A generalized Luria–Delbrück model

A. Dewanji [a], E.G. Luebeck [b], S.H. Moolgavkar [b,*]

[a] *Indian Statistical Institute, Applied Statistics Division, 203, B.T. Road, Kolkata 700 108, India*
[b] *Fred Hutchinson Cancer Research Center, Modeling and Methods, 1100 Fairview Ave. N., M2-B500, Seattle, WA 98109-1024, USA*

Abstract

We develop extensions of the Luria–Delbrück model that explicitly consider non-exponential growth of normal cells and a birth–death process with mean exponential or Gompertz growth of mutants. Death of mutant cells can be important in clones arising during cancer progression. The use of a birth–death process for growth of mutant cells, as opposed to a pure birth process as in previous work on the Luria–Delbrück model, leads to a large increase in the extra Poisson variation in the size of the mutant cell populations, which needs to be addressed in statistical analyses. We also discuss connections with previous work on carcinogenesis models.

Keywords: Luria–Delbrück distribution; Fluctuation analysis; Gompertz growth; Birth–death process; Carcinogenesis; Mutant clones

1. Introduction

The fluctuation test proposed by Luria and Delbrück in 1943 [1] is an elegantly simple method to address a fundamental question in biology. Do mutations occur spontaneously and randomly

* Corresponding author. Tel.: +1 206 667 4145; fax: +1 206 667 7004.
E-mail addresses: dewanjia@isical.ac.in (A. Dewanji), gluebeck@fhcrc.org (E.G. Luebeck), smoolgav@fhcrc.org (S.H. Moolgavkar).

or do they occur in response to pressure in a hostile environment? The specific question considered by Luria and Delbrück concerned the appearance of phage resistant clones of bacteria. When a bacterial culture is attacked by a phage, most of the bacteria die; however a few bacteria that have acquired resistance as a result of mutation survive and give rise to clones. The question then is the following. Had the mutation that protects the bacterium arisen by chance in the bacterial culture before the phage was added or did the mutation occur directly as a result of the hostile environment created by the phage? For details of the fluctuation test addressing this specific question, the reader is referred to the pertinent literature. For a review and relevant references, see [2].

Over the years, the Luria–Delbrück problem has attracted periodic attention from a number of mathematicians and statisticians. In their original paper, Luria and Delbrück adopted a very simple probabilistic model for cell growth and the process of mutation. They assumed deterministic and exponential growth with the same rate of growth for both normal and mutant bacteria. The process of mutation was stochastic with each normal bacterium having a small probability, of the order $o(dt)$, of giving rise to a mutant in a small time interval of length dt. Under these modeling assumptions, they obtained the mean and variance of $Y(t)$, the number of mutant bacteria at time t, which is the variable of interest. The distribution of $Y(t)$ has over the years come to be known as the Luria–Delbrück distribution, and is different from the distribution obtained by assuming that mutations occurred in response to the addition of phage. In a slight generalization of the original model, Lea and Coulson [3] allowed the mutant bacteria to grow stochastically according to a linear birth process (also known as the Yule process), and derived an approximate probability generating function for $Y(t)$. The exact probability generating function for the Lea–Coulson model was first published by Armitage in 1952 [4]. In that paper, Armitage considered several variations of the assumptions of the Lea–Coulson model and derived the corresponding cumulants. Crump and Hoel [5] viewed the accumulation of mutants as a filtered Poisson process (FPP) and derived a simple formula for the probability generating function of $Y(t)$ under both deterministic and stochastic growth of the mutant bacteria. Ma et al. [6] and Sarkar et al. [7] developed a recursive formula for the probability distribution. Kepler and Oprea [8] considered inference of mutation rates based on an integral representation for the Luria–Delbrück distribution, conditional on the total size of the bacterial culture. In a second paper, Oprea and Kepler [9] suggested an approximate method of inference based on simulation of the non-Markovian case, when the waiting time to cell division may not be exponentially distributed, as is commonly assumed. In these publications (except [9]), the growth rates for normal and mutant bacteria were assumed to be equal. Zheng [2] provided an excellent and comprehensive review, and in a later paper [10] developed a software package, appropriately called SALVADOR after Salvador Luria, for computing probability distributions under various scenarios and for statistical analyses of data. In a recent paper, Natarajan et al. [11] developed a discrete time stochastic model for estimation of mutation rates in cancer progression, and illustrated its usefulness by application to data derived from a colorectal cell line.

The set of assumptions underlying the Luria–Delbrück distribution need to be generalized to make it applicable to a broader range of problems. There are many possible generalizations. We consider some of these here. First, the rates of growth of normal and mutant cells might be different. That is, even in a non-selective medium, the mutation might not be neutral, conferring either a growth advantage or disadvantage on the mutant cell. This is particularly likely with cancer cells, in which clonal evolution is thought to depend on selective growth advantage for specific mutations. Second, the growth of both normal and mutant cells may involve not only cell

divisions but also cell death. Again, in cancer cells, both cell division and cell death occur. Third, the rate parameters could be time dependent. Fourth, particularly in the case of mutations occurring in cancer, many distinct mutations might be of interest. Then the focus of analyses would be the size distribution of clones carrying distinct mutations, rather than the size distribution of a single mutant clone. Stewart et al. [12] developed a recursive method of obtaining the Luria–Delbrück distribution (conditional on the total size of the culture) in very general terms covering a broad class of Markovian situations, including the one being studied here. This general representation, however, does not explicitly provide the size distribution of individual clones, although it introduces several related concepts. Angerer [13] also attempted some generalizations incorporating cell death.

The mathematical and statistical development of a general Luria–Delbrück process has parallels in the development of a stochastic model for carcinogenesis by Moolgavkar and Venzon [14] and Moolgavkar and Knudson [15]. Specifically, expressions for the number and size distribution of intermediate lesions on the pathway to cancer, based on this model and developed by Dewanji et al. [16] and Luebeck and Moolgavkar [17], are directly relevant to the Luria–Delbrück process. In this paper, we exploit this connection to develop the mathematical theory of an extension of the Luria–Delbrück process in which cell death is explicitly considered in addition to cell division. Such a process would be particularly helpful in analyzing mutations in populations of cancer cells. We introduce the general framework in Section 2 and then discuss the FPP approach in Section 3. Section 4 works on the probability distribution of $Y(t)$ and its moments, while Section 5 investigates the effect of cell death. Section 6 focuses on the distinct mutations leading to the number and size distribution of the mutant clones. Section 7 ends with a discussion.

2. A general framework

In the most general framework, both normal and mutant cells would grow stochastically with possibly different rates for both cell division and death. The mutations would occur randomly. If both normal and mutant cells grow via linear birth–death processes with possibly different rates, then a Riccati differential equation for the joint probability generating function for normal and mutant cells can easily be derived as in [14]. Specifically, let $X(t)$ and $Y(t)$ be the number of normal and mutant cells at time t, respectively, and suppose that these populations grow according to linear birth–death processes with cell division rates α_1 and α_2 and cell death rates β_1 and β_2, respectively. Suppose further that the mutation rate per cell division is $\nu/(\alpha_1+\nu)$. Then, using an argument analogous to the one in [14] a Riccati differential equation for the joint probability generating function for $(X(t), Y(t))$, $\Psi(x,y;t)=\sum_{i,j}P_{ij}(t)x^iy^j$, where $P_{ij}(t)=Pr(X(t)=i, Y(t)=j)$, is given by the following expression

$$\Psi'(x,y;t)=\alpha_1\Psi^2(x,y;t)+\{\nu\phi(y;t)-(\alpha_1+\beta_1+\nu)\}\Psi(x,y;t)+\beta_1,$$

where $\phi(Y;t)$ is the probability generating function of the linear birth–death process with parameters α_2 and β_2 (see, e.g., [18]). This rather general framework for an extended Luria–Delbrück process is not particularly useful for data analyses. In what follows we develop a generalized Luria–Delbrück process based on a FPP approach as described in [5,16,17].

3. A filtered Poisson process approach

Let $X(t)$ denote the number of normal cells at time t described by an arbitrary deterministic growth curve. In earlier papers, $X(t)$ was restricted to be exponential. Let the random mutation rate per normal cell per unit of time at time t be denoted by $\nu(t)$. That is, mutations occur according to a Poisson process with rate $\nu(t)X(t)$. Once a mutation occurs at time s (resulting in one mutant cell and one normal cell), the mutant cell grows according to a linear birth and death process with rates $\alpha(t,s)$ and $\beta(t,s)$, respectively, depending most generally on both the current time t and the time of its origin s. In particular, the rates may depend on the difference $(t-s)$, or may be constants. Specifically, when $\nu(t)$ is a constant ν, $X(t)$ an exponential given by $X(t)=\exp(\gamma t)$, $\alpha(t,s)=\alpha$, a constant, and $\beta(t,s)=0$, the general model reduces to the model of Lea and Coulson [3].

From Luebeck and Moolgavkar [17], the probability generating function of the number of mutant cells, $Y(t,s)$, in a colony at time t originating from a single mutant arising at time s and growing according to a linear birth and death process with rates $\alpha(t,s)$ and $\beta(t,s)$, respectively, is given by

$$\phi(y;t,s)=1-\frac{y-1}{(y-1)G(t,s)-g(t,s)}, \tag{1}$$

with the initial condition $\phi(y;s,s)=y$, where

$$g(t,s)=\exp\left[-\int_s^t(\alpha(u,s)-\beta(u,s))\,\mathrm{d}u\right], \quad \text{and} \tag{2}$$

$$G(t,s)=\int_s^t \alpha(u,s)g(u,s)\,\mathrm{d}u. \tag{3}$$

In the special case when the birth and death rates are constants α and β, respectively, the probability generating function reduces to the well-known probability generating function for the linear birth–death process with constant parameters (e.g., [18]).

$$\phi(y;t,s)=\frac{\beta(1-y)-(\beta-\alpha y)\mathrm{e}^{-(\alpha-\beta)(t-s)}}{\alpha(1-y)-(\beta-\alpha y)\mathrm{e}^{-(\alpha-\beta)(t-s)}}. \tag{4}$$

For the linear birth process (with $\beta=0$), used in the Lea and Coulson [3] model (referred to as LC model hereafter), the above probability generating function reduces further to

$$\phi(y;t,s)=\frac{y\mathrm{e}^{-\alpha(t-s)}}{1-y(1-\mathrm{e}^{-\alpha(t-s)})}. \tag{5}$$

Note that, even in the most general case as described above, $Y(t)$ can be written as a filtered Poisson process

$$Y(t)=\sum_{i=1}^{M(t)}Y(t,s_i),$$

where $M(t)$ denotes the number of Poisson mutations (from normal bacteria) by time t with rate $v(s)X(s)$ at times s. From the corresponding theory (see, e.g., [19]), we have the probability generating function of $Y(t)$ as

$$\psi(y;t) = \exp\left\{\int_0^t v(s)X(s)[\phi(y;t,s)-1]\,\mathrm{d}s\right\}. \tag{6}$$

The integral in (6) does not simplify in general; but for the special case when $Y(t,s)$ is described by a linear birth process (with $\beta = 0$), using (5), we have

$$\psi(y;t) = \exp\left\{-\int_0^t v(s)X(s)\,\mathrm{d}s + \sum_{k=1}^{\infty} q_k(t)y^k\right\}, \tag{7}$$

where

$$q_k(t) = \int_0^t \left[1-\mathrm{e}^{-\alpha(t-s)}\right]^{k-1}\mathrm{e}^{-\alpha(t-s)}v(s)X(s)\,\mathrm{d}s.$$

See Eq. (49) of [2] for a special case of this when $v(s) = v$ and $X(s) = \exp(\gamma s)$, the LC model. This representation (7) of the probability generating function is useful to obtain the probability distribution of $Y(t)$ using the recursive formula in Lemma 2 of [2]. A similar recursive formula can be derived in the most general case and used to compute the probability distribution of $Y(t)$. This will be described in the following section.

4. Probability distribution of $Y(t)$

Recall that the coefficients of the powers of y in the expansion of $\psi(y;t)$ represent the probability distribution of $Y(t)$. Thus, we get $P_0(t) = P[Y(t)=0] = \psi(0;t)$ and, for $m \geqslant 1$,

$$P_m(t) = P[Y(t)=m] = \frac{1}{m!}\psi^{(m)}(0;t), \tag{8}$$

where $\psi^{(m)}(0;t)$ is $\psi^{(m)}(y;t)$, the mth order partial derivative of $\psi(y;t)$ with respect to y, evaluated at $y=0$. Using (6), we get

$$\begin{aligned} P_0(t) &= \exp\left\{\int_0^t v(s)X(s)[\phi(0;t,s)-1]\,\mathrm{d}s\right\} \\ &= \exp\left\{-\int_0^t v(s)X(s)[G(t,s)+g(t,s)]^{-1}\,\mathrm{d}s\right\}, \quad \text{using (1)}, \\ &= \exp\{-\Lambda(t)\}, \end{aligned}$$

where $\Lambda(t)$ is the expected number of bacterial clones at time t (see Section 6). Note that $\phi(0;t,s)$ is the probability of extinction at time t for the mutant clone $Y(t,s)$ originating at time s. A simple expression for this probability is available when $Y(t,s)$ grows according to a linear birth and death process with constant rates (see, e.g., [18]).

For the probability distribution of $Y(t,s)$, we need the mth order partial derivative of $\phi(y;t,s)$ with respect to y, denoted by $\phi^{(m)}(y;t,s)$ as before, for $m \geqslant 1$. We have

$$P[Y(t,s)=m] = \frac{1}{m!}\phi^{(m)}(0;t,s) = \frac{1}{G(t,s)}H(t,s)[1-H(t,s)]^m, \tag{9}$$

where $H(t,s) = g(t,s)[G(t,s) + g(t,s)]^{-1}$. See [17] for a similar result. Therefore, $\phi^{(m)}(0;t,s)$ can be obtained in closed form as

$$\frac{m!}{G(t,s)} H(t,s)[1 - H(t,s)]^m.$$

Write $\xi(y;t) = \int_0^t v(s)X(s)\phi(y;t,s)\,ds$ so that the mth order partial derivative of $\xi(y;t)$ with respect to y can be written as $\xi^{(m)}(y;t) = \int_0^t v(s)X(s)\phi^{(m)}(y;t,s)\,ds$. Then, by simple calculation of successive derivatives of both sides of (6) with respect to y and using the method of induction, one can prove the following Lemma.

Lemma. *For* $m \geqslant 1$,

$$\psi^{(m)}(y;t) = \sum_{i=0}^{m-1} \binom{m-1}{i} \psi^{(i)}(y;t)\xi^{(m-i)}(y;t),$$

where $\psi^{(0)}(y;t) = \psi(y;t)$.

From (8), we need $\psi^{(m)}(0;t)$'s for the probability distribution of $Y(t)$. For each m, $\xi^{(m)}(0;t)$ can be obtained from (9) as

$$m! \int_0^t v(s)X(s) \frac{1}{G(t,s)} H(t,s)[1 - H(t,s)]^m\,ds.$$

These $\xi^{(m)}(0;t)$'s are closely related to the probability distribution of individual clone sizes (see Section 6 for more detail), as in Eq. (10) of [17]. From the above Lemma and using (8), the probability distribution of $Y(t)$ in general can be obtained by the recursive formula

$$P_m(t) = \sum_{i=0}^{m-1} \frac{m-i}{m} P_i(t)p_{m-i}(t), \tag{10}$$

where $p_k(t) = \int_0^t v(s)X(s)G^{-1}(t,s)H(t,s)[1 - H(t,s)]^k\,ds$, which is proportional to the probability that an individual mutant clone is of size k at time t. This $p_k(t)$ is identical to $q_k(t)$ in (7) for the special case mentioned there. The integral in $p_k(t)$ can be analytically worked out for a large class of growth parameters (called Gompertz growth; see Section 6) but with at most piecewise constant $v(t)$ and $X(t)$ (see [20]). For the LC model, the above Lemma yields Lemma 2 of [2]. The probability distribution of $Y(t)$ under the LC model, given by Eqs. (46), (47) and Lemma 2 of [2], can also be obtained by using the Lemma above. See also [7] for a similar recursive result. One can also obtain the commonly available recursive formulae for cumulants and moments of different orders, by using the above Lemma, under the most general case.

It is useful to derive the mean and variance of $Y(t)$ in the most general case. Note that $E[Y(t)] = \psi^{(1)}(1;t)$ and $E[Y(t)(Y(t)-1)] = \psi^{(2)}(1;t)$. Using the above Lemma, we have

$$\begin{aligned} E[Y(t)] = \psi^{(1)}(1;t) = \xi^{(1)}(1;t) &= \int_0^t v(s)X(s)\phi^{(1)}(1;t,s)\,ds \\ &= \int_0^t v(s)X(s)g^{-1}(t,s)\,ds, \quad \text{using (1)} \\ &= \int_0^t v(s)X(s)\exp\left[\int_s^t (\alpha(u,s) - \beta(u,s))\,du\right] ds, \quad \text{using (2).} \end{aligned} \tag{11}$$

Similarly, one can calculate the variance of $Y(t)$ as

$$V[Y(t)] = \psi^{(2)}(1;t) + E[Y(t)] - E^2[Y(t)] = 2\int_0^t v(s)X(s)G(t,s)g^{-2}(t,s)\,ds + E[Y(t)]$$
$$= \int_0^t v(s)X(s)[2G(t,s) + g(t,s)]g^{-2}(t,s)\,ds. \tag{12}$$

In the special case when the growth of normal bacteria $X(t)$ is described by the exponential curve $\exp(\gamma t)$, $v(t) = v$, $\alpha(t,s) = \alpha$ and $\beta(t,s) = \beta$ (LC model with death of mutants, referred to as LC-D model hereafter), the mean and variance of $Y(t)$ are, using (11) and (12),

$$E[Y(t)] = \frac{ve^{(\alpha-\beta)t}}{\gamma - (\alpha-\beta)}\left[e^{(\gamma-(\alpha-\beta))t} - 1\right], \tag{13}$$

and

$$V[Y(t)] = E[Y(t)] + \frac{2v\alpha}{\alpha-\beta}\left\{\frac{e^{2(\alpha-\beta)t}}{\gamma - 2(\alpha-\beta)}\left[e^{(\gamma-2(\alpha-\beta))t} - 1\right] - \frac{e^{(\alpha-\beta)t}}{\gamma-(\alpha-\beta)}\left[e^{(\gamma-(\alpha-\beta))t} - 1\right]\right\}. \tag{14}$$

As expected, the mean depends on α and β only through the growth rate $(\alpha - \beta)$ of the mutant bacteria. However, there is considerable extra-Poisson variation, as evidenced in (14). In fact this extra variation depends on the birth rate α. For $(\alpha - \beta)$ fixed, the extra-Poisson variation increases with α (see the next section for more detail).

For the LC model, the above mean and variance reduce to

$$E[Y(t)] = \frac{ve^{\alpha t}}{\gamma - \alpha}\left[e^{(\gamma-\alpha)t} - 1\right] \tag{15}$$

and

$$V[Y(t)] = \frac{ve^{\alpha t}}{(\gamma-2\alpha)(\gamma-\alpha)}\left\{\gamma\left[e^{(\gamma-\alpha)t} - 2e^{\alpha t} + 1\right] + 2\alpha[e^{\alpha t} - 1]\right\}, \tag{16}$$

which are the same as those in Eqs. (52) and (53) of [2].

5. Effect of incorporating death of mutant cells

In this section we consider incorporation of the death of mutant cells and compare the resulting distributions with those generated by the Lea–Coulson model (cell death is zero). The probability distributions of $Y(t)$, given by (10), for the LC-D model, are shown in Fig. 1. The effective cell division rate, $(\alpha - \beta)$, is kept fixed. The parameter values for the LC model are the same as those in Fig. 1 of [2]. With $\alpha - \beta$ kept fixed at 2.5, Fig. 1 shows four distributions corresponding to $\alpha = 100$, 10, 5, and $\alpha = 2.5$, which corresponds to the LC model. With positive death rates, the mode is shifted to the left when compared with the LC model. With larger values of α and β, the spread is larger, as expected. However, as seen from (13), since $(\alpha - \beta)$ is fixed (at 2.5) in all three distributions, the distributions have the same mean.

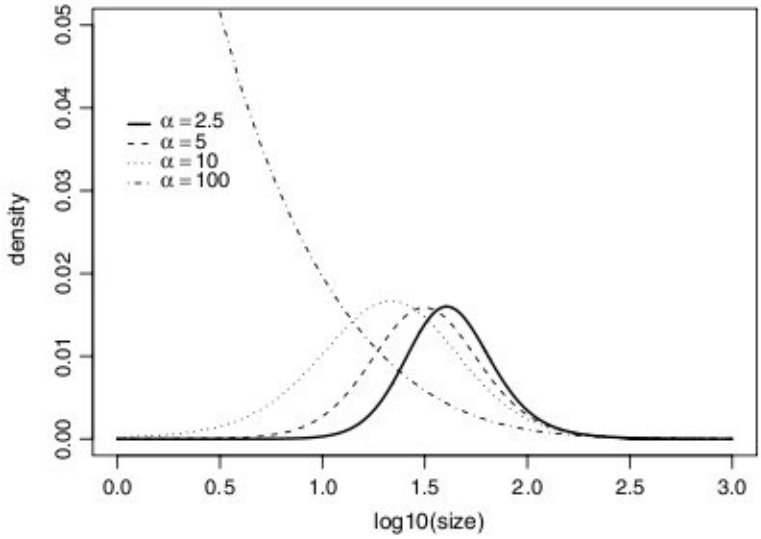

Fig. 1. Probability distribution of $Y(t)$ for $t = 6.7$, $\gamma = 3$, and $\alpha - \beta = 2.5$, for different values of α as indicated. The solid line ($\alpha = 2.5$) corresponds to the LC model, also shown in Fig. 1 of [2].

In Fig. 2, we plot $P[Y(t) = 0]$, the probability of no mutant bacteria at time t, against time t for the LC-D model. As expected, for a given t, this probability is larger with positive death rates than for the LC model. Moreover, it is larger with larger β values (death rate), even though $(\alpha - \beta)$ is fixed. This is also quite intuitive by noting that a larger β implies a larger asymptotic probability of extinction for any one mutant colony, which is given by β/α. As a consequence, the expected number of non-extinct mutant clones (given by $A(t)$ in the next section) also decreases.

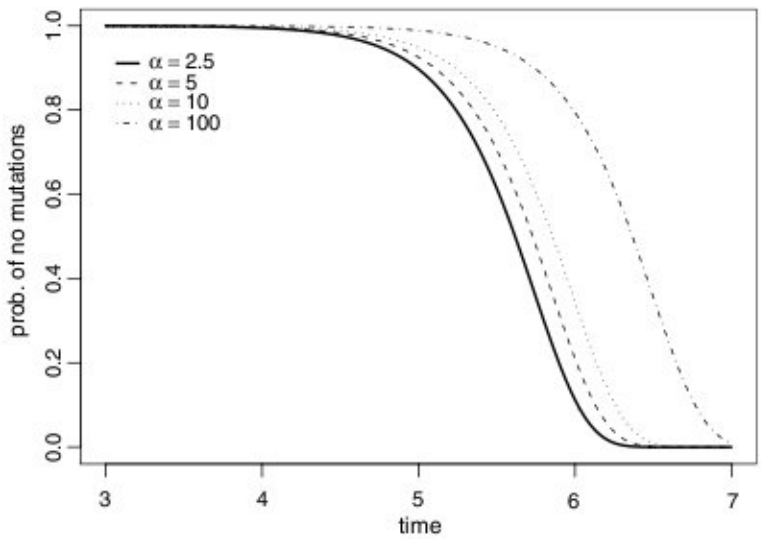

Fig. 2. Probability of no mutants, $P_0(t)$, as a function of time t for the same parameter values as used for Fig. 1.

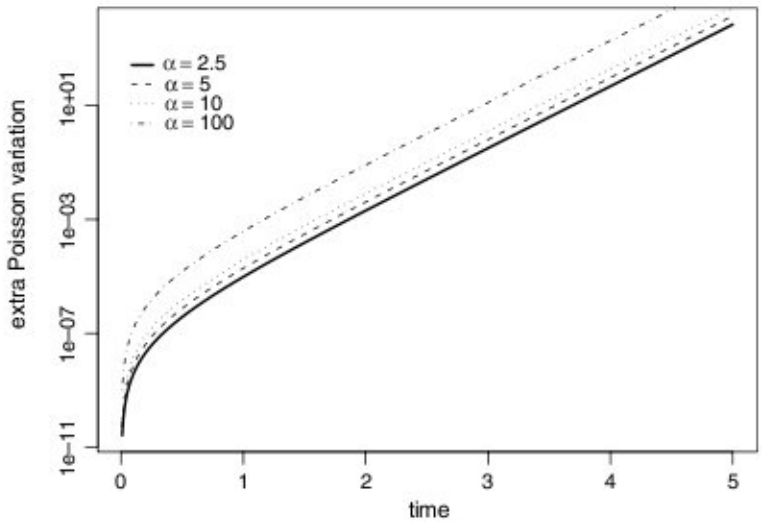

Fig. 3. Extra-Poisson variation, $V[Y(t)] - E[Y(t)]$, as a function of time t for the same parameter values as used for Fig. 1.

It may be noticed, from (14), that the extra-Poisson variation $V[Y(t)] - E[Y(t)]$ has a factor $\alpha/(\alpha - \beta)$, which is unity in the LC model and is otherwise greater than one (assuming $\alpha > \beta$). Therefore, for a fixed $(\alpha - \beta)$, the extra-Poisson variation increases with increasing α and β. This phenomenon is illustrated in Fig. 3 in which the extra-Poisson variation is plotted against time t for various values of α with $\alpha - \beta$ fixed. The parameters used in Fig. 3 are identical to those in Fig. 2. As expected, the extra variation increases with time and α. The presence of extra-Poisson variation has been explicit in the Luria–Delbrück distribution. However, the fact that its magnitude increases with time and birth rate (in the presence of cell death) of mutant cell, with fixed growth rate, has not been recognized before. This recognition is critical for appropriate statistical analyses.

6. Number and size distribution of individual clones of mutant cells

The classical Luria–Delbrück problem addresses the distribution of the total number of mutants in the colony of cells. As mentioned in Section 1, there is scope for the Luria–Delbrück distribution to be applicable in a wider range of applications. In some of them, information may be available on the surviving progeny of a single mutant arising from normal cells. Then this information should be used to estimate rates of mutation and of cell division and cell death. Such information is available, for example, in rodent hepatocarcinogenesis experiments. Moreover, fluctuation tests could be designed to specifically exploit this information. The requisite mathematical theory has been developed in a couple of papers (see [16,17]). We briefly review it here.

For the mathematical development, we follow Luebeck and Moolgavkar [17]. Note that, given that a mutation occurs at time s, the probability that the corresponding clone is extinct at time $t > s$ is given by $\phi(0; t, s)$, which turns out to be $1 - [G(t,s) + g(t,s)]^{-1}$, from (1). From Luebeck

and Moolgavkar [17], the number of non-extinct mutant clones at time t, denoted by $N(t)$, say, follows a Poisson distribution with mean $\Lambda(t) = \int_0^t \nu(s)X(s)[G(t,s) + g(t,s)]^{-1}\,ds$. For the LC-D model, $\Lambda(t)$ reduces to

$$\begin{aligned}\Lambda(t) &= \frac{\nu(\alpha-\beta)}{\alpha}\int_0^t e^{\gamma s}\left[1-\frac{\beta}{\alpha}e^{-(\alpha-\beta)(t-s)}\right]^{-1} ds \\ &= \nu\left(1-\frac{\beta}{\alpha}\right)\sum_{i=0}^{\infty}\left(\frac{\beta}{\alpha}\right)^i \frac{e^{-i(\alpha-\beta)t}}{i(\alpha-\beta)+\gamma}\left[e^{(i(\alpha-\beta)+\gamma)t}-1\right].\end{aligned}$$

When the death rate is zero, i.e., with LC model, we have

$$\Lambda(t) = \frac{\nu}{\gamma}[e^{\gamma t}-1].$$

Let us denote the sizes of the $N(t)$ non-extinct mutant clones by $W_i(t)$, $i = 1,\ldots,N(t)$. From Luebeck and Moolgavkar [17], the $W_i(t)$'s are independent and identically distributed with the probability distribution given by $p_k(t)/\Lambda(t)$, $k = 1,2,\ldots$, where $p_k(t)$ is as given after Eq. (10). For LC and LC-D models, the $p_k(t)$ reduces to

$$p_k(t) = \nu e^{-\alpha t}\int_0^t e^{(\alpha+\gamma)s}\left[1-e^{-\alpha(t-s)}\right]^{k-1} ds$$

(see Eq. (47) of [2]) and

$$p_k(t) = \nu\left(1-\frac{\beta}{\alpha}\right)^2 e^{-(\alpha-\beta)t}\int_0^t e^{(\alpha-\beta+\gamma)s}\frac{\left[1-e^{-(\alpha-\beta)(t-s)}\right]^{k-1}}{\left[1-\frac{\beta}{\alpha}e^{-(\alpha-\beta)(t-s)}\right]^{k+1}} ds,$$

respectively. These two forms of $p_k(t)$ have been used in (10) to obtain the probability distribution plots of Fig. 1.

A further generalization is possible to reflect the eventual slowing down of the growth of mutant clones as the total population approaches the carrying capacity of the environment. In that case, the growth of mutant colonies is better described by a Gompertz rather than an exponential growth curve. Then a stochastic Gompertz birth–death process [21] can be used to model $Y(t,s)$. In this model, the net growth rate $\alpha(t,s) - \beta(t,s)$ is modeled as

$$\alpha(t,s) - \beta(t,s) = b\exp[-a(t-s)] = \delta(t-s), \quad \text{say},$$

where $\delta(0) = b \geqslant 0$ and a is the shape parameter with $-\infty < a < \infty$. Taking $a = 0$, we recover the exponential model with net proliferation rate b.

After modeling the growth rate $\alpha(t,s) - \beta(t,s)$, it becomes natural to work with the ratio $\beta(t,s)/\alpha(t,s)$, which is the asymptotic extinction probability for a linear birth and death process with constant rates. For algebraic and computational simplicity, we assume this ratio to be a constant (η, say) independent of both t and s. We can then recover the birth and death rates as

$$\alpha(t,s) = \alpha(t-s) = \frac{\delta(t-s)}{1-\eta}$$

and

$$\beta(t,s) = \beta(t-s) = \frac{\eta\delta(t-s)}{1-\eta}.$$

We can then use the results of Sections 3 and 4 with these expressions for $\alpha(t,s)$ and $\beta(t,s)$ and obtain the quantities of interest. Most of these expressions are available in Luebeck and Moolgavkar [17]. We however give here expressions for the two important quantities:

$$\Lambda(t) = \int_0^t v(s)X(s)\frac{1-\eta}{1-\eta g(t-s)}\,\mathrm{d}s$$

and

$$p_k(t) = (1-\eta)\int_0^t v(s)X(s)\frac{g(t-s)}{1-\eta g(t-s)}\left(\frac{1-g(t-s)}{1-\eta g(t-s)}\right)^k \mathrm{d}s,$$

with

$$g(t,s) = g(t-s) = \exp\left[-\frac{b}{a}(1-\mathrm{e}^{-a(t-s)})\right].$$

7. Discussion

In this paper, we have provided the mathematical theory for a generalization of the Luria–Delbrück process to arbitrary (but still deterministic) growth of normal cells and with mutant cells growing according to a linear birth–death process with either mean exponential or Gompertz growth. A completely stochastic model to include stochastic growth of normal cells appears to be mathematically cumbersome and not particularly useful at the moment. We have also derived expressions for the number and size distribution of individual mutant clones following Dewanji et al. [16] and Luebeck and Moolgavkar [17]. These expressions are needed in analyses of data in which information on individual clones is available as is often the case with experimental carcinogenesis studies, in which the focus is on understanding early lesions on the pathway to cancer. Likelihood methods for such analyses using results on the number and the size distribution of individual clones are given in a number of publications [22–24]. In general, one can hope to have observations from a number of bacterial cultures, or tumor-bearing animals, at different time points. Using the theory developed in this paper, likelihood methods for inference on the model parameters are readily applicable to such data.

It is clear from the derivations in previous sections that similar results can be obtained, at least in theory, when the birth and death rates differ from those of exponential and Gompertz growth, as long as the Markovian assumptions are valid. More generally, as long as the mutations occur according to a Poisson process, if the growth of an individual clone $Y(t,s)$ can be described by a probability distribution $p(n;t,s) = p[Y(t,s) = n]$ (mimicking the notation of Stewart et al. [12]), Markovian or not, the FPP approach of Section 3 goes through. However, we need the assumption that the individual clones, given their mutation times, grow independently of each other.

With $p(n;t,s)$ replacing the right hand side of (9), the $p_k(t)$ in (10) have the form $p_k(t) = \int_0^t v(s)X(s)p(k;t,s)\,\mathrm{d}s$ and $\Lambda(t)$ has the form $\int_0^t v(s)X(s)(1-p(0;t,s))\,\mathrm{d}s$. Hence, this FPP approach, which is mathematically simpler, is capable of covering the most general situations including those of Stewart et al. [12] and Oprea and Kepler [9]. This approach has the added capability of considering other observables, like number and size distribution of individual clones, as described in Section 6. Even the consideration of different types of mutations can be easily accommodated in the FPP approach simply by introducing the corresponding mutation rates, leading to a generalization of the formulation in Stewart et al. [12].

When both X and v are constant, the distribution of sizes of mutant clones is described by a logarithmic series distribution [16]. The total number of mutant cells is the Poisson sum of individual clone sizes and has a negative binomial distribution. This version of the Luria–Delbrück problem is formally identical to a problem on the relative abundance of species of insects addressed by Fisher et al. in the 1940s [25].

Acknowledgments

The authors acknowledge support from NIH Grant R01CA047658.

References

[1] S.E. Luria, M. Delbrück, Mutations of bacteria from virus sensitivity to virus resistance, Genetics 28 (1943) 491.
[2] Q. Zheng, Progress of a half century in the study of the Luria–Delbrück distribution, Math. Biosci. 162 (1999) 1.
[3] D.E. Lea, C.A. Coulson, The distribution of the numbers of mutants in bacterial populations, J. Genet. 49 (1949) 264.
[4] P. Armitage, The statistical theory of bacterial populations subject to mutation, J. Roy. Stat. Soc. B 14 (1952) 1.
[5] K.S. Crump, D.G. Hoel, Mathematical models for estimating mutation rates in cell populations, Biometrika 61 (1974) 237.
[6] W.T. Ma, G.v.H. Sandri, S. Sarkar, Analysis of the Luria–Delbrück distribution using discrete convolution powers, J. Appl. Prob. 29 (1992) 255.
[7] S. Sarkar, W.T. Ma, G.v.H. Sandri, On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants, Genetica 85 (1992) 173.
[8] T.B. Keplar, M. Oprea, Improved inference of mutation rates: I. An integral representation for the Luria–Delbrück distribution, Theor. Popul. Biol. 59 (2001) 41.
[9] M. Oprea, T.B. Kepler, Improved inference of mutation rates: II. Generalization of the Luria–Delbrück distribution for realistic cell-cycle time distributions, Theor. Popul. Biol. 59 (2001) 49.
[10] Q. Zheng, Statistical and algorithmic methods for fluctuation analysis with SALVADOR as an implementation, Math. Biosci. 176 (2002) 237.
[11] L. Natarajan, C.C. Berry, C. Gasche, Estimation of spontaneous mutation rates, Biometrics 59 (2003) 555.
[12] F.M. Stewart, D.M. Gordon, B.R. Levin, Fluctuation analysis: the probability distribution of the number of mutants under different conditions, Genetics 124 (1990) 175.
[13] W.P. Angerer, An explicit representation of the Luria–Delbrück distribution, J. Math. Biol. 42 (2001) 145.
[14] S.H. Moolgavkar, D.J. Venzon, Two event models for carcinogenesis: incidence curves for childhood and adult tumors, Math. Biosci. 47 (1979) 55.
[15] S.H. Moolgavkar, A. Knudson, Mutation and cancer: a model for human carcinogenesis, J. Natl. Canc. Inst. 66 (1981) 1037.

[16] A. Dewanji, D.J. Venzon, S.H. Moolgavkar, A stochastic two-stage model for cancer risk assessment. The number and size of premalignant clones, Risk Anal. 9 (1989) 179.
[17] E.G. Luebeck, S.H. Moolgavkar, Stochastic analysis of intermediate lesions in carcinogenesis experiments, Risk Anal. 11 (1991) 149.
[18] N.T.J. Bailey, The Elements of Stochastic Processes with Applications to the Natural Sciences, Wiley, New York, 1964.
[19] E. Parzen, Stochastic Processes, Holden-Day, San Francisco, 1962.
[20] M.C.M. de Gunst, A. Dewanji, E.G. Luebeck, Exploring heterogeneity in tumor data using Markov chain Monte Carlo, Stat. Med. 22 (2003) 1691.
[21] W.Y. Tan, A stochastic Gompertz birth–death process, Stat. Prob. Lett. 4 (1986) 25.
[22] S.H. Moolgavkar, E.G. Luebeck, M. de Gunst, R.E. Port, M. Schwarz, Quantitative analysis of enzyme-altered foci in rat hepatocarcinogenesis experiments – I. Single agent regimen, Carcinogenesis 11 (1990) 1271.
[23] S.H. Moolgavkar, E.G. Luebeck, E.G. Buchmann, K.W. Boch, Quantitative analysis of enzyme-altered foci initiated with diethylnitrosamine and promoted with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin or 1,2,3,4,6,7,8-heptachlorodibenzo-*p*-dioxin, Toxicol. Appl. Pharmacol. 138 (1996) 31.
[24] E.G. Luebeck, A. Buchmann, S. Stinchcombe, S.H. Moolgavkar, M. Schwarz, Effects of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin on initiation and promotion of GST-P-positive foci in rat liver: a quantitative analysis using a stochastic model, Toxicol. Appl. Pharmacol. 167 (2000) 63.
[25] R.A. Fisher, A.S. Corbet, C.B. Williams, The relation between the number of species and the number of individuals in a random sample of an animal population, J. Anim. Ecol. 12 (1943) 42.