Robust fluctuation analyses dealing with the plating efficiency and fluctuating final counts

Adrien Mazoyer

Robust fluctuation analyses dealing with the plating efficiency and fluctuating final counts

Adrien Mazoyer

February 20, 2018

Abstract

Two specific extensions for fluctuation analysis are considered here: mutants are sampled from only a fraction of the final cultures, also called plating efficiency, and the final number of cells is no longer assumed to be constant from one culture to another. We are focusing in this paper to the extension of three specific robust methods: the classic P0 method of Luria and Delbrück, the Maximum Likelihood and a method based on the generating function of the mutant count. Unbiased estimators are thus proposed. Their asymptotic variances are computed. These statistical properties are illustrated with simulation experiments. The methods are also applied to real data sets. In particular, the results are compared with those obtained using previous methods.

Introduction

Consider a culture of cells, which initially contains a homogeneous population. Make grow this population and assume that mutations appear spontaneously upon cell division, at any time during the growth process. A mutation which appears early during the growth process will produce a large final number of mutants. These large values, called *jackpots*, are evidence of a heavy-tail distribution for the final mutant count. This feature has been first proposed by Luria and Delbrück [18], introducing *fluctuation analyses*. The main purpose of fluctuation analysis is the estimation of the probability π for a mutation to appear upon any cell division, given samples of integers, interpreted as final numbers of mutant cells. These numbers may be coupled with final numbers of cells or a mean final number of cells. Since the pioneer article of Luria and Delbrück, a large number of studies about fluctuation analysis have been led: see for example [29, 36, 1] for different reviews. The probabilistic description of the appearance of mutations during a growth of a cell population constitutes the *mutation models*. Any classic mutation model can be interpreted as the result of the following ingredients [10]:

1. a random number of mutations occurring with small probability among a large number of cell divisions. Due to the law of small numbers, the number of mutations approximately follows a Poisson distribution. The expectation of that distribution, denoted by m, is the product of the mutation probability π with the total number of divisions;

2. from each mutation, a *clone* of mutant cells growing for a random time. Due to exponential growth, most mutations occur close to the end of the experiment, and the developing time of a random clone is exponentially distributed. The rate of that distribution, denoted by ρ, is the relative *fitness*, i.e. the ratio of the growth rate of normal cells to that of mutants;

3. the number of mutant cells that any clone developing for a given time will produce. The distribution of this number depends on the distribution of division times of mutants.

The definition of the growth rate (also called "Malthusian parameter") mentioned in the second ingredient can be found in [3, Chap. IV Sec. 4] or [10].

The estimation of mutation probabilities is of crucial importance in several domains, such as the appearance of multidrug resistance of *Mycobacterium Tuberculosis*. Estimates are realizations of an estimator which is a random variable depending on the considered sample. Therefore, the considered estimator have to satisfy two properties: consistency and explicit asymptotic distribution. In other words, an estimator $\hat{\theta}_n$ of a theoretical value θ, built from a sample of size n must satisfy first:

$$\lim_{n\to+\infty} \hat{\theta}_n = \theta\,,$$

where the considered limit may be *weak* (convergence in probability) or *strong* (almost sure convergence). Moreover, the study of the asymptotic distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ must be possible. Luria and Delbrück have proposed two estimators [18, eqs. (5); (8)]. The first is obtained taking the negative logarithm of the relative frequency of null counts inf the sample. Thus, it is a consistent and asymptotically normal estimator of the mean number of mutation. The second estimator uses the relation between the mean number of mutations and the mean number of mutants. However, it can be shown that the number of mutants does not have an expectation. Thus, the obtained estimator is not consistent and should not be used. Several years after, Luria and Delbrück, Lea and Coulson have proposed a method based on the empirical median and an approximation of the mutant count distribution [15, eq. (25)]. However, the properties of consistency and explicit asymptotic distribution are not satisfied by the empirical median in the discrete case. Moreover, according to the authors, this method provides satisfying results only for theoretical values of m larger than 4 and smaller than 15, when the typical values of m are unit order. Thereafter, a wide panel of estimation methods based on empirical median

or other quantiles have been proposed. Most of them are described in [8]. All these methods do not satisfy the required properties and other methods should be considered instead. In particular, under appropriate modelling assumptions, the distribution of the final mutant counts is explicit. Therefore, the Maximum Likelihood (ML) seems to be an obvious optimal choice [19, 37]. However, because of the jackpots, likelihood computation can be numerically unstable. There are several ways to reduce tail effects [32, Sec. 2.2], among which "Winsorization" consists in truncating the sample beyond some maximal value. Another robust estimation method uses the probability generating function [26, 10]. The estimates obtained with Generating Function (GF) method proved to be close to optimal efficiency, with a broad range of calculability, a good numerical stability, and a negligible computing time.

All the mentioned methods compute estimates of the mean number of mutations m, and possibly the relative fitness ρ. However, the true parameter of interest is the mutation probability π. Usually, its estimate is computed dividing the estimate of m by the final number of cells. Therefore, it requires that the final number of cells is the same in each considered culture. However, even under careful monitoring, it is not possible to constrain the final number of cells to be constant [14]. Theoretical models considering variations in the population size have previously been proposed by Angerer [2] and Komarova et al. [13]. The bias induced by ignoring the fluctuation of the final number of cells has been studied for the P0 method by Ycart and Veziris [35]. They have proposed a bias correction for the estimation of π. In this paper, this study will be extended to the GF and ML method.

Another bias source will be considered in this article: the fully efficient plating. In practice, only a fraction of the whole population is actually observed. Then, some of the mutants will not be observed in the sample. The assumption has already been studied by several authors (see for e.g. [29, 11, 1, 9]). In particular, Stewart et al. [29] have proposed a correction which takes into account a plating efficiency smaller than 100%. However, as it will be described later, this correction can be applied only under specific modelling assumption. In this paper, extensions of the P0, ML and GF methods taken into account the plating efficiency under general assumptions are proposed.

The paper is organized as follows. Section 1 is devoted to the probabilistic settings: main assumptions are described and distribution of the final mutant count is given. The P0, ML and GF estimation methods are shortly described in Section 2. The extensions of these methods to not fully efficient plating and fluctuating final number of cells are also proposed in Section 2. These extensions have already been illustrated in [23] with simulation experiments performed on R with the package `flan`. Then, only applications to the data sets of will be exposed in this paper, in Section 3.

1 Probabilistic description

In this section, probabilistic mutation models are described. In previous articles [22, 21], mutation models with birth date dependence have been described. Since the construction of the estimators of interest is similar in both inhomogeneous and homogeneous case [21, Sec. 5], the latter will be considered. The basic modelling hypotheses are the following:

- at time 0, n normal cells are present;
- the lifetime of any normal cell is a random variable with distribution function F_ν;
- upon completion of the lifetime of a normal cell:
  - with probability π one normal and one mutant cell are produced;
  - with probability γ the cell dies out;
  - with probability $1-\gamma-\pi$ two normal cells are produced;
- the lifetime of any mutant cell is a random variable with distribution function F_μ;
- upon completion of the lifetime of a mutant cell:
  - with probability δ the cell dies out;
  - with probability $1-\delta$ two mutant cells are produced;
- all random variables and events (division times, mutations, and deaths) are mutually independent.

Usually, the scale of time is supposed to be adjusted so that the growth rate of mutants is 1; thus the growth rate of normal cells is ρ. Let $(\tau_n)_{n\in\mathbb{N}}$ be a sequence of observation instants, tending to infinity as n tends to infinity. Let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of mutation probabilities, tending to 0 as n tends to infinity. At large instant τ_n, a proportion ε of the clones stemming from the n initial cells will have died out: the final number of normal cells will be asymptotically equivalent to $n(1-\varepsilon)C\mathrm{e}^{\rho\tau_n}$. The expected number of mutations before τ_n is then asymptotically equal to $n\pi_n(1-\varepsilon)C\mathrm{e}^{\rho\tau_n}$, where C is a constant depending on F_ν [34]. The asymptotic context is assumed to be such that

$$\lim_{n\to+\infty} n\pi_n(1-\varepsilon)C\mathrm{e}^{\rho\tau_n} = m\,,$$

with m positive and finite.

Under the above hypotheses, as n tends to infinity, the pgf (probability generating function) of the number of mutants at time τ_n starting with n normal cells tends to the pgf

$$\phi(z) = \exp\left(-m\left(1-\mathcal{I}(z)\right)\right)\,, \tag{1}$$

with

$$\mathcal{I}(z) = \int_0^{+\infty} \psi(z,t)\rho \mathrm{e}^{-\rho t} \mathrm{d}t \,, \tag{2}$$

where $\psi(z,t)$ is the pgf of the number of cells at time t in a mutant clone, starting from a single cell at time 0. Observe that it depends on the lifetime distribution of normal cells F_ν only through ρ. Moreover, expressions (1) and (2) illustrate the three ingredients decomposition described in the introduction:

1. the Poisson distribution with expectation m models the total number of mutations;
2. the exponential distribution with rate ρ is that of the time during which clone develops;
3. the pgf $\psi(\cdot,t)$ models the number of cells in a random clone developing during a time interval of length t. It is the solution of the following Bellman-Harris [4, eq. (2)] equation:

$$\psi(z,t) = \int_0^t \delta + (1-\delta)\psi(z,u)^2 \mathrm{d}F_\mu(u) + z(1-F_\mu(t)) \,. \tag{3}$$

Therefore, any mutation model is a Poisson compound of an exponential mixture. Note that the pgf ψ has expressible form in two cases:

- Exponentially distributed lifetimes, i.e.

  $$F_\mu(t) = \left(1 - \mathrm{e}^{-t}\right) \mathbb{1}_{t \geqslant 0} \,.$$

  In this case, (3) has an explicit solution [3, p. 109]:

  $$\psi(z,t) = \frac{\delta(1-z) + \mathrm{e}^{-t}((1-\delta)z - \delta)}{(1-\delta)(1-z) + \mathrm{e}^{-t}((1-\delta)z - \delta)} \,;$$

- Constant lifetimes, i.e.

  $$F_\mu(t) = \mathbb{1}_{t \geqslant a} \,,$$

  where $a = \log(2(1-\delta))$ (so the growth rate of mutants is 1). In that case, (3) does not make sense. However ψ remains expressible [22, p. 2938]:

  $$\psi(z,t) = \sum_{i \geqslant 0} b_i(z) \mathbb{1}_{t \in [ia\,;\,(i+1)a)} \,,$$

  where b_i is the pgf of the size in the time interval $[ia\,;(i+1)a)$ of a mutant clone started at time 0, i.e.

  $$b_0(z) = z \,, \quad b_i(z) = \delta + (1-\delta)\left(b_{i-1}(z)\right)^2 \,, \quad i \geqslant 1 \,.$$

In practice, the plating process can be less than 100% efficient. In that case, a random number of mutants will not be counted: if only a proportion ζ of the final population is plated, then each cell will be observed with probability ζ. Denote by M_{tot} the total numbers of mutants. Given $M_{\text{tot}} = k$, M follows the binomial distribution with parameters k and ζ. Thus, the pgf ϕ of M is given by:

$$\begin{aligned}\phi(z) &= \mathbb{E}\left[\mathbb{E}\left[z^M \mid M_{\text{tot}}\right]\right] \\ &= \exp\left(-m\left(1-\mathcal{I}^{(\zeta)}(z)\right)\right) . \end{aligned} \tag{4}$$

where

$$\begin{aligned}\mathcal{I}^{(\zeta)}(z) &= \mathcal{I}(1-\zeta+\zeta z) \\ &= \int_0^{+\infty} \psi(1-\zeta+\zeta z, t)\rho \mathrm{e}^{-\rho t}\mathrm{d}t\,, \end{aligned} \tag{5}$$

The pgf (4) defines a family of distributions, denoted hereafter by $MM(m,\rho,\delta,\zeta,F_\mu)$ (Mutation Model). The mutation model where the pgf (2) is explicit will be denoted by $LD(m,\rho,\delta,\zeta)$ for the exponential case (Luria-Delbrück), and $H(m,\rho,\delta,\zeta)$ for constant case (Haldane model). Here the computation of probabilities of the LD models with $\zeta \leqslant 1$ is exposed. Define for any $z \in (0\,;1)$

$$\psi^{(\zeta)}(z,t) = \psi(1-\zeta+\zeta z, t)\,.$$

In the case of a LD formulation, $\psi^{(\zeta)}$ is given by

$$\psi^{(\zeta)}(z,t) = \frac{\delta(1-(1-\zeta+\zeta z)) + \mathrm{e}^{-t}((1-\delta)(1-\zeta+\zeta z)-\delta)}{(1-\delta)(1-(1-\zeta+\zeta z)) + \mathrm{e}^{-t}((1-\delta)(1-\zeta+\zeta z)-\delta)}\,.$$

Denote by $\left(r_k^{(\zeta)}(t)\right)_{k\in\mathbb{N}}$ the probabilities associated with $\psi^{(\zeta)}(\cdot,t)$. Let us rewrite $\psi^{(\zeta)}$ as

$$\begin{aligned}\psi^{(\zeta)}(z,s,t) &= \frac{\delta\left(1-\mathrm{e}^{-t}\right) - (1-\zeta)\left(\delta - \mathrm{e}^{-t}(1-\delta)\right) - z\zeta\left(\delta - \mathrm{e}^{-t}(1-\delta)\right)}{1-\delta-\delta\mathrm{e}^{-t} - (1-\zeta)(1-\delta)\left(1-\mathrm{e}^{-t}\right) - z\zeta(1-\delta)\left(1-\mathrm{e}^{-t}\right)} \\ &= \frac{n_0(s,t)+zn_1(s,t)}{d_0(s,t)+zd_1(s,t)}\,, \end{aligned}$$

where

$$\begin{aligned} n_0(s,t) &= \delta\left(1-\mathrm{e}^{-t}\right) - (1-\zeta)\left(\delta-\mathrm{e}^{-t}(1-\delta)\right)\,, \\ n_1(s,t) &= -\zeta\left(\delta-\mathrm{e}^{-t}(1-\delta)\right)\,, \\ d_0(s,t) &= 1-\delta-\delta\mathrm{e}^{-t} - (1-\zeta)(1-\delta)\left(1-\mathrm{e}^{-t}\right)\,, \end{aligned}$$

and

$$d_1(s,t) = -\zeta(1-\delta)\left(1-\mathrm{e}^{-t}\right)\,.$$

The $r_k^{(\zeta)}(t)$'s can then be defined as follows:

$$r_0^{(\zeta)}(t) = \frac{n_0(t)}{d_0(t)}\,, \quad r_1^{(\zeta)}(t) = \frac{n_1(t)}{d_0(t)} - \frac{d_1(t)}{d_0(t)} r_0^{(\zeta)}(t)$$

and for any $k \geqslant 2$:

$$r_k^{(\zeta)}(t) = -\frac{d_1(t)}{d_0(t)} r_{k-1}^{(\zeta)}(t)\,.$$

Now, (5) can be written as

$$\mathcal{I}^{(\zeta)}(z) = \sum_{k \geqslant 0} q_k^{(\zeta)} z^k\,,$$

where for any $k \geqslant 0$:

$$q_k^{(\zeta)} = \int_0^{+\infty} r_k^{(\zeta)}(t) \rho \mathrm{e}^{-\rho t} \mathrm{d}t\,.$$

Finally the probabilities $(p_k)_{k\in\mathbb{N}}$ of the mutant count after plating M are computed with the following recursive algorithm [6]:

$$\begin{aligned} p_0 &= \mathrm{e}^{-m(1-q_0^{(\zeta)})}\,, \\ p_k &= \frac{m}{k} \sum_{i=1}^{k} i q_i^{(\zeta)} p_{k-i} \quad k \geqslant 1\,. \end{aligned} \tag{6}$$

Until now, there is no expression for the probabilities of the mutant count for Haldane formulation when $\zeta < 1$.

As mentioned in the introduction, the classic approach assumes that the total number or cells N is constant. Therefore, if a reliable estimation of m is available, the mutation probability π is estimated dividing the estimate of m by N. However, even under close experimental monitoring, assuming that the final number of cells is a constant is quite unrealistic. Thus, N must be considered as a random variable with a certain probability distribution function K on $[0, +\infty)$. Therefore, the conditional pgf of the number of mutants given $N = k$ is given by

$$\phi(z \,|\, N = k) = \exp\left(-\pi k \left(1 - \mathcal{I}(z)\right)\right)\,.$$

Hence, the conditional distribution of the number of mutants given $N = k$ is the distribution $MM(\pi k, \rho, \delta, \zeta, F_\mu)$. Assume that K is known. Therefore:

$$\begin{aligned} \phi(z) &= \int_0^{+\infty} \phi(z \,|\, N = k) \mathrm{d}K(k) \\ &= \mathcal{L}[\pi(1 - \mathcal{I}(z))]\,, \end{aligned} \tag{7}$$

where $\mathcal{L}$ is the Laplace transform of K:

$$\mathcal{L}(z) = \int_0^{+\infty} \mathrm{e}^{-zk} \mathrm{d}K(k)\,. \tag{8}$$

The pgf (7) defines a family of distributions, denoted hereafter by $MMF(m, \rho, \delta, \zeta, F_\mu, K)$ (Mutation Model with Fluctuating number of cells). A realization of these distribution is a couple (M, N) such that conditionally to $N = k$ (according to the distribution K), M follows the distribution $MM(m, \rho, \delta, \zeta, F_\mu)$. The two particular cases for distribution F_μ will be denoted by $LDF(m, \rho, \delta, \zeta, K)$ for the exponential case (Luria-Delbrück with Fluctuating number of cells), and $HF(m, \rho, \delta, \zeta, K)$ for constant case (Haldane model with Fluctuating number of cells).

2 Robust estimation methods

Here the three estimation methods of interest (P0, ML and GF methods) are shortly described and extended to the case where the plating is not fully efficient and when the final number of cells is random. The methods exposed in this section perform estimation only for m and ρ. Indeed, the fluctuations of the distribution of final mutant counts with respect to δ are very small [34]. Therefore, the death parameter δ is assumed to be known from now. Remark that this assumption is quite unrealistic: in practice, only the magnitude of δ can be measured [27, 7]. Note that the three methods of interest are fully implemented in the R package `flan` [23], which is available on the CRAN (https://cran.r-project.org/package=flan). Moreover, a web-tool based on `flan` and implemented with the R package `shiny` is also available (https://toltex-shiny.u-ga.fr/RodaShiny/ShinyFlan/).

Before the definition of the estimators, let us recall why these methods should be preferred to the median methods. The P0, ML and GF method satisfy the following asymptotic properties: consistency and explicit asymptotic variance; whereas the other methods not. To illustrate the fact that estimation based on the empirical median should not be used, simulation experiments have been performed using R [24]. Figure 1 shows the results for the following estimators:

- P0: the estimator proposed by Luria and Delbrück [18, eq. (5)];
- GF: the Generating Function estimator proposed by Hamon and Ycart [10];
- ML: the Maximum Likelihood estimator;
- LC: the median estimator proposed by Lea and Coulson [15, eq. (25)];
- JM: the median estimator proposed by Jones et al. [11, eq. (6)];

- KQ: the quartile estimator proposed by Koch [12, eqs. (3)-(5)];
- AC: accumulation of clones method proposed by Luria [17] (see also [8, eqs (14); (15)]).

For each $m = (0.5, 1, 2, 4)$, 10^4 samples of size 100 of the $LD(m, 1, 0)$ distribution were simulated. Estimates of m were calculated using the above methods. Each boxplot represents the distribution of the 10^4 ratio $\hat{m}/m$ obtained for each estimation method. According to the visual results, the GF and ML methods provide good estimates, whatever the theoretical value of m: for these methods, most of the estimates have a relative bias smaller than 10%. The P0 estimates seem to be less robust: when m increases. Remark that for $m = 4$ the probability of null count is very small. Therefore, the P0 method cannot be applied systematically. In that case, the ratio $\hat{m}/m$ has been set to 0 (hence the number of outliers for the P0 method). The median methods provide good estimates when m is large. However, the visual observations seem to show that these methods should not be used. Indeed, the estimates are not relevant when m decreases: for example, the estimates obtained with LC and JM methods are almost deterministic and are not centered on m.

Recall that the parameter of interest is the mutation probability π. For P0, ML and GF methods, the estimate of π is computed dividing the estimate of m by the mean final number of cells. In the introduction, the mean method of [18, eq. (8)] (denoted by LDM) has been mentioned. In this method, π is estimated by solving an equation depending on the mean number of mutants and the final number of cells. As for the median methods, simulation studies have been performed to show the lack of robustness of the LDM method. These methods requires to find the root of an increasing function, so a finite domain of research is required. Since the GF method provides precise estimates, the research interval has been set to $[0.01 * \hat{\pi}_{GF} ; 100 * \hat{\pi}_{GF}]$. Table 1 shows the summary of the simulation study.

According to Table 1, the variance of LDM method is much larger than the three other methods. Moreover, the LDM estimates are not centered on the true value of π. These observations have been made by Rosche and Foster [25]: this method is not recommended, although it still appears in recent studies (see for example [31]).

Therefore, only P0, ML and GF methods will be considered in this paper. Two kinds of samples are considered from now:

1 a sample $\mathbf{X} = (M_i)_{i=1,\dots,n}$ of a MM distribution, with the mean κ and the standard deviation σ of the final number of cells;

2 a sample of couples $\mathbf{X} = (M_i, N_i)_{i=1,\dots,n}$ of a MMF distribution, where there is no particular assumption on the distribution of the N_i's.

Remark that when $\zeta < 1$, the M_i's correspond to mutant counts *after* plating, when the information on the final number of cells (i.e. κ and σ, or the N_i's) are observed *before* the plating.

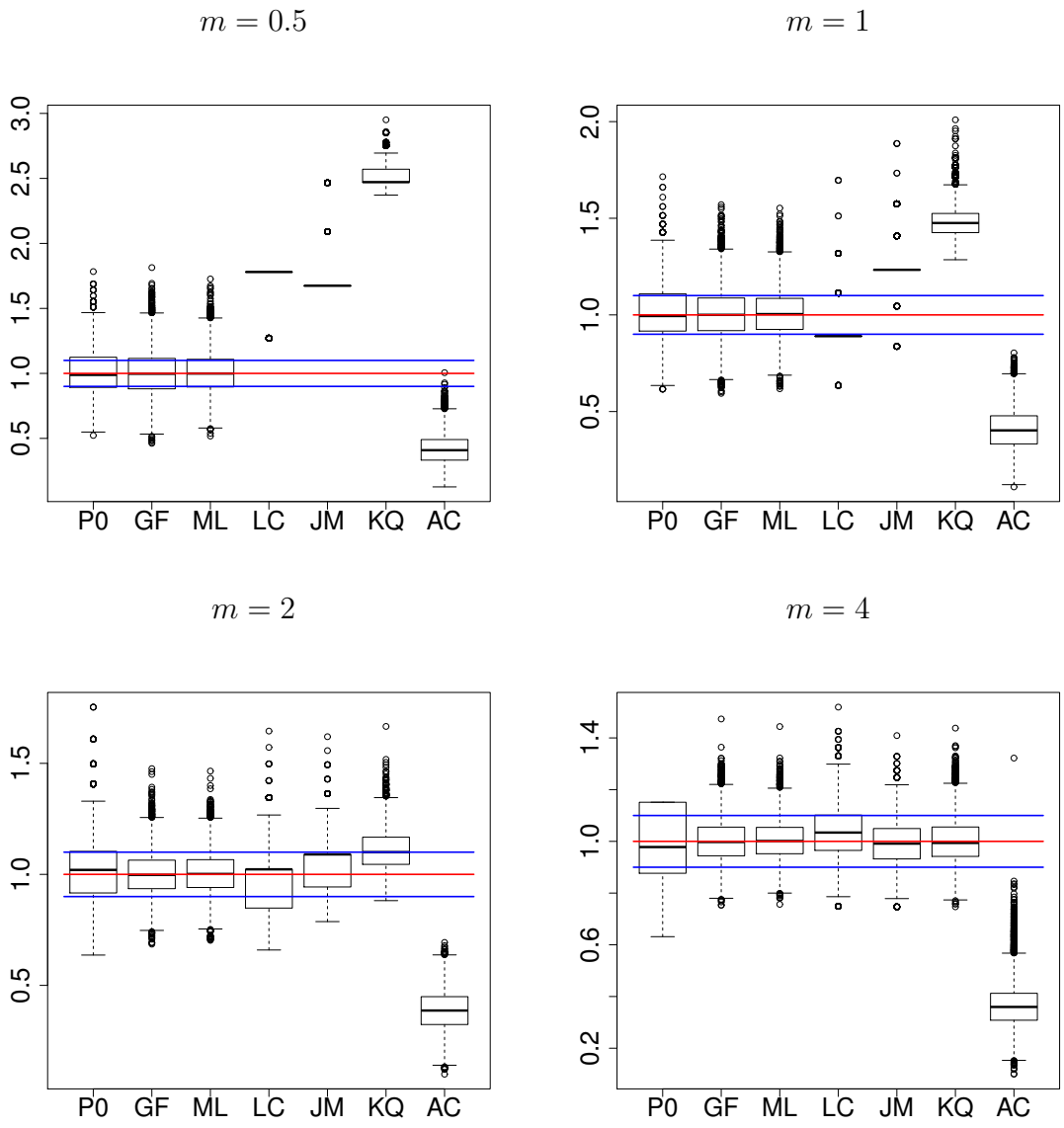

Figure 1: **Comparison of different estimation methods of m.** Red horizontal lines mark unit. Blue horizontal lines mark relative bias of 0.9 and 1.1. For each $m = (0.5, 1, 2, 4)$, 10^4 samples of size 100 of the $LD(m, 1, 0)$ distribution were simulated. Estimates of m were calculated using the methods described earlier. Each boxplot represents the distribution of the 10^4 ratio $\hat{m}/m$ obtained for each estimation method.

$m = 0.5$

	P0	GF	ML	LDM
Minimum	0.4969	0.4525	0.4839	0.4004
First quartile	0.8926	0.8851	0.8973	0.9648
Median	0.9886	0.9974	1.0019	1.2805
Mean	1.0077	1.0049	1.0062	2.1454
Last quartile	1.1242	1.1169	1.1093	1.8589
Maximum	1.8326	1.8643	1.7287	182.3220

$m = 1$

	P0	GF	ML	LDM
Minimum	0.5978	0.5255	0.5838	0.4452
First quartile	0.9163	0.9123	0.9185	0.9647
Median	0.9943	0.9958	0.9971	1.2422
Mean	1.0078	1.0011	1.0030	2.0966
Last quartile	1.0788	1.0844	1.0812	1.7669
Maximum	1.5606	1.5312	1.5104	127.472

$m = 2$

	P0	GF	ML	LDM
Minimum	0.6189	0.6744	0.6938	0.5226
First quartile	0.9163	0.9371	0.9041	0.9646
Median	1.0201	0.9976	1.0013	1.12159
Mean	1.0176	1.0012	1.0048	2.1193
Last quartile	1.1036	1.0605	1.0659	1.6787
Maximum	1.9560	1.3820	1.3700	2059.1089

$m = 4$

	P0	GF	ML	LDM
Minimum	0.6314	0.7459	0.7639	0.5879
First quartile	0.8766	0.9433	0.9512	0.9670
Median	0.9780	0.9965	1.0004	1.1835
Mean	$+\infty$	1.0004	1.0033	1.7693
Last quartile	1.1513	1.0545	1.0526	1.6041
Maximum	$+\infty$	1.3727	1.3050	110.0355

Table 1: **Comparison of different estimation methods of parameter $\boldsymbol{\pi}$.** For each $m = (0.5, 1, 2, 4)$, 10^4 samples of size 100 of the $LDF(\pi, 1, 0, K)$ distribution were simulated, where $\pi = m/\kappa$, $\kappa = 10^9$ and K is the cumulative distribution function of Dirac measure located at κ. Each column contains the main statistics of the 10^4 ratio $\hat{\pi}/\pi$ for each estimation method.

The P0 method can be used whether the growth model of mutants is known. However, the ML and GF methods require having an explicit expression of the distribution of the mutant count. Hence, only LD, LDF, H and, HF models will be considered for ML and GF methods. From now, the probabilities of the mutant count M will be denoted by $(p_k)_{k\in\mathbb{N}}$.

2.1 P0 method

The P0 method was introduced by Luria and Delbrück [18] in the case where $\delta = 0$ and $\zeta = 1$. In that case, the probability of null count in the sample is $p_0 = \mathrm{e}^{-m}$, whatever the distribution F_μ. Hence m can be estimated by

$$\hat{m}_0 = -\log\left(\hat{p}_0\right) , \tag{9}$$

where $\hat{p}_0$ is the relative frequency of null counts among the sample $\mathbf{X}$. By definition, $\hat{p}_0$ is a consistent and asymptotically normal estimator of the probability p_0. By the Δ-method [30, p. 79], $\hat{m}_0$ is a consistent and asymptotically normal estimator of m. Its asymptotic variance is given by

$$v_{\hat{m}_0} = \frac{1-\hat{p}_0}{n\hat{p}_0} .$$

Of course, this method cannot be applied if $\mathbf{X}$ does not contain null count. An extension of the P0 method to the case where $\delta > 0$ has been proposed by Ycart [34] (Fixed Point estimator). It requires that $\delta < 1/2$ (i.e. supercritical process), and is based on the fact that

$$\delta_* = \frac{\delta}{1-\delta} ,$$

is a fixed point of the pgf $\psi(\cdot, t)$ [3, Chap. I; p. 141]. Then, δ_* is also a fixed point of the pgf (2), and a consistent and asymptotically normal estimator of m is given by

$$\hat{m}_0 = \frac{-\log\left(\hat{\phi}_n(\delta_*)\right)}{1-\delta_*} , \tag{10}$$

where $\hat{\phi}_n$ denotes the empirical pgf of the sample $\mathbf{X}$. For any $z \in (0,1)$, $\hat{\phi}_n$ is a consistent and asymptotically normal estimator of $\phi(z)$. By the Δ-method, $\hat{m}_0$ is then a consistent and asymptotically normal estimator of m. Its asymptotic variance is given by

$$v_{\hat{m}_0} = \frac{1}{n(1-\delta_*)^2}\left(\frac{\hat{\phi}_n(\delta_*^2)}{\hat{\phi}_n(\delta_*)^2} - 1\right) . \tag{11}$$

Remark that $\hat{m}_0$ does not depend on ρ and distribution F_μ. The P0 method can then be used on any mutation model, but does not directly yield an estimator of ρ.

Assume now that $\zeta < 1$. According to (6), the probability of null count in the sample *after* plating is given by

$$p_0 = \exp\left(-m\left(1 - \mathcal{I}(1-\zeta)\right)\right) .$$

Consider first the case where $\mathbf{X}$ is a sample of $LD(m, 1, 0)$. The pgf (2) is given by [15]:

$$\mathcal{I}(z) = 1 + \frac{1-z}{z}\log(1-z) ,$$

and then

$$p_0 = \exp\left(-m\frac{\zeta}{1-\zeta}\log(\zeta)\right) .$$

Therefore, if $\hat{m}_0$ is given by (9), a consistent and asymptotically normal estimator of m is given by

$$\hat{m}_0^{(\zeta)} = \frac{-\hat{m}_0(1-\zeta)}{\zeta\log(\zeta)} .$$

This corresponds to the correction proposed by Stewart et al. [29, eq. (41)]. However, this correction can be applied only in this specific case. For example, if $\rho \neq 1$, the probability of null count after plating is given by

$$p_0 = \exp\left(-m\left(1 - \rho\int_0^1 \frac{(1-\zeta)v^\rho}{1-(1-\zeta)(1-v)}\mathrm{d}v\right)\right) .$$

Hence, the estimation of m requires knowing ρ: the P0 method cannot be used when $\zeta < 1$.

However, the estimator (10) can be extended to the case where $\zeta < 1$. Since δ_* is a fixed point of $\psi(\cdot, t)$:

$$\phi\left(\delta_*^{(\zeta)}\right) = \exp(-m(1-\delta_*)) ,$$

where

$$\delta_*^{(\zeta)} = \frac{\delta_* - (1-\zeta)}{\zeta} .$$

Therefore, a consistent and asymptotically normal estimator of m is given by

$$\hat{m}_0^{(\zeta)} = \frac{-\log\left(\hat{\phi}_n\left(\delta_*^{(\zeta)}\right)\right)}{1-\delta_*} . \tag{12}$$

Its asymptotic variance is

$$v_{\hat{m}_0^{(\zeta)}} = \frac{1}{n(1-\delta_*)^2}\left(\frac{\hat{\phi}_n\left(\delta_*^{(\zeta)^2}\right)}{\hat{\phi}_n\left(\delta_*^{(\zeta)}\right)^2} - 1\right) .$$

Remark that (12) is well defined only if $\delta_*^{(\zeta)}$ belongs to the unit disk. In particular, if $\delta = 0$, ζ has to be larger than $1/2$. Therefore, ζ will be assumed to be 1 for P0 method.

The estimate of π is then computed dividing the estimate of m by the final number of cells. If the final number of cells N is random, its fluctuations are thus not taken into account. It has been showed [35] that ignoring these fluctuations induced a bias on the estimate of π. The authors have proposed the following correction when data is a sample of type 1:

$$\hat{\pi}_{\text{ub}} = \hat{\pi}_0 \left(1 + \frac{\hat{\pi}_0 C^2}{2}\right),$$

with $\hat{\pi}_0 = \hat{m}_0/\kappa$ (where $\hat{m}_0$ is given by (9)) and C is the coefficient of variation of N, i.e.:

$$C = \frac{\sigma}{\kappa}.$$

This correction is now extended to the case where $\delta > 0$. Conditionally to $N = k$, the mutant count M follow the MM distribution with $m = \pi k$, and:

$$\phi\left(\delta_* \,|\, N = k\right) = \mathrm{e}^{-\pi k(1-\delta_*)},$$

and then

$$\phi(\delta_*) = \mathcal{L}[\pi(1-\delta_*)],$$

where $\mathcal{L}$ denotes the Laplace transform (8). Therefore, the estimator $\hat{m}_0$ defined by (10) is a consistent estimator of

$$\frac{-\log\left(\mathcal{L}[\pi(1-\delta_*)]\right)}{1-\delta_*}.$$

By Jensen inequality, $\hat{m}_0$ underestimates m, and $\hat{\pi}_0 = \hat{m}_0/\kappa$ underestimates π. Assume that the distribution of N is known and that the inverse $\mathcal{L}^{-1}$ of $\mathcal{L}$ is well defined. A consistent and asymptotically normal estimator of π is then given by

$$\frac{\mathcal{L}^{-1}\left[\hat{\phi}_n(\delta_*)\right]}{1-\delta_*}.$$

Its asymptotic variance is given by

$$\frac{\hat{\phi}_n(\delta_*^2) - \hat{\phi}_n(\delta_*)^2}{\left((1-\delta_*)\mathcal{L}'[\pi(1-\delta_*]\right)^2}.$$

However, the distribution of N is not known in practice. Usually, estimates of the expectation and variance of N are available at best. Consider now the series expansion of $\mathcal{L}$

$$\mathcal{L}[\pi] = 1 - \kappa\pi + \frac{\mathbb{E}[N^2]\pi^2}{2} + \ldots$$

Then:

$$\frac{-\log\left(\mathcal{L}[\pi(1-\delta_*)]\right)}{\kappa(1-\delta_*)} = \pi - \frac{\pi^2 C^2 \kappa(1-\delta_*)}{2}$$

$$= \pi\left(1 - \frac{\pi C^2 \kappa(1-\delta_*)}{2}\right).$$

Therefore, a consistent and asymptotically normal estimator of π is given by

$$\hat{\pi}_0 = \frac{\hat{m}_0}{\kappa}\left(1 + \frac{\hat{m}_0 C^2(1-\delta_*)}{2}\right).$$

Its asymptotic variance is given by

$$v_{\hat{\pi}_0} = \left(1 + \hat{m}_0 C^2(1-\delta_*)\right)^2 \frac{v_{\hat{m}_0}}{\kappa^2},$$

where $v_{\hat{m}_0}$ is the variance (11).

In a case of a sample of type 2 with $\delta = 0$ and $\zeta = 1$, it is possible to compute directly estimate of π [35]. Instead of the couples $(M_i, N_i)_{i=1\ldots n}$, consider the couples $(X_i, N_i)_{i=1\ldots n}$ where $X_i = 1$ if M_i is null, 0 else. Therefore, the log-likelihood of the sample $(X_i, N_i)_{i=1\ldots n}$ is given by

$$\ell(\pi) = -\sum_{i=1}^{n} -\pi N_i X_i + (1 - X_i)\log\left(1 - \mathrm{e}^{-\pi N_i}\right).$$

Thus, the maximum $\hat{\pi}_{ML0}$ of ℓ is a consistent and asymptotically normal estimator of π [16, Corollary 3.11, Chap. 6] with asymptotic variance

$$v_{\hat{\pi}_{ML0}} = \left(\sum_{i=1}^{n}\left(-N_i X_i + \frac{(1-X_i)N_i}{\mathrm{e}^{\pi N_i} - 1}\right)^2\right)^{-1}.$$

Note that this estimator does not depend on the distribution of N.

The P0 estimators do not depend on the modelling assumption for mutants. Then, if an estimate of the relative fitness is desired, the Maximum Likelihood can be used for ρ only.

2.2 ML method

From now, the data $\mathbf{X}$ is assumed to be a sample distributed according to an expressible formulation (i.e. LD, LDF, H or HF). Since algorithms for the computation of the probabilities of LD [6, 37, 10, 34] and H [33, 22] models are available, Maximum Likelihood

seems to be an obvious choice for estimation of m and ρ. Consider first data of type 1. Then, the ML method can be used to estimate m and ρ maximizing the log-likelihood

$$\ell(m,\rho) = \sum_{i=1}^{n} \log\left(p_{M_i}\right)$$

$$= \sum_{i=0}^{\max_j M_j} \left[\log\left(p_i\right) \sum_{k=1}^{n} \mathbb{1}_{M_k=i}\right] . \tag{13}$$

The couple of estimators $(\hat{m}_{ML}, \hat{\rho}_{ML})$ obtained by maximizing (13) is thus consistent, asymptotically normal [16, Chap. 6., Theo. 5.1]. As for the P0 method, ignoring the fluctuations of the final number of cells N will induces a bias on the estimate of π. Assume first that the Laplace transform (8) is known. Then, the parameter m of a formulation MM can be related to the parameter π of a MMF formulation:

$$m = \frac{-\log\left(\mathcal{L}[\pi(1-\mathcal{I}(z))]\right)}{1-\mathcal{I}(z)} .$$

Hence, a consistent and asymptotically normal estimator of π is given by

$$\hat{\pi}_{ML} = \frac{\mathcal{L}^{-1}\left[\mathrm{e}^{-\hat{m}_{ML}(1-\mathcal{I}(z))}\right]}{1-\mathcal{I}(z)} ,$$

where $\mathcal{I}_x$ is the pgf (2) with $\rho = x$. By the Δ-method, its asymptotic variance is

$$v_{\hat{\pi}_{ML}} = \left(\frac{\mathrm{e}^{-\hat{m}_{ML}(1-\mathcal{I}(z))}}{\mathcal{L}'[\pi(1-\mathcal{I}(z))]}\right)^2 v_{\hat{m}_{ML}} .$$

As mentioned earlier, $\mathcal{L}$ is unknown in practice. However, the following approximation is deduced from the series expansion of $\mathcal{L}$:

$$\frac{-\log\left(\mathcal{L}[\pi(1-\mathcal{I}(z))]\right)}{\kappa(1-\mathcal{I}(z))} \approx \pi\left(1 - \frac{m(1-\mathcal{I}(z))C^2}{2}\right) .$$

Consequently, the following estimator of π is approximately consistent and asymptotically normal:

$$\hat{\pi}_{ML} = \frac{\hat{m}_{ML}}{\kappa}\left(1 + \frac{\hat{m}_{ML}(1-\mathcal{I}(z))C^2}{2}\right) .$$

Its asymptotic variance is given by

$$v_{\hat{\pi}_{ML}} = \left(1 + m(1-\mathcal{I}(z))C^2\right)^2 v_{\hat{m}_{ML}} .$$

Remark that this correction depends directly on the value of z. In theory, this value should be set such that it minimizes the above variance. However, the variance depends on the

unknown parameter m. In the R package `flan`, the value of z has been set according to simulation studies (see [10] for more details).

Consider now data of type 2. The ML method can be directly applied to estimate π and ρ [35] by maximizing the following log-likelihood

$$\ell(\pi, \rho) = \sum_{i=1}^{n} \log \left(p_{M_i|N_i} \right) , \tag{14}$$

where for any $k, j \geqslant 0$, $p_{k|j}$ corresponds to probabilities (6) with $m = \pi j$. The couple of estimators $(\hat{\pi}_{ML}, \hat{\rho}_{ML})$ obtained by maximizing ℓ is thus consistent and asymptotically normal.

The ML methods has been widely recommended for the case $\delta = 0$ (see for example [19, 28, 37]). However, when the sample maximum is large, sums of products of small terms must be computed [10]. The procedure can then be very long and numerically unstable. In particular, the ML estimators may fail for large m and/or small ρ. The computing time is also directly influenced by the maximal value of the sample and by the choice of the model: when $\delta > 0$, the computation of the log-likelihood is more expensive for a H formulation than for a LD formulation; the computation of (14) is also more expensive than that of (13). For LD formulation, the computational time of (13) can be impoved using equivalents of the probabilities p_k's for large values of k are available [34]. In practice, these queue issues can be avoided using Winsorization [32, Sec. 2.2], which consists in replacing any value of the sample that pass a certain bound by the bound itself. All information above the bound is lost, and in an extreme case where the sample minimum is greater than the bound, irrelevant estimates will be obtained. Therefore, the ML method is not adapted for sample with large jackpots.

2.3 GF method

Using pgf to estimate the parameter of a Poisson compound was already known [26, 20]. The GF method has been exposed in [10] for LD formulation with $\delta = 0$ and $\zeta = 1$. However, it can be easily adapted to any explicit formulation (see [34] for $\delta > 0$). Consider a sample $\mathbf{X}$, whatever its type.

Let $z_1, z_2, z_3 \in (0\,;1)$, with $z_1 \neq z_2$. GF estimators of m and ρ are defined by

$$\hat{m}_{GF}(z_3) = \frac{\log\left(\hat{\phi}_n(z_3)\right)}{\mathcal{I}_{\hat{\rho}_{GF}(z_1,z_2)}(z_3) - 1} \quad \text{and} \quad \hat{\rho}_{GF}(z_1, z_2) = g^{-1}(\hat{y}_n) , \tag{15}$$

where $\mathcal{I}_x$ is the pgf (2) with $\rho = x$, and

$$g(x) = \frac{\mathcal{I}_x(z_1) - 1}{\mathcal{I}_x(z_2) - 1} \quad \text{and} \quad \hat{y}_n = \frac{\log\left(\hat{\phi}_n(z_1)\right)}{\log\left(\hat{\phi}_n(z_2)\right)} .$$

By Theorem (3.4) of [26] and the Δ-method, it can be proved that the couple of estimators $(\hat{m}_{GF}(z_3), \hat{\rho}_{GF}(z_1, z_2))$ is consistent and asymptotically normal, with explicit asymptotic variance [10, Prop. 4.1].

Consider now the case $\zeta \leqslant 1$. GF estimators of m and ρ are now defined by

$$\hat{m}_{GF}(z_3) = \frac{\log\left(\hat{\phi}_n(z_3)\right)}{\mathcal{I}^{(\zeta)}_{\hat{\rho}_{GF}(z_1,z_2)}(z_3) - 1} \quad \text{and} \quad \hat{\rho}_{GF}(z_1, z_2) = g^{-1}(\hat{y}_n)\,, \tag{16}$$

where $\mathcal{I}^{(\zeta)}_x$ is the pgf (5) with $\rho = x$, and

$$g(x) = \frac{\mathcal{I}^{(\zeta)}_x(z_1) - 1}{\mathcal{I}^{(\zeta)}_x(z_2) - 1} \quad \text{and} \quad \hat{y}_n = \frac{\log\left(\hat{\phi}_n(z_1)\right)}{\log\left(\hat{\phi}_n(z_2)\right)}\,.$$

Therefore, it can be shown that the couple of estimators is still consistent and asymptotically normal. Moreover, its asymptotic variance is explicit.

Proposition 2.1. *Let $z_1, z_2, z_3 \in (0\,;1)$, such that $z_1 \neq z_2$. Denote by $C = (c(z_i, z_j))_{i,j=1,2,3}$ the asymptotic covariance matrix of*

$$\sqrt{n}\left(\left(\hat{\phi}_n(z_1), \hat{\phi}_n(z_2), \hat{\phi}_n(z_3)\right) - \left(\phi(z_1), \phi(z_2), \phi(z_3)\right)\right)\,,$$

i.e.

$$c(z_i, z_j) = \phi(z_i z_j) - \phi(z_i)\phi(z_j)$$

Then the couple of random variables

$$\sqrt{n}\Big((\hat{m}_{GF}, \hat{\rho}_{GF}) - (m, \rho)\Big)$$

converges in distribution to the bivariate centered normal distribution with covariance

matrix A^TCA, *where the matrix* $A = (a_{i,j})_{\substack{i=1,2,3\\ j=1,2}}$ *is such that*

$$a_{1,1} = \frac{m a_{1,2}}{\mathcal{I}^{(\zeta)}(z_3) - 1} \frac{\partial \mathcal{I}^{(\zeta)}(z_3)}{\partial \rho} ;$$

$$a_{1,2} = \frac{\mathcal{I}^{(\zeta)}(z_2) - 1}{m\phi(z_1)\left(\frac{\partial \mathcal{I}^{(\zeta)}(z_1)}{\partial \rho}(\mathcal{I}^{(\zeta)}(z_2) - 1) - \frac{\partial \mathcal{I}^{(\zeta)}(z_2)}{\partial \rho}(\mathcal{I}^{(\zeta)}(z_1) - 1)\right)} ;$$

$$a_{2,1} = \frac{m a_{2,2}}{\mathcal{I}^{(\zeta)}(z_3) - 1} \frac{\partial \mathcal{I}^{(\zeta)}(z_3)}{\partial \rho} ;$$

$$a_{2,2} = \frac{\mathcal{I}^{(\zeta)}(z_1) - 1}{m\phi(z_2)\left(\frac{\partial \mathcal{I}^{(\zeta)}(z_2)}{\partial \rho}(\mathcal{I}^{(\zeta)}(z_1) - 1) - \frac{\partial \mathcal{I}^{(\zeta)}(z_1)}{\partial \rho}(\mathcal{I}^{(\zeta)}(z_2) - 1)\right)} ;$$

$$a_{3,1} = \frac{1}{\phi(z_3)(\mathcal{I}^{(\zeta)}(z_3) - 1)} ;$$

$$a_{3,2} = 0 .$$

Remain that P0 and ML methods cannot be used under H models when $\zeta < 1$. Since it requires only a close expression of the pgf $\mathcal{I}$ and its derivative with respect to ρ, GF method is available for both LD and H models. The GF estimators depend mainly on the choice of the arbitrary values of z_1, z_2 and z_3. Their value should be such that the variance in Proposition 2.1 is minimized. However, the optimal values depend on the unknown parameter m and ρ. Therefore, the values of z_1, z_2 and z_3 have to be set before performing estimation. The main advantage of the GF method is to allow a rescaling of the sample, which makes the method applicable to large values of m. It is possible to replace z by $z^{1/b}$ in the definition of $\hat{\phi}_n(z)$. Then, b is a fourth control parameter. Hamon and Ycart [10] proposed values for z_1, z_2, z_3 and b based on simulation studies.

Again, if π is estimated by the ratio of $\hat{m}_{GF}$ to the mean final number of cells, a bias will appear. As for the ML method, it is possible to adapt the correction exposed for the P0 method to reduce this bias. As defined by (15), $\hat{m}_{Gf}$ is a consistent and asymptotically normal estimator of

$$\frac{-\log\left(\mathcal{L}[\pi(1 - \mathcal{I}(z_3))]\right)}{1 - \mathcal{I}(z_3)} .$$

By Jensen inequality, $\hat{m}_{GF}/\kappa$ underestimates π. If $\mathcal{L}^{-1}$ is well defined and known, a consistent and asymptotically normal estimator of π is given by

$$\frac{\mathcal{L}^{-1}\left(\hat{\phi}_n(z_3)\right)}{1 - \mathcal{I}(z_3)} .$$

Its asymptotic variance is given by

$$\frac{\hat{\phi}_n(z_3^2) - \hat{\phi}_n(z_3)^2}{\Big((1-\mathcal{I}(z_3))\mathcal{L}'[\pi(1-\mathcal{I}(z_3))]\Big)^2}$$

Since the distribution of N is unknown, the above expression cannot be used. However, by considering the series expansion of $\mathcal{L}$, the following estimator of π can be constructed:

$$\hat{\pi}_{GF} = \frac{\hat{m}_{GF}}{\kappa}\left(1 + \frac{\hat{m}_{GF}(1-\mathcal{I}(z_3))C^2}{2}\right).$$

Its asymptotic variance is given by

$$v_{\hat{\pi}_{GF}} = \big(1 + m(1-\mathcal{I}(z_3))C^2\big)^2 \, v_{\hat{m}_{GF}}.$$

Simulation studies [23, Fig. 1, 2] have shown that the 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. Therefore, the optimization procedure of the ML method is initialized with GF estimates, ensuring numerical stability and computational economy.

3 Application on real data sets

Final number data are rarely reported in fluctuation analysis experiments, although exceptions exist such as two well-known data sets [5, 31]. In these two references, the authors study the resistance to some antibiotics of *Mycobacterium tuberculosis* and use the LDM method to perform estimation of π. As mentioned in Section 2, this method should not be used. Moreover, the fluctuations of the final number of cells and the plating efficiency are not rightly taking into account. In this section, these data sets are studied using R package `flan`. The results obtained by the authors using the LDM method are first compared with the estimates of P0, ML and GF methods, under $LD(m, 1, 0, 1)$ model. Therefore, estimates and statistical tests will be computed and performed taking into account the fluctuation of the final counts and the plating efficiency (still under LD formulation).

Consider first the study of David [5]. Table 1 of [5] contains 10 samples of mutant counts. The author has decided to cluster the mutant counts according to some ranges (for example, mutant count between 11 and 20). In this paper, these intervals have been replaced by the median of their bounds (for example, 15 for the range 11-20). Each sample of mutant counts is associated with a final number of cells. Table 2 of [5] is a single sample of size 10 of couples (mutant count - final count of cells). Because of the size of this sample, only the first data set will be considered here. Table 2 compares the estimates obtained by the author with those of the P0, ML and GF methods under

	Author	P0		ML		GF	
Antibio.	$\hat{\pi}(\times 10^{-8})$	$\hat{\pi}(\times 10^{-8})$	**CI (95%) of** $\pi(\times 10^{-8})$	$\hat{\pi}(\times 10^{-8})$	**CI (95%) of** $\pi(\times 10^{-8})$	$\hat{\pi}(\times 10^{-8})$	**CI (95%) de** $\pi(\times 10^{-8})$
Ison.	1.84	1.85	[0.67 ; 3.03]	1.74	[1.07 ; 2.42]	1.78	[1.06 ; 2.50]
	3.50	0.943	[0.514 ; 1.37]	2.62	[2.27 ; 2.98]	2.68	[2.01 ; 3.35]
	1.70	1.17	[0.527 ; 1.82]	1.00	[0.545 ; 1.47]	1.01	[0.555 ; 1.46]
	3.20	0.746	[0.451 ; 1.04]	2.05	[1.81 ; 2.29]	2.09	[1.67 ; 2.50]
Strept.	0.900	0.493	[0.194 ; 0.792]	0.376	[0.0857 ; 0.666]	0.241	[0.035 ; 0.447]
	5.00	0.591	[0.375 ; 0.807]	1.89	[1.68 ; 2.09]	1.38	[1.08 ; 1.69]
Rifam.	0.180	0.0317	[0 ; 0.0937]	0.0309	[0 ; 0.0929]	0.0154	[0 ; 0.0665]
	0.0270	0.0289	[0 ; 0.0572]	0.028	[0 ; 0.0563]	0.0166	[0 ; 0.0418]
Etham.	7.00	0.315	[0.0944 ; 0.536]	0.381	[0.165 ; 0.598]	0.606	[0.267 ; 0.944]
	13.0	0.356	[0.224 ; 0.488]	0.649	[0.522 ; 0.777]	0.776	[0.578 ; 0.975]

Table 2: **Estimates of π with data of David [5, Tab.1] under *LD* model**. The estimates of π by P0, ML and GF methods are computed dividing the estimates of m obtained under LD model by the final number of cells. The relative fitness ρ and the plating efficiency are set to 1, and the death parameter is set to 0. For each method, the 95% confidence intervals are also given.

the LD model. Their 95% confidence intervals are also given. Remark that the data of David [5] corresponds to the mutant counts in 0.1 mL of solution, when the cultures are contained in 2 mL of solution. The author has exposed his results in terms of "mean number of mutants by mL" and "final number of cells by mL". In particular, the mean number of mutants are obtained by multiplying by 10 those et in the samples. Of course, this is not the right way to take into account the plating efficiency. This fact will be first ignored: the current purpose is to show that the LDM method is not relevant. Indeed, Table 2 illustrates the empirical observations of Table 1: the LDM method can provide precise estimates (1st and 8th lines), but most of the estimates are biased, with some times outliers (last line). Among all the estimates, only 3 belongs to the corresponding confidence intervals (1st, 3rd and 8th lines). Recall that the one of the main conclusion of this article was a high mutation rate to Ethambutol resistance.

Let us consider now that the relative fitness ρ is unknown and that the plating efficiency is equal to $\zeta = 0.05$. The estimates of the author has been computed with his reasoning, but considering the "mean number of mutants in 2 mL" (resp. "final number of cells in 2 mL") instead of the "mean number of mutants by mL" (resp. "final number of cells by mL"). These estimates are compared with GF estimates in Table 3. Remark that the fitness of the 7th sample cannot be estimated. In theory, that means that the fitness ρ is larger than the upper bound of the research domain (set to $[0\,;100]$ in `flan`). The associated estimate of m is thus irrelevant. The 5th and 8th estimates are the only that belong t the corresponding confidence intervals. Remark that the size of 5th confidence interval of π is very large, which is due to the large estimate of ρ. With these observations, we could affirm that only the conclusion about the 8th sample conclusion in [5] seems relevant.

Consider now the data set published by Werngren and Hoffner [31, Tab. 1]. In this paper, the authors study the mutation rate of *Beijing* strains of *Mycobacterium tuberculosis* to antibiotic resistance. The data set includes 13 samples of mutant counts (7 non-*Beijing* strains, 6 *Beijing* strains). Each sample is associated with a final number of cells. Table 4 compares the estimates of the authors with those of P0, ML and GF methods under LD model. Their 95% confidence intervals are also given. The estimates obtained by the authors are much closer to that of P0, ML and GF methods than in the previous data set. Only the P0 method excludes some estimates (2nd, 4th, 5th and 9th strains). Remark that the 1st and 7th samples do not contain null count: the P0 method cannot be applied to these samples Consider now that the fitness ρ is unknown, and that the plating efficiency is equal to $\zeta = 0.2$. As for the previous data set, the LDM estimates of π have been recalculated by replacing " mean number of mutants by mL" (resp. "final number of cells by mL") by "mean number of mutants in 5 mL" (resp. "final number of cells in 5 mL"). These estimates are compared with GF estimates in Table 5. Remark that the GF method provides the mean number of mutations in the whole population (i.e. in 5 mL of solution): then the final number of cells has to be multiplied by 5 in this method.

	Author		GF under LD model with $\zeta = 0.05$			
Antibio.	**Estimates of $\pi(\times 10^{-8})$**	**Recomputed estimates of $\pi(\times 10^{-8})$**	**Estimates of $\pi(\times 10^{-8})$**	**CI (95%) of $\pi(\times 10^{-8})$**	**Estimates of ρ**	**CI (95%) of ρ**
Ison.	1.84	1.68	0.432	[0.143 ; 0.721]	0.719	[0.471 ; 0.967]
	3.50	3.21	0.781	[0.461 ; 1.1]	0.789	[0.649 ; 0.93]
	1.70	1.52	0.249	[0.0468 ; 0.451]	0.72	[0.383 ; 1.06]
	3.20	3.04	0.0803	[0.0327 ; 0.128]	0.222	[0.106 ; 0.339]
Strept.	0.900	0.791	0.456	[0 ; 3.15]	16.3	[0 ; 1480]
	5.00	4.39	0.0364	[0.0194 ; 0.0534]	0.0595	[0 ; 0.130]
Rifam.	0.0180	0.0140	0.00667	[0 ; 0.0255]	/	/
	0.0270	0.0234	0.0148	[0 ; 0.0358]	1.69	[0 ; 3.87]
Etham.	7.00	6.26	0.0236	[0.00299 ; 0.0442]	0.151	[0 ; 0.323]
	13.0	12.2	0.0236	[0.0127 ; 0.0346]	0.0929	[0.0157 ; 0.170]

Table 3: **Estimates of π and ρ with data sets of David [5, Tab.1] under LD model taking into account the plating efficiency**. The LDM estimates have been recomputed for the whole culture (2nd column). The estimates of m and ρ are computed using GF method under LD model with a plating efficiency equal to $\zeta = 0.05$ and a zero death parameter. The estimates of π are obtained by dividing the estimates of m by the final number of cells. The 95% confidence intervals of π and ρ are also given.

	Authors	P0		ML		GF	
Strains	$\hat{\pi}(\times 10^{-8})$	$\hat{\pi}(\times 10^{-8})$	**CI (95%) of** $\pi(\times 10^{-8})$	$\hat{\pi}(\times 10^{-8})$	**CI (95%) of** $\pi(\times 10^{-8})$	$\hat{\pi}(\times 10^{-8})$	**CI (95%) of** $\pi(\times 10^{-8})$
H37Rv	0.860	/	/	1.54	[0.484 ; 2.60]	0.998	[0.548 ; 1.45]
E865/94	2.40	6.27	[2.44 ; 10.1]	3.99	[1.79 ; 6.19]	3.03	[1.60 ; 4.46]
E729/94	0.960	1.94	[0.92 ; 2.97]	1.45	[0.759 ; 2.15]	1.16	[0.638 ; 1.69]
E740/94	1.10	2.53	[1.20 ; 3.86]	1.93	[0.868 ; 2.99]	1.35	[0.687 ; 2.01]
E1221/94	0.650	1.33	[0.662 ; 1.99]	0.921	[0.447 ; 1.40]	0.761	[0.401 ; 1.12]
E1449/94	1.50	3.18	[1.26 ; 5.10]	2.45	[1.22 ; 3.67]	1.8	[0.975 ; 2.62]
Harl.	1.40	/	/	2.53	[0.898 ; 4.16]	1.72	[0.933 ; 2.50]
E26/95	1.30	2.29	[1.17 ; 3.41]	1.73	[0.846 ; 2.60]	1.51	[0.823 ; 2.20]
E80/95	0.790	2.10	[0.997 ; 3.21]	1.41	[0.648 ; 2.17]	1.00	[0.500 ; 1.51]
E55/94	1.00	1.83	[0.888 ; 2.77]	1.49	[0.639 ; 2.35]	1.21	[0.500 ; 1.91]
E26/94	0.940	3.16	[1.50 ; 4.82]	1.76	[0.565 ; 2.96]	1.10	[0.461 ; 1.74]
E3942/94	1.50	2.81	[1.33 ; 4.28]	2.31	[1.28 ; 3.33]	1.90	[1.09 ; 2.72]
E47/94	1.20	1.59	[0.811 ; 2.36]	1.48	[0.849 ; 2.11]	1.46	[0.815 ; 2.10]

Table 4: **Estimates of π with data set of Werngren and Hoffner [31, Tab.1] under *LD* model**. The estimates of π by P0, ML and GF methods are computed by dividing those of m by the associated final number of cells. The relative fitness ρ and the plating efficiency are set to 1 and the death parameter δ is null. For each method, 95% confidence interval of π are also given.

	Authors		GF sous modèle LD avec $\zeta = 0.2$			
Souches	**Estimates of $\pi(\times 10^{-8})$**	**Estimations recalculées de $\pi(\times 10^{-8})$**	**Estimates of $\pi(\times 10^{-8})$**	**IC (95%) de $\pi(\times 10^{-8})$**	**Estimates of ρ**	**IC (95%) de ρ**
H37Rv	0.860	0.172	2.95	[0 ; 6.50]	8.59	[0 ; 87.0]
E865/94	2.40	0.482	4.05	[1.50 ; 6.59]	1.99	[0.433 ; 3.54]
E729/94	0.960	0.193	1.17	[0.552 ; 1.79]	1.45	[0.737 ; 2.17]
E740/94	1.10	0.223	2.11	[0.817 ; 3.40]	2.32	[0.233 ; 4.40]
E1221/94	0.650	0.128	0.739	[0.311 ; 1.17]	1.52	[0.636 ; 2.39]
E1449/94	1.50	0.300	2.35	[1.16 ; 3.55]	1.73	[0.815 ; 2.64]
Harl.	1.40	0.278	4.28	[0 ; 9.90]	8.01	[0 ; 81.9]
E26/95	1.30	0.256	1.29	[0.502 ; 2.07]	1.42	[0.555 ; 2.28]
E80/95	0.790	0.161	1.42	[0.549 ; 2.30]	2.16	[0.346 ; 3.98]
E55/94	1.00	0.202	1.12	[0.39 ; 1.85]	1.57	[0.478 ; 2.66]
E26/94	0.940	0.188	0.701	[0.338 ; 1.06]	/	/
E3942/94	1.50	0.302	1.75	[0.837 ; 2.66]	1.34	[0.725 ; 1.96]
E47/94	1.20	0.252	0.756	[0.320 ; 1.19]	0.874	[0.450 ; 1.30]

Table 5: **Estimates of π and ρ with data set of Werngren and Hoffner [31, Tab.1] under LD model taking into account the plating efficiency**. The LDM estimates have been recalculated for the whole population. The GF estimates of m and ρ are computed under LD model with a plating efficiency $\zeta = 0.2$ and a death parameter equal to 0. The estimates of π are then obtained by dividing the estimates of m by the final number or cells. The 95% confidence intervals are also given.

Remark that the fitness cannot be estimated for the 11th sample. Also, its estimates for the 1st and the 7th strains are ridiculously large. The estimates obtained by the authors are very far from the GF estimates. However, since the samples are small, the sizes of the confidence intervals are such that they contain the author's estimates.

In this paper, the authors classify the strains according to their genotype (*Beijing* and non-*Beijing*) to check if the mutation probability is higher for a particular strain. Without performing a statistical test, they claim that there is no significative difference between the *Beijing* and the non-*Beijing*. Denote by π_{nB} an π_B the mutation probability for non-*Beijing* and *Beijing* strains. Performing the following bilateral Student test

$$H_0 \ : \ \pi_{nB} = \pi_B \text{ vs } H_1 \ : \ \pi_{nB} \neq \pi_B;$$

using authors' estimates gives a p-value equal to 0.56: thus it seems there is no significant difference between the two genotypes. However, performing now the following unilateral test

$$H_0 \ : \ \pi_{nB} = \pi_B \text{ vs } H_1 \ : \ \pi_{nB} > \pi_B;$$

with the GF estimates taking into account the plating efficiency gives a p-value equal to 0.0192: therefore, the mutation probability of the non-*Beijing* strains seems to be significantly higher than that of the *Beijing* strains.

It is also possible to perform two sample test on all the strains, without considering their genotype. Table 6 shows the p-values of the statistical tests

$$H_0 \ : \ \pi_i = \pi_j \text{ vs } H_1 \ : \ \pi_i > \pi_j;$$

which compare the probability mutation of the i-th strain (rows) with that of the j-th strain (columns). Table 7 exposes the p-values of the statistical tests

$$H_0 \ : \ \pi_i = \pi_j \text{ vs } H_1 \ : \ \rho_i > \rho_j;$$

which compare the fitness parameter of the i-th strain (rows) with that of the j-th strain (columns). The LD model with plating efficiency is considered, with $\zeta = 0.2$ and $\delta = 0$. The type I error is set to 5%.

First of all, only E1449/94 and E47/94 seems to be significantly different in term of relative fitness. Moreover, since the GF method cannot estimate the relative fitness of the E26/94, the statistical tests have not been performed for this strain. Remain that the previous tables contain unusually high values of fitness for some strains. Concerning the mutation probability, the results are more precise than previously: the mutation probability of the E865/94 strain seems to be significantly higher the *Beijing* strains. However, other non-*Beijing* strains are not significantly different from *Beijing* strain: for example, no significant difference is observed between the E729/94 or the Harlingen strains and the *Beijing* strains. The assertion "the mutation probability of non-*Beijing* strains is significantly higher than that of *Beijing* strains" has been previously accepted. According to the results, it seems that this difference is mainly caused the E865/94 and E1449/94 strains. Moreover, significant differences are also observed between strains with the same genotype: see for example strains E1449/94 and E1221/94.

		non-*Beijing*							*Beijing*					
		H37Rv	**E865/94**	**E729/94**	**E740/94**	**E1221/94**	**E1449/94**	**Harl.**	**E26/95**	**E80/95**	**E55/94**	**E26/94**	**E3942/94**	**E47/94**
non-*Beijing*	**H37Rv**	—	0.689	0.166	0.331	0.113	0.377	0.653	0.185	0.207	0.161	0.108	0.261	0.115
	E865/94	0.311	—	**0.0156**	0.0917	**0.00599**	0.119	0.530	**0.0211**	**0.0281**	**0.0151**	**0.00537**	**0.0479**	**0.00624**
	E729/94	0.834	0.984	—	0.901	0.131	0.958	0.860	0.592	0.68	0.461	0.100	0.849	0.142
	E740/94	0.669	0.908	0.099	—	**0.0242**	0.607	0.770	0.143	0.195	0.0959	**0.0199**	0.328	**0.0259**
	E1221/94	0.887	0.994	0.869	0.976	—	0.994	0.891	0.885	0.916	0.812	0.448	0.975	0.522
	E1449/94	0.623	0.881	**0.0421**	0.393	**0.00632**	—	0.745	0.0721	0.110	**0.0423**	**0.00475**	0.216	**0.00692**
	Harl.	0.347	0.470	0.140	0.23	0.109	0.255	—	0.151	0.163	0.137	0.107	0.192	0.110
Beijing	**E26/95**	0.815	0.979	0.408	0.857	0.115	0.928	0.849	—	0.590	0.380	0.0921	0.774	0.123
	E80/95	0.793	0.972	0.320	0.805	0.0838	0.890	0.837	0.410	—	0.301	0.0672	0.693	0.0901
	E55/94	0.839	0.985	0.539	0.904	0.188	0.958	0.863	0.620	0.699	—	0.157	0.854	0.200
	E26/94	0.892	0.995	0.900	0.980	0.552	0.995	0.893	0.908	0.933	0.843	—	0.982	0.575
	E3942/94	0.739	0.952	0.151	0.672	**0.0246**	0.784	0.808	0.226	0.307	0.146	**0.0182**	—	**0.0270**
	E47/94	0.885	0.994	0.858	0.974	0.478	0.993	0.890	0.877	0.910	0.800	0.425	0.973	—

Table 6: ***p*-values of the two sample tests comparing the mutation probabilities of the strains of Werngren and Hoffner [31, tab. 1] using GF method under *LD* model.** The hypothesis is the following: "the mutation probability of the i-th strain (rows) is significantly higher than that of the j-th strain (columns)". The plating efficiency is set to $\zeta = 0.2$ and the death parameter is assumed to be zero.

		non-*Beijing*							*Beijing*					
		H37Rv	**E865/94**	**E729/94**	**E740/94**	**E1221/94**	**E1449/94**	**Harl.**	**E26/95**	**E80/95**	**E55/94**	**E26/94**	**E3942/94**	**E47/94**
non-*Beijing*	**H37Rv**	⁄	0.434	0.429	0.438	0.430	0.432	0.496	0.429	0.436	0.430	⁄	0.428	0.423
	E865/94	0.566	⁄	0.271	0.598	0.302	0.388	0.563	0.266	0.558	0.334	⁄	0.224	0.0878
	E729/94	0.571	0.729	⁄	0.778	0.542	0.677	0.569	0.476	0.762	0.570	⁄	0.407	0.0860
	E740/94	0.562	0.402	0.222	⁄	0.244	0.306	0.560	0.218	0.457	0.267	⁄	0.190	0.0918
	E1221/94	0.570	0.698	0.458	0.756	⁄	0.628	0.568	0.439	0.736	0.531	⁄	0.375	0.0988
	E1449/94	0.568	0.612	0.323	0.694	0.372	⁄	0.566	0.316	0.663	0.415	⁄	0.246	**0.0482**
	Harl.	0.504	0.437	0.431	0.44	0.432	0.434	⁄	0.431	0.438	0.432	⁄	0.43	0.425
Beijing	**E26/95**	0.571	0.734	0.524	0.782	0.561	0.684	0.569	⁄	0.766	0.584	⁄	0.442	0.133
	E80/95	0.564	0.442	0.238	0.543	0.264	0.337	0.562	0.234	⁄	0.292	⁄	0.200	0.0877
	E55/94	0.570	0.666	0.430	0.733	0.469	0.585	0.568	0.416	0.708	⁄	⁄	0.360	0.122
	E26/94	⁄	⁄	⁄	⁄	⁄	⁄	⁄	⁄	⁄	⁄	⁄	⁄	⁄
	E3942/94	0.572	0.776	0.593	0.810	0.625	0.754	0.570	0.558	0.8	0.64	⁄	⁄	0.110
	E47/94	0.577	0.912	0.914	0.908	0.901	0.952	0.575	0.867	0.912	0.878	⁄	0.890	⁄

Table 7: ***p*-values of the two sample tests comparing the fitness parameters of the strains of Werngren and Hoffner [31, tab. 1] using GF method under *LD* model.** The hypothesis is the following: "the fitness parameter of the i-th strain (rows) is significantly higher than that of the j-th strain (columns)". The plating efficiency is set to $\zeta = 0.2$ and the death parameter is assumed to be zero.

Conclusion

This paper has first proposed an extension to classic mutation model to the case where the plating is not fully efficient. The distribution of the final mutant count has been explicited: it depends on the mean number of mutations m, the relative fitness ρ, the death parameter δ, and the plating efficiency ζ. The estimation problem has been treated: assuming that δ and ζ are known, and that the mutant lifetimes are exponentially i.i.d., the three methods of interest (P0, Maximum Likelihood and Generating Function method) can be applied, as in the case $\zeta = 1$. When the mutant lifetimes are assumed to be constant, only the Generating Function method can be used. The choice of these three specific methods is justified: simulation studies had been performed to compare their results with that of other methods which are still used despite their lack of robustness. An other bias source studied in this paper is the varying final number of cells: it has been previously shown that ignoring these fluctuations induces a negative bias on the mutation probability estimate. Two types of data sets can be considered: for couples (mutation count – final number), Maximum Likelihood can be directly used; in the case where data are mutant counts associated with empirical informations (mean and standard deviation), the correction previously proposed for the P0 method can be adapted to the Maximum Likelihood and the Generating Function method. Remark that the best mutation probabilities estimates are obtained when first type data are used. The R package `flan` including the methods and extensions exposed in this paper is available on the CRAN. It had been used on well-known real data sets, comparing the obtained results with those of authors.

References

[1] W.P. Angerer. "A note on the evaluation of fluctuation experiments". In: *Mutation Research* 479 (2001), pp. 207–224.

[2] W.P. Angerer. "An explicit representation of the Luria-Delbrück distribution". In: *J. Math. Biol.* 42.2 (2001), pp. 145–174.

[3] K.B. Athreya and P.E. Ney. *Branching Processes*. Berlin Heidelberg: Springer, 1972.

[4] R. Bellman and T. Harris. "On age-dependent binary branching processes". In: *Ann. Math.* 55.2 (1952), pp. 280–295.

[5] H.L. David. "Probability distribution of drug-resistant mutants in unselected populations of *Mycobacterium tuberculosis*". In: *Appl. Microbiol.* 20.5 (1970), pp. 810–814.

[6] P. Embrechts and J. Hawkes. "A limit theorem for tails of discrete infinitely divisible laws with applications to fluctuation theory". In: *J. Austral. Math. Soc. Series A* 32 (1982), pp. 412–422.

[7] F. Fontaine, E.J. Stewart, A.B. Lindner, and F. Taddei. "Mutations in two global regulators lower individual mortality in *Escherichia Coli*". In: *Mol. Microbio.* 67.1 (2008), pp. 2–14.

[8] P.L. Foster. "Methods for Determining Spontaneous Mutation Rates". In: *Method. Enzymol.* 409 (2006), pp. 195–213.

[9] P. Gerrish. "A simple formula for obtaining markedly improved mutation rate estimates". In: *Genetics* 180.3 (2008), pp. 1773–1778.

[10] A. Hamon and B. Ycart. "Statistics for the Luria-Delbrück distribution". In: *Elect. J. Statist.* 6 (2012), pp. 1251–1272.

[11] M.E. Jones, S.M. Thomas, and A. Rogers. "Luria-Delbrück Fluctuation Experiments: Design and Analysis". In: *Genetics* 136 (1994), pp. 1209–1216.

[12] A.L. Koch. "Mutation and growth rates from Luria-Delbrück fluctuation tests". In: *Mutation Res.* 95 (1982), p. 129.

[13] N.L. Komarova, L. Wu, and P. Baldi. "The fixed-size Luria-Delbrück model with a nonzero death rate". In: *Math. Biosci.* 210.1 (2007), pp. 253–290.

[14] K.P. Koutsoumanis and A. Lianou. "Stochasticity in colonial growth dynamics of individual bacterial cells". In: *Appl. Environ. Microbiol.* 79.7 (2013), pp. 2294–2301.

[15] D.E. Lea and C.A. Coulson. "The distribution of the number of mutants in bacterial populations". In: *J. Genet.* 49.3 (1949), pp. 264–285.

[16] E.L. Lehmann and G. Casella. *Theory of Point Estimation*. 2e. Springer Texts in Statistics. New York: Springer, 2003.

[17] S.E. Luria. "The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction". In: *Cold Spring Harbor Symp. Quant. Biol.* 16 (1951), pp. 463–470.

[18] S.E. Luria and M. Delbrück. "Mutations of bacteria from virus sensitivity to virus resistance". In: *Genetics* 28.6 (1943), pp. 491–511.

[19] W.T. Ma, G.v.H. Sandri, and S. Sarkar. "Analysis of the Luria-Delbrück distribution using discrete convolution powers". In: *J. Appl. Probab.* 29.2 (1992), pp. 255–267.

[20] M. Marcheselli, A. Baccini, and L. Barabesi. "Parameter estimation for the discrete stable family". In: *Commun. Statist. Theory Methods* 37.6-7 (2008), pp. 815–830.

[21] A. Mazoyer. "Fluctuation analysis on mutation models with birth-date dependence". (submitted). 2017. URL: https://hal.archives-ouvertes.fr/hal-01637808/.

[22] A. Mazoyer. "Time inhomogeneous mutation models with birth-date dependence". In: *B. Math. Biol.* 79.12 (2017), 2929—2953.

[23] A. Mazoyer, R. Drouilhet, S. Despréaux, and B. Ycart. "flan: An R package for inference on mutation models". In: *The R Journal* (2017). URL: https://journal.r-project.org/archive/2017/RJ-2017-029/index.html.

[24] R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2008.

[25] W.A. Rosche and P.L. Foster. "Determining mutation rates in bacterial populations". In: *Methods* 20.1 (2000), pp. 1–17. DOI: 10.1006/meth.1999.0901.

[26] B. Rémillard and R. Theodorescu. "Inference based on the empirical probability generating function for mixtures of Poisson distributions". In: *Statist. Decisions* 18 (2000), pp. 349–366.

[27] E.J. Stewart, R. Madden, G. Paul, and F. Taddei. "Aging and death in an organism that reproduces by morphologically symmetric division". In: *PLoS Biology* 3.2 (2005), pp. 295–300.

[28] F.M. Stewart. "Fluctuation Tests: How Reliable Are the Estimates of Mutation Rates?" In: *Genetics* 137.4 (1994), pp. 1139–1146.

[29] F.M. Stewart, D.M. Gordon, and B.R. Levin. "Fluctuation analysis: the probability distribution of the number of mutants under different conditions". In: *Genetics* 124.1 (1990), pp. 175–185.

[30] L. Wasserman. *All of Statistics: a concise course in statistical inference*. New York: Springer, 2004.

[31] J. Werngren and S.E. Hoffner. "Drug susceptible *Mycobacterium tuberculosis Beijing* genotype does not develop motation-conferred resistance to Rifampin at an elevated rate". In: *J. Clin. Microbiol.* 41.4 (2003), pp. 1520–1524.

[32] R. Wilcox. *Introduction to Robust Estimation and Hypothesis Testing*. 3^{e}. Amsterdam: Elsevier, 2012.

[33] B. Ycart. "Fluctuation analysis: can estimates be trusted?" In: *PLoS One* 8.12 (2013), pp. 1–12. URL: http://dx.doi.org/10.1371%2Fjournal.pone.0080958.

[34] B. Ycart. "Fluctuation analysis with cell deaths". In: *J. Appl. Probab. Statist* 9.1 (2014), pp. 13–29.

[35] B. Ycart and N. Veziris. "Unbiased estimates of mutation rates under fluctuating final counts". In: *PLoS One* 9.7 (2014), pp. 1–10. URL: http://dx.doi.org/10.1371/journal.pone.0101434.

[36] Q. Zheng. "Progress of a half century in the study of the Luria-Delbrück distribution". In: *Math. Biosci.* 162 (1999), pp. 1–32.

[37] Q. Zheng. "New algorithms for Luria-Delbrück fluctuation analysis". In: *Math. Biosci.* 196.2 (2005), pp. 198–214.