Electronic Journal of Statistics
Vol. 6 (2012) 1251–1272
ISSN: 1935-7524
DOI: 10.1214/12-EJS711

Statistics for the Luria-Delbrück distribution

Agnès Hamon and Bernard Ycart

Laboratoire Jean Kuntzmann
Université de Grenoble and CNRS
51 rue des Mathématiques 38041 Grenoble cedex 9, France

Abstract: The Luria-Delbrück distribution is a classical model of mutations in cell kinetics. It is obtained as a limit when the probability of mutation tends to zero and the number of divisions to infinity. It can be interpreted as a compound Poisson distribution (for the number of mutations) of exponential mixtures (for the developing time of mutant clones) of geometric distributions (for the number of cells produced by a mutant clone in a given time). The probabilistic interpretation, and a rigourous proof of convergence in the general case, are deduced from classical results on Bellman-Harris branching processes. The two parameters of the Luria-Delbrück distribution are the expected number of mutations, which is the parameter of interest, and the relative fitness of normal cells compared to mutants, which is the heavy tail exponent. Both can be simultaneously estimated by the maximum likehood method. However, the computation becomes numerically unstable when the maximal value of the sample is large, which occurs frequently due to the heavy tail property. Based on the empirical probability generating function, robust estimators are proposed and their asymptotic variance is given. They are comparable in precision to maximum likelihood estimators, with a much broader range of calculability, a better numerical stability, and a negligible computing time.

AMS 2000 subject classifications: Primary 92D25; secondary 60J28.
Keywords and phrases: Luria-Delbrück distribution, fluctuation analysis, Bellman-Harris branching process, probability generating function estimator.

Received March 2012.

1. Introduction

Luria and Delbrück (1943) reported an experiment on virus resistant bacteria: cultures of the same strain of *Escherichia Coli* having grown up independently for several generations, the cells were plated onto selective medium and surviving bacteria counted. The major feature of the data was a surprisingly high fluctuation, with frequent appearance of large counts. This experiment, adapted since on many different types of cell cultures, founded *fluctuation analysis*, whose objective is to estimate probabilities of mutations. Mathematical models were introduced by Lea and Coulson (1949), and Bartlett in the discussion following Armitage (1952): see chap. II p. 59 of Kendall (1952) for an early review, and Zheng (1999, 2010) for more recent ones. Classical modeling hypotheses are the following:

- at time 0 a homogeneous culture of n normal cells is given;
- the generation time of any normal cell is a random variable with distribution function G;
- when a division of a normal cell occurs, it is replaced by:
  - one normal and one mutant cell with probability p,
  - two normal cells with probability $1-p$;
- the generation time of any mutant cell is exponentially distributed with parameter μ;
- when a division of a mutant cell occurs, it is replaced by two mutant cells;
- all random variables and events (division times and mutations) are mutually independent.

In Kendall's notations (Kendall, 1952, p. 61), the model considered here is the G/M/0 (general distribution for normal cells, Markovian evolution of mutants, no phenotypic lag). The particular case where generation times of normal cells follow the exponential distribution (M/M/0 or "fully stochastic" model), appeared for the first time in the discussion following (Armitage, 1952, p. 37). There Bartlett obtained the asymptotics in the equal growth rates case ($\mu = \nu$), for large n and t and small p; he later generalized to differential growth rates (Bartlett, 1978, p. 134). In between, Mandelbrot (1974) had proposed similar asymptotics, but the paper had several errors (pointed out by Pakes (1993) and Zheng (2002)), and no mathematical proof nor interpretation was given (see Zheng (2008)). Several extensions of the M/M/0 case have been proposed, notably by Oprea and Kepler (2001) and Angerer (2010). Yet, to the best of our knowledge, no rigourous proof of convergence, and no clear probabilistic interpretation has been given for the G/M/0 model.

Our first objective is to establish the convergence in distribution of the number of mutants in the general case, and give a probabilistic interpretation of the result, based on the theory of branching processes. To state our main result, we need to introduce the growth rate ν and the proportionality constant C associated to the distribution G, for a binary division dynamics in a Bellmann-Harris process (Harris (1963); Athreya and Ney (1972)). The growth rate ν (also called *Malthusian parameter*) is defined as the unique root of the equation:

$$2\int_0^{+\infty} \mathrm{e}^{-\nu s}\,\mathrm{d}G(s) = 1\,. \tag{1.1}$$

The proportionality constant C is:

$$C = \left(4\int_0^{+\infty} s\mathrm{e}^{-\nu s}\,\mathrm{d}G(s)\right)^{-1}\,. \tag{1.2}$$

Theorem 1.1. *Consider the model G/M/0 described above. Let $p = p_n$ and $t = t_n$ be two sequences, and α a positive real such that:*

$$\lim_{n\to\infty} p_n = 0\,, \quad \lim_{n\to\infty} t_n = +\infty\,, \quad \lim_{n\to\infty} p_n n C \mathrm{e}^{\nu t_n} = \alpha\,.$$

As n tends to $+\infty$, the distribution of the number of mutants at time t_n, starting with n normal cells at time 0, converges to the distribution with probability generating function:

$$g_{\alpha,\rho}(z) = \exp(\alpha(h_\rho(z) - 1)) \,, \tag{1.3}$$

where h_ρ is the probability generating function of the Yule distribution with parameter $\rho = \nu/\mu$:

$$h_\rho(z) = \rho z \int_0^1 \frac{v^\rho}{1 - z + zv} \, \mathrm{d}v \,. \tag{1.4}$$

Following Zheng (2002) we call the distribution described by (1.3) and (1.4) the *Luria-Delbrück distribution* with parameters α and ρ and we denote it by $\mathrm{LD}(\alpha, \rho)$. Observe that it depends on G (the generation time distribution of normal cells) only through its growth rate ν, thus matching the conclusions of Jones, Wheldrake and Rogers (1993). The parameter α is the mean number of mutations and $\rho = \nu/\mu$ is the relative *fitness* of normal cells compared to mutants. Theorem 1.1 has a very simple interpretation that can be summarized in 3 points (precise justifications from the theory of branching processes will be given in section 2):

1. the number of divisions of normal cells before time t_n is equivalent in probability to $nC\mathrm{e}^{\nu t_n}$. Therefore the expected number of mutations that can lead to mutants at time t_n is equivalent to $p_n \times nC\mathrm{e}^{\nu t_n} \simeq \alpha$. Since the number of divisions is large and the probability of mutation is small, the total number of mutations approximately follows the Poisson distribution with parameter α, by the law of small numbers (this remark had already been made by (Luria and Delbrück, 1943, p. 499));
2. mutations happen at random instants, but due to exponential growth, the vast majority of divisions occur rather close to the end of the observation interval. Actually, the time between the occurrence of a typical mutation and the end of the observation, asymptotically follows the exponential distribution with parameter ν. This is the time for which any given mutant clone (population stemming from a single cell) will develop;
3. a mutant clone develops according to a Yule process with parameter μ. Its size at time s is geometric with parameter $\mathrm{e}^{-\mu s}$.

Indeed, h_ρ is the probability generating function of an exponential mixture of geometric distributions (changing v into $\mathrm{e}^{-\mu s}$ in (1.4)):

$$h_\rho(z) = \int_0^{+\infty} \frac{z\mathrm{e}^{-\mu s}}{1 - z + z\mathrm{e}^{-\mu s}} \, \nu \mathrm{e}^{-\nu s} \, \mathrm{d}s \,. \tag{1.5}$$

Therefore, the Luria-Delbrück distribution is a compound Poisson of an exponential mixture of geometric distributions. It is a heavy tail distribution, with tail exponent ρ: the higher the fitness of mutants compared to normal cells, the heavier the tail. This explains the appearance of unusually high counts of mutants in cell growth experiments.

The main goal of fluctuation analysis is to estimate the mutation probability p, from a sample of mutant counts. Using theorem 1.1, mutant counts can be

considered as realizations of the LD(α, ρ), α and ρ being unknown. If an estimate of α has been calculated, then an estimate of p can be obtained, dividing by the total number of cells at the end of the experiment. Many methods of estimation for α have been proposed: see Foster (2006). The simplest consists in estimating the probability of observing no mutant: $\mathrm{e}^{-\alpha}$; this is the original method used by Luria and Delbrück (1943). The relative fitness is not taken into account, and it can only be used if α is small. Any other consistent estimator of α has to be sensitive to the value of ρ: even if only the parameter α is of interest, its estimation cannot be decoupled from that of ρ. As far as we know, no thorough statistical study of the coupled estimation problem has been made. Moreover, even though fluctuation analysis with differential growth rates has been advocated by several authors (Koch (1982), Jones (1994), Jaeger and Sarkar (1995), and Zheng (2002, 2005)), most studies are still being made using the LD$(\alpha, 1)$ without questionning the equal rate hypothesis (e.g. Wu et al. (2009), Jean et al. (2010)). The main objective of this work is to propose statistically founded estimation procedures for α and ρ.

Maximum Likelihood (ML) seems to be the obvious choice: see Zheng (2005). Indeed, the likelihood and its derivatives can be computed by iterative algorithms: theoretically at least, the problem could be considered as solved. This is not so in practice, mainly because the multiple sums that must be computed by the optimization algorithm make it quite unstable. According to the numerous tests that we have made, the ML estimates cannot be reliably computed for samples whose maximum exceeds 1000. Therefore a robust estimation procedure must be used (see Wilcox (2012) as a general reference). A first approach is Winsorization, that consists of replacing any value of the sample that pass a certain bound, by the bound itself. As an alternative robust estimation technique, we propose to transform the data through $X \mapsto z^X$, with $0 < z < 1$, i.e. consider the empirical probability generating function. This method, particularly adapted to compound Poisson distributions, has already been considered in analogous cases by Rémillard and Theodorescu (2000) and Marcheselli, Baccini and Barabesi (2008). We have derived consistent Generating Function (GF) estimators for α and ρ, for which the asymptotic covariance matrix has been calculated. They proved to be close to optimal efficiency, with a broad range of calculability, a good numerical stability, and a negligible computing time. We have developed in R (R Development Core Team (2008)) a set of functions that perform the usual operations on the LD distributions, output ML and GF estimates, confidence regions and p-values for hypothesis testing. These functions have been made available online[1].

The rest of the paper is organized as follows. In section 2, some standard results of branching process theory are reviewed, and the justification of theorem 1.1 is given. The implementation of the ML algorithm is described in section 3, and its drawbacks are discussed. Section 4 is devoted to the Generating Function estimators, their asymptotic variance, and their implementation. The two methods have been compared on a simulation study, whose results are reported in section 5.

[1] http://www.ljk.imag.fr/membres/Bernard.Ycart/LD/

2. Bellman-Harris processes

This section reviews a few aspects of supercritical continuous time branching processes, adapting them to the context of cell division: see Harris (1963); Athreya and Ney (1972) as general references. The proof of theorem 1.1, as well as the probabilistic interpretation of the LD(α, ρ) as a compound Poisson of exponential mixtures of geometric distributions, rely upon the following three assertions (the hypotheses are those of theorem 1.1).

A1: the number of mutations converges in distribution to the Poisson distribution with parameter α.
A2: the joint distribution of the developing times of a fixed number k of mutant clones converges in distribution to the product of k independent copies of the exponential distribution with parameter ν.
A3: the size at time s of a mutant clone has geometric distribution with parameter $\mathrm{e}^{-\mu s}$.

From assertions A2 and A3, the size of any mutant clone is an exponential mixture of geometric distributions, the probability generating function of which is given by (1.5). Moreover, these sizes on a fixed number of mutant clones are asymptotically independent. From assertion A1, the number of mutants asymptotically follows a compound Poisson distribution, hence (1.3). A first by-product of this interpretation is a fast simulation algorithm that we have used extensively for our simulation study.

We shall not insist on A3 which is well known: the geometric distribution of a Yule process taken at time s is formula (5) p. 35 of Yule (1925) (see also (Athreya and Ney, 1972, p. 109)). Assertion A1 follows from the law of small numbers, provided we prove that the number of mutation occasions (divisions of normal cells before t_n) is equivalent in probability to $nC\mathrm{e}^{\nu t_n}$. Consider n independent copies of the Bellman-Harris branching process with generation time distribution G and binary divisions (without mutations), each starting with one single cell at time 0. We assume the usual hypotheses on G to ensure that the number of cells has finite mean and variance at each instant (see Athreya and Ney (1972)). In each copy, the number of divisions before time t is equal to the number of cells living at time t minus one. Let $N_1(t)$ be the number of cells living at time t of one copy. Then:

$$\lim_{t\to+\infty} \mathbb{E}[N_1(t)]\mathrm{e}^{-\nu t} = C \ ,$$

from Theorem 17.1 p. 142 of Harris (1963). From there, it is easy to deduce that the total number of cells living at time t_n for the n independent copies, is equivalent in probability to $nC\mathrm{e}^{\nu t_n}$, since t_n tends to infinity. Hence the total number of divisions in the n copies, $N_n(t_n) - n$, is also equivalent in probability to the same quantity. That number does not have the same distribution as the number of divisions of normal cells in the G/M/0 model: when a mutation occurs, the subsequent mutant clone develops according to a different dynamics. However, mutations are rare, and the difference between the number of divisions in the G/M/0 model and in the n independent copies of non mutating

normal clones, remains negligible. To prove that assertion, a coupling will be constructed. Start with the n independent copies above, and mark independently each division as potentially mutant with probability p_n, or non mutant with probability $1-p_n$. Since $p_n nC\mathrm{e}^{\nu t_n}$ tends to α, the number of marked divisions converges in distribution to the Poisson distribution with parameter α, hence remains bounded in probability. Moreover, with probability tending to 1, at most one division per copy has been marked. To deduce the G/M/0 model from the marked independent copies, replace clones after divisions marked as mutants, by the mutant dynamics. Hence with probability tending to 1, the number of mutation occasions in the G/M/0 model is less than the number of marked divisions in the independent copies, and differs from it by a bounded number. This proves that the number of division occasions in the G/M/0 model is equivalent in probability to $nC\mathrm{e}^{\nu t_n}$, as requested.

Assertion A2 can be rephrased into a well known statement about the empirical distribution of split times in a continuous time branching process. The split times sequence, denoted by (τ_i), is the increasing sequence of instants at which divisions occur. Actually, a much stronger result than needed here has been given on the asymptotic distribution of the sequence (τ_i). Theorem 2.1 below (Kuczek, 1982, p. 669) states the almost sure convergence of the empirical distribution of split lags, i.e. the differences $t-\tau_i$, to the distribution function of the exponential with parameter ν.

Theorem 2.1 (Kuczek (1982))**.** *As t tends to $+\infty$, and for any fixed $s>0$, the random variable*

$$\left(\sum_{i=1}^{+\infty}\mathbb{I}_{[0,s]}(t-\tau_i)\right)\Big/\left(\sum_{i=1}^{+\infty}\mathbb{I}_{[0,+\infty)}(t-\tau_i)\right)$$

converges almost surely to the constant $1-\mathrm{e}^{-\nu s}$.

The almost sure convergence implies convergence in distribution: if a split time τ_i is chosen at random among those before t, then the distribution of $t-\tau_i$ converges to the exponential with parameter ν; this is one part of assertion A2. The other part is the asymptotic independence of a fixed number of split times chosen at random; it follows from Theorem 3.1 p. 673 of Kuczek (1982). Notice that in the particular case where the generation times of normal cells are exponentially distributed, assertion A2 is exact, and not asymptotic.

Proposition 2.1. *Let $\{N_t\,,\ t\geqslant 0\}$ be a Yule process and $(\tau_n)_{n\in\mathbb{N}}$ its sequence of split times. Conditionally on $N_t=i+1$, $t-\tau_1,\ldots,t-\tau_i$ are distributed as i independent random variables ranked in decreasing order, each following the exponential distribution with parameter ν, truncated to $[0,t]$.*

This is Reed's statement, from (Reed, 2006, p. 8). Different formulations of the same "order statistic" property have been given by Kendall (1966) and Neuts and Resnick (1971).

3. Maximum Likelihood estimators

A priori, the Maximum Likekihood method should be the obvious choice for estimating α and ρ: under some mild assumptions it is asymptotically optimal (Lehmann and Casella (1999)), and the asymptotic bias can be improved in different ways (Eldar (2006)). For the LD(α, ρ), it has been recommended by several authors: Ma, Sandri and Sarkar (1992), Jones, Wheldrake and Rogers (1993), Zheng (2002, 2005). Its main features and drawbacks are discussed in this section.

As pointed out by Pakes (1993), the compound Poisson interpretation yields both computation algorithms and asymptotic results on the LD(α, ρ), even though its probabilities do not have an explicit form (see also Möhle (2005), Dewanji, Luebeck and Moolgavkar (2005), and Zheng (2002, 2005)). Indeed, let p_k denote the probabilities of the Yule distribution with parameter $\rho = \mu/\nu$ (its probability generating function being h_ρ). For $k \geqslant 1$:

$$p_k = \int_0^{+\infty} \mathrm{e}^{-\mu t}(1 - \mathrm{e}^{-\mu t})^{k-1}\, \nu \mathrm{e}^{-\nu t}\, \mathrm{d}t = \rho B(\rho + 1, k)\ ,$$

where B is the Beta function. The probabilities q_k of the LD(α, ρ) can be computed by the following recursive formula, easily deduced from the probability generating function (1.3) (see Pakes (1993) and references therein):

$$q_0 = \mathrm{e}^{-\alpha} \text{ and for } k \geqslant 1,\ q_k = \frac{\alpha}{k} \sum_{i=1}^{k} i p_i q_{k-i}\ . \tag{3.1}$$

The log-likelihood and its derivatives with respect to the parameters also have explicit algorithms; they have been implemented by Zheng (2005).

Let $(X_1, \ldots, X_n)$ be n independent identically distributed random variables with common distribution LD(α, ρ), and $M = \max_j X_j$. For $i \geqslant 0$ we define $c_i = \sum_{j=1}^{n} 1_{(X_j = i)}$. The log-likelihood is:

$$\ell = \sum_{j=1}^{M} c_j \log q_j. \tag{3.2}$$

Using (3.1), the log-likelihood can be calculated iteratively. Its derivatives are expressed in terms of the derivatives of q_k with respect to α and ρ. These derivatives satisfy the following equations. For $k \geqslant 1$:

$$\frac{\partial q_k}{\partial \alpha} = \left(\sum_{h=1}^{k} p_h q_{k-h} \right) - q_k \quad \text{and} \quad \frac{\partial q_k}{\partial \rho} = \alpha \sum_{h=1}^{k} \frac{\partial p_h}{\partial \rho} q_{k-h}\ , \tag{3.3}$$

where the derivative of p_h with respect to ρ can be computed by a recursive formula (see Zheng (2005)) and the derivatives of q_0 in α and ρ initialize the algorithms:

$$\frac{\partial q_0}{\partial \alpha} = -\mathrm{e}^{-\alpha}\ , \quad \frac{\partial^2 q_0}{\partial \alpha^2} = \mathrm{e}^{-\alpha}\ , \quad \frac{\partial q_0}{\partial \rho} = \frac{\partial^2 q_0}{\partial \alpha \partial \rho} = \frac{\partial^2 q_0}{\partial \rho^2} = 0\ . \tag{3.4}$$

The second derivatives are computed by:

$$\begin{aligned}
\frac{\partial^2 q_k}{\partial \alpha^2} &= \left(\sum_{h=1}^{k} p_h \frac{\partial q_{k-h}}{\partial \alpha}\right) - \frac{\partial q_k}{\partial \alpha} \\
\frac{\partial^2 q_k}{\partial \alpha \partial \rho} &= \sum_{h=1}^{k} \frac{\partial p_h}{\partial \rho}\left(q_{k-h} + \alpha \frac{\partial q_{k-h}}{\partial \alpha}\right) \\
\frac{\partial^2 q_k}{\partial \rho^2} &= \alpha\left(\sum_{h=1}^{k} \frac{\partial^2 p_h}{\partial \rho^2} q_{k-h} + \frac{\partial p_h}{\partial \rho}\frac{\partial q_{k-h}}{\partial \alpha}\right)
\end{aligned} \tag{3.5}$$

where the second derivative of p_h with respect to ρ can be expressed as a function of the Beta, digamma, and trigamma functions. The gradient optimization procedure at iteration $i+1$ computes:

$$\begin{pmatrix} \hat{\alpha}_{i+1} \\ \hat{\rho}_{i+1} \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_i \\ \hat{\rho}_i \end{pmatrix} - H_i^{-1} D_i \ , \tag{3.6}$$

where D_i denotes the gradient, and H_i the Hessian of the log-likelihood evaluated at $(\hat{\alpha}_i, \hat{\rho}_i)$. So formulas (3.1), (3.3) and (3.5) must be applied iteratively for vectors as large as the sample maximum M. The convolution products involve sums of products of small terms for large values of k, and numerical errors accumulate along iterations. In practice, the procedure is very long and may become numerically unstable as soon as the maximal value of the sample exceeds 1000. To give an example, the probability to get at least one value larger than 1000 on a 100-size sample of the $\mathrm{LD}(2, 0.8)$ is 0.53.

Moreover, it is quite difficult to calculate dispersion regions, hence to output confidence intervals and p-values for hypothesis testing. Indeed, the Fisher information matrix $I(\alpha, \rho)$ is the following:

$$I(\alpha,\rho) = \begin{pmatrix} \sum_{k=0}^{+\infty} \left(\frac{\partial q_k}{\partial \alpha}\right)^2 \frac{1}{q_k} & \sum_{k=0}^{+\infty} \frac{\partial q_k}{\partial \alpha}\frac{\partial q_k}{\partial \rho}\frac{1}{q_k} \\ \sum_{k=0}^{+\infty} \frac{\partial q_k}{\partial \alpha}\frac{\partial q_k}{\partial \rho}\frac{1}{q_k} & \sum_{k=0}^{+\infty} \left(\frac{\partial q_k}{\partial \rho}\right)^2 \frac{1}{q_k} \end{pmatrix} .$$

Just like the series $\sum q_k$ itself, the series defining $I(\alpha, \rho)$ converge very slowly: depending on ρ, hundreds of thousand terms may be necessary to get an acceptable precision, unless a convergence acceleration method is used. Fortunately the partial sums increase, so that when computing the inverse $I^{-1}(\alpha, \rho)$ the sum of the first m terms yields conservative confidence intervals; yet we do not consider it satisfactory.

An obvious answer to the instability problem is to Winsorize the sample (see e.g. Wilcox (2012)). Here this consists of replacing any value of the sample that pass a certain bound, by the bound itself. It seems to have been adopted by experimentalists: in the largest fluctuation experiment ever conducted, Boe et al. (1994) had 4 data above 512, and they did not give a precise count for them. Indeed, Winsorization outputs acceptable estimates, as long as the value of α remains small. We argue that such a limitation is not acceptable. On

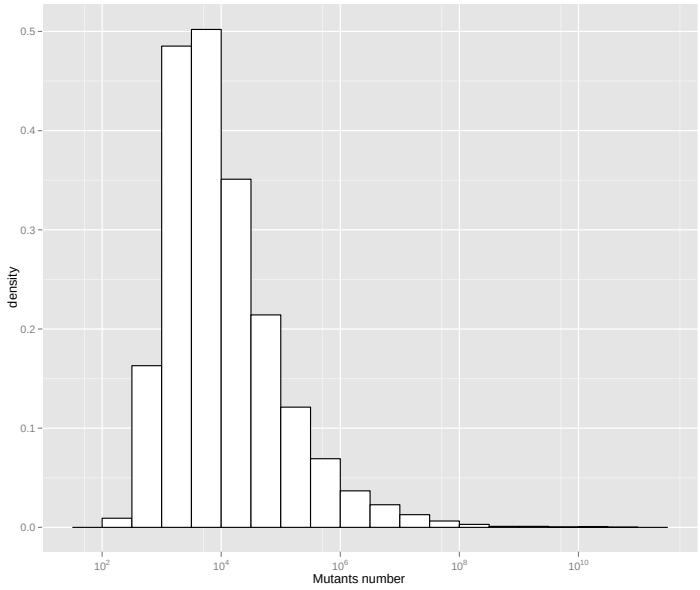

FIG 1. *Histogram of a* 10^5*-sample from the LD*$(50, 0.5)$ *on a logarithmic scale. The class intervals are* $(10^n\,;\,10^{n+1})$ *for* $n = 2$ *to* $n = 15$.

the one hand, progress in cell counting, particularly using flow cytometry, will probably lead in the near future to much higher cell counts. On the other hand, the limitation in the size of data is a theoretical paradox. For a given type of cell, increasing the initial number of cells leads to a proportional increase of α. The probabilistic translation is that $\mathrm{LD}(\alpha, \rho)$, as any other compound Poisson distribution, is infinitely divisible. One could think that increasing α should lead to a more precise estimate of p. Due to the heavy tail property, this is only true if $\rho > 1$. In any case, increasing α is possible only if estimates can be computed for samples with very high values. As an example, figure 1 shows a histogram of a 10^5-size sample of the $\mathrm{LD}(50, 0.5)$, on a logarithmic scale. The range of values was $[128, 1.32 \times 10^{15}]$, the quartiles were 2292, 6798, and 30737. There is no hope to calculate the ML estimates on such a sample, even Winsorized.

4. Generating function estimators

The idea of using the probability generating function to estimate the parameter of a compound Poisson distribution is not new: see Rémillard and Theodorescu (2000) and Marcheselli, Baccini and Barabesi (2008). It turns out that for the $\mathrm{LD}(\alpha, \rho)$, GF estimators are quite comparable in precision to ML estimators, with a much broader range of calculability, a better numerical stability, and a negligible computing time. They are described in this section, and we refer

the reader to the R functions that have been made available on line for further experiments: they include estimation, confidence region, and testing procedures.

Let $(X_1, \ldots, X_n)$ be a sample of independent identically distributed random variables, each with probability generating function $g_{\alpha,\rho}$. Define the empirical probability generating function $\hat{g}_n(z)$ as:

$$\hat{g}_n(z) = \frac{1}{n} \sum_{i=1}^{n} z^{X_i} \; .$$

The random variables z^{X_i} are bounded and mutually independent: by the strong law of large numbers, $\hat{g}_n(z)$ is a strongly consistent estimator of $g_{\alpha,\rho}(z)$, for any z in $[0,1]$. Estimates of α and ρ can be obtained by solving $\hat{g}_n(z) = g_{\alpha,\rho}(z)$ for different values of z. The implementation is described below.

Recall the probability generating function of the LD(α, ρ):

$$g_{\alpha,\rho}(z) = \exp(\alpha(h_\rho(z) - 1)) \; ,$$

with:

$$h_\rho(z) = \rho z \int_0^1 \frac{v^\rho}{1 - z + zv} \, \mathrm{d}v \; .$$

The derivative of $h_\rho(z)$ with respect to ρ will be denoted by $h^1_\rho(z)$.

$$h^1_\rho(z) = \frac{\partial h_\rho(z)}{\partial \rho} = z \int_0^1 \frac{v^\rho (1 + \rho \log(v))}{1 - z + zv} \, \mathrm{d}v \; .$$

All these functions are easily computed by standard numerical procedures. For $0 < z_1 < z_2 < 1$, consider the following ratio:

$$f_{z_1,z_2}(\rho) = \frac{h_\rho(z_1) - 1}{h_\rho(z_2) - 1} \; .$$

The function that maps ρ onto $y = f_{z_1,z_2}(\rho)$ is continuous and strictly monotone, hence one-to-one. Therefore the inverse, that maps y onto $\rho = f^{-1}_{z_1,z_2}(y)$, is well defined. It is, as well, easily computed by standard numerical methods.

For $0 < z_1 < z_2 < 1$, let $\hat{y}_n(z_1, z_2)$ denote the following log-ratio.

$$\hat{y}_n(z_1, z_2) = \frac{\log(\hat{g}_n(z_1))}{\log(\hat{g}_n(z_2))} \; .$$

An estimator of ρ is obtained by:

$$\hat{\rho}_n(z_1, z_2) = f^{-1}_{z_1,z_2}(\hat{y}_n)$$

Then an estimator of α by:

$$\hat{\alpha}_n(z_1, z_2, z_3) = \frac{\log(\hat{g}_n(z_3))}{h_{\hat{\rho}_n(z_1,z_2)}(z_3) - 1} \; ,$$

where $z_3 \in (0\,;1)$ is a new control, possibly different from z_1 and z_2. They will be referred to as Generating Function (GF) estimators. By the strong law of large numbers, as n tends to infinity, the random vector $(\hat{g}_n(z_1), \hat{g}_n(z_2), \hat{g}_n(z_3))$ converges a.s. to $(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_3))$. Since the GF estimators are continuous functions of $(\hat{g}_n(z_1), \hat{g}_n(z_2), \hat{g}_n(z_3))$, the following limits hold with probability 1.

$$\lim_{n\to\infty} \hat{\rho}_n(z_1, z_2) = \rho\ , \quad \lim_{n\to\infty} \hat{\alpha}_n(z_1, z_2, z_3) = \alpha\ .$$

Therefore the GF estimators are strongly consistent. Observe that $\hat{\alpha}_n(z_1, z_2, z_3)$ depends on $\hat{\rho}_n(z_1, z_2)$, whereas $\hat{\rho}_n(z_1, z_2)$ only depends on the arbitrary choice of the couple (z_1, z_2).

Of course the question arises of the variance of the GF estimators and of their use in hypothesis testing, i.e. of computing confidence regions. Asymptotic variances are obtained through the central limit theorem, applied to the vector $(\hat{g}_n(z_1), \hat{g}_n(z_2), \hat{g}_n(z_3))$. Indeed,

$$\sqrt{n}\Big((\hat{g}_n(z_1), \hat{g}_n(z_2), \hat{g}_n(z_3)) - (g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_3)) \Big)$$

converges in distribution to the trivariate centered normal distribution, with covariance matrix $C = (c(z_i, z_j))_{i,j=1,2,3}$, where:

$$c(z_i, z_j) = g_{\alpha,\rho}(z_i z_j) - g_{\alpha,\rho}(z_i) g_{\alpha,\rho}(z_j)\ . \tag{4.1}$$

(Rémillard and Theodorescu, 2000, Proposition 3.1) give a stronger result, stating the functional convergence of $\hat{g}_n(z)$ to a Gaussian process.

The following proposition is deduced from Slutsky's theorem, in the formulation of (Rémillard and Theodorescu, 2000, Theorem 3.4). Details of the calculation are given in the appendix.

Proposition 4.1. *For any z_1, z_2, z_3 in $(0\,;1)$ such that $z_1 \neq z_2$, the couple of random variables*

$$\sqrt{n}\Big((\hat{\alpha}_n, \hat{\rho}_n) - (\alpha, \rho) \Big)$$

converges in distribution to the bivariate centered normal distribution with covariance matrix $M^t C M$, where $M = \begin{pmatrix} A_1 & R_1 \\ A_2 & R_2 \\ A_3 & R_3 \end{pmatrix}$, with

1. $R_1 = \frac{h_\rho(z_2)-1}{\alpha g_{\alpha,\rho}(z_1)((h_\rho(z_2)-1)h^1_\rho(z_1)-(h_\rho(z_1)-1)h^1_\rho(z_2))}$,
2. $R_2 = \frac{h_\rho(z_1)-1}{\alpha g_{\alpha,\rho}(z_2)((h_\rho(z_1)-1)h^1_\rho(z_2)-(h_\rho(z_2)-1)h^1_\rho(z_1))}$,
3. $R_3 = 0$,
4. $A_1 = \frac{\alpha h^1_\rho(z_3)}{1-h_\rho(z_3)} R_1$,
5. $A_2 = \frac{\alpha h^1_\rho(z_3)}{1-h_\rho(z_3)} R_2$,
6. $A_3 = \frac{1}{g_{\alpha,\rho}(z_3)(h_\rho(z_3)-1)}$.

Using proposition 4.1 to compute confidence regions and intervals, or p-values for hypothesis testing is standard and we shall not develop that aspect (see e.g. Anderson (2003)).

The GF estimators such as they have been described so far, depend on the three arbitrary values of z_1, z_2 and z_3. Another tuning parameter will be added. In the LD(α, ρ) the parameter ρ (the heavy tail exponent) determines the size and frequency of much larger values than usual (called "jackpots" in Luria and Delbrück (1943)). For $\rho < 1$, some very large values can be obtained, even for a small α. Using the empirical probability generating function is a simple way to damp down jackpots, and get robust estimates. The variable z can be seen as a tuning parameter for the damping. At the limit case $z = 0$, $\hat{g}_n(0)$ is simply the frequency of null values, and $\hat{\alpha}_n(0) = -\log(\hat{g}_n(0))$ is the so called p_0-estimator of α, already considered by Luria and Delbrück (1943) (it does not depend on ρ). For $z_1 = 0.1$, only small observations will be taken into account, whereas for $z_2 = 0.9$, much larger values will influence the sum: $0.9^{174} \simeq 0.1^8$. Thus the empirical probability generating function damps down jackpots in a differential way according to z_1 and z_2. Choosing $z_1 = 0.1$ and $z_2 = 0.9$ will contrast small values compared to jackpots, which explains why $\hat{\rho}_n$ can efficiently estimate ρ for small α's. However, for large values of α (say $\alpha > 5$), even $z_2 = 0.9$ will output very small values, below the machine precision. This will make the estimates numerically unstable. A natural way to stabilize them is to rescale the sample, dividing all values by a common factor b. This amounts to replacing z by $z^{1/b}$ in the definition of $\hat{g}_n(z)$:

$$\frac{1}{n}\sum_{i=1}^{n} z^{X_i/b} = \frac{1}{n}\sum_{i=1}^{n} (z^{1/b})^{X_i} = \hat{g}_n(z^{1/b}) \; .$$

We propose to set b to the q^{th} quantile of the sample, where q is another control. In theory, z_1, z_2, z_3, q should be chosen so as to minimize the asymptotic variances from proposition 4.1. Numerical evidence showed little dependence of asymptotic variances on the choice of the tuning parameters. Also, the optimal values depend on the (unknown) values of α and ρ. We have run simulation experiments from 1000 samples of size 100 of the LD(α, ρ), looking for those values of z_1, z_2, z_3, q that minimized the Mean Squared Error (MSE); this was repeated for values of α from 0.5 to 5 and ρ ranging from 0.5 to 2. Our best compromise is $z_1 = 0.1$, $z_2 = 0.9$, $z_3 = 0.8$, $q = 0.1$. In our implementation of the GF estimators, the scaling factor b is set to the q^{th} quantile of the sample, and all data are divided by that scaling factor (which amounts to replacing z_1, z_2, z_3 by $z_1^{1/b}, z_2^{1/b}, z_3^{1/b}$). The estimators $\hat{\alpha}_n$ and $\hat{\rho}_n$ are computed with these values. We are quite aware that in doing so, $z_1^{1/b}, z_2^{1/b}, z_3^{1/b}$ become functions of the sample, hence the theoretical result of proposition 4.1 does not apply. Nevertheless, the simulation experiments showed a correct match between theoretical confidence regions and the experimental ones, so we have used them for computing confidence intervals and p-values. Observe that rescaling the sample does not improve the ML method in any way.

GF estimates are good, even for extreme values. As an example, on the sample of size 100.000 of the LD$(50, 0.5)$ from figure 1, the 95% confidence intervals were $(48.2\,;\,51.7)$ for α and $(0.49\,;\,0.51)$ for ρ. On the many repetitions of

simulation experiments that we have made, we have consistently observed the following:

- GF estimators output, in virtually null computer time, reliable values even in cases where the ML estimators fail (large α, small ρ);
- on samples where both GF and ML estimates can be computed, the results are quite close;
- ML estimates (when they can be computed, i.e. for small values of α), are slightly better than the GF ones, so they should definitely be preferred.

However, initializing the optimization procedure of the ML estimators by a close enough starting point, is both a guaranty of numerical stability and economy in computer time. So we have made the obvious choice of initializing our ML procedures by GF estimates.

5. Simulation study

Experimental results are reported in this section. We first checked on published data sets that the ML and GF methods give coherent results: results are reported in table 1. Luria and Delbrück (1943) (table 2, p. 504) had data under 3 different experimental conditions. We have grouped in sample A experiments numbers 1, 10, 11 and 21b; in sample B experiments 16 and 17. We have also used data published in Boe et al. (1994); Rosche and Foster (2000); Zheng (2002). For each data set the ML and GF 95% confidence intervals on α and ρ are given. The data set from Rosche and Foster (2000) has a high frequency of zeros, and no jackpot; this explains why ρ cannot be reliably estimated. For the data sets from Boe et al. (1994) and from Luria and Delbrück (1943) B, the value of ρ is significantly smaller than 1: this is a strong argument in favor of jointly estimating α and ρ rather than assuming $\rho = 1$ as in most fluctuation analyses studies so far.

TABLE 1
Maximum Likelihood (first line) and Generating Function (second line) 95% confidence intervals on α and ρ for published data sets

Reference	size	α	ρ
Luria and Delbrück (1943) A	42	(5.24 ; 8.74)	(0.83 ; 1.33)
		(5.22 ; 8.89)	(0.82 ; 1.35)
Luria and Delbrück (1943) B	32	(0.36 ; 1.00)	(0.23 ; 0.84)
		(0.35 ; 1.04)	(0.18 ; 0.81)
Boe et al. (1994)	1102	(0.65 ; 0.77)	(0.76 ; 0.92)
		(0.65 ; 0.77)	(0.73 ; 0.91)
Rosche and Foster (2000)	52	(1.00 ; 1.80)	(1.15 ; 6.22)
		(1.03 ; 1.98)	(0.00 ; 12.12)
Zheng (2002)	30	(6.76 ; 12.94)	(0.67 ; 1.11)
		(6.65 ; 12.78)	(0.66 ; 1.11)

In order to assess the relative efficiency of the GF method, we simulated 1000 samples of size $n = 100$ for α in $(1, 2, 4, 6, 8, 10)$ and ρ in $(2, 1.5, 1, 0.8, 0.5)$, and calculated on each sample the estimates of α and ρ by the GF, ML, and Winsorized ML methods (the Winsorization parameter was 500); the output was the Mean Squared Errors (MSEs) on α and ρ, over the 1000 samples. Apart from the extensive study of Boe et al. (1994), usual fluctuation experiment samples have size of order a few tens, which motivated our choice for the sample size. The range of values for ρ covers all practical situations. For α, very small values were not considered as significant: if $\alpha < 1$, a large part of the information is contained in the frequency of zeros: the p_0-method gives almost as good results on α as the ML or GF methods. For $\alpha > 10$, the ML estimator fails in most case and its Winsorized version is strongly biased. The results are reported in table 2 (MSEs on estimates of α) and table 3 (MSEs on estimates of ρ). In all cases where the ML estimator can be computed, the errors are quite comparable. As α increases, and for $\rho < 1$, the ML method fails more and more often, and the bias of the Winsorized version increases.

As already said, the GF estimates can be computed for much larger values of α. Table 4 shows the MSEs obtained on α and ρ for α in $(50, 100, 150, 200)$. As expected, the relative precision (quotient of the MSE by the value) on α improves as ρ increases, and the relative precision on ρ improves as α increases. Due to the heavy tail property, the relative precision on α decreases as α increases if $\rho < 1$. We did not try to adapt the GF estimators to much larger values: we believe that an approximation of the $\mathrm{LD}(\alpha, \rho)$ in terms of stable distributions should be used instead. This is the approach of Angerer (2010) for the case $\rho = 1$.

In order to evaluate the bias induced by the equal growth rate hypothesis, we have compared GF and ML estimates of α, to the ML estimates obtained by assuming the $\mathrm{LD}(\alpha, 1)$ as a model. Figure 2 shows boxplots of estimates on 1000 samples of size 100 of the $\mathrm{LD}(\alpha, \rho)$ for α in $(1, 4, 8)$ and ρ in $(0.5, 1, 2)$. Not surprisingly if $\rho = 1$, asuming the $\mathrm{LD}(\alpha, 1)$ as a model slightly improves the quadratic error. When the true value of ρ is less than 1, assuming $\rho = 1$ leads to overestimating α (due to larger and more frequent jackpots); the larger α, the higher the bias.

The effect of either α or ρ on the estimate of the other can be seen on the dispersion regions, obtained through proposition 4.1 or through the Fisher information matrix. Figure 3 displays dispersion ellipses for different values of α and ρ and samples of size 100. When α is small, the information is concentrated on the value of the distribution at 0, which does not depend on ρ, hence the bad precision on the estimate of ρ. When ρ is small, frequent jackpots make the estimates on α less precise. In all cases, the estimates of α and ρ are positively correlated.

None of the other estimation methods is comparable in quality to the ML or GF estimators. We have included in our script a comparison function for 6 different methods. Figure 4 shows a typical output: boxplots of estimates of $\alpha = 2$, using the GF estimator and 5 other methods, computed on 1000 samples

TABLE 2
Mean Squared Errors on estimates of α (GF on first line, ML on second line and Winsorized ML on third line), for 1000 samples of size 100 of the LD(α, ρ). For $\rho = 0.5$, ML estimates could not be computed for $\alpha > 1$

	$\rho = 0.5$	$\rho = 0.8$	$\rho = 1$	$\rho = 1.5$	$\rho = 2$
	0.13	0.13	0.13	0.13	0.12
$\alpha = 1$	0.09	0.13	0.13	0.13	0.12
	0.14	0.13	0.13	0.13	0.12
	0.23	0.21	0.22	0.22	0.20
$\alpha = 2$	–	0.20	0.20	0.21	0.19
	0.26	0.21	0.21	0.21	0.19
	0.45	0.40	0.39	0.39	0.38
$\alpha = 4$	–	0.43	0.39	0.35	0.32
	0.63	0.40	0.37	0.34	0.32
	0.73	0.62	0.56	0.51	0.50
$\alpha = 6$	–	0.71	0.58	0.46	0.45
	1.38	0.62	0.53	0.46	0.45
	1.00	0.75	0.70	0.63	0.61
$\alpha = 8$	–	1.59	0.80	0.60	0.58
	2.55	0.79	0.68	0.60	0.58
	1.30	0.96	0.88	0.73	0.72
$\alpha = 10$	–	–	1.09	0.71	0.68
	4.28	1.10	0.87	0.69	0.68
	1.61	1.15	1.01	0.86	0.82
$\alpha = 12$	–	–	1.23	0.84	0.78
	6.81	1.37	1.00	0.83	0.78

TABLE 3
Mean Squared Errors on estimates of ρ (GF on first line, ML on second line and Winsorized ML on third line), for 1000 samples of size 100 of the LD(α, ρ). For $\rho = 0.5$, ML estimates could not be computed for $\alpha > 1$

	$\rho = 0.5$	$\rho = 0.8$	$\rho = 1$	$\rho = 1.5$	$\rho = 2$
	0.09	0.35	0.55	1.09	1.72
$\alpha = 1$	0.12	0.39	0.56	1.09	1.73
	0.07	0.35	0.55	1.08	1.73
	0.08	0.33	0.53	1.07	1.64
$\alpha = 2$	–	0.38	0.56	1.07	1.61
	0.07	0.34	0.53	1.07	1.61
	0.05	0.32	0.53	1.05	1.62
$\alpha = 4$	–	0.39	0.56	1.04	1.59
	0.08	0.33	0.53	1.04	1.59
	0.05	0.32	0.52	1.04	1.58
$\alpha = 6$	–	0.41	0.57	1.04	1.57
	0.1	0.34	0.53	1.03	1.57
	0.04	0.31	0.51	1.04	1.57
$\alpha = 8$	–	0.43	0.56	1.04	1.56
	0.12	0.34	0.52	1.03	1.56
	0.04	0.31	0.52	1.03	1.56
$\alpha = 10$	–	–	0.57	1.04	1.56
	0.15	0.34	0.53	1.03	1.56
	0.02	0.31	0.51	1.02	1.55
$\alpha = 12$	–	–	0.58	1.03	1.55
	0.18	0.35	0.52	1.02	1.55

TABLE 4
Mean Squared Errors on GF estimates of α (first line) and ρ (second line), on 1000 samples of size 100 of the $LD(\alpha, \rho)$

	$\rho = 2$	$\rho = 1.5$	$\rho = 1$	$\rho = 0.8$	$\rho = 0.5$
$\alpha = 50$	3.006	3.630	4.764	6.045	8.888
	0.161	0.102	0.055	0.045	0.030
$\alpha = 100$	5.966	8.003	10.598	13.036	22.511
	0.150	0.100	0.052	0.041	0.030
$\alpha = 150$	9.146	12.423	17.395	20.447	38.234
	0.146	0.098	0.051	0.038	0.030
$\alpha = 200$	12.992	18.850	27.258	29.854	53.048
	0.150	0.105	0.053	0.038	0.030

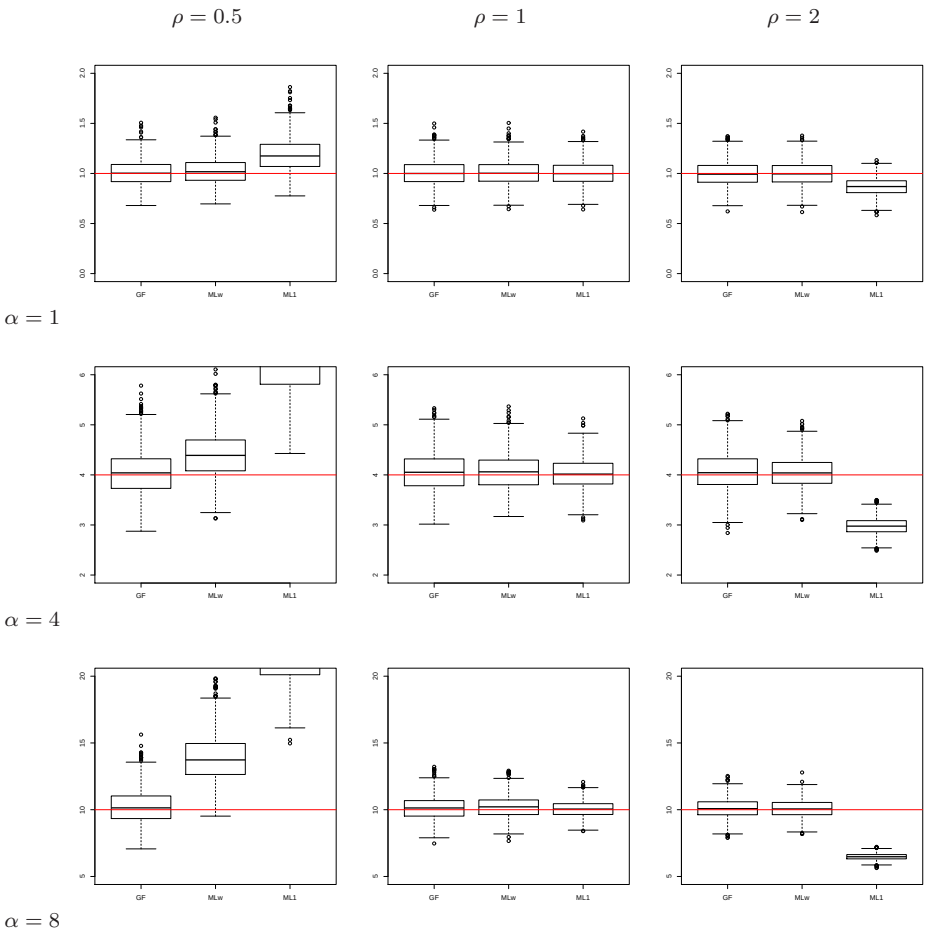

FIG 2. *Boxplots of estimates of α based on 1000 samples of size* 100. *The GF estimates (left plot) and ML estimates (middle plot) assume the $LD(\alpha, \rho)$ as a model, the ML*1 *estimate (right plot) assumes the $LD(\alpha, 1)$.*

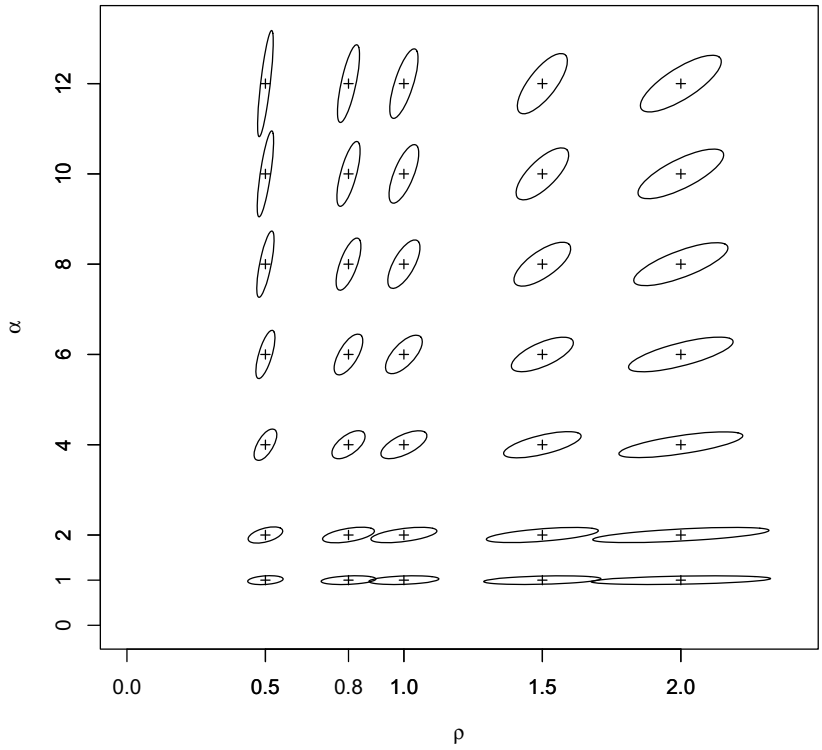

FIG 3. *Dispersion ellipses for GF estimators of α and ρ at level 95%. The computation is based on the asymptotic variance of proposition 4.1, for samples of size* 100.

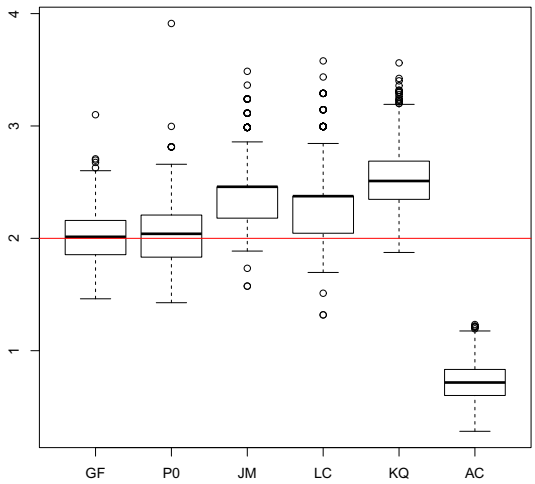

FIG 4. *Boxplots of estimates of α computed by 6 different methods on 1000 samples of size 100 of the $LD(2, 0.8)$ distribution.*

of size 100 of the LD$(2, 0.8)$ distribution. The 5 other methods are the following (see Foster (2006) for descriptions and references):

P0: p_0-estimate;
JM: Lea-Coulson median estimate;
LC: Jones median estimate;
KQ: Koch quartiles estimate;
AC: accumulation of clones estimate.

Appendix A: Proof of Proposition 4.1

We follow (Rémillard and Theodorescu, 2000, section 3). Let z_1, z_2, z_3 be three reals in $(0, 1)$, two by two distinct. Recall that the mapping f_{z_1,z_2}, from $\mathbb{R}$ to $\mathbb{R}$, maps ρ onto:

$$f_{z_1,z_2}(\rho) = \frac{h_\rho(z_1) - 1}{h_\rho(z_2) - 1} .$$

Define the mapping ϕ, from $\mathbb{R}^2$ to $\mathbb{R}$ by:

$$\phi(g_1, g_2) = f^{-1}_{z_1,z_2}\left(\frac{\log(g_1)}{\log(g_2)}\right) ,$$

then ψ, from $\mathbb{R}^3$ to $\mathbb{R}$ by:

$$\psi(g_1, g_2, g_3) = \frac{\log(g_3)}{h_{\phi(g_1,g_2)}(z_3) - 1} .$$

Finally define H, from $\mathbb{R}^3$ to $\mathbb{R}^2$, by:

$$H(g_1, g_2, g_3) = (\psi(g_1, g_2, g_3), \phi(g_1, g_2)) .$$

The GF estimator $(\hat{\alpha}_n, \hat{\rho}_n)$ was defined as:

$$(\hat{\alpha}_n, \hat{\rho}_n) = H(\hat{g}_n(z_1), \hat{g}_n(z_2), \hat{g}_n(z_3)) .$$

Applying Theorem 3.4 of Rémillard and Theodorescu (2000), $\sqrt{n}((\hat{\alpha}, \hat{\rho}) - (\alpha, \rho))$ converges to the bivariate centered Gaussian vector with covariance matrix $M^t C M$, where C is the asymptotic covariance of $\sqrt{n}((\hat{g}_n(z_1), \hat{g}_n(z_2), \hat{g}_n(z_3)) - (g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_3)))$, and M is the Jacobian matrix of H, evaluated at $(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_3))$:

$$M = \begin{pmatrix} \frac{\partial\psi}{\partial g_1} & \frac{\partial\phi}{\partial g_1} \\ \frac{\partial\psi}{\partial g_2} & \frac{\partial\phi}{\partial g_2} \\ \frac{\partial\psi}{\partial g_3} & \frac{\partial\phi}{\partial g_3} \end{pmatrix} (g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_3)) .$$

The partial derivatives of ϕ and ψ in g_1, g_2, and g_3 are computed as follows.

$$\begin{aligned} \frac{\partial\phi}{\partial g_1} &= \frac{\partial f^{-1}_{z_1,z_2}(\log(g_1)/\log(g_2))}{\partial g_1} \\ &= \frac{1}{g_1 \log(g_2)} \frac{(h_\rho(z_2) - 1)^2}{h^1_\rho(z_1)(h_\rho(z_2) - 1) - h^1_\rho(z_2)(h_\rho(z_1) - 1)} \end{aligned}$$

Taking the value at $g_1 = g_{\alpha,\rho}(z_1)$ and $g_2 = g_{\alpha,\rho}(z_2)$ gives:

$$\begin{aligned} & \frac{\partial \phi}{\partial g_1}(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2)) \\ = \quad & \frac{1}{g_{\alpha,\rho}(z_1)\alpha(h_\rho(z_2)-1)} \frac{(h_\rho(z_2)-1)^2}{h^1_\rho(z_1)(h_\rho(z_2)-1) - h^1_\rho(z_2)(h_\rho(z_1)-1)} \\ = \quad & \frac{h_\rho(z_2)-1}{\alpha g_{\alpha,\rho}(z_1)((h_\rho(z_2)-1)h^1_\rho(z_1) - (h_\rho(z_1)-1)h^1_\rho(z_2))} = R_1 \ . \end{aligned}$$

The partial derivative of ϕ in g_2 is obtained by swapping indices 1 and 2:

$$\begin{aligned} & \frac{\partial \phi}{\partial g_2}(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2)) \\ = \quad & \frac{h_\rho(z_1)-1}{\alpha g_{\alpha,\rho}(z_2)((h_\rho(z_1)-1)h^1_\rho(z_2) - (h_\rho(z_2)-1)h^1_\rho(z_1))} = R_2 \ . \end{aligned}$$

The derivative of ϕ in g_3 is null. The derivative of ψ in g_1 is:

$$\frac{\partial \psi}{\partial g_1} = h^1_\rho(z_3) \frac{\log(g_3)}{(h_{\phi(g_1,g_2)}-1)^2} \frac{\partial \phi}{\partial g_1} \ .$$

The value at $(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_3))$ is:

$$\frac{\partial \psi}{\partial g_1}(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_2)) = \frac{\alpha h^1_\rho(z_3)}{1-h_\rho(z_3)} R_1 = A_1 \ .$$

Replace R_1 by R_2 to get A_2:

$$\frac{\partial \psi}{\partial g_2}(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_2)) = \frac{\alpha h^1_\rho(z_3)}{1-h_\rho(z_3)} R_2 = A_2 \ .$$

Finally, the derivative of ψ in g_3 is:

$$\frac{\partial \psi}{\partial g_3} = \frac{1}{g_3(h_{\phi(g_1,g_2)}(z_3)-1)} \ ,$$

Hence:

$$\frac{\partial \psi}{\partial g_3}(g_{\alpha,\rho}(z_1), g_{\alpha,\rho}(z_2), g_{\alpha,\rho}(z_3)) = \frac{1}{g_{\alpha,\rho}(z_3)(h_\rho(z_3)-1)} = A_3 \ .$$

Acknowledgements

The authors are grateful to Anestis Antoniadis, Joël Gaffé, Alain le Breton, and Dominique Schneider for helpful and pleasant discussions. They are indebted to the two anonymous referees for insightful suggestions.

Supplementary Material

Additional material: R script
(http://ljk.imag.fr/membres/Bernard.Ycart/LD/). Set of R functions for statistical computation with Luria-Delbrück distributions, including random sample simulation, estimation of parameters with ML and GF methods, asymptotic variance matrices, confidence intervals and p-values for hypothesis testing.

References

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*, 3rd ed. Wiley, New-York. MR1990662

Angerer, W. P. (2010). Proliferation model dependence in fluctuation analysis: the neutral case. *J. Math. Biol.* **61** 55–93. MR2643883

Armitage, P. (1952). The statistical theory of bacterial populations subject to mutation. *J. R. Statist. Soc. B* **14** 1–40. MR0050864

Athreya, K. B. and Ney, P. E. (1972). *Branching processes*. Springer-Verlag, Berlin. MR0373040

Bartlett, M. S. (1978). *An introduction to stochastic processes, with special reference to methods and applications*, 3rd ed. Cambridge University Press. MR0475536

Boe, L., Tolker-Nielsen, T., Eegholm, K. M., Spliid, H. and Vrang, A. (1994). Fluctuation analysis of mutations to nalidixic acid resistance in *Escherichia Coli*. *J. Bacteriol.* **176** 2781–2787.

Dewanji, A., Luebeck, E. G. and Moolgavkar, S. H. (2005). A generalized Luria-Delbrück model. *Math. Biosci.* **197** 140–152. MR2186954

Eldar, Y. C. (2006). Uniformly improving the Cramér-Rao bound and maximum likelihood estimation. *IEEE Trans. Signal Proc.* **54** 2943–2956.

Foster, P. L. (2006). Methods for determining spontaneous mutation rates. *Methods Enzymol.* **409** 195–213.

Harris, T. E. (1963). *The theory of branching processes*. Springer-Verlag, Berlin. MR0163361

Jaeger, G. and Sarkar, S. (1995). On the distribution of bacterial mutants: the effects of differential fitness of mutants and non-mutants. *Genetica* **96** 217–223.

Jean, L. W., Suchorolski, M. T., Jeon, J. and Luebeck, E. G. (2010). Multiscale estimation of cell kinetics. *Comput. Math. Meth. Med.* **11** 239–254. MR2678237

Jones, M. E. (1994). Luria-Delbrück fluctuation experiments; accounting simultaneously for plating efficiency and differential growth rate. *J. Theo. Biol.* **166** 355–363.

Jones, M. E., Wheldrake, J. and Rogers, A. (1993). Luria-Delbrück fluctuation analysis: estimating the Poisson parameter in a compound Poisson distribution. *Comput. Biol. Med.* **23** 525–534.

Kendall, D. G. (1952). Les processus stochastiques de croissance en biologie. *Ann. IHP* **13** 43–108. MR0057526

Kendall, D. G. (1966). Branching processes since 1873. *J. London Math. Soc.* **41** 385–406. MR0198551

Koch, A. L. (1982). Mutation and growth rates from Luria-Delbrück fluctuation tests. *Mutat. Res.* **95** 129–143.

Kuczek, T. (1982). Almost sure limit results for the supercritical Bellman-Harris process. *J. Appl. Probab* **19** 668–674. MR0664851

Lea, D. E. and Coulson, C. A. (1949). The distribution of the number of mutants in bacterial populations. *J. Genetics* **49** 264–285.

Lehmann, E. L. and Casella, G. (1999). *Theory of point estimation*, 2nd ed. Springer, New York. MR1639875

Luria, D. E. and Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28** 491–511.

Ma, W. T., v. H. Sandri, G. and Sarkar, S. (1992). Analysis of the Luria-Delbrück distribution using discrete convolution powers. *J. Appl. Probab.* **29** 255–267. MR1165212

Mandelbrot, B. (1974). A population birth-and-mutation process, I: explicit distributions for the number of mutants in an old culture of bacteria. *J. Appl. Probab.* **11** 437–444. MR0359035

Marcheselli, M., Baccini, A. and Barabesi, L. (2008). Parameter estimation for the discrete stable family. *Commun. Statist. Theory Methods* **37** 815–830. MR2408282

Möhle, M. (2005). Convergence results for compound Poisson distributions and applications to the standard Luria-Delbrück distribution. *J. Appl. Probab.* **42** 620–631. MR2157509

Neuts, M. F. and Resnick, S. I. (1971). On the times of birth in a linear birth process. *Natural Resource Modeling* **12** 473–475. MR0293732

Oprea, M. and Kepler, T. B. (2001). Improved inference of mutation rates II: generalization of the Luria-Delbrück distribution for realistic cell-cycle time distributions. *Theor. Pop. Biol.* **59** 49–59.

Pakes, A. G. (1993). Remarks on the Luria-Delbrück distribution. *J. Appl. Probab.* **30** 991–994. MR1242029

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reed, W. J. (2006). A brief introduction to scale free networks. *Natural Resource Modeling* **19** 3–14. MR2216129

Rémillard, B. and Theodorescu, R. (2000). Inference based on the empirical probability generating function for mixtures of Poisson distributions. *Statist. Decisions* **18** 349–366. MR1820334

Rosche, W. A. and Foster, P. L. (2000). Determining mutation rates in bacterial populations. *Methods* **20** 1–17.

Wilcox, R. (2012). *Introduction to robust estimation and hypothesis testing*, 3rd ed. Elsevier, Amsterdam.

Wu, X., Strome, E. D., Meng, Q., Hastings, P. J., Plon, S. E. and Kimmel, M. (2009). A robust estimator of mutation rates. *Mut. Res.* **661** 101–109.

Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Phil. Trans. Roy. Soc. London Ser. B* **213** 21–87.

Zheng, Q. (1999). Progress of a half century in the study of the Luria-Delbrück distribution. *Math. Biosc.* **162** 1–32. MR1726870

Zheng, Q. (2002). Statistical and algorithmic methods for fluctuation analysis with SALVADOR as an implementation. *Math. Biosci.* **176** 237–252. MR1899877

Zheng, Q. (2005). New algorithms for Luria-Delbrück fluctuation analysis. *Math. Biosci.* **196** 198–214. MR2158817

Zheng, Q. (2008). On Bartlett's formulation of the Luria-Delbrück mutation model. *Math. Biosci.* **215** 48–54. MR2459528

Zheng, Q. (2010). The Luria-Delbrück distribution: early statistical thinking about evolution. *Chance* **23** 15–18.