Old Dominion University
ODU Digital Commons

Computer Science Faculty Publications | Computer Science

2012

On a Versatile Stochastic Growth Model

Samiur Arif
Old Dominion University

Ismail Khalil

Stephan Olariu
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_fac_pubs

Part of the Artificial Intelligence and Robotics Commons

On a Versatile Stochastic Growth Model

Samiur Arif [1] Ismail Khalil [2] Stephan Olariu[1]

[1] *Computer Science, Old Dominion University,*
4700 Elkhorn Avenue,
Norfolk, VA 23508, USA

E-mail: {sarif,olariu}@cs.odu.edu

[2] *Institute of Telecooperation, Johannes Kepler Linz University,*
Altenberger Strasse 69,
Linz, Austria

E-mail: ismail.khalil@jku.at

Abstract

Growth phenomena are ubiquitous and pervasive not only in biology and the medical sciences, but also in economics, marketing and the computer and social sciences. We introduce a three-parameter version of the classic pure-birth process growth model when suitably instantiated, can be used to model growth phenomena in many seemingly unrelated application domains. We point out that the model is computationally attractive since it admits of conceptually simple, closed form solutions for the time-dependent probabilities.

Keywords: Stochastic Growth Models, Pure Birth Process, Time-Dependent Probabilities, Continuous Markov Chain

1. Introduction

Over the centuries, the task of understanding the dynamics of various growth phenomena noticed in nature and society had captured the interest of scholars and philosophers. However, with the exception of D. Bernoulli's attempt at modeling the outbreak of a smallpox epidemic [7], T. R. Malthus' population growth model [28] and P. Verhulst formulation of the logistic growth model [39], most of the early efforts, including Gompertz's human mortality model [17], were empirical, the data available was highly unreliable and the general sense of evidence was lacking by today's standards. All this were to change in 1874 when Watson and Galton undertook the first systematic attempt at understanding growth and extinction phenomena noticed in the social sciences that, nonetheless, could not be convincingly explained [16, 41]. Their work was among the earliest systematic attempts at enlisting the help of probability theory in modeling, and thus, understanding, the dynamics of population growth. While their pioneering work had focused on a rather narrow problem, namely that of accounting for the wholesale extinction of family names in Great Britain, their mathematical methods (and, in particular, the by-now classic *Galton-Watson process*) turned out to be surprisingly powerful and general [18].

Unfortunately, Galton and Watson's work and their mathematical model was neglected for many

years, more precisely until 1924, when Yule [46] applied similar probabilistic machinery to the study of the dynamics of the proliferation of new species and genera. Yule's contribution, a *linear pure-birth process*, was rediscovered, a few years later by W. H. Furry [15] in the context of electron physics and by Feller and Lotka in population biology [14, 24].

The next major landmark event in modeling growth model was provided by Kolmogorov and Dimitriev's seminal work [27] where they generalized the Galton-Watson model, proposing *branching processes* as a powerful modeling tool for a large class of growth phenomena that, as it turned out, sub-summed much of the previous work cited above and set the stage for a systematic look at growth models.

Not surprisingly, in the following decades a plethora of mathematical growth models intended to capture the essence of natural and social phenomena ranging from the social sciences to genetics, to biology, to epidemiology, to physics, to astronomy, to computer science and macro-economics have been proposed in the literature [1, 3, 8, 9, 11, 18, 20, 26, 31].

Some of these models are intended to capture the essence of high-powered, unhindered growth as witnessed, for example, in particle physics and astronomy (e.g. the Big Bang theory). By contrast, most of the growth phenomena that we encounter in biology and medicine, economics and the social sciences involve a close interaction between the phenomenon under study and its surrounding environment. For example in economics, the merger of companies is subject to internal stimuli and to external pressure (inhibition) coming from the marketplace and competition [34]. In the biological sciences, when resources are plentiful and environmental conditions appropriate, bacteria population can increase rapidly. However, in most instances resources are not unlimited and environmental conditions are far from optimal. Climate, food, habitat, water availability, and other similar factors conspire to keep population growth in check. Indeed, the environment can only support a limited number of individuals in a population before some resource runs out and endangers the very survival of those individuals. Population models are used to determine

maximum harvest for agriculture, to understand the dynamics of biological invasions, and have numerous environmental conservation implications. Population models are also used to understand the spread of parasites, viruses, and disease. The realization of our dependence on environmental health has created a need to understand the dynamic interactions of the earths flora and fauna. Methods in population modeling have greatly improved our understanding of ecology and the natural world [22, 26, 33, 36].

1.1. Our contributions

Among the numerous growth models proposed over the decades, two main general-purpose models stand out: the *deterministic* growth models and the *stochastic* growth models [1, 23, 29, 31, 33, 37]. The deterministic models, including the well-known *exponential*, *logistic*, *Gompertz* and *Bass* models are well understood and in widespread use, chiefly because they are mathematically tractable and, most of the time, fairly accurate. On the other hand, the stochastic growth models, while often more realistic, typically lead to complicated or even intractable mathematics. We note that for particular instances of the stochastic model closed forms for the various transition probabilities can be obtained by using bivariate probability generating functions and by solving a Fokker-Planck partial differential equation. However, it is common knowledge that, in general, the resulting partial differential equation does not admit of a closed form solution and the only recourse is to employ various approximation schemes [3, 13]. This is, no doubt, one of the main reasons that stochastic growth models have been less widely used, especially in the biological and biomedical communities [29, 30].

With this in mind, the main contribution of this work is draw attention to a simple stochastic growth model, namely, the pure birth process, that was somehow overlooked by the research community. We point out that the pure-birth model admits of conceptually simple, closed form solutions for the time-dependent probabilities of the system being in a given state. Importantly, these closed-form solutions can be obtained through a straightforward iterative approach that lends itself to efficient com-

puter implementation. Our extensive simulations have shown that, suitably instantiated, the pure-birth process can be used to model growth phenomena in many, seemingly unrelated, application domains ranging from yeast production to the spread of Tasmanian sheep population, to the adoption of durable consumer goods and to software reliability.

In summary, we do not to advocate replacing the aforementioned, well-established, growth models. Our goal is to argue that being simple and versatile, the pure-birth model is a worthwhile addition to the panoply of tools that researchers have at their disposal and, consequently, deserves a place in the growth model toolbox.

The remainder of the paper is organized as follows: Section 2 offers a quick review of modeling the interaction between population growth and the surrounding environmental attributes and limitations and looks briefly at the exponential, logistic Gompertz, Bass and Yule growth models. Next, in Section 3 we describe the generic pure-birth model that we deal with and for which we derive closed form for the various transition probabilities. In Section **??** we discuss a simple algorithmic implementation of the task of computing efficiently the state probabilities $P_k(t)$. Further, in Section 4 we show that the parameters of the pure-birth model can be fine-tuned to track very closely various growth phenomena. Section 5 offers concluding remarks and directions for future work.

2. State of the art

In many biological situations the finiteness of available resource ensure that an isolated population cannot grow without limit and in due time after achieving carrying capacity a violent downward fluctuation in population size is bound to occur which will drive the population to extinction [3, 33]. Population growth in the natural world does not occur smoothly up to the carrying capacity and remain there. In other words it fluctuates before it reaches the carrying capacity.

Self-reproduction is the main feature of all living organisms. This is what distinguishes them from non-living things. Any model of population dy-

namics must include reproduction. In Subsections 2.1 and 2.2 we will discuss two important models of population growth based on reproduction of organisms: the exponential and the logistic models. In Subsection 2.3 we look briefly at the Gompertz model used in a variety of situations ranging from reliability to cancer research [29, 31]. Further in Subsection 2.4 we discuss Bass's growth model, the immensely popular model used in marketing research to predict, among other things, the time of adoption of durable consumer goods. Finally, to set the stage for our further discussion, in Subsection 2.5 we look at a simple stochastic growth model proposed by Yule, that turns out to be a linear growthrate pure birth process.

2.1. *The exponential growth model*

The exponential population growth model is usually associated with the work of T. R. Malthus (1766-1834) [42] who first realized that any species can potentially increase in numbers according to a geometric series. If a species has non-overlapping populations (e.g., annual plants), and each organism produces R offspring then the number, N_t, of individual in generation $t = 0, 1, 2, \cdots\cdots$ is given by

$$N_t = N_0 R^t, \tag{1}$$

where N_0 is the initial size of the population (i.e., at the 0-th generation). By writing $r = \ln R$, (1) becomes

$$N_t = N_0 e^{rt}, \tag{2}$$

where, depending on the application domain, r is referred to as Malthusian parameter, intrinsic rate of increase, instantaneous rate of natural increase, or population growth rate.

A glance at (2) reveals that depending on the parameter r, there are three possible outcomes:

- if $r < 0$, the population declines exponentially;
- if $r > 0$, the population increases exponentially;
- if $r = 0$, the population does not change.

Some of the applications of the exponential model and equation (2) are in microbiology (growth of bacteria), conservation biology (restoration of

disturbed populations), insect rearing (prediction of yield), plant or insect quarantine (population growth of introduced species), fishery management (prediction of fish population dynamics) [1, 11, 26].

In spite of its widespread use, a serious shortcoming of the exponential model is that it ignores the limitations to unbounded growth imposed by environmental factors and conditions [3].

2.2. The logistic growth model

The logistic growth model was proposed by the Belgian mathematician P. Verhulst around 1838. Verhulst suggested that the rate of population increase may be limited, as it may depend on the population size, $N = N(t)$, and on the growth rate, r, defined as

$$r = r_0 \left(1 - \frac{N}{K} \right),$$

where the parameter K represents the upper limit on population growth and is called *carrying capacity*. The carrying capacity of a biological species in a given environment is the largest population size of the species that the environment can sustain indefinitely, given the food, habitat, water and other necessities required.

Observe that when N is small relative to K, the population growth rate r is maximal and is nearly r_0. The parameter r_0 can be interpreted as the population growth rate in the absence of competition. The population growth rate declines with N and reaches 0 when $N = K$. If N exceeds K, the population growth rate becomes negative and the population declines, eventually becoming extinct. It can be easily seen population size N is a function of time t. The dynamics of population growth are captured by the following differential equation, with boundary condition $N = N_0$ at $t = 0$,

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right). \tag{3}$$

Equation (3) tells us that when N is small, the early unimpeded growth rate can be approximated by rN. Later, as N grows, $\frac{rN^2}{K}$, becomes larger and larger, as some members of the population interfere with each other by competing for critical resource,

such as food or living space. The competition diminishes the combined growth rate, until the value of N ceases to grow which indicated the saturation of the population in the given environment.

As it turns out, taking into account the boundary condition, equation (3) has the solution

$$N(t) = \frac{N_0 K}{N_0 + (K - N_0) e^{-rt}}. \tag{4}$$

Observe that

$$\lim_{t \to +\infty} N(t) = K$$

which is to say that the limiting value of N is K, the highest value that the population can reach given infinite time or come close to reaching in finite time.

2.3. The Gompertz growth model

Gompertz function is used to approximate growth up to certain limit.It is the most popular function to estimate growth,probability and proportion along with logistic function. It is more general type than logistic function class.The right-hand or future value asymptote of the function is approached much more gradually by the curve than the left-hand or lower valued asymptote, in contrast to the logistic function in which both asymptotes are approached by the curve symmetrically.

$$G(x) = ae^{be^{cx}} \tag{5}$$

where $a \sim (0, \infty), b \sim (-\infty, 0)$ and $c \sim (-\infty, 0)$. $G(t)$ usually denotes the number of individuals at time t.

The 3 coefficients control the shape of the function . The upper limit of the curve is controlled by the coefficient a. Whereas b is the growth rate and c is the acceleration rate .

Gompertz model have been used in various fields, for example

- Mobile phone uptake, where costs were initially high (so uptake was slow), followed by a period of rapid growth, followed by a slowing of uptake as saturation was reached. [44]

- Population in a confined space, as birth rates first increase and then slow as resource limits are reached. [19]
- Modeling of growth of tumors [31, 32]
- Reliability growth model [38]

2.4. *The Bass growth model*

The Bass growth model [4] was developed by Frank Bass and describes the process of how new products get adopted as an interaction between users and potential users. It has been described as one of the most famous empirical generalizations in marketing. Based on very simple behavioral observations, Bass's model became immensely popular. In spite of its deceiving simplicity, the Bass model is an excellent predictor of new-product sales and is still heavily used.

The model is based on number of empirical observations. First, that there exists an intrinsic tendency of some individuals to make a purchase, independent of the number of previous adopters. Bass calls these individuals the *innovators*. In contrast to the innovators, some people buys a product due media and social pressure. Bass refers to these folks as *imitators*. Bass's second major assumption is that the probability that an initial purchase will be made at time t is a linear function of the number of previous buyers. Under these assumptions Bass postulates that

$$\frac{dN(t)}{dt} = [M - N(t)][p + \frac{q}{M}N(t)] \qquad (6)$$

where

- p and q be two parameters that quantify the extent of the influence of innovators and imitators, respectively;
- M – the market potential of a given product, that is, the size of the potential consumer population;
- $N(t)$ – the cumulative number of adopters in $[0,t]$
- $M - N(t)$ is the size of the remaining population, and
- $p + q/MN(t)$ is the instantaneous adoption rate of every individual in the remaining population.

The differential equation (6) with boundary condition $N(0) = 0$) has solution

$$N(t) = M\frac{e^{(p+q)t} - 1}{e^{(p+q)t} + \frac{q}{p}}.$$

It is easy to confirm that

$$\lim_{t \to \infty} N(t) = M$$

and that $N(t)$ has an S-shaped graph characteristics of many other growth models for example the logistic growth model.

It was recently pointed out by Yan *et al.* [45] that Bass' growth model is a common generalization of the exponential and logistic models. Indeed,

- for $q = 0$, Bass' model reduces to the exponential model;
- for $p = 0$, Bass' model reduces to the logistic model.

As was pointed out by Bass himself in [4] as well as in later work [5, 6], the accuracy of his model depends on the three parameters, M, p, and q. In the case of a new product, these parameters are obtained by using existing sales data for previous versions of the same product or by extrapolating from sales data for similar products.

2.5. *The Yule model*

No doubt one of the earliest stochastic processes was proposed in 1924 by G. U. Yule [46] in his mathematical theory of evolution. In Yule's model $\{Y(t) \mid t \geqslant 0\}$ represents the number of species in some genus of plants or animals [35]. Yule's model assumes that evolution begins with a single species at time $t = 0$ and that species do not die out. If there are $Y(t)$ species at time t, then $\lambda Y(t)\delta$ is the probability that a new species will be created in the time interval $[t, t + \delta]$ for some very small $\delta > 0$.

More formally, the process $\{Y(t) \mid t \geqslant 0\}$ has boundary conditions $P_1(0) = 1$ and $P_i(0) = 0$ for $i \neq 1$ and transition rate λ_n such that:

$$\forall n \in \mathbb{N},, \ \lambda_n = n\lambda > 0.$$

An easy inductive argument confirms that $\forall t \geqslant 0$

$$P_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$$

It follows that In Yule's growth model the events $\{Y(t) = n\}$ are geometrically distributed with success probability $e^{-\lambda t}$.

It is clear that because of its underlying assumptions, the Yule model is not sufficiently flexible to deal with growth phenomena encountered in real-life application. In addition to the assumed linear growth rate, the Yule model assumes unbounded growth and, consequently, cannot be used to model growth in limited-capacity situations.

3. A generic pure-birth growth model

Consider a stochastic process $\{N(t) \mid t \geqslant 0\}$ of continuous parameter t, where for every positive integer k, $(1 \leqslant k \leqslant N)$, the event $\{N(t) = k\}$ occurs if the population contains k individuals at time t.

For simplicity, we assume that at time $t = 0$ the population starts with a single individual and assume an upper bound N on the *carrying capacity*, i.e., the total size of the population. It is worth noting that in many, if not most, practical situations this assumption is violated. However, it is rather straightforward to see that the case of n_0 initials reduces to the n_0-fold convolution of independent simple processes.

When the population contains k individuals, λ_k captures the growth rate of the population. In the generic pure-birth process that we investigate the λ_ks are subject to the mild restriction

$$\lambda_i \neq \lambda_j \qquad (7)$$

for all $1 \leqslant i, j \leqslant N$; moreover, since N is the largest size the population may assume, we have

$$\lambda_N = 0. \qquad (8)$$

As it turns out, and will become apparent in Section 4, the condition (7) is not unduly restrictive since we can instantiate the λ_ks in such a way that (7) is satisfied.

Further, we let $P_k(t)$ denote the probability that the event $\{N(t) = k\}$ occurs. In other words,

$$P_k(t) = \Pr[\{N(t) = k\}].$$

In this section we outline a possible derivation of a closed form for $P_k(t)$ for all k, $(1 \leqslant k \leqslant N)$ subject to (7) and (8).

3.1. Deriving a closed form for $P_k(t)$, $1 \leqslant k < N$

It is fairly obvious that for $k > 1$ and for a small $h > 0$, $P_k(t+h)$ has the following components:

- $P_k(t)[1 - h\lambda_k + o(h)]$, which describes the probability of staying at state k.
- $P_{k-1}(t)[h\lambda_{k-1} + o(h)]$, which describes the probability of reaching to the state k from state $k-1$.

This allows us to write

$$
\begin{aligned}
P_k(t+h) &= P_k(t)[1 - h\lambda_k + o(h)] \\
&+ P_{k-1}(t)[h\lambda_{k-1} + o(h)] + o(h) \\
&= P_k(t)[1 - h\lambda_k] + P_{k-1}(t)h\lambda_{k-1} + o(h).
\end{aligned}
$$

Transposing $P_k(t)$ and dividing with h yields

$$\frac{P_k(t+h) - P_k(t)}{(t+h) - t} = -\lambda_k P_k(t) + \lambda_{k-1} P_{k-1}(t) + \frac{o(h)}{h}.$$

Taking limits on both sides of this equality as $h \to 0$ yields the differential equation

$$\frac{dP_k(t)}{dt} = -\lambda_k P_k(t) + \lambda_{k-1} P_{k-1}(t) \qquad (9)$$

with the boundary condition $P_k(0) = 0$.

Proceeding similarly, when $k = 1$, we write

$$
\begin{aligned}
P_1(t+h) &= P_1(t)[1 - h\lambda_1 + o(h)] + o(h) \\
&= P_1(t)[1 - h\lambda_1] + o(h).
\end{aligned}
$$

Transposing $P_1(t)$ and dividing with h yields

$$\frac{P_1(t+h) - P_1(t)}{(t+h) - t} = -\lambda_1 P_1(t) + \frac{o(h)}{h}.$$

Taking limits on both sides of this equality as $h \to 0$ yields the differential equation

$$\frac{dP_1(t)}{dt} = -\lambda_1 P_1(t) \qquad (10)$$

with the boundary condition $P_1(0) = 1$. This latter equation can be easily solved to obtain

$$P_1(t) = e^{-\lambda_1 t}. \qquad (11)$$

Having obtained $P_1(t)$, we use the differential equation (9) to obtain a closed for $P_2(t)$. This process is then continued, iteratively, until a closed form is obtained for $P_k(t)$ for $(1 \leqslant k \leqslant N)$. The interested reader can find the details of the derivation, as well as the proof of Theorem 1, on the first author's webpage [2]. To summarize, we state the following general result.

Theorem 1. *For all $t \geqslant 0$,*

$$P_k(t) = \begin{cases} \lambda_1 \lambda_2 \cdots \lambda_{k-1} \sum_{i=1}^{k} \dfrac{e^{-\lambda_i t}}{k \prod\limits_{\substack{j=1 \\ j \neq i}} (\lambda_j - \lambda_i)}, & 1 \leqslant k \leqslant N-1 \\[2em] 1 - \sum_{i=1}^{N-1} \left(\prod\limits_{\substack{j=1 \\ j \neq i}}^{N-1} \left[\dfrac{\lambda_j}{\lambda_j - \lambda_i} \right] e^{-\lambda_i t} \right), & k = N. \end{cases}$$

(12)

The detail derivation of 12 can be found at [2].

4. Applications

The main goal of this section is to show that by a suitable instantiation of the λ_ks, the generic pure-birth process can be used to model various growth phenomena.

In Section 3 we have derived simple formulas for the state probabilities for a *generic* pure-birth process growth model. Specifically, (12) allow us to compute the probability of having k, $(1 \leqslant k \leqslant N)$, individuals at any given time t for all λ_ks subject to the condition (7). In turn, once the state probabilities are computed, we are in a position to derive the expected number, $E[N(t)]$, of individuals at any given time t, by the well-known formula

$$E[N(t)] = \sum_{k=1}^{N} k P_k(t). \tag{13}$$

In order to model various flavors of growth processes, we instantiate λ_k, $(1 \leqslant k \leqslant N)$, subject to (7), as described below:

$$\lambda_k = k^\alpha (N-k)^\beta \tag{14}$$

where

- α and β are *distinct* positive real numbers, assumed to be of infinite precision. This allows us to always "tweak" α and *beta* in such a way that $\lambda_i \neq \lambda_j$ for $i \neq j$, thereby satisfying the condition (7);

- we observe that k^α is the driving force of population growth where k is the current size of the population. In most applications we shall take $\alpha = 1.0$. Interestingly, there are growth models applications (such as software engineering) where $\alpha = 0.0$;

- notice also that $(N-k)^\beta$ acts as an *inhibitive force* imposed by environmental limitations.

4.1. Selecting the parameters

One of the key issues in using a growth model is the use of the parameters. In our case, the pure-birth process has three parameters, namely, N, α, and β. We now offer guidelines for selecting these parameters in a given application domain. First, the *finite capacity*, N, can be estimated quite accurately from available domain knowledge, along the lines that Bass uses to estimate the market potential of a given product. It is easy to see both analytically and empirically, that the prediction of the expected size of the population at time t is rather insensitive to N, especially if N turns out to be very large.

We have adopted Bass' strategy detailed in [4] for choosing the parameters p and q to selecting α and β. As we discussed already, these values are application-specific and involve a two-stage process:

- we begin by a standard learning process that involves selecting the best fit of α and β to existing data;

- we extrapolate the values of α and β obtained by learning to similar situations. In case no good match is available between the situations on which α and β were trained, a weighted average is employed.

4.2. Growth of yeast population

Many laboratory populations have been followed as they increase in size and the success of the deter-

ministic logistic curve in summarizing the resulting data sets depends on the circumstances surrounding each particular experiment. Classic yeast growth experiments are described by [10, 21, 25] and subsequent analyzes of Pearl [36]. For the prediction of yeast population growth we have used $N = 670,000$, $\alpha = 1$ and $\beta = 0.2$ and verified that (7) holds for the chosen parameters. Moreover, for $1 \leqslant i \neq j \leqslant N$, $\lambda_i \neq \lambda_j$ and $\lambda_N = 0$.
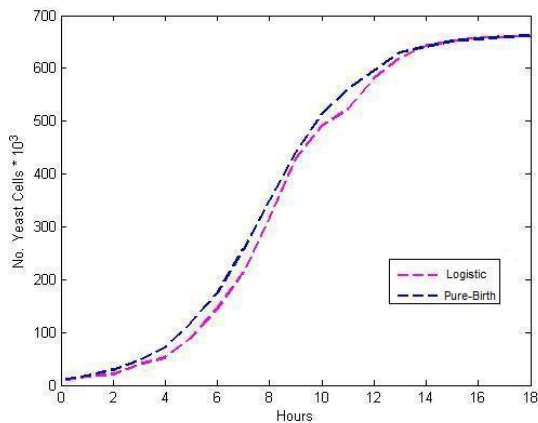

Fig. 1. Growth of Yeast Population.

The above figure shows amount of yeast cell produced against time t in hours [10]. With the choice of α and β as above, the pure-birth model closely tracks the Logistic model.

4.3. Growth of Tasmanian sheep population

As the experiments on Yeast growth were performed under ideal laboratory conditions, the assumptions underlying the derivation of the logistic model can be expected to hold true. The growth of sheep population in [12] which shows good agreement with logistic model. In Figure 2 below we compare the pure-birth model with the observed data and logistic model predication. It is fair to say pure-birth model predicts the growth of sheep population closely with other models and observed real data for the whole period. For the prediction of sheep population we have used $N = 1600,000$, $\alpha = 1$ and $\beta = 0.13$ and again, for $1 \leqslant i \neq j \leqslant N$, $\lambda_i \neq \lambda_j$ and $\lambda_N = 0$.
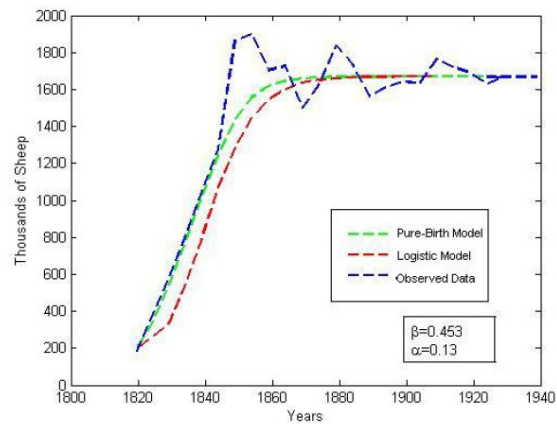

Fig. 2. Growth of Tasmanian sheep population.

4.4. Predicting color television sales from 1964-1970

So far we have shown results of growth models in population biology. However, in this subsection we would present that our growth model to predict sales of durable goods using Bass's model assumption. In Bass model we have two groups of consumers innovators and imitators. Some individuals decide to adopt an innovation independently of the decisions of other individuals in a social system. They are the innovators. Imitators or adopters are the ones that are influenced in the timing of adoption by the pressures of the social system. From [4] we have gathered data of actual sales of color televisions from 1964-1970. The total population size would be around 32 million and in the simulation results reported in Figure 3 we took $N = 31450000$, $\alpha = 1$ and $\beta = 0.28$. Moreover, for $1 \leqslant i \neq j \leqslant N$, $\lambda_i \neq \lambda_j$ and $\lambda_N = 0$. On the horizontal axis we have time in years and on the vertical axis we have total sales of the product during that period. From the figure we can see that with α and β as above, the pure-birth model very closely follows Bass model's prediction of total sales of color television from 1964-1970.
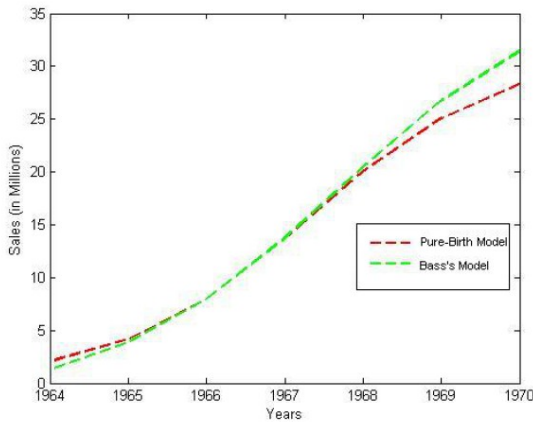
Fig. 3. Prediction of color television sales

4.5. *Software reliability growth model*

Software reliability is a critical component of computer system availability, so it is important that software customers experience a small number of software failures in their production environments. Software reliability growth models [40] can be used as an indication of the number of failures that may be encountered after the software has shipped and thus as an indication of whether the software is ready to ship. These models use system test data to predict the number of defects remaining in the software. Software reliability growth models have been applied to portions of several releases at Tandem over the past few years. One variation of our model can address and predict remaining software defects efficiently as we will show in this subsection. If we take $N = 40$, $\alpha = 0$, $\beta = 1$ then we have the following form for λ_k,

$$\lambda_k = (N - k) \tag{15}$$

It is not hard to see [2] that (15) leads to

$$P_k(t) = \binom{N-1}{k-1} \left[e^{-t}\right]^{N-k} \left[1 - e^{-t}\right]^{k-1}. \tag{16}$$

It is clear that (16) indicates that $N(t)$ is governed by negative-binomial transition probabilities, which is somewhat surprising.

Now, imagine for a certain software release the QA team predicts there are N bugs so for each state

with new discovery of bugs we would be left with one less bug. So it at when we have discovered k bugs we would be left with $N - k$ bugs. It is right to incorporate this fact in the growth model with $\lambda_k = (N - k)$. In this following simulation we compare the findings from the pure-birth model instantiated as above with the results reported in [40, 43]. The horizontal axis tracks time (in days), while the vertical axis keeps track of the number of new bugs found. So, for a particular project this graph represent number of bugs found at certain point of software testing. As shown by Figure 4, the pure-birth model instantiated as described closely tracks the actual bugs found represented in red from a particular project debugging data found at [43].
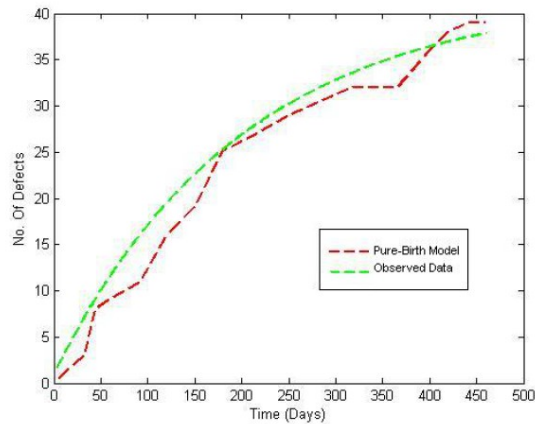

Fig. 4. Prediction of software bugs.

5. Concluding remarks and directions for future work

Growth models have captured the attention of researchers worldwide for various applications. In this paper we tried to sketch some possible growth phenomenas, tried to predict them with stochastic tools. Most of the growth phenomena are deterministic and we showed that, suitably instantiated, the well-known pure-birth model depicts a far more realistic picture of different growth phenomenas than the deterministic models. However, in this paper we have barely scratched the surface of a large field which has numerous applications not only in population bi-

ology and biomedical arena but also in vast array of social, ecological and financial phenomenas. Even in the medical field important incidents like cancer cell growth have not been touched in this paper.

In future work, we would like to understand how and why different parameters work from our model for different applications and have a better feel for the upper and lower bound of the values which depicts the characteristics of the growth. We would also like to investigate growth phenomena in geological surveys such as mine reserve estimation, estimation of ground-water reserves for aquifers. Also in population biology we would want to incorporate latent period to reproduction in our model.

References

1. R. M. Anderson and R. M. May, Infectious Diseases of Humans: Dynamics and Control,*Oxford University Press*, (1991)
2. S. Arif, PhD Thesis, Department of Computer Science, Old Dominion University, 2012, http://www.cs.odu.edu/~sarif/
3. N. T. J. Bailey, The Elements of Stochastic Processes with Applications to the Natural Sciences, John Wiley, New York, (1964)
4. F. M. Bass, A new product growth for model consumer durables, **Management Science**, **15(5)**, 215–227 (1969)
5. F. M. Bass, A new product growth for model consumer durables, *Management Science*, **50(12)**, 1825–1832, (2004)
6. F. M. Bass, Comments on "A New Product Growth for Model Consumer Durables", *Management Science*,**50(12)**, 1833–1840, (2004)
7. D. Bernoulli, Essai dune nouvelle analyse de la mortalit cause par la petite vrole., Mm. Math. Phys. Acad. Roy. Sci., Paris, 1766. (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, p. 1. (2005)
8. L. Billings and I. B. Schwartz, A unified prediction of computer virus spread in connected networks, *Physics Letters*, A, **297**, 261–262 (2005)
9. P. J. Boland and H. Singh, A birth-process approach to Moranda's geometric software-reliability model, *IEEE Transactions on Reliability*, **22(2)**, 168–174 (2005)
10. T. Carlson, Ueber Geschwindigkeit und Grsse der Hefevermehrung in Wrzei, *Biochem. Z.*, **57**, 313–334 (1913)
11. D. J. Daley and J. Gani, Epidemic Modelling: An Introduction, *Cambridge University Press*, Cambridge, (1999)
12. J. Davidson, On the ecology of the growth of the sheep population in South Australia, *Transactions of the Royal Society of Australia*, 141-148 (2005)
13. C. Doering, K. Sargsyan and L. Sander, Extinction times for birth-death processes: exact results, continuum asymptotics and the failure of the Fokker-Planck approximation, *Multiscale Modeling and Simulation*, **39(2)**, 283–299 (2005)
14. W. Feller, Die Grundlagen der Volterrashen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretiker Behandlung, *Acta Biotheoretica*, **5**, 11–40, (1939)
15. W. H. Furry, On fluctuation phenomena in the passage of high energy electrons through lead, *The Physical Review*, **52**, 569–581.
16. F. Galton, Natural Inheritance, MacMillan and Co., London, (1889)
17. B.Gompertz, On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies, *Philosophical Transactions of the Royal Society of London*, vol. **115**, 513–583 (1825)
18. T. E. Harris, The Theory of Branching Processes, *Springer-Verlag, Berlin*, (1963)
19. S. Gamito, Growth models and their use in ecological modelling: an application to a fish population, *Ecological Modelling*, **113(1-3)**, 83–94 (1998)
20. H. W. Hethcote, The mathematics of infectious diseases, *SIAM Review*, **42(4)**, 599–652 (2000)
21. C. L'Homme, C. Bizeau, J. F. Arthaud and P. Galzy, Design of a growth model for yeast, *PubMed*, **25(5)**, 418–423 (1980)
22. M. Iosifescu and P. Tautu, Stochastic processes and applications in biology and medicine, vol. 1,*Springer-Verlag, Berlin*. (1973.)
23. M. Iosifescu and P. Tautu, Stochastic processes and applications in biology and medicine, vol. **2**, *Springer-Verlag, Berlin*. (1973)
24. A. J. Lotka, Theorie Analytique des Associations Biologiques, *Actualities Scientifiques et Industrielles*, Hermann, Paris, (1939)
25. H. Jonsson and A. Levchenko, An explicit spatial model of yeast microcolony, *SIAM Society for Industrial and Applied Mathematics*,**3(2)**, 346-361 (2005)

26. M. J. Keeling, Ecology, Genetics and Evolution, *Elsevier*, New York, (2004)

27. A. Kolmogorov and N. A. Dimitriev, Branching stochastic processes, *Doklady Akademii Nauk SSSR*, **56**, 5–8, (1947)

28. T. R. Malthus, An Essay on the Principle of Population, 1803; edited with an introduction by Geoffrey Gilbert. *Oxford University Press* (1993)

29. J. D. Murray, Mathematical Biology: I. An Introduction. *Springer-Verlag*, Berlin, (2003)

30. B. Nicholas, Essential Mathematical Biology, *Springer-Verlag*, Berlin, 235–249 (2009)

31. L. A. Norton, A Gompertzian model of human breast cancer growth, *Cancer Research*, **48**, 7067–7071 (1988)

32. C. Obcemea, Chaotic Dynamics of Tumor Growth and Regeneration, *Springer-Verlag, Berlin*, 349- 354 (2006)

33. A. Okubo, Diffusion and Ecological Problems: Mathematical Models, *Springer-Verlag, Berlin*, (1980)

34. S. Olariu and J. Nickerson, A probabilistic model of integration, *Decision Support Systems*, **45(4)**, 746–763 (2008)

35. E. Parzen, Stochastic Processes, Holden-Day, Inc., San Francisco, 1962.

36. R. Pearl, The growth of populations, *Quarterly Review of Biology*, **2**, 532–548 (1927)

37. D. Pfeifer, The structure of elementary pure-birth processes, *Journal of Applied Probability*, **19(3)**, 664-667 (1982)

38. E.P. Virene, Reliability growth and its upper limit,*Proceedings of 1968 Annual Symposium on Reliability*, 265–270.

39. P. F. Verhulst, Notice sur la loi que la population poursuit dans son accroissement, *Correspondance mathmatique et physique*, **10**, 113-121 (1838)

40. A. Wood, Software Reliability Growth Models: Assumptions vs. Reality, *Eighth International Symposium on Software Reliability Engineering* , (1997)

41. H. W. Watson and F. Galton, On the probability of extinction of families, *Journal of the Antrolological Institute of Great Britain and Ireland*, **4**, 138–144 (1874)

42. A. Sharov, Quantitative population ecology online lecture notes, Department of Entomology, Virginia Tech, Blacksburg VA, `http://home.comcast.net/~sharov/PopEcol/lec5/exp.html`

43. A. Wood, Software reliability growth models, *Technical Report*, **96.1** , Tandem Computer Corporate Information Center, (1996)

44. Z. Yan, Prediction model based on Gompertz function, *Proc. 2nd IEEE International Conference on Broadband Network & Multimedia Technology*, 893-898, 18-20 Oct. (2009)

45. G. Yan, S. Rizvi and S. Olariu, A time-critical information diffusion model in vehicular ad-hoc networks, *Proc. ACM MoMM'2010*, Paris, France, November (2010)

46. G. U. Yule, A mathematical theory of evolution, based on the theory of Dr. J. C. Willis, *Philosophical Transactions of the Royal Society of London*, **213**, 21–87 (1924)