



**HAL**  
open science

# General formulation of Luria-Delbrück distribution of the number of mutants

Bahram Houchmandzadeh

► **To cite this version:**

Bahram Houchmandzadeh. General formulation of Luria-Delbrück distribution of the number of mutants. 2015. hal-01148693v1

**HAL Id: hal-01148693**

**<https://hal.science/hal-01148693v1>**

Preprint submitted on 5 May 2015 (v1), last revised 16 Jul 2015 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# General formulation of Luria-Delbrück distribution of the number of mutants.

Bahram Houchmandzadeh  
CNRS, LIPHY, F-38000 Grenoble, France  
Univ. Grenoble Alpes, LIPHY,  
F-38000 Grenoble, France

The Luria-Delbrück experiment is a cornerstone of evolutionary theory, demonstrating the randomness of mutations before selection. The distribution of the number of mutants in this experiment has been the subject of intense investigation during the last 70 years. Despite this considerable effort, most of the results have been obtained under the assumption of constant growth rate, which is far from the experimental condition. We derive here the properties of this distribution for arbitrary growth function, for both the deterministic and stochastic growth of the mutants. The derivation we propose is surprisingly simple and versatile, allowing many generalizations to be taken easily into account.

## I. INTRODUCTION.

The Darwin-Wallace theory of evolution rests upon mutation of living organisms and their selection. In their seminal article [1], Luria and Delbrück (LD) described an experiment demonstrating the randomness of mutational events *before* the selection process. The experiment consists of growing  $C$  cultures of bacteria in parallel in identical environments, beginning with a small number  $N_0$  (typically  $10^3$ ) in each batch. After a sufficient growth period, the cultures saturate and the number of bacteria reaches  $N$  (typically  $10^9 - 10^{10}$ ). Each culture is then tested against an antibacterial agent, a phage virus in the LD case, and the number of surviving bacteria arising from mutation in the cultures is counted by a plating method. If  $C$  is large, the probability  $P(m)$  of having  $m$  mutants can be experimentally determined; in practice,  $C$  cannot be large and therefore only statistical quantities such as the mean and the variance of the number of mutants can be estimated.

The LD experiment has spurred a large interest and many authors have developed increasingly refined models to estimate statistical properties of the random variable  $X$  of the number of mutants, such as its cumulant/probability/moment generating functions (cgf, pgf, mgf), from which the mutation rate or probability can be estimated. The pioneering authors were Lea and Coulson[2], Armitage [3], Bartlett[4], Crump and Hoel[5], Mandelbrot[6], Sarkar, Ma and Sandri[7] who set the LD distribution on a solid mathematical ground, generalized the model to take into account stochastic growth of the mutant and the wild type, and developed algorithms to estimate the mutation rate. The works of these and other authors have been reviewed in an elegant article by Zheng[8] which also contains original results and corrections of some of the errors contained in the previous works. These investigations have been extended during the last 15 years by authors such as Angerer[9], Dewanji et al[10] and Ycart[11]. A description of some of these more recent works will be given in the following sections.

The fundamental LD experiment is now currently used to estimate mutation rates in various setups such as an-

tibiotic resistance or experimental investigation of the evolutionary process[12–14].

However, nearly all of the existing computations have focused on the exponential growth (either deterministic or stochastic) of the WT and mutant bacteria, although Dewanji et al.[10] have extended these results to Gompertz growth. The reason behind this choice is that in these formulations, an explicit expression for the number of wild type (WT) cells as a function of time is needed in order to compute the statistical properties of the number of mutants.

The assumption of constant growth rate is however too restrictive. Experimentally, the growth is never exponential but follows a Monod curve[15] : the growth rate is not constant, but begins with a value close to zero (called the lag phase), increases gradually to a maximum value and then decreases as the number of bacteria increases, to finally reach zero when the culture is saturated. Various functions (logistic, Gompertz, Richards, Stannard, ...) are used in the literature to model the growth curve and their relevance has been studied in depth by Zwietering et al. [16].

The *real time* however is not the relevant independent variable in terms of which the system may be described. The WT population grows from an initial number of cells  $N_0$  to reach a final value  $N$ . Each time a WT cell divides, there is a small probability  $\nu$  that a mutant having the desired trait (phage or antibiotic resistance) appears. It does not matter how much time the system spends between WT population size  $n$  and  $n + 1$ , but only the fact that once a division has taken place, a mutation may appear. For the mutants growing in the same environment as the WT, their growth curve will be similar (but not necessarily equal) to that of the WT. The only quantities that are indeed measured in an experiment are  $N_0$ ,  $N$  and the number of mutants. Even though the growth curve can theoretically be measured, its determination is cumbersome and, as we will see, not relevant.

In this paper, we shall use the WT population size  $n$  as the independent variable. It appears that this formulation of the LD distribution is surprisingly simple and applies to any growth curve for the bacteria, including of course exponential growth. The formulation is versatile

and can take into account many generalizations, some of which are considered in this article.

This article is organized as follows: in the next section, we present the basic concept for the simple case of deterministic growth of both WT and mutant bacteria, where only mutations apparitions are random. We present some generalizations, such as different growth rates for WT and mutants and non-constant mutation probability. In the following section, we generalize the model to the case of stochastic growth of the mutant, where we consider (i) a linear birth process for the mutant and (ii) a random relative growth rate for the mutant. We stress that in the following computations, the growth rate is not constant, but can have an arbitrary form. The last section is devoted to a discussion of possible extensions of this work and conclusions. An appendix, containing a straightforward mathematical derivation is included in order to make this article self sufficient.

## II. DETERMINISTIC MODEL.

### A. Equal growth of WT and mutant.

Consider a culture of bacteria growing from size  $N_0$  to size  $N$ . The growth curve can be as general as possible assuming that no death event takes place. Let  $X_n$  be the random variable describing the occurrence of a mutant when the WT population increases from  $n$  to  $n + 1$ . Throughout this paper, we use the term mutant to designate an individual which acquire a trait (e.g. phage or antibody resistance) which will be tested once the growth is stopped.

Denoting the mutation *probability* by  $\nu$ ,

$$\Pr(X_n = 0) = 1 - \nu ; \Pr(X_n = 1) = \nu \quad (1)$$

This may seem to be an approximate description, because if during a cell division, a mutation has occurred, obviously the number of wild type cells has not increased from  $n$  to  $n + 1$ . A more precise formulation would be  $\Pr(X_n = k) = (1 - \nu)\nu^k$ , *i.e.*, the number of mutants when the WT population increases by one unit is geometrically distributed. However, as  $\nu \ll 1$  ( $\nu$  is usually of the order of  $10^{-8}$ ), we will use the relation (1) to describe the random variable  $X_n$ . The generalization to the geometric distribution of  $X_n$  is straightforward (see appendix 1). Note that most formulations of LD distribution use the above approximation.

We assume in this subsection that both mutant and the wild type population follow a deterministic, equal growth. We do not assume the growth rate to be a constant. As the mutants are similar to wild types, a mutant appearing in one copy at WT population size  $n$  will contribute  $N/n$  to the number of mutants when the wild type population reaches size  $N$  (figure 1). In other words, the proportion of the number of this mutant to the number of WT population will remain constant.

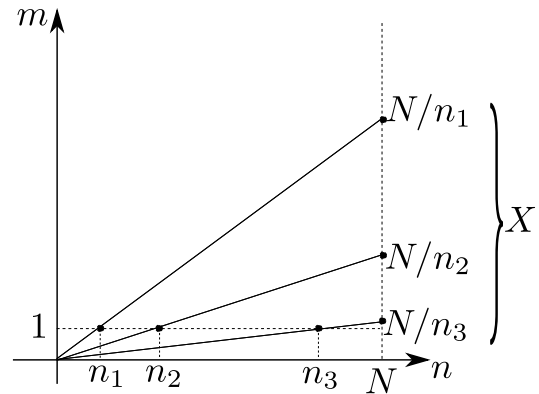


Figure 1: The number of mutants  $X$  at population size  $N$  is the sum of the contributions of mutants appearing at population size  $n_i$ . The  $i$ -th mutant, appearing at size 1 at WT population  $n_i$  will be present at size  $N/n_i$  in the final population.

Let  $Y_n^N = (N/n)X_n$  be the contribution of this mutant to the final number of mutants  $X$ , when the WT population reaches size  $N$ . Then

$$X = \sum_{n=N_0}^N Y_n^N$$

As the mutant occurrences are independent random variables, the moment (mgf) and cumulant generating functions (cgf) are

$$\begin{aligned} \phi(s) &= \langle e^{sX} \rangle = \prod_{n=N_0}^N \langle e^{sY_n^N} \rangle \\ \psi(s) &= \log(\langle e^{sX} \rangle) = \sum_{n=N_0}^N \log(\langle e^{sY_n^N} \rangle) \end{aligned}$$

On the other hand, by its very definition,

$$\langle e^{sY_n^N} \rangle = 1 - \nu + \nu e^{sN/n}$$

which gives the cgf as

$$\begin{aligned} \psi(s) &= \sum_{n=N_0}^N \log(1 - \nu + \nu e^{sN/n}) \\ &= N \int_{x_0}^1 \log(1 - \nu + \nu e^{s/x}) dx \quad (2) \end{aligned}$$

where in the second expression, we have used the continuous approximation for the sum,  $x = n/N$  and  $x_0 = N_0/N$ .

Note that this derivation is analogous to the filtered Poisson process derivation used when the problem is formulated in real time ([8], eq. 14). However, because the problem is formulated in terms of WT population size, the propagator is simply a straight line regardless of the growth rate function (figure 1).

Expanding the expression (2) to the first order in  $\nu$  and restricting the domain of definition to  $s \lesssim -x_0 \log \nu$ ,

$$\psi(s) = -\theta\phi + \theta \int_{x_0}^1 e^{s/x} dx + O(\nu^2) \quad (3)$$

Where  $\theta = N\nu$  and  $\phi = 1 - x_0$ . For the particular case of exponential growth, the expression (3) for the cgf has been obtained by Crump and Hoel [5] and in closed form by Zheng ([8],eq.14). Note that for Zheng,  $N = \exp(\beta t)$ , *i.e.* the initial number of WT bacteria is 1.

The first two cumulant coefficients  $\kappa_p = \psi^{(p)}(0)$  are then

$$\kappa_1 = -\theta \log x_0 \quad (4)$$

$$\kappa_2 = \frac{\theta\phi}{x_0} \quad (5)$$

These expressions for the average ( $\kappa_1$ ) and the variance ( $\kappa_2$ ) have been obtained originally by LD for the exponential growth of bacteria [1]. As we see here, the hypothesis of exponential growth is superfluous and the expressions (4-5) are valid for arbitrary growth curves. To the leading order in  $\nu$  and  $x_0$ , the general expression for cumulant coefficients is

$$\kappa_{p+1} = \frac{\theta}{px_0^p} \quad p > 0 \quad (6)$$

These expressions are known for the special case of exponential growth ([8],eq.9 for equal growth rate).

### B. Different growth of WT and mutant.

Let us now consider the case where WT ( $n$ ) and mutants ( $m$ ) (once they have appeared) have similar but different growth rates:

$$\frac{dn}{dt} = \alpha(n, t)ndt ; \quad \frac{dm}{dt} = c\alpha(n, t)mdt$$

where  $c$  is a constant. We do not specify any particular form for the growth rate, but we suppose that the mutant follows the same law as the WT, within a constant multiplicative factor. This is the case for example where the resources are depleted by the growth of the bacteria, and the mutant is inferior to the WT for its duplication.

Time can be eliminated between the above equations:  $dm/dn = c(m/n)$ . A mutant appearing at one copy when the population size is  $n$  will contribute  $(N/n)^c$  to the final number of mutants (figure 2). The computation of the preceding section can be repeated and leads to

$$\psi(s) = N \int_{x_0}^1 \log \left( 1 - \nu + \nu e^{s/x^c} \right) dx$$

Keeping only the leading term in  $\nu$  and  $x_0$ , for  $c \neq 1/p$ , the  $p$ -th cumulant coefficient is given by

$$\kappa_p = \frac{\theta}{cp-1} (x_0^{1-cp} - 1)$$

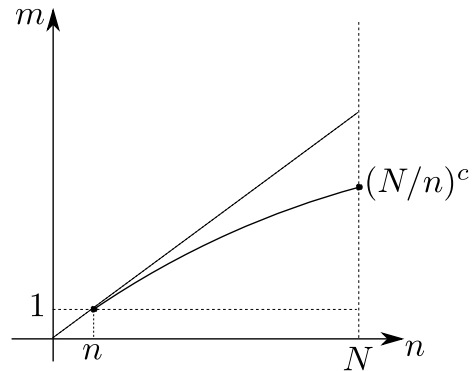


Figure 2: Contribution of mutants with different growth rate.

The case  $c = 1/p$  can be recovered from the above formula by taking the limiting value for  $c \rightarrow 1/p$  and reads

$$\kappa_p = -\theta \log x_0$$

For the particular case of exponential growth ( $\alpha = Cte$ ), this is the expression given by Zheng ([8],eq.9).

### C. Variable mutation probability.

In most models the mutation probability  $\nu$  is considered to be a constant and independent of time, *i.e.*, population size. This is a sound hypothesis when the mutation involves only point-mutations on the chromosome. However, traits such as virus or antibiotic resistance may involve many point mutations before the trait is functional. Bacteria at the end of the growth process, having achieved more divisions, may be more prone to mutate to the given trait than bacteria at the early stage of the growth. A crude approximation of the above phenomena will be a mutation rate that depends on the population size  $\nu = \nu(n)$ . The formulation for the number of mutants of the previous section does not suppose a constant rate of mutation and the relation 2 for  $\psi(s)$  remains valid. The first two cumulants are given by

$$\kappa_1 = N \int_{x_0}^1 \frac{\nu(x)}{x} dx$$

$$\kappa_2 = N \int_{x_0}^1 \frac{\nu(x) - \nu^2(x)}{x^2} dx$$

## III. STOCHASTIC GROWTH OF MUTANTS.

### A. General discussion.

Until now, we have considered the deterministic propagation of the mutant from its appearance at population size  $n$  to its final value at population size  $N$ . We will denote the propagator as  $K_n^N$ . The contribution of the

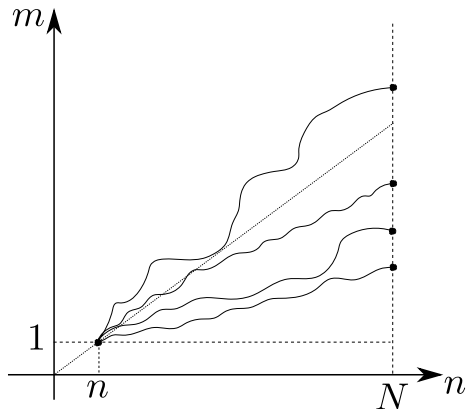


Figure 3: The stochastic propagator  $K_n^N$  of the mutant appearing at WT population size  $n$

mutant appearing at WT population size  $n$  to the final population  $N$  was expressed as

$$Y_n^N = X_n K_n^N$$

For the deterministic case,  $K_n^N = N/n$  and the mgf of  $Y_n^N$  was simply

$$\langle e^{sY_n^N} \rangle = 1 - \nu + \nu e^{sN/n}$$

We will now consider for the mutant a stochastic propagator  $K_n^N$  (figure 3). Because  $X_n$  takes only the values 0 or 1,

$$\langle e^{sY_n^N} \rangle = \langle e^{sX_n K_n^N} \rangle = 1 - \nu + \nu \langle e^{sK_n^N} \rangle \quad (7)$$

(see appendix 2). Therefore, all the discussion of the preceding section naturally generalizes to stochastic propagation and the cgf for the number of mutants at population size  $N$  is

$$\psi(s) = \sum_{n=N_0}^N \log \left( 1 - \nu + \nu \langle e^{sK_n^N} \rangle \right). \quad (8)$$

Knowing the statistical properties of the propagator gives access directly to the statistical properties of the total number of mutants. Before applying this concept to specific cases, let us compute the first two cumulant coefficients:

$$\kappa_1 = \nu \sum_{n=N_0}^N \langle K_n^N \rangle \quad (9)$$

$$\kappa_2 = \nu \sum_{n=N_0}^N \left( \langle (K_n^N)^2 \rangle - \nu^2 \sum_{n=N_0}^N \langle K_n^N \rangle^2 \right) \quad (10)$$

The mean  $\kappa_1$  is what we already had in the deterministic case, where  $\langle K_n^N \rangle = N/n$ . Let us express the second

moment of  $K_n^N$  as a function of its mean and variance  $V_n^N$

$$\langle (K_n^N)^2 \rangle = V_n^N + \langle K_n^N \rangle^2$$

.Then

$$\kappa_2 = \nu(1 - \nu) \sum_{n=N_0}^N \langle K_n^N \rangle^2 + \nu \sum_{n=N_0}^N V_n^N$$

The first term on the RHS of the above relation is what we already had in the case of deterministic growth. The second term is the contribution of the stochasticity of the propagator to the variance of the number of mutants at population size  $N$ .

## B. Linear birth process.

Consider the case where the growth of the WT is deterministic and continuous

$$\frac{dn}{dt} = \alpha(n, t)n$$

while the mutant, once it has appeared, follows a stochastic growth with transition probability density

$$\tilde{W}(m \rightarrow m+1) = \alpha(n, t)m$$

where  $\tilde{W}(m \rightarrow m+1)dt$  is the probability that this lineage of the mutant has increased its size by one unit in the interval  $[t, t+dt]$ . This model was first introduced by Lea and Coulson[2] for exponential growth case  $\alpha(n, t) = \alpha$ .

A note of caution should be made here. Although widely used, the linear birth model may not be very realistic, as bacterial division times are not exponentially distributed. Indeed, after a division, a bacterium needs to elongate again to its original size before being able to divide again, so that the next division time cannot be smaller than a finite time  $\tau$ . In fact, the distribution of division times around the time  $\tau$  is fairly narrow and the division process is much less random than a linear birth process. The phenomenon has been experimentally investigated by a microfluidic device by Wang et al [17] ; the overestimation of the mutation rate by a linear birth model, for the exponential growth case, has been investigated by Ycart [11].

Let us now come back to the linear birth model. As in the previous section, the real time is not the best choice of independent variable and we can write the stochastic growth of the mutant in terms of WT population size: noting  $W(m \rightarrow m+1)dn$  the probability that a mutant has divided when the WT population size  $\in [n, n+dn]$ , we have

$$W(m \rightarrow m+1) = \tilde{W}(m \rightarrow m+1) \frac{dt}{dn} = \frac{m}{n}$$

The master equation governing the growth of the mutant is

$$\frac{\partial P(m; n)}{\partial n} = \frac{W(m-1 \rightarrow m)P(m-1; n) - W(m \rightarrow m+1)P(m; n)}{W(m \rightarrow m+1)P(m; n)} \quad (11)$$

and the mgf of the propagator  $K_n^N$  is given by (see appendix 3):

$$\langle e^{sK_n^N} \rangle = \frac{ne^s}{(n-N)e^s + N} \quad (12)$$

Using now the expression (8) and the continuous variable  $x = n/N$ , we obtain the cumulant generating function of the number of mutants

$$\psi(s) = N \int_{x_0}^1 \log \left( 1 - \nu + \frac{\nu e^{sx}}{e^s(x-1) + 1} \right) dx \quad (13)$$

Note that the above integral can be expressed in an analytical, albeit cumbersome, form. The expressions for the two first cumulant coefficients are

$$\begin{aligned} \kappa_1 &= -\theta \log x_0 \\ \kappa_2 &= \frac{\theta \phi}{x_0} (2 - \nu) + \theta \log x_0 \\ &\approx 2 \frac{\theta}{x_0} \end{aligned}$$

Where as before,  $\theta = N\nu$  and  $\phi = 1 - x_0$ . The variance of the number of the mutants is now approximately twice what we had for the deterministic case. The exponential growth case can be recovered from the above expression and is equal to the expressions given by Lea and Coulson and Zheng ([8], eq.52-53).

Other cumulant coefficients can be readily recovered by multiple derivation of expression (13). Restricting the computations to the leading order of  $\nu$  and  $x_0$ , the expression for the cumulant coefficients is (see appendix 4):

$$\kappa_p = \theta \frac{p!}{(p-1)x_0^{p-1}} \quad p \geq 2 \quad (14)$$

Note that even in the case of exponential growth, no general expression for the cumulant coefficients could be obtained by classical methods ([8]). Comparing the above expression with the deterministic case where  $\kappa_p = N\nu/(p-1)x_0^{p-1}$ , we see that indeed the linear birth process induces large amplification of the  $p$ -th cumulant coefficient by a factor of  $p!$ .

*The probabilities.* To compute the probabilities, it is more advantageous to use the probability generating function (pgf)

$$G(z) = \langle z^X \rangle = e^{\psi(z)}$$

where  $\psi(z)$  is defined from (13) by setting  $z = e^s$ , *i.e.*:

$$\psi(z) = N \int_{x_0}^1 \log \left( 1 - \nu + \frac{\nu z x}{z(x-1) + 1} \right) dx \quad (15)$$

Approximating the above expression to the first order in  $\nu$ , we have

$$\psi(z) = \theta \left( \frac{1}{z} - 1 \right) \log(1 - \phi z) + O(\nu^2) \quad (16)$$

where  $\theta = N\nu$  and  $\phi = 1 - x_0$ . We therefore obtain a simple expression for the probability generating function

$$G(z) = (1 - \phi z)^{\theta(1/z-1)} \quad (17)$$

For the case of exponential growth, the above expression (without the  $\phi$  factor) was first discovered by Lea and Coulson[2]; omitting the  $\phi$  factor however results in divergent moments. The correct expression can be seen in the Zheng review ([8], eq 65). We stress again that relation (17) is very general and does not depend on the assumption of exponential growth. Note that in the case of exponential growth,  $\theta \sim \exp(\beta t)$  and all the moments diverge as  $t \rightarrow \infty$ . This divergence, discussed by Bartlett and later by Zheng [8], cannot be cured within the framework of the exponential growth model. No such divergence exists in the present formulation, as the WT population size, following any realistic growth curve, will remain finite.

In order to evaluate the probabilities, we have first to compute  $\psi^{(p)}(0)$ . Expanding expression (16) in powers of  $z$  we have

$$\psi(z) = -\theta \phi + \theta \sum_{n=1}^{\infty} \left( \frac{\phi^n}{n} - \frac{\phi^{n+1}}{n+1} \right) z^n$$

Keeping only the first leading terms in  $\theta$  and  $x_0$  we have

$$\psi(z=0) = -\theta \phi \quad (18)$$

$$\frac{1}{p!} \psi^{(p)}(z=0) = \frac{\theta}{p(p+1)} - \frac{\theta x_0^2}{2} \quad (19)$$

where the second term in  $x_0^2$  can be neglected for  $p \ll 1/x_0$  and  $\theta \phi \approx \theta$ . We can now use the Faà Di Bruno Formula [18] to compute the  $k$ -th derivative of  $G(z)$  at  $z = 0$  and obtain the probabilities (see appendix 5):

$$\begin{aligned} P(k) &= \frac{1}{k!} G^{(k)}(0) \\ &= e^{-\theta} \sum_{\{m_j\}} \prod_{j=1}^k \frac{1}{(m_j)!} \left( \frac{\theta}{j(j+1)} \right)^{m_j} \end{aligned}$$

where the sum is taken over all  $n$ -tuples  $\{m_j\}$  such that  $\sum_{j=1}^k j m_j = k$ . The above formula can be easily programmed to compute numerically the probabilities. The only linear term in  $\theta$  in the above formula is for  $m_k = 1$ ,  $m_{i < k} = 0$ . Therefore, for  $\theta \ll 1$ , the probabilities take the simple form of

$$\begin{aligned} P(0) &= e^{-\theta} \\ P(k) &= e^{-\theta} \frac{\theta}{k(k+1)} \quad k > 0 \end{aligned}$$

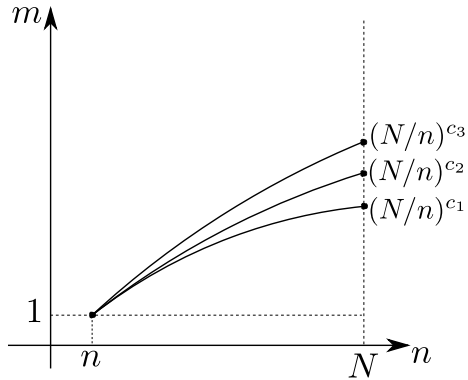


Figure 4: Deterministic growth, random relative growth rate.

The explicit expression for the probabilities in the linear birth model has been intensely investigated by many authors, and reviewed by Zheng ([8], 5.3) and the above expressions are known for the constant linear birth model.

Experimentally, obtaining the probabilities implies a very large number of parallel cultures and the above expressions may not be of great practical use.

*Different growth rate.* The discussion of the preceding subsection can be extended to take into account a different relative growth rate for the mutant compared to the wild type:

$$\frac{dn}{dt} = \alpha(n, t)n$$

$$\tilde{W}(m \rightarrow m + 1) = c\alpha(n, t)m$$

Repeating the discussion of the preceding sections, the cumulant generating function in this case is (see appendix 3)

$$\psi(s) = N \int_{x_0}^1 \log \left( 1 - \nu + \nu \frac{x^c e^s}{(x^c - 1)e^s + 1} \right) \quad (20)$$

from which the cumulant coefficients can be deduced as before.

### C. random relative growth rate.

The traditional formulation of LD distribution assumes that all mutants are similar in their growth function. As many different mutations can bring a bacteria to the same phage resistance, this assumption may seem too restrictive and can be easily relaxed. For example, a comprehensive study of this phenomenon has recently been published [19] where the growth rate of *all* mutants in the gene TEM-1, conferring resistance to the antibiotic cefotaxime, were measured and shown to be variable in some conditions.

Consider the case where the growth of both mutants and WT are deterministic as in subsection II B

$$\frac{dn}{dt} = \alpha(n, t)ndt ; \quad \frac{dm}{dt} = c\alpha(n, t)mdt$$

but the relative growth rate  $c$  is now a random variable (figure 4): when it appears, a mutant picks a relative growth rate  $c$  from a given distribution, which is transmitted to its progeny.

Following the discussion of subsection III A, the propagator this time is

$$K_n^N = (N/n)^c$$

where  $c$  is now a random variable. Let us denote  $\rho(s)$  with its moment generating function

$$\rho(s) = \langle e^{sc} \rangle$$

Then according to relations (9-10), the first two cumulants are now, to the first order in  $\nu$ :

$$\kappa_1 = \theta \int_0^{-\log x_0} \rho(z) e^{-z} dz$$

$$\kappa_2 = \theta \int_0^{-\log x_0} \rho(2z) e^{-z} dz$$

For example, if  $c$  follows a Normal distribution  $c = \mathcal{N}(\mu, \sigma)$ , then  $\rho(s) = \exp(\mu s + (1/2)\sigma^2 s^2)$  and these expressions can be evaluated in terms of the error function. For the specific case of  $\mu = 1$  and  $\sigma \ll 1$ , restricting the computation to the leading orders of  $\nu$  and  $x_0$ , we have

$$\kappa_1 = -\theta \log x_0 \left( 1 + \frac{\sigma^2}{6} (\log x_0)^2 \right)$$

$$\kappa_2 = \frac{\theta}{x_0} (1 + 2\sigma^2 (1 + (1 + \log x_0)^2))$$

## IV. DISCUSSION AND CONCLUSION.

In the previous sections we have given the general solution of the LD distribution through the derivation of its cumulant generating function. The key point to this derivation is to change the independent variable from time to WT population size, which is indeed the relevant variable: whatever the growth function, a mutation can occur only when a WT cell divides and WT population size changes from  $n$  to  $n + 1$ . This consideration considerably simplifies the solution of the problem and allows us to extend the solution to arbitrary growth curves for the WT population. The mathematical formulation applies equally well to the case of deterministic and stochastic growth.

This mathematical formulation is sufficiently simple to allow for many generalizations, some of which have been considered in this article: for example variable mutation probability (subsection II C) or random relative growth rate (subsection III C) have been investigated.

Other generalizations that we have not developed can be considered. For example, for the stochastic growth case, we have only considered the linear birth process. Other, more realistic cases can be envisaged where the

distribution of the division times are not exponential, along the lines developed by Ycart[11]. One can also consider the case where both mutants and WT grow stochastically. These generalizations would be straightforward if the moment generating function of the propagator  $K_n^N$  can be derived in explicit form. Another possible generalization would be to take into account the experimental uncertainty on the initial and final value of the WT population, and its influence of the estimation of mutation probability, as has been considered by Ycart and Veziris[20].

To summarize, we have developed in this article a versatile method for investigating the Luria Delbrück distribution with an arbitrary growth function. The method uses only very few measurable parameters, namely the initial and final number of the WT population. We believe that the method we propose here can be used as a simple basis for further investigations of the LD distribution.

## Appendix: Mathematical details.

### 1. Geometric distribution of mutants appearing at WT population size $n$ .

We have noted  $\nu$  the probability of apparition of a mutant when a wild type cell divides, and  $X_n$  the random variable tracking the number of mutants appearing when WT population changes its size from  $n$  to  $n + 1$ . The probability of producing  $k$  mutant during this change is geometrically distributed:

$$\Pr(X_n = k) = (1 - \nu)\nu^k$$

As  $\nu \ll 1$ , we have approximated  $X_n$  by a binary process ( $X_n = 0, 1$ ) in the article. This constraint can be relaxed. Following notations of subsection II A,

$$\langle e^{sY_n^N} \rangle = \frac{1 - \nu}{1 - \nu e^{sN/n}}$$

and

$$\begin{aligned} \psi(s) &= \sum_{n=N_0}^N \log \left( \langle e^{sY_n^N} \rangle \right) \\ &= N \int_{x_0}^1 \log \left( \frac{1 - \nu}{1 - \nu e^{s/x}} \right) dx \\ &= -\theta\phi - N \int_{x_0}^1 \log \left( 1 - \nu e^{s/x} \right) dx \end{aligned}$$

where  $\theta = -N \log(1 - \nu)$  and  $\phi = 1 - x_0$ . The above expression is equal to expression (2) to the first order in  $\nu$ .

### 2. Moment generating function of $Y_n^N$ in the stochastic case.

Consider the random variable  $Y_n^N = X_n K_n^N$  where  $X_n$  is a boolean variable  $P(X_n = 0) = 1 - \nu$  and  $P(X_n = 1) = \nu$  (subsection III A); For simplicity, we suppose that  $K_n^N$  is a positive discrete random variable. Then,

$$\begin{aligned} P(Y_n^N = 0) &= 1 - \nu + \nu P(K_n^N = 0) \\ P(Y_n^N = k \neq 0) &= \nu P(K_n^N = k) \end{aligned}$$

The mgf is then

$$\begin{aligned} \langle e^{sY_n} \rangle &= \sum_{k=0}^{\infty} P(Y_n = k) e^{sk} \\ &= 1 - \nu + \nu P(K_n^N = 0) + \\ &\quad \sum_{k=1}^{\infty} \nu P(K_n^N = k) e^{sk} \\ &= 1 - \nu + \nu \langle e^{sK_n^N} \rangle \end{aligned}$$

which is the expression (7).

### 3. Moment generating function of the propagator $K_n^N$ .

A mutant appearing in one copy when the WT population size is  $n_0$  will reach size  $K_{n_0}^n$  when the WT population reaches size  $n$ . The forward master equation governing the probabilities of the propagator,  $\Pr(K_{n_0}^n = m)$  is given by equation (11). The mgf of the propagator,  $\phi(s, n)$  therefore obeys the following equation :

$$n \frac{\partial \phi}{\partial n} + (1 - e^s) \frac{\partial \phi}{\partial s} = 0 \quad (\text{A.1})$$

with the boundary conditions

$$\begin{aligned} \phi(s, n = n_0) &= e^s \\ \phi(s = 0, n) &= 1 \end{aligned}$$

Equation (A.1) is a linear first order partial differential equation that can be solved by the method of characteristics:

$$\phi(s, n) = \frac{(n_0/n)e^s}{(n_0/n - 1)e^s + 1}$$

Changing now the notation to denote  $n$  as the WT population size of the mutant occurrence, and  $N$  the final size of the WT population, we obtain the expression (12).

If the relative growth rate of the mutant is not 1 but  $c$ , where  $c$  is an arbitrary constant, the transition probability for the mutant once it has appeared is

$$W(m \rightarrow m + 1) = c \frac{m}{n}$$



and  $\phi(s, n)$  obeys the following equation:

$$\frac{1}{c}n\frac{\partial\phi}{\partial n} + (1 - e^s)\frac{\partial\phi}{\partial s} = 0 \quad (\text{A.2})$$

This equation can be transformed into equation (A.1) by a simple scaling  $n \rightarrow n^c$  and the mgf is therefore given by

$$\phi(s, n) = \frac{(n_0/n)^c e^s}{((n_0/n)^c - 1)e^s + 1}$$

from which the cumulant generating function of the number of mutants can be deduced.

#### 4. Cumulant coefficients for the stochastic growth.

To the first order in  $\nu$ , the cgf for the linear birth model (subsection III B) is given by

$$\frac{1}{\theta}\psi(s) = (e^{-s} - 1) \log(1 - e^s + x_0 e^s)$$

Expanding the above function into powers of  $(1 - e^{-s})$ , we have

$$\psi(s) = (s + \log x_0)(e^{-s} - 1) + \sum_{n=2}^{\infty} \frac{1}{n-1} \frac{(1 - e^{-s})^n}{x_0^{n-1}}$$

Evaluating the  $p$ -th derivative at  $s = 0$ , we have

$$\psi^p(0) = \frac{p!}{(p-1)x_0^{p-1}} + O\left(\frac{1}{x_0^{p-2}}\right)$$

which is the expression given in equation (14).

#### 5. Probabilities.

The Faà Di Bruno Formula for chain derivation of the function  $f(g(z))$  is [18]

$$\frac{1}{k!} \frac{d^k}{dz^k} f(g(z)) = \sum_{\{m_j\}} f^{(\ell)}(g(z)) \prod_{j=1}^k \frac{1}{m_j!} \left( \frac{g^{(j)}(z)}{j!} \right)^{m_j}$$

where the sum is taken over all  $\{m_j\}$  such that

$$\sum_{j=1}^n j m_j = n$$

and  $k = \sum_{j=1}^n m_j$ . In the present case,  $g(z) = \psi(z)$  and  $f(u) = \exp(u)$ . Specifically,  $g(0) = -N\nu = -\theta$ , and therefore  $f^{(\ell)}(g(0)) = \exp(-\theta)$ . On the other hand, according to relation (19),

$$\frac{g^{(j)}(0)}{j!} = \frac{\theta}{j(j+1)}$$

*Acknowledgements.* I sincerely thank Marcel Vallade, Olivier Rivoire and Erik Geissler for the critical reading of the manuscript.

- 
- [1] S E Luria and M Delbrück. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics*, 28(6):491–511, November 1943.
- [2] D. E. Lea and C. A. Coulson. The distribution of the numbers of mutants in bacterial populations. *Journal of Genetics*, 49(3):264–285, December 1949.
- [3] P. Armitage. The statistical theory of bacterial populations subject to mutations. *J. Roy. Stat.Soc. B*, 14:1–33, 1952.
- [4] M. S. Bartlett. *An introduction to stochastic processes*. 1955.
- [5] Kenny S. Crump and David G. Hoel. Mathematical models for estimating mutation rates in cell populations. *Biometrika*, 61(2):237–252, August 1974.
- [6] B. Mandelbrot. A population birth and mutation process I. *J. Applied Probability*, 1974:437, 1974.
- [7] S. Sarkar, W. T. Ma, and G. v. H. Sandri. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica*, 85(2):173–179, 1992.
- [8] Qi Zheng. Progress of a half century in the study of the Luria-Delbrück distribution. *Mathematical Biosciences*, 162(1-2):1–32, November 1999.
- [9] Wolfgang P. Angerer. An explicit representation of the Luria-Delbrück distribution. *Journal of Mathematical Biology*, 42(2):145–174, February 2001.
- [10] A Dewanji, E G Luebeck, and S H Moolgavkar. A generalized Luria-Delbrück model. *Mathematical biosciences*, 197(2):140–52, October 2005.
- [11] Bernard Ycart. Fluctuation analysis: can estimates be trusted? *PloS one*, 8(12):e80958, January 2013.
- [12] W A Rosche and P L Foster. Determining mutation rates in bacterial populations. *Methods (San Diego, Calif.)*, 20(1):4–17, January 2000.
- [13] P D Sniegowski, P J Gerrish, and R E Lenski. Evolution of high mutation rates in experimental populations of E. coli. *Nature*, 387(6634):703–5, June 1997.
- [14] Csaba Pal, María D Maciá, Antonio Oliver, Ira Schachar, and Angus Buckling. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, 450(7172):1079–81, December 2007.
- [15] Jacques Monod. The Growth of Bacterial Cultures. *Annual review of microbiology*, 3:371–394, 1949.
- [16] M H Zwietering, I Jongenburger, F M Rombouts, and

- K van 't Riet. Modeling of the bacterial growth curve. *Applied and environmental microbiology*, 56(6):1875–81, June 1990.
- [17] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of *Escherichia coli*. *Current biology : CB*, 20(12):1099–103, June 2010.
- [18] Frank W J Olver, Daniel W Lozier, Ronald F Boisvert, and Charles W Clark. *NIST handbook of mathematical functions*. Cambridge University Press, 2010.
- [19] Michael A. Stiffler, Doeke R. Hekstra, and Rama Ranganathan. Evolvability as a Function of Purifying Selection in TEM-1  $\beta$ -Lactamase. *Cell*, 160(5):882–892, February 2015.
- [20] Bernard Ycart and Nicolas Veziris. Unbiased estimation of mutation rates under fluctuating final counts. *PloS one*, 9(7):e101434, January 2014.