A Modified Epidemic Chain Binomial Model (MECBM) and Its 2,3-Introductory Probabilities

Dilip C. Nath[1], Kishore K. Das[2], Tandrima Chakraborty[2]

[1]Administration, Assam University, Silchar, India
[2]Department of Statistics, Gauhati University, Guwahati, India
Email: Dilipc.nath@gmail.com, daskkishore@gmail.com, tandrimac@yahoo.com

Abstract

Through the ages, human life and happiness are affected much by wide spread diseases, an unpredictable phenomenon. And infectious disease spread is one of the most unsolvable problems. How infection evolves, how it spreads from person to person and mainly which pattern it follows are some questions which are always unanswerable. Out of many techniques, discrete epidemic models like the chain binomial model are ones which are applied to describe the physical phenomena of spreading infectious diseases in a household. In this paper, an attempt has been made to develop a modified epidemic chain model by assuming a beta distribution of third kind for the probability of being infected by contact with a given infection from the same household with closed population. This paper emphasized mainly on developing the probabilities of all possible epidemic chains with two introductory cases for three-, four- and five-member household and three introductory cases for four- and five-member household. The key phenomenon towards developing this paper was to provide an extension of the modified chain binomial model and its possible probabilities.

Keywords

1. Introduction

In a time when it has become unfathomable to breathe fresh air, drink clean water and live in an era, technological and industrial prosperity with all imaginable boons of life also brings with it the bane of a degraded environment and various ensuing health hazards. In the words of the Secretary General of UN, Mr. Ban Ki-Moon, "*We must connect the dots between climate change, water scarcity, energy shortages, global health, food security and women's empowerment. Solu-*

tions to one problem must be solutions for all." In present days, apart from terrorism, increasing health issues, variants of fatal diseases are probably the greatest threat to the human race, irrespective of their nationality, ethnicity, religion or language. A country with a sick mass of people as its citizens can never imagine prospering, whatsoever might be its wealth. In that direction, each nation spends a huge chunk of its money towards the research and development in the health sector. In such a scenario, an investigation into the infection and propagation of infectious disease with a proper statistical methodology is worthy.

Some data from the past and some data on which Bailey (1957) [1] focused in his paper have ultimately shown us the exact scenario or status of infectious diseases in the world. For example, in 14th century, Black death killed almost 50 million (60 percent) of Europe's entire population; the disastrous mortal disease known as Black Death spread across Europe in the years 1346-1353. In 1520, the Aztecs lost about half of their population, 3.5 million from the smallpox epidemic and the downfall of their empire in 1521 was more due to smallpox than that of any other causes. It was estimated that Russia suffered about 25 million cases of epidemic typhus in the years 1918 to 1921 with a death rate of approximate 10 percent. The influenza pandemic of 1918-1919 killed more people than the Great War, known today as World War I, at somewhere between 20 and 40 million people.

Infectious diseases continue to be a serious health issue throughout the globe. Progress in controlling, eliminating or eradicating infectious diseases is a key part of the international health agenda. In fact, changing lifestyles, patterns of behaviour and several such complex factors has led to the emergence and spread of disease in India. And, people in India have been infected with infectious diseases like SARS, dengue, chickungunya, malaria and bird flu in recent times. And day by day, education and information become the key factors in the management of infectious diseases. The distinction between the disease and the epidemic is the epidemic of the disease and the epidemic of anxiety. The later was significantly more infectious. In infectious disease the uncertainties can broadly be divided into two completely different but related issues.

1) The Disease Process: What will happen to me if I get sick?

2) The Epidemic Process: How likely am I to get sick now?

Will that change in the future?

This section describes the foundations and development of epidemic modelling, starting with an early model for smallpox. Daniel Bernoulli (1766) [2] was one of the first mathematicians to attempt to model the effects of disease in a population. He used a deterministic model to show that inoculation with a mild form of the smallpox virus would reduce the death rate of the population of France. An early reference to the non-linearity of epidemic models is made in a paper by Hamer (1906) [3]. Hamer postulated that the probability of an infection in the next period of time (in a discrete time model) was proportional to the number of infectious individuals multiplied by the number of susceptible individuals. This idea is called the mass action principle and has been used in many

areas of science, in particular to determine the rate of chemical reactions, in work as early as Boyle's *c.* 1674 (Daley and Gani, 1999) [4]. Kermack and McKendrick (1927) [5] incorporated this idea into the Deterministic General Epidemic Model.

McKendrick (1927) [5] suggested one of the first epidemic models to incorporate the randomness observed in real life outbreaks. This model is a stochastic continuous time version of the Deterministic General Epidemic Model. Heasman and Reid (1961) [6], for example, demonstrated that the Reed-Frost chain binomial model can provide an adequate fit to data on outbreaks of the common cold in households of size five. By comparing observed with expected frequencies for the total number of cases, they also demonstrated that the stochastic version of the Kermack-McKendrick epidemic model (Bailey, 1975) [7] may provide an even better fit. Ridler-Rowe (1967) [8] obtained some limit results for the duration time for an epidemic process with immigration of the susceptible and infective. And this work addresses itself to the question of obtaining the mean duration time for the general epidemic process.

Another early discrete-time model is the chain Binomial model of Reed and Frost (Bailey, 1975) [7] in which the number of infectives to appear in the next time unit follows a binomial distribution, with the probability of infection dependent on the number of infectives in the current time unit. It was not until Bartlett (1949) [9] studied McKendrick's model, that stochastic models in continuous time were examined more extensively.

Afterwards, in connection with applications in reliability theory and communication nets, Keilson (1974a) [10] considered stationary processes which could be modelled as Markov chains in continuous time. His underlying state space was partitioned into good and bad states. Among other results, Keilson derived mean sojourn times on the good set until the process entered the bad set. For the general epidemic process, Keilson identified the good set with the continuance of the epidemic and the bad set with the completion of the epidemic. And using these ideas only, he derived the mean duration time for the general epidemic process.

A general epidemic process is said to be completed whenever the number of susceptibles or the number of infectives reduces to zero. Billard (1977) [11] derived the mean duration time for this process to be completed. This work was an extension of Bailey's work where the derivation of the probability distribution of the duration time of the epidemic process requires complicated and apparently intractable mathematics. Using generating functions, Bailey (1975) [7] has obtained expressions for the mean duration time for the simple epidemic process.

Classical epidemic models have invariably proved to be mathematically intractable. By considering the distribution of the infectives in a simple epidemic process as a convolution of exponential waiting times, the solution to the classical model is obtained easily by Billard *et al.* (1978) [12] by giving more insight into the underlying structures. Further the idea has been extended to other simple epidemic models.

Again Becker (1980) [13] had developed an epidemic chain model by assuming a beta distribution for the probability of being infected by contact with a given infective from same household, and also formulated a multi-parameter chain binomial model to describe outbreaks of an infectious disease in household.

Again in the later part, Jacquez (1987) [14] had analyzed the derivations of the Reed-Frost model in terms of the assumptions about the probabilistic process used and in terms of internal consistency. Demiris (2004) [15] had developed the statistical methodology for the analysis of stochastic SIR (Susceptible → Infective → Removed) epidemic models by adopting the Bayesian paradigm and also developed a suitable tailored Markov chain Monte Carlo (MCMC) algorithms. And the focus was mainly given on the methods that were easy to generalize in order to accommodate epidemic models with complex population structures. At first, Barbour and Utev (2004) [16] had done the refining of two well known approximations to the Reed-Frost epidemic process. The first was the branching process approximation in the early stages of the epidemic; and extended its range of validity, and sharpen the estimates of the error incurred. And the second is the normal approximation to the distribution of the final size of a large epidemic, which they complement with a detailed local limit approximation, where he found the latter, in particular, to be relevant, if the approximations are to be used for statistical inference.

Since then, research has been directed towards the study of a wide variety of models, and their statistical analysis. This paper aims to develop a solution to the problem of epidemic processes of infectious diseases. Specifically, the discrete time models are considered for the study. The study is entirely based on the spread of infectious diseases.

The main objective of this paper is to give more emphasize on the theoretical development of a new discrete-time model named as modified epidemic chain model and its extensions. The model was developed from the Becker's epidemic chain model (1980) [13], by assuming a beta distribution of third kind for the probability of being infected by contact with a given infective from the same household in our previous paper by Nath *et al.* (2015) [17]. Also, the paper aims in developing the general formula for the number of epidemic chains possible for the household of size three, four and five along with their chain probabilities for households with two and three introductory cases.

2. Theory of Epidemics

One of the most important applications of Stochastic Processes in the area of Biology and Medicine has been to the Mathematical theory of Epidemics. This is of interest, not only because of the biological and epidemiological implications, but also because a more complicated type of process is involved than those considered therein.

Bailey (1975) [7] discussed that, in general an epidemic process can be characterized as a time-dependent process of transition by the members of a popu-

lation, where the state transitions are caused by exposure to some influence called infectious material. The members of the population can belong to one of three basic states at a given point in time: 1) *Infective*, those members of the population who are host to the infectious material; 2) *Susceptible*, those members of the population who can become infectives given effective contact with infectious material; 3) *Removal*, those members of the population who have been removed from circulation for one of a variety of reasons such as death, immunity, hospitalization, etc.

The epidemic processes is further classified into continuous-time and discrete-time processes, where the continuous-time processes involved models in which the transition probabilities were at most linear functions of the population size. On the other hand, the discrete-time processes are those, for which the transition probabilities are usually non-linear functions of the population size. The two processes are elaborated as below in Section A and B:

1) Continuous-time epidemic models (Bailey, 1975) [7];

The simplest continuous-time treatments are covered below in (a), (b) and (c)

a) *Simple epidemics*: The simple epidemics is the most simplest possible kind of continuous-time epidemic model, in which we have the susceptibles in a group liable to catch current infection, but in which there is no removal from circulation by death, recovery, or isolation. Such a model might well be approximately true for certain mild infections of the upper respiratory tract, where there is a comparatively long interval of time between the infection of the individual and his actual removal from circulation. The bulk of the epidemic would then take place before anyone was removed.

There occurs two versions for such situations one is a deterministic case and the other one is the stochastic case. It is more convenient to examine the approximate deterministic version of such a situation than that of the stochastic version because for a group with sufficiently large numbers, we can sometime use a deterministic component as a first approximation than that of a full stochastic model and also because it is instructive to see how the stochastic mean in small groups differs from the corresponding deterministic value.

b) *General epidemics*: It has been explained by Bailey that, there exists considerable difficulties in theoretical handling of even a simple epidemic in which we have only infection and no removal. But in case of general epidemics, the more complicated general situation in which both of these latter possibilities are realized were looked upon thoroughly by the author. This theory somehow helped to find a way out for handling the distribution of the total number of cases of disease that may occur. But little is known about the exact form of the epidemic curve, or the distribution of the duration time.

c) *Recurrent Epidemics*: It is a characteristics feature of many infectious diseases that each outbreak has a kind of epidemic behavior investigated in the last part, but in addition these outbreaks tend to recur with certain regularity. The disease is then, in a sense, endemic as well as epidemic. And such situations are of considerable interest to explore mathematically, and there exists a fundamen-

tal distinction between the properties of deterministic and stochastic models. And, in general, the stochastic formulation of the model for recurrent epidemics leads to a permanent succession of undamped outbreaks of disease, although not exhibiting a strict sequence of oscillations.

2) Discrete-time epidemic model;

This is the main area of the paper. It has many such models under this, but the most important discrete—time epidemic model was developed by Bailey in 1975 which is explained in Section 2.1. And the further extension of the chain binomial model, *i.e.*, epidemic chain binomial model developed by Becker in 1980 is explained in Section 2.2

To start with, we assumed that, we have a group of *susceptible* individuals all mixing homogeneously together. One or more, of this group then contracts a certain disease which may in due course be passed on to the other susceptibles. In general, we assume that after the receipt of the infectious materials, there is *latent period* during which the disease develops purely internally within the infected person. The latent period is followed by an infectious period, during which the infected person, or *infective* as he/she is then called, is able to discharge infectious matter in some way and possibly communicate the disease to other susceptible *i.e.*, the time during which the disease may be transmitted to other members of the population. This time period is contracted to a single point. The infected person may spread the disease upon "adequate contact" to susceptibles in the population. This adequate contact is the probability of contact at any time between an infective and a susceptible sufficient to transmit the infection. Denote this parameter as "p" where $0 < p - l - q < l$, and q is the probability of no contact with the infection.

Sooner or later actual symptoms appear in the infective and he/she is removed (isolated) from circulation amongst the susceptibles until he/she either recovers or dies *i.e.*, removed from the rest of the population until recovery. This removal brings the infectious period effectively to an end (at least so far as the possibility of spreading the disease is concerned). The time interval between the receipt of infection and the appearance of symptoms is of course the *incubation period*. At each time step, a new generation or set of cases following a binomial distribution depending on the parameter "p" is presented. The epidemic continues until at some stage there are no new cases generated. An epidemic is defined as the transient outbreak of a disease which is terminated when there are no new infectives.

As viewed by Bailey (1975) [17] in his book that, there exist certain special cases of the above general situation. In the simplest continuous-time models, the latent period is assumed to be zero, so that the infected individual becomes infectious to others immediately after the receipt of the infection. On the other hand, in the simplest discrete-time models like chain-binomial model, the latent period is considered to be constant, and an infectious period is assumed to be short.

Cairoli (1988) [18] in his work stated that, the mathematical formulation of discrete time epidemic models flows from attempts by several investigators to

present models which realistically describe the progress of a disease through a population. The usual starting point in model building is the set of assumptions about those factors which control the spread of a disease. These assumptions should create a model which describes actual disease patterns. The epidemic model is then useful as a predictive tool for epidemiologists.

2.1. Chain-Binomial Models

The chain binomial models (Bailey, 1975) [7] have met with reasonable accomplishment, when fitted to data on communicable diseases for households, for example diseases like common cold or influenza. Also, Heasman and Reid (1961) [6] have demonstrated that, the Reed-Frost chain binomial model can provide an adequate fit to data on outbreaks of the common cold in households of size five. And, by comparing the observed frequencies with the expected frequencies for the total number of cases, they also demonstrated that, the stochastic version of the Kermack-McKendrick epidemic model (Bailey, 1975) [7] may provide an even better fit. In the later stage, a detailed comparison of the fits provided by these two models was attempted by Becker (1980) [13] by formulating an epidemic chain model, that was developed by assuming a beta distribution of first kind, for the probability of being infected by contact with a given infective from the same household. This model includes, as a particular case, the epidemic chain model corresponding to the stochastic version of the Kermack-McKendrick epidemic model (Bailey, 1975) [7] and, as a limiting case, the Reed-Frost chain binomial model. The advantages of the more general model were also illustrated with an application to household data for the common cold. Also the assumptions made were similar in many ways to those used by Ludwig (1975) [19] in his derivations of the final size distributions for epidemics with arbitrary time-dependent infectiousness.

The model should be relatively simple and mathematically accurate in describing essential features of the epidemic. Chain binomial models satisfy both these criteria. These models have been useful in describing viral diseases such as measles, chicken pox, influenza, and the common cold. Modeling the spread of these diseases among individuals in a population is a complex task. It is necessary to make several mathematical and biological assumptions about the factors which control the disease process. Mathematically, the population under consideration is assumed to be closed and homogeneously mixed.

In this paper, we concentrated mainly upon the discrete-time type of epidemic model. In certain circumstances, we may prefer to employ a discrete-time model, and represent the epidemic process by some suitably defined Markov Chain. Such models have been used quite successfully in the statistical fitting of certain epidemic theories to data relating to small groups such as families in statistical theory. However, so far as the analysis of epidemic processes in large groups is concerned, the discrete-time models are rather difficult, and it is easy to rely on the insights provided by continuous-time models for understanding the behavior of epidemics in reasonably large groups.

It is perhaps worth taking a quick look at the way in which the discrete-time models can be constructed, as future developments may enable them to be used as a basis for the investigation of the corresponding stochastic processes. The basic idea is that, the latent period is fixed, which may be used as a unit of time, and an infectious period is contracted to a single point. As the population consists of two classes of individuals like the *infectives*, and *susceptibles*. The models assume that all individuals have equal susceptibility, capability to transmit the disease, and the ability to be removed from observation when the transmitting period is over.

2.2. Epidemic Chain Binomial Model

Becker (1980) [13] developed an epidemic chain model by assuming beta distribution of first kind for the probability of being infected by contact with a given infective from the same household. This model includes, as a particular case, the epidemic chain model corresponding to the stochastic Kermack-McKendrick model and, as a limiting case, the Reed-Frost chain binomial model.

A more detailed comparison of the fits provided by the two models namely, Reed-Frost chain binomial model and the stochastic version of the Kermack-McKendrick epidemic model was not attempted by Becker for any epidemic chain model developed by assuming any other kind of Beta distribution for the probability of being infected by contact with a given infective from the same household. In order to make a more exhaustive comparison, we formulate a modified epidemic chain model by assuming a beta distribution of third kind for the probability of being infected by contact with a given infective from the same household. Also, in Becker's paper only the chain probabilities of the possible chains for a household of size three, four and five with one-introductory case was shown, but in this paper we have made an attempt to develop the chain probabilities of the possible chains for a household of size three, four and five with two and three introductory cases for the modified epidemic chain model.

3. Probability of Escaping Infection

As reviewed in the previous paper by Nath *et al.* (2015) [17], let us consider a disease say, influenza, which is able to spread from person in a household. Let the time at which the disease is introduced to the household as the time origin and suppose that the outbreak within the household is over by time t^*. Assume that during the time interval $(0, t^*)$ the chance of infection from outside the household is negligible compared with the chance of infection from within the household. Following a latent period of random duration, an infected person becomes infectious and remains so until his removal by isolation, death or recovery, with immunity for the duration of the outbreak. The probability that a given infected person *A*, say, transmits the disease to any given susceptible during the time increment $(t, t+h)$ is assumed to be

$$\Lambda(t)h + o(h). \tag{1}$$

So, $\Lambda(t)$ indicates how infectious A is at time t. By partitioning the interval $\left(0, t^*\right)$ into n small time increment $\left(t_{i-1}, t_i\right)$ of length $h_i = t_i - t_{i-1}$, $i = 1, 2, \cdots, n$, the probability that any given susceptible escapes infection by A during the interval $\left(0, t^*\right)$ is

$$\prod_{i=1}^{n}\left\{1 - \Lambda\left(t_{i-1}\right)h_i + o\left(h_i\right)\right\} \tag{2}$$

which tends in the limit as $n \to \infty$ and the partition becomes finer, to

$$\varepsilon = \exp(-I) \tag{3}$$

where $I = \int_{0}^{t^*} \Lambda(t)\mathrm{d}t$.

In particular case when Λ assumes the constant value λ when A is infectious, but assumes the value zero otherwise , we find that $I = \lambda T$, where T is the duration of A's infectious period and λ is A's infection rate , so I indicates the potential that A has for transmitting the disease to any given susceptible of the household.

ε, the probability that any given susceptible escapes infection by any given infected person is constant. If both λ and T are constants then ε is a constant.

4. Chains of Infection

According to the chain of infection so developed by Becker (1980) [13], he explained that it is not always possible to determine which infective is responsible for a certain infection. It is easier by making use of the gaps between cases, to partition the cases of a household into generations: the susceptible infected by direct contact with the introductory cases are said to make up the first generation of cases; the susceptibles infected by direct contact with first generation cases are said to make up the second generation and so forth. By an epidemic chain we mean the enumeration of the number of cases in each generation.

Let us take an example of a five member household with one introductory case, as per the formula (Equation (13)) explained in this paper, the total no. of possible epidemic chains are

$$2^{5-1} = 16.$$

And out of the sixteen (16) possible epidemic chains, let 1-2-1-0 be one such chains out of all sixteen (16) combinations, where this chain 1-2-1-0 denotes as the chain consisting of one introductory case, two first generation cases, one second generation case and no cases in later generation.

$$1\text{-}2\text{-}1\text{-}0.$$

1: Introductory case;

2: First generation case;

1: Second generation case;

0: Third generation case.

Corresponding to a given infective A, the conditional probability that r out of k susceptibles of the household escape infection by A is

$$^kC_r\varepsilon^r\left(1-\varepsilon\right)^{k-r} \qquad (4)$$

given the infection potential *I* of infective *A*.

Corresponding to a given infective *A*, unconditional probability that *r* out of *k* susceptibles of the household escape infection *A* is given by

$$^kC_rE\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}. \qquad (5)$$

Becker (1980) [13] has considered ε being a beta distribution of first kind given by the density function

$$f\left(\varepsilon\right)=\left\{B\left(\alpha,\beta\right)\right\}^{-1}\varepsilon^{\alpha-1}\left(1-\varepsilon\right)^{\beta-1}, \quad o<\varepsilon<1,\alpha,\beta>0, \qquad (6)$$

and $E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}$ is given by Becker (1980) as

$$E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}=\frac{B\left(\alpha+r,\beta+k-r\right)}{B\left(\alpha,\beta\right)}. \qquad (7)$$

Again in the paper of Nath *et al*. (2015) [17], ε was considered to follow beta distribution of third kind by Nagar and Ramirez-Venagas (2012) [20] given by the density function

$$f\left(\varepsilon\right)=\frac{2^\alpha\varepsilon^{\alpha-1}\left(1-\varepsilon\right)^{\beta-1}}{B\left(\alpha,\beta\right)\left(1+\varepsilon\right)^{\alpha+\beta}}, \quad o<\varepsilon<1,\alpha,\beta>0. \qquad (8)$$

Then $E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}$ is given by $\dfrac{1}{B\left(\alpha,\beta\right)}\displaystyle\int_0^1\frac{2^\alpha\varepsilon^{\alpha+r-1}\left(1-\varepsilon\right)^{\beta+k-r-1}}{\left(1+\varepsilon\right)^{(\alpha+r)+(\beta+k-r)}}\sum_{i=0}^k{}^kC_i\varepsilon^i\mathrm{d}\varepsilon$.

Since, ε is the probability of being infected by contact with a given infective from the same household. So the higher herms of ε can be neglected.

$E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}$ can further found to be as

$$\frac{B\left(\alpha+r,\beta+k-r\right)}{B\left(\alpha,\beta\right)}+\frac{k}{B\left(\alpha,\beta\right)}\sum_{j=1}^\infty B\left(\alpha+r+j,\beta+k-r\right), \quad k\geq r. \qquad (9)$$

For the practical application the term $E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}$ can be considered as

$$E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}=\frac{B\left(\alpha+r,\beta+k-r\right)}{B\left(\alpha,\beta\right)}+k\frac{B\left(\alpha+r+1,\beta+k-r\right)}{B\left(\alpha,\beta\right)}, \quad \text{if} \quad \beta>r+1. \qquad (10)$$

The above term is resulted after applying the test for convergence of the infinite series $\displaystyle\sum_{j=1}^\infty B\left(\alpha+r+j,\beta+k-r\right)$. In this process Raabe's test was proved to be stronger than the D'Alembert's Ratio test and succeed when the Ratio test fails. For the test of convergence of the infinite beta series, the Raabe's test is applied when the test fails for the Ratio test.

Further, $E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}$ can also be expressed as

$$E\left\{\varepsilon^r\left(1-\varepsilon\right)^{k-r}\right\}=\frac{\alpha^{(r)}\beta^{(k-r)}}{\left(\alpha+\beta\right)^{(k)}}\left\{1+k\frac{\left(\alpha+r\right)}{\left(\alpha+\beta+k\right)}\right\} \qquad (11)$$

where $\alpha^{(r)}=\alpha\left(\alpha+1\right)\left(\alpha+2\right)\cdots\left(\alpha+r-1\right)$.

Then expression (5) using Equation (11) is given by

$$^{k}C_{r}E\left\{\varepsilon^{r}\left(1-\varepsilon\right)^{k-r}\right\}=\,^{k}C_{r}\frac{\alpha^{(r)}\beta^{(k-r)}}{\left(\alpha+\beta\right)^{(k)}}\left\{1+k\frac{\left(\alpha+r\right)}{\left(\alpha+\beta+k\right)}\right\}. \qquad (12)$$

In our study, a theory has been developed for the number of household members *i.e.*, for say the household size be *m* and the number of introductory cases be *j*.

In our case, let us consider that, *m* and *j* takes the values as *m* = 3, 4 or 5 and *j* = 1, 2 or 3.

Considering the expression (5), *k*, the number of susceptibles of the household who escape infection by *A* is also known as the number of final cases possible for a *m*-member household containing *j*-introductory cases. Therefore, *k* takes values as $0,1,\cdots,\left(m-j\right)$.

In general, for a binomial distribution, the total number of epidemic chains possible for the households of size *m* containing *j*-introductory cases is given as

$$\sum\nolimits_{i=0}^{m-j}{}^{m-j}C_{i}=2^{m-j}. \qquad (13)$$

Some particular cases:

1) Number of epidemic chains possible for household of size, *m* = 5 containing *j* = 2 introductory cases are $2^{5-2}=8$.

2) Number of epidemic chains possible for household of size, *m* = 5 containing *j*= 3 introductory cases are $2^{5-3}=4$.

To illustrate the computation of the probabilities associated with the different possible epidemic chains we consider the chain 1-1-2-0 in a household of size five including one introductory case. The probability of this chain, conditional on the probabilities $\varepsilon_{1},\varepsilon_{2},\varepsilon_{3}\,\&\,\varepsilon_{4}$ that a given susceptible escape infection by each of the four infected individuals, respectively found to be

$$^{4}C_{3}\varepsilon_{1}^{3}\left(1-\varepsilon_{1}\right)\,^{3}C_{1}\varepsilon_{2}^{1}\left(1-\varepsilon_{2}\right)^{2}\,^{1}C_{1}\left(\varepsilon_{3}\varepsilon_{4}\right)^{1}\left(1-\varepsilon_{3}\varepsilon_{4}\right)^{0}. \qquad (14)$$

The unconditional probability so obtained by taking the expectation of this conditional probability and using the fact that $\varepsilon_{1},\varepsilon_{2},\varepsilon_{3}\,\&\,\varepsilon_{4}$ are independent random variables having the same beta distribution of third kind. Using the form of Equation (11), the probability of the chain 1-1-2-0 in a household of size five including one introductory case was found by Nath *et al.* (2015) [17] as,

$$E\left\{^{4}C_{3}\varepsilon_{1}^{3}\left(1-\varepsilon_{1}\right)\,^{3}C_{1}\varepsilon_{2}^{1}\left(1-\varepsilon_{2}\right)^{2}\,^{1}C_{1}\left(\varepsilon_{3}\varepsilon_{4}\right)^{1}\left(1-\varepsilon_{3}\varepsilon_{4}\right)^{0}\right\}$$
$$=12\left[\frac{\alpha^{(3)}\beta^{(1)}}{\left(\alpha+\beta\right)^{(4)}}\left\{1+\frac{\left(4\alpha+12\right)}{\left(\alpha+\beta+4\right)}\right\}\right]\left[\frac{\alpha^{(1)}\beta^{(2)}}{\left(\alpha+\beta\right)^{(3)}}\left\{1+\frac{3\alpha+3}{\left(\alpha+\beta+3\right)}\right\}\right] \qquad (15)$$
$$\left[\frac{\alpha^{(1)}\beta^{(0)}}{\left(\alpha+\beta\right)^{(1)}}\left\{1+\frac{\alpha+1}{\left(\alpha+\beta+1\right)}\right\}\right].$$

Since, $\alpha^{(0)}=\beta^{(0)}=1$, $\alpha^{(1)}=\alpha$ and $\beta^{(1)}=\beta$, so putting the values in the above equation we have

$$E\left\{^{4}C_{3}\varepsilon_{1}^{3}\left(1-\varepsilon_{1}\right)\,^{3}C_{1}\varepsilon_{2}^{1}\left(1-\varepsilon_{2}\right)^{2}\,^{1}C_{1}\left(\varepsilon_{3}\varepsilon_{4}\right)^{1}\left(1-\varepsilon_{3}\varepsilon_{4}\right)^{0}\right\}$$
$$=12\frac{\alpha^{(3)}\alpha^{2}\beta^{(2)}\beta\left(5\alpha+\beta+16\right)\left(4\alpha+\beta+6\right)\left(2\alpha+\beta+2\right)}{\left(\alpha+\beta\right)^{(5)}\left(\alpha+\beta\right)^{(4)}\left(\alpha+\beta\right)^{(2)}}. \qquad (16)$$

5. Chain Probabilities for Two and Three Introductory Cases

The probabilities of the possible chains given for household of sizes three, four and five with two introductory cases and for household of sizes four and five with three introductory cases has been developed in this paper are shown in the below Tables 1-5.

The number of possible cases, *i.e.*, the combination of the epidemic chains was calculated by using the formula (Equation (13)) and the probabilities are calculated by using the Equation (12). Using real life data, *i.e.*, for some observed frequencies of the possible chains, expected frequencies can be generated and goodness of fit can be tested. As this paper is restricted to development of theories and data collection for the application to real-life data is in progress, therefore it is planned for our next communication.

6. Conclusions

The paper aims to develop an alternative approach to the Becker's epidemic chain binomial model (1980) [13] of infectious diseases and its extensions. Since, the modified epidemic chain binomial model developed in the Nath *et al.* (2015) [17] paper was a more complicated model than that of the epidemic chain binomial model of Becker (1980) [13], therefore, the chain probabilities so developed as a extension of the previous work of Nath *et al.* (2015) [17] for household of size three, four and five for two and three introductory cases in a closed population

Table 1. Epidemic chain probabilities for households of size three with 2-introductory cases.

Type of chain	Probability assuming beta type III
2-0	$\dfrac{\alpha\left(2\alpha+\beta+2\right)}{\left(\alpha+\beta\right)^{(2)}}$
2-1	$\dfrac{\beta\left(2\alpha+\beta+1\right)}{\left(\alpha+\beta\right)^{(2)}}$

Table 2. Epidemic chain probabilities for households of size four with 2-introductory cases.

Type of chain	Probability assuming beta type III
2-0	$\dfrac{\alpha^{(2)}\left(3\alpha+\beta+6\right)}{\left(\alpha+\beta\right)^{(3)}}$
2-1-0	$\dfrac{2\alpha^2\beta\left(3\alpha+\beta+4\right)\left(2\alpha+\beta+2\right)}{\left(\alpha+\beta\right)^{(3)}\left(\alpha+\beta\right)^{(2)}}$
2-1-1	$\dfrac{2\alpha\beta^2\left(3\alpha+\beta+4\right)\left(2\alpha+\beta+1\right)}{\left(\alpha+\beta\right)^{(3)}\left(\alpha+\beta\right)^{(2)}}$
2-2	$\dfrac{\beta^{(2)}\left(3\alpha+\beta+2\right)}{\left(\alpha+\beta\right)^{(3)}}$

Table 3. Epidemic chain probabilities for households of size five with 2-introductory cases.

Type of chain	Probability assuming beta type III
2-0	$\dfrac{\alpha^{(3)}(4\alpha+\beta+12)}{(\alpha+\beta)^{(4)}}$
2-1-0	$\dfrac{3\left\{\alpha^{(2)}\right\}^{2}\beta(4\alpha+\beta+9)(3\alpha+\beta+6)}{(\alpha+\beta)^{(4)}(\alpha+\beta)^{(3)}}$
2-1-1-0	$\dfrac{6\alpha^{(2)}\alpha^{2}\beta^{2}(4\alpha+\beta+9)(3\alpha+\beta+4)(2\alpha+\beta+2)}{(\alpha+\beta)^{(4)}(\alpha+\beta)^{(3)}(\alpha+\beta)^{(2)}}$
2-1-1-1	$\dfrac{6\alpha^{(2)}\alpha\beta^{3}(4\alpha+\beta+9)(3\alpha+\beta+4)(2\alpha+\beta+1)}{(\alpha+\beta)^{(4)}(\alpha+\beta)^{(3)}(\alpha+\beta)^{(2)}}$
2-1-2	$\dfrac{3\alpha^{(2)}\beta^{(2)}\beta(4\alpha+\beta+9)(3\alpha+\beta+2)}{(\alpha+\beta)^{(4)}(\alpha+\beta)^{(3)}}$
2-2-0	$\dfrac{3\alpha^{2}\beta^{(2)}(4\alpha+\beta+6)(2\alpha+\beta+2)}{(\alpha+\beta)^{(4)}(\alpha+\beta)^{(2)}}$
2-2-1	$\dfrac{3\alpha\beta\beta^{(2)}(4\alpha+\beta+6)(2\alpha+\beta+2)}{(\alpha+\beta)^{(4)}(\alpha+\beta)^{(2)}}$
2-3	$\dfrac{\beta^{(3)}(4\alpha+\beta+3)}{(\alpha+\beta)^{(4)}}$

Table 4. Epidemic chain probabilities for households of size four with 3-introductory cases.

Type of chain	Probability assuming beta type III
3-0	$\dfrac{\alpha(2\alpha+\beta+2)}{(\alpha+\beta)^{(2)}}$
3-1	$\dfrac{\beta(2\alpha+\beta+1)}{(\alpha+\beta)^{(2)}}$

Table 5. Epidemic chain probabilities for households of size five with 3-introductory cases.

Type of chain	Probability assuming beta type III
3-0	$\dfrac{\alpha^{(2)}(3\alpha+\beta+6)}{(\alpha+\beta)^{(3)}}$
3-1-0	$\dfrac{2\alpha^{2}\beta(3\alpha+\beta+4)(2\alpha+\beta+2)}{(\alpha+\beta)^{(3)}(\alpha+\beta)^{(2)}}$
3-1-1	$\dfrac{2\alpha\beta^{2}(3\alpha+\beta+4)(2\alpha+\beta+1)}{(\alpha+\beta)^{(3)}(\alpha+\beta)^{(2)}}$
3-2	$\dfrac{\beta^{(2)}(3\alpha+\beta+2)}{(\alpha+\beta)^{(3)}}$

also came out as complicated expressions. Though the probabilities come out to be complicated and application to real life data may be laborious, but in modern era, such difficulties may be easily sorted out with the help of the advanced computer techniques and various statistical softwares like, *R*-programming, SPSS/ PASW, SAS, etc.

Hence, the application of this purely academic investigation can be pursued further with proper thinking and dedication by using various statistical methods and techniques, and a more in-depth inference of the theory can be drawn. And this phenomenon may greatly help to alleviate and prevent at least some of the human sufferings which exist in the present time due to infectious diseases. However, our work in this paper restricts to the development of the theoretical concepts only, but we are in a process of illustrating the application of this method to the real life data on common cold for three-, four-, and five-member household with closed population. And we are expecting to present that in our next communication.

Acknowledgements

References

[1] Bailey, N.T.J. (1957) The Mathematical Theory of Epidemics. Chap. 6, Charles Griffin & Co., Ltd., London, 75-108.

[2] Bernoulli, D. (1766) Essai d'une nouvelle analyse de la mortalite causee par la petite verole. Mem. Math. Phys. Acad. Roy. Sci., Paris (1766) 1. (Reprinted in: L.P. Bouckaert, B.L. van der Waerden (Eds.), Die Werke von Daniel Bernoulli, Bd. 2 Analysis und Wahrscheinlichkeitsrechnung, Birkhauser, Basel, 1982, p. 235. English translation entitled "An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it" in: L. Bradley, Smallpox Inoculation: An Eighteenth Century Mathematical Controversy, Adult Education Department, Nottingham, 1971, p. 21. Reprinted in: S. Haberman, T.A. Sibbett (Eds.), History of Actuarial Science, Vol. VIII, Multiple Decrement and Multiple State Models, William Pickering, London, 1995, p. 1.)

[3] Hamer, W.H. (1906) The Milroy Lectures on Epidemic Disease in England—The Evidence of Variability and Persistence of Type. *The Lancet*, **1**, 733-739.

[4] Daley, D.J. and Gani, J. (1999) Epidemic Modelling: An Introduction. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511608834

[5] Kermack, W.O. and McKendrick, A.G. (1927) A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society, Series A*, **115**, 700-721. https://doi.org/10.1098/rspa.1927.0118

[6] Heasman, H.A. and Reid, D.D. (1961) Theory and Observation in Family Epidemics of the Common Cold. *British Journal of Preventive and Social Medicine*, **15**, 12-16. https://doi.org/10.1136/jech.15.1.12

[7] Bailey, N.T.J. (1975) The Mathematical Theory of Infectious Diseases and Its Application. Griffin, London.

[8] Ridler-Rowe, C.J. (1967) On a Stochastic Model of an Epidemic. *Journal of Applied*

Probability, **4**, 19-33. https://doi.org/10.1017/S0021900200025201

[9]   Bartlett, M.S. (1949) Some Evolutionary Stochastic Processes. *Journal of the Royal Statistical Society*, **B11**, 211-229.

[10]   Keilson, J. (1974) Sojourn Times, Exit Times and Jitter in Multivariate Markov Processes. *Advances in Applied Probability*, **6**, 747-756. https://doi.org/10.1017/S000186780002855X

[11]   Billiard, L. (1977) Mean Duration Time for a General Epidemic Process. *Journal of Applied Probability*, **14**, 232-240. https://doi.org/10.2307/3212994

[12]   Billiard, L., Lacayo, H. and Langberg, N. (1978) A New Look at the Simple Epidemic Process. Florida State University, Tallahassee, FL.

[13]   Becker, N. (1980) An Epidemic Chain Model. *Biometrics*, **36**, 249-254. https://doi.org/10.2307/2529976

[14]   Jacquez, J.A. (1987) A Note on Chain-Binomial Models of Epidemic Spread: What Is Wrong with the Reed-Frost Formulation? *Mathematical Biosciences*, **87**, 73-82.

[15]   Demiris, N. (2004) Bayesian Inference for Stochastic Epidemic Models Using Markov Chain Monte Carlo Methods. University of Nottingham, Nottingham.

[16]   Barbour, A.D. and Utev, S. (2004) Approximating the Reed-Frost Epidemic Process. University of Nottingham, Nottingham.

[17]   Nath, D.C., Das, K.K. and Chakraborty, T. (2015) A Modified Epidemic Chain Binomial Model. *Open Journal of Statistics*, **6**, 1-6.

[18]   Cairoli, L.H. (1988) Chain Binomial Epidemic Models. An Abstract of Master's Report, Department of Statistics, Kansas State University, Manhattan, KS.

[19]   Ludwig, D. (1975) Final Size Distributions for Epidemics. *Mathematical Biosciences*, **23**, 33-46. https://doi.org/10.1016/0025-5564(75)90119-4

[20]   Nagar, D.K. and Ramirez-Venagas, Y.A. (2012) Non-Central Beta Type 3 Distribution. *Progress in Applied Mathematics*, **4**, 29-43.

Scientific Research Publishing