1 = 0$, where $R_0 = N/(\gamma v)$. This equation has the roots 1 and $1/R_0$. If $R_0 > 1$ then the fraction of susceptible population that escapes the disease is $1/R_0$.

Comparing the results obtained for the heterogeneous SIR model (1)–(3) with the well known results for the simple homogeneous SIR model, we can conclude that the questions whether the disease can or cannot invade the heterogeneous population with distributed susceptibility can be studied in the framework of the homogeneous model. The reason is that the considered population heterogeneity in susceptibility does not impact the basic reproductive number (28) (this holds, obviously, if we identify the mean value of $\beta(\omega)$ over the population of susceptibles at the initial time moment with the usual constant β in the homogeneous model). Here we note that this result is pertinent only to our particular
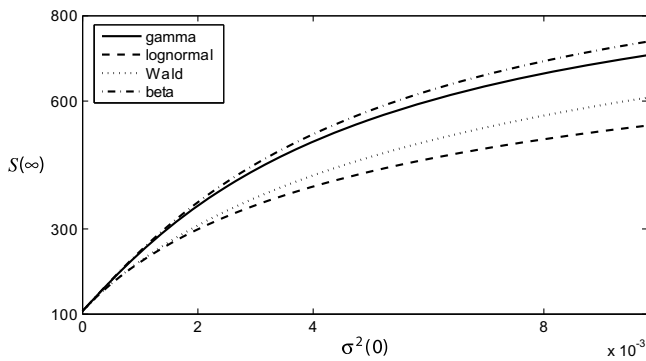
Fig. 2. The size of the susceptible population that never gets infected, $S(\infty)$ versus the initial variance of the parameter distribution $\sigma^2(0)$ for four different initial distributions, $\bar{\beta}(0)$ is the same in all cases. The parameters are $S(0) = 999, I(0) = 1$, $\gamma = 20, \bar{\beta}(0) = 0.05$.

model (1)–(3), since it is well known that in general population heterogeneity impacts R_0 (e.g., [6,18]).

From the other hand, the heterogeneity of the population has direct impact on the final size of the disease since Eq. (27) depends on the initial distribution in contrast to the homogeneous analogue $z = \exp\{-R_0(1 - z)\}$ (the last formula is valid under variety of different conditions [30]).

In Fig. 2 the final size of the susceptible population versus the initial variance of the parameter distribution is shown for four different initial distributions that have the same means at $t = 0$. From Fig. 2 it can be seen that the more heterogeneous the population of susceptibles, the less severe the disease not only in the short run (Proposition 3) but also globally (see [31] for the same result for frequency-dependent transmission). At the same time, it is worth emphasizing that the conditions $\bar{\beta}_1(0) = \bar{\beta}_2(0), \sigma_1^2(0) > \sigma_2^2(0)$ for two different initial distributions do not imply that $z_1 > z_2$, where z_i are the solutions of (27). A counterexample can be easily found (e.g., taking the gamma-distribution with parameters $k = 2, v = 4$ and the uniform distribution on the interval [0,1], $N/\gamma = 20$, we find that $z_1 = 0.093 < z_2 = 0.112$ whereas $\sigma_1(0) = 1/8 > \sigma_2(0) = 1/12$).

If we compare two distributions of the same family then sometimes it is possible to prove rigorously that, on the assumption of equal initial means and different second moments, more heterogeneous population (i.e., the one that has larger initial variance) experiences less severe disease. This is true, e.g., for two gamma-distributions:

Proposition 4. *Assume that we have model (1)–(3) with two initial gamma-distributions with parameters k_1, v_1 and k_2, v_2 such that $\bar{\beta}_1(0) = \bar{\beta}_2(0)$ and $\sigma_1^2(0) > \sigma_2^2(0)$. Assume that $R_0 > 1$. Then solutions of (27) that belong to (0,1) are always satisfy $z_1 > z_2$.*

Proof. We need to prove that

$$\frac{(k_2 + R_0(1 - z))^{k_2}}{(k_1 + R_0(1 - z))^{k_1}} \frac{k_1^{k_1}}{k_2^{k_2}} > 1$$

for any $z \in (0, 1)$. This follows from the fact that $f(r) = (k_2 + r)^{k_2} / (k_1 + r)^{k_1}$ is monotonically increasing function for $r > 0$, i.e., $f'(r) > 0$. □

Remark. The last proposition can be extended to other initial distributions, e.g., it holds for the Wald distribution, for the uniform distribution and some others. However, for any distribution, for which more than two first moments are needed to specify all the parameters, an analogous proposition is no longer true.

6. Discussion and conclusions

We have presented several results concerning the course of a disease in a closed heterogeneous population, where heterogeneity is mediated by invariable traits whose distributions are determined by population structure at every time moment. One of the purposes of the present text is to introduce in the area of epidemiological modeling the general techniques of the theory of heterogeneous populations [21,23] which allow reduction of the initial infinite-dimensional dynamical system to an ODE system of a low dimension.

A usual strategy in the literature when analyzing systems similar to (1)–(3) is to consider an infinite-dimensional system of ODEs where the variables can be moments or cumulants of the corresponding distributions (e.g., [11,33,39]) and then analyze this system, or consider one of many possible moment-closure strategies to extract valuable information. Having the general theory outlined in Section 3 and the known results from the literature, we can critically discuss the latter and be prepared to possible pitfalls.

For example, in [11] the equation for the final epidemic size was obtained by using two different strategies: first, an initial gamma-distribution was assumed and analytical treatment was applied, and second, using the procedure suggested in [8], an approximation method was used in which the infinite-dimensional system of ODEs was replaced with two equations under the assumption that the coefficient of variation is constant. It is not surprising that Dwyer et al. obtained identical results because the gamma-distribution, according to the theory of heterogeneous populations, is the only continuous distribution which does not change its shape during the system evolution, and keeps the coefficient of variation constant (see Appendix). Therefore, the conclusion that '…the assumption of gamma-distributed susceptibility is not strictly necessary to derive equation [for the final epidemic size]' is not valid in many situations. Another initial distribution, how it is shown by (25), can yield another equation for the final epidemic size and, consequently, can produce significant discrepancy with the moment-closure approximation suggested by Dushoff [8].

The well known fact from the theory of heterogeneous populations that to model the system dynamics for a substantial time period we need to know the *exact* initial distribution implies that any results obtained for an epidemic in a heterogeneous population on the ground of knowledge of only several first moments of the initial distribution have to taken with extreme care. One, two, or more first moments of the initial distribution can be insufficient or even misleading. We also note that the short run behavior can be predicted when we have information only on several first moments (Section 5).

Another initial distribution which was used in the literature is the normal distribution [33] (see also Appendix, Eq. (34)). For the normal distribution we have that $\kappa_i = 0, i \geq 3$, where κ_i is the ith cumulant. Combining the last property with the fact that the initial normal distribution remains normal within the framework of heterogeneous models we obtain that $\kappa_i(t) = 0, i \geq 3$, holds for any t (see Appendix). This was used in [33] to obtain an explicit solution to the equation

$$\frac{\partial}{\partial t} n(t, r) = K_g n(t, r) - r n(t, r).$$

Here $n(t, r)$ is the number of cells at time moment t, which are killed by antimicrobial agents with the kill rate r, and K_g is a constant (notations are changed from the original). First we note that, using Theorem 1, we obtain explicit solution of this equation for an arbitrary initial distribution of the kill rate:

$$N(t) = N_0 M(0, -t) \exp\{K_g t\},$$

where $N(t) = \int_R n(t, r)\,dr$. Second, the results from Sections 4 and 5 cannot be applied to the normal distribution because this distribution is defined from $-\infty$ to ∞ and thus the corresponding mgf does not have a unique inverse. Which is more important, however, the total system dynamics can be influenced by these negative kill rates even if they occur with vanishingly small probability (we note that this issue is discussed in [33]). An example of such influence can be found in [24] where infinitely large growth rates occurring with small probabilities drive the population to explosion. Therefore, any approximations based on an initial normal distribution in a situation where parameter can take only non-negative values should be taken with care.

Summarizing the main results we can assert that the theory of heterogeneous populations can be successfully applied to many different mathematical models in epidemiology. Examples are given in Section 3. In many simple cases the original model can be reduced to a model described by ODEs, which simplifies the analysis. The law of mass action for a distributed susceptibility model implies a nonlinear incidence function in a homogeneous model. Moreover, one of the well known transmission functions, power relationship, follows in exact form from the initial gamma-distribution, at least in the case when exponents exceed one (Section 4). Therefore, a mechanistic derivation has been given to the transmission power function, which was shown previously to approximate real data with high accuracy. The short term behavior of the models considered can be approximately described knowing only two first moments of the initial distribution, whereas the long-term behavior depends on the exact initial distribution and can vary significantly (Section 5) even for the distributions whose several first moments are identical (Fig. 2).

It is a tempting challenge to include various demography processes to the analyzed models. The main obstacle is the need to specify the function $\phi(\omega, \omega')$ similar to the one used in (5). The delta-function yields the models that can be analyzed using the general approach from Section 3 but it is usually difficult to interpret the underlying assumptions. These problems are the subject of ongoing research.

Acknowledgments

Appendix A

In Appendix we collect the definitions of the distributions used throughout the main text. The definitions are taken from [19,20]. In addition to that we list some facts concerning evolution of these distributions if they are used as the initial distributions for the models studies in the text. Everywhere below it is assumed that $\beta(\omega) = \omega$. Inasmuch as we are interested in characteristics of distributions depending on time, the following formula is very useful (see [23]):

$$M(t, \lambda) = \frac{M(0, \lambda + q(t))}{M(0, q(t))},\qquad(29)$$

where $q(t)$ in the solution of the corresponding auxiliary differential equation (see Theorem 1), $M(t, \lambda)$ is the mgf of the parameter distribution at time t. Eq. (29) shows that the mgf at any time instant can be expressed using the initial mgf.

A.1. Gamma-distribution

The pdf of gamma-distribution with parameters k and v is given by

$$p(0, \omega) = \frac{v^k}{\Gamma(k)}\omega^{k-1}e^{-v\omega},\qquad \omega \geqslant 0,\quad k > 0,\quad v > 0.\qquad(30)$$

The mgf of gamma-distribution is $M(0, \lambda) = (1 - \lambda/v)^{-k}$.

It follows from (29) that for $t > 0$ the distribution does not change its form, i.e., it is gamma-distribution with parameters k and $s - q(t)$. The mean and variance of the distribution are given by

$$\bar{\beta}(t) = \frac{k}{v - q(t)},\quad \sigma^2(t) = \frac{k}{(v - q(t))^2}.$$

Note that at any time moment the coefficient of variation is constant: $cv = \sigma(t)/\bar{\beta}(t) = 1/\sqrt{k}$. Actually, gamma-distribution is the only continuous distribution whose coefficient of variation remains constant with time within the outlined in Section 3 theory of heterogeneous models.

A.2. Wald distribution

The pdf of inverse gaussian (Wald) distribution with parameters μ and v is

$$p(0, \omega) = \left[\frac{v}{2\pi\omega^3}\right]^{1/2}\exp\left\{-\frac{v}{2\mu^2\omega}(\omega - \mu)^2\right\},$$
$$\omega > 0,\quad \mu, v > 0.\qquad(31)$$

The mgf of Wald distribution is given by

$$M(0, \lambda) = \exp\left\{\frac{v}{\mu}\left(1 - \left[1 - \frac{2\mu^2\lambda}{v}\right]^{1/2}\right)\right\}.$$

Again the distribution remains the Wald distribution with parameters

$$\mu\left[1 - \frac{2\mu^2 q(t)}{v}\right]^{-\frac{1}{2}}, v,$$

and temporal characteristics of the distribution are

$$\bar{\beta}(t) = \mu\left[1 - \frac{2\mu^2 q(t)}{v}\right]^{-\frac{1}{2}},\quad \sigma^2(t) = \frac{\bar{\beta}^3(t)}{v},\quad cv^2 = \frac{\bar{\beta}(t)}{v}.$$

A.3. Beta-distribution

Sometimes it is useful to study evolution of distribution with compact support. A good candidate in this case is the family of beta-distributions with pdf

$$p(0, \omega) = \frac{1}{B(r_1, r_2)}\frac{(\omega - a)^{r_1-1}(b - \omega)^{r_2-1}}{(b - a)^{r_1+r_2-1}},$$
$$a \leqslant \omega \leqslant b,\quad r_1 > 0,\quad r_2 > 0,\qquad(32)$$

where $B(r_1, r_2)$ is the beta-function.

The initial mean and variance are

$$\bar{\beta}(0) = \frac{r_1}{r_1 + r_2},\quad \sigma^2(0) = \frac{r_1 r_2}{(r_1 + r_2)^2(r_1 + r_2 + 1)}.$$

Unfortunately in the case of beta-distribution it is impossible to write down the mgf, and, correspondingly, the temporal characteristics of the distribution. In a special case $r_1 = r_2 = 1$ we have a uniform distribution on $[a, b]$. Eq. (29) shows that in this case for $t > 0$ the distribution is no longer uniform but turns into truncated exponential distribution.

In the text we used beta-distribution on $[0, 1]$.

A.4. Log-normal distribution

The pdf of the log-normal distribution defined on the non-negative half-axis $\omega \geqslant 0$ with parameters μ and v is

$$p(0,\omega) = \left[\omega\sqrt{2\pi v}\right]^{-1} \exp\left\{-\frac{1}{2}\frac{(\ln\omega - \mu)^2}{v^2}\right\},$$
$$\omega \geqslant 0, \quad \mu > 0, \quad v > 0. \tag{33}$$

The initial characteristics are

$$\beta(0) = \exp\{\mu + v^2/2\}, \quad \sigma^2(0) = \exp\{2\mu + v^2\}(\exp\{v^2\} - 1).$$

As in the case of beta-distribution we cannot present explicit formulas for the mgf and other characteristics. We note that this distribution can be used only if $q(t) < 0$ for $t > 0$, otherwise the integral in the mgf diverges. This is the case, e.g., for the model (1)–(3), but not the case for (6), (7), for which the log-normal distribution cannot be used (see also discussion in Section 3).

A.5. Normal distribution

The pdf is

$$p(0,\omega) = \left[\sqrt{2\pi}\sigma\right]^{-1} \exp\left\{-\frac{(\omega - \mu)^2}{2\sigma^2}\right\}, \quad \sigma > 0. \tag{34}$$

The mgf is $M(0,\lambda) = \exp\{\frac{1}{2}\lambda(2\mu + \lambda\sigma^2)\}$. The temporal characteristics are

$$\bar{\beta}(t) = \mu + 2q(t)\sigma^2, \quad \sigma^2(t) = \sigma^2.$$

We present here the normal distribution because it was used in a number of papers. Using the normal distribution for the considered models is not always appropriate because the parameter takes the values from $-\infty$ to ∞ whereas usually all the parameters have definite signs. To overcome this difficulty a truncated normal distribution can be used, albeit without simple analytical expressions.

A.6. Poisson distribution

In the same spirit discrete distributions can be managed. Consider, for example, the Poisson distribution with parameter θ, i.e,

$$p(0,\omega) = \frac{\theta^\omega}{\omega!}e^\theta, \quad \theta > 0, \quad \omega = 0, 1, \ldots \tag{35}$$

then $M(0,\lambda) = \exp\{\theta(e^\lambda - 1)\}$, and

$$\bar{\beta}(t) = \sigma(t) = \theta\exp\{q(t)\}, \quad cv = 1.$$

Other possible initial distribution can be considered in a similar vein.

References

[1] A.S. Ackleh, Estimation of rate distributions in generalized Kolmogorov community models, Nonlinear Anal. 33 (7) (1998) 729.

[2] A.S. Ackleh, D.F. Marshall, H.E. Heatherly, Extinction in a generalized Lotka–Volterra predator–prey model, J. Appl. Math. Stochastic Anal. 13 (3) (2000) 287.

[3] A.S. Ackleh, D.F. Marshall, H.E. Heatherly, B.G. Fitzpatrick, Survival of the fittest in a generalized logistic model, Math. Models Methods Appl. Sci. 9 (9) (1999) 1379.

[4] R.M. Anderson, R.M.C. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University, New York, 1991.

[5] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, John Wiley, 2000.

[6] O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations, J. Math. Biol. 28 (4) (1990) 365.

[7] O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, The legacy of Kermack and McKendrick, in: D. Mollison (Ed.), Epidemic Models: Their Structure and Relation to Data, Cambridge University, 1993, p. 95.

[8] J. Dushoff, Host heterogeneity and disease endemicity: a moment-based approach, Theor. Popul. Biol. 56 (3) (1999) 325.

[9] J. Dushoff, S. Levin, The effects of population heterogeneity on disease invasion, Math. Biosci. 128 (1–2) (1995) 25.

[10] G. Dwyer, J. Dushoff, J.S. Elkinton, J.P. Burand, S.A. Levin, Host heterogeneity in susceptibility: lessons from an insect virus, in: U. Diekmann, J.A.J. Metz, M. Sabelis, K. Sigmund (Eds.), Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management, Cambridge University, 2002, p. 74.

[11] G. Dwyer, J. Dushoff, J.S. Elkinton, S.A. Levin, Pathogen-driven outbreaks in forest defoliators revisited: building models from experimental data, Am. Nat. 156 (2) (2000) 105.

[12] G. Dwyer, J.S. Elkinton, J.P. Buonaccorsi, Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model, Am. Nat. 150 (6) (1997) 685.

[13] V.J. Haas, A. Caliri, M.A.A. da Silva, Temporal duration and event size distribution at the epidemic threshold, J. Biol. Phys. 25 (4) (1999) 309.

[14] J.A.P. Heesterbeek, The law of mass-action in epidemiology: a historical perspective, in: B.E. Beisner (Ed.), Ecological Paradigms Lost: Routes of Theory Change, Academic, 2005, p. 81.

[15] M.E. Hochberg, Non-linear transmission rates and the dynamics of infectious disease, J. Theor. Biol. 153 (3) (1991) 301.

[16] S. Hsu Schmitz, Effects of genetic heterogeneity on HIV transmission in homosexual populations, in: C. Castillo-Chavez (Ed.), Mathematical Approaches for Emerging and Reemerging Infectious Diseases: Models, Methods, and Theory, vol. 126, IMA, 2002, p. 245.

[17] J.M. Hyman, J. Li, Differential susceptibility epidemic models, J. Math. Biol. 50 (6) (2005) 626.

[18] J.A. Jacquez, C.P. Simon, J. Koopman, Core groups and the R_0s for subgroups in heterogeneous SIS and SI models, in: D. Mollison (Ed.), Epidemic Models: Their Structure and Relation to Data, Cambridge University, 1995, p. 279.

[19] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, vol. 1, John Wiley, New York, 1994.

[20] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, vol. 2, John Wiley, New York, 1995.

[21] G.P. Karev, Heterogeneity effects in population dynamics, Doklady Math. 62 (1) (2000) 141.

[22] G.P. Karev, Inhomogeneous models of tree stand self-thinning, Ecol. Model. 160 (1–2) (2003) 23.

[23] G.P. Karev, Dynamics of heterogeneous populations and communities and evolution of distributions, Discrete Contin. Dyn. Syst. (Suppl.) (2005) 487.

[24] G.P. Karev, Dynamics of inhomogeneous populations and global demography models, J. Biol. Syst. 13 (1) (2005) 83.

[25] G.P. Karev, A.S. Novozhilov, E.V. Koonin, Mathematical modeling of tumor therapy with oncolytic viruses: effects of parametric heterogeneity on cell dynamics, Biol. Direct 1 (30) (2006) 19.

[26] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. Ser. A 115 (772) (1927) 700.

[27] R.J. Knell, M. Begon, D.J. Thompson, Transmission dynamics of *Bacillus thuringiensis* infecting *Plodia interpunctella*: a test of the mass action assumption with an insect pathogen, Proc. R. Soc. Lond. Ser. B Biol. Sci. 263 (1366) (1996) 75.

[28] W.M. Liu, H.W. Hethcote, S.A. Levin, Dynamical behavior of epidemiological models with nonlinear incidence rates, J. Math. Biol. 25 (4) (1987) 359.

[29] W.M. Liu, S.A. Levin, Y. Iwasa, Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models, J. Math. Biol. 23 (2) (1986) 187.

[30] J. Ma, D.J.D. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease, Bull. Math. Biol. 68 (3) (2006) 679.

[31] R.M. May, R.M. Anderson, The transmission dynamics of human immunodeficiency virus, Proc. R. Soc. Lond. Ser. B Biol. Sci. 321 (1207) (1988) 565.

[32] H. McCallum, N. Barlow, J. Hone, How should pathogen transmission be modelled?, Trends Ecol Evol. 16 (6) (2001) 295.

[33] M. Nikolaou, V.H. Tam, A new modeling approach to the effect of antimicrobial agents on heterogeneous microbial populations, J. Math. Biol. 52 (2) (2006) 154.

[34] A.S. Novozhilov, Analysis of a generalized population predator–prey model with a parameter distributed normally over the individuals in the predator population, J. Comput. Syst. Sci. Int. 43 (3) (2004) 378.

[35] R. Ross, The Prevention of Malaria, J. Murray, London, 1910.

[36] M. Roy, M. Pascual, On representing network heterogeneities in the incidence rate of simple epidemic models, Ecol. Complexity 3 (1) (2006) 80.

[37] N.C. Severo, Generalizations of some stochastic epidemic models, Math. Biosci. 4 (1969) 395.

[38] P.D. Stroud, S.J. Sydoriak, J.M. Riese, J.P. Smith, S.M. Mniszewski, P.R. Romero, Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing, Math. Biosci. 203 (2) (2006) 301.

[39] V.M. Veliov, On the effect of population heterogeneity on dynamics of epidemic diseases, J. Math. Biol. 51 (2) (2005) 123.