

STATISTICAL STUDY OF CELLULAR AGING

JULIEN GUYON¹, ARIANE BIZE¹, GRÉGORIY PAUL², ERIC STEWART², JEAN-FRANCOIS DELMAS¹ AND FRANCOIS TADDÉI²

Abstract. Whereas aging is obvious in macroscopic organisms, it is not in single-celled ones, such as the model organism *E. Coli*, where one has the best chances of describing and quantifying the molecular process involved. To determine if *E. Coli* experiences aging related to the inheritance of the old pole, E. J. STEWART *et al.* [1] followed 94 individual exponentially growing cells through up to nine generations, determining the complete lineage and the growth rate of each cell. Averaging over these 94 experiments leads the authors to the conclusion that the two supposedly identical cells produced during cell division are functionally asymmetric.

This preliminary study, avoiding inter-experiment averaging, aims at reasoning experiment-wise and describing the structure of dependency within a lineage of *E. Coli*. We propose a single-experiment model for the dynamics of such a lineage of growth rates. We fit it to the data and perform experiment-wise tests which lead us to accept that the old pole progeny cell experiences slowed growth rate and hence should be considered an aging parent repeatedly producing rejuvenated offspring. Eventually, as a first step towards answering statistical questions about markovian models on binary trees, we run simulations of the model and conjecture results. At this preliminary stage, we have not used the results on bifurcative autoregressive models [2] which seem to be an appropriate background for the gaussian model.

Résumé. S'il est facile d'identifier le vieillissement chez les organismes macroscopiques, il en va autrement chez les êtres unicellulaires tel l'organisme modèle *E. Coli* où l'on a pourtant les meilleures chances de décrire et de quantifier les processus moléculaires qu'il met en jeu. Afin de savoir si l'héritage du vieux pôle parental est source de vieillissement chez *E. Coli*, E. J. STEWART *et al.* [1] ont suivi 94 cellules dans leur développement exponentiel pendant 9 générations, construisant 94 arbres généalogiques et y reportant les taux de croissance de chacun des descendants. En moyennant ces arbres, les auteurs concluent à l'assymétrie fonctionnelle des deux cellules filles produites à chaque division et jusque là supposées identiques.

Dans cette étude préliminaire, nous évitons le recours au moyennage inter-expériences et nous raisonnons expérience par expérience, décrivant la structure de dépendance au sein d'un lignage de *E. Coli*. Nous proposons pour chaque expérience un modèle de dynamique d'un tel lignage de taux de croissance. Nous le calons aux données et réalisons pour chaque expérience prise individuellement un test qui nous conduit à accepter l'hypothèse selon laquelle que la cellule fille qui hérite du vieux pôle de sa mère croît moins vite que sa soeur et doit être considérée comme un parent vieillissant pouvant donner naissance à une progéniture rajeunie. Enfin, des simulations du modèle nous permettent de conjecturer des résultats sur les modèles markoviens sur les arbres binaires. A ce stade préliminaire de l'étude, nous n'avons pas utilisé les résultats sur les modèles autorégressifs avec bifurcation [2] qui semblent fournir un cadre adapté au modèle gaussien.

¹ ENPC-CERMICS, 6 av. Blaise Pascal, Champs-sur-Marne, 77455 Marne La Vallée, France; e-mail: julien.guyon@cermics.enpc.fr & ariane.bize@cermics.enpc.fr & delmas@cermics.enpc.fr

² Génétique Moléculaire Evolutive et Médicale, INSERM U571, Faculté de Médecine, Université René Descartes-Paris V, 156 rue de Vaugirard, 75730 Paris Cedex 15, France; e-mail: gregory.paul@necker.fr & stewart@necker.fr & taddei@necker.fr

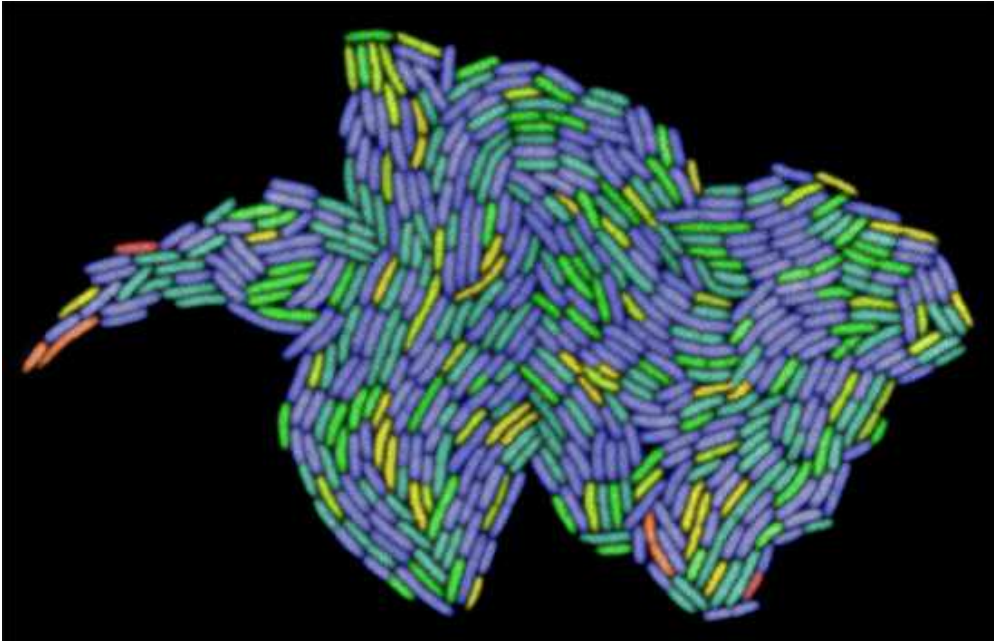


FIGURE 1. "What is true for *E. coli* is true for the elephant" As aging requires a differential distribution of aged and young components between parent and offspring, it has been postulated that organisms that reproduce by symmetrical division do not age. It has now been shown, however, that in the symmetrically dividing bacterium *Escherichia coli* aging can occur even in the absence of both a visible asymmetry and a juvenile phase. This false-color image shows a growing microcolony of *E. coli*, where the cells are colored by age (in numbers of divisions). The oldest cells are red, and the youngest blue. Original quote attributed to Jacques Monod. Cover credit: Photograph provided by Eric Stewart and Stefanie Timmermann, Inserm U571, Faculté de Médecine Necker Enfants-Malades, Paris, France.

INTRODUCTION

E. J. STEWART, G. PAUL, F. TADDÉI, from the Laboratoire de Génétique moléculaire, évolutive et médicale (INSERM U571, Faculté de Médecine Necker) and R. MADDEN, from the Institut des Hautes Etudes Scientifiques, have been working on *Escherichia Coli*'s aging [1]. *E. Coli* is a single-celled, model organism. It has been widely studied by the biologists who have gathered great information on it. Whereas aging is obvious in macroscopic organisms, it is not in single-celled ones, where one has the best chances of describing and quantifying the molecular process involved. It is especially hard to identify in *E. Coli*, which reproduces without a juvenile phase and with an apparently symmetric division. E. J. STEWART *et al.* have designed an experimental protocol which brings evidence of aging in *E. Coli* and we propose a statistical study of the data they get.

In Section 1, we describe the biological experiment and present the data. Inter-experiment averaging shows a clear segregation between the new and old poles (see Subsection 1.1), whereas single-experiment data does not. In Section 2, we propose a mathematical model for this data and perform single-experiment statistical tests that lead us to accept aging in *E. Coli*. The model consists of a markovian dynamics on a binary tree. It raises interesting statistical questions. In Section 3, as a first step towards answering them, we run simulations of the model and conjecture some results.

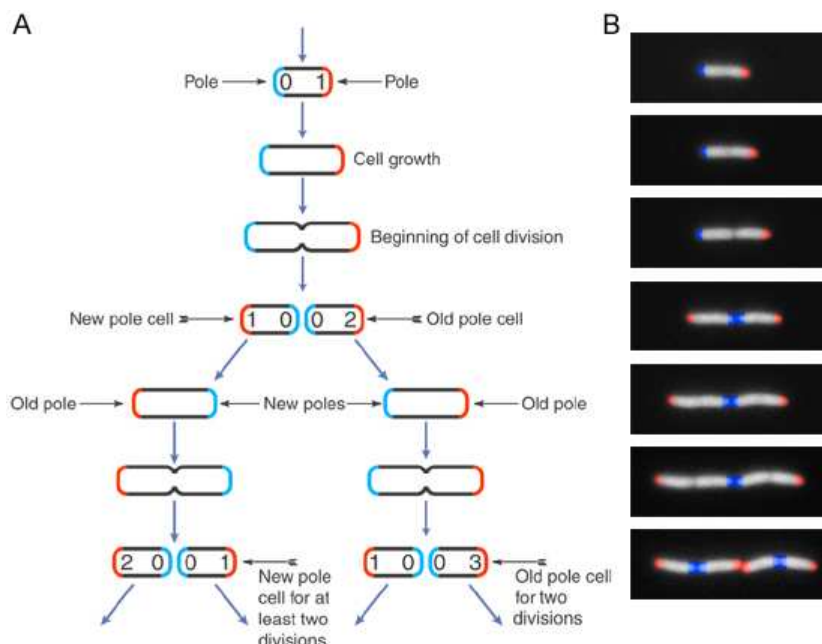


FIGURE 2. The life cycle of *E. coli*, from E. J. STEWART *et al.* [1]

1. THE BIOLOGY

Here we briefly describe *E. Coli*'s life cycle, the experiment designed by E. J. STEWART *et al.* and the data they get. Figures 2, 5 and 6 are taken from [1] where one can find further information.

1.1. The experiment

E. Coli is a rod-shaped bacterium. It reproduces by dividing in the middle, producing a new end per progeny cell, see Figure 2. This new end is called new pole, whereas the other end is pre-existing and is called old pole.

This defines an age in divisions for each pole, and hence for each cell. One expects any cell component formed in the poles and with limited diffusion to accumulate at the old pole, so that there might be a physiological asymmetry between the old and new poles. To determine if *E. Coli* experiences aging related to the inheritance of the old pole, E. J. STEWART *et al.* followed 94 individual exponentially growing cells through up to nine generations in an automated fluorescence microscopy system which allowed them to determine the complete lineage, the identity of each pole and, among other physical parameters, the growth rate of each cell. In the next two sections, we present their results.

1.2. Original data

Each of the 94 films give rise to a genealogical tree such as the ones in Figure 3. The new poles are in blue and the old poles in red. On the y -axis appears the growth rate, whereas the x -axis displays time in divisions. There is no striking evidence for reproductive asymmetry between the progeny cells visible to the naked eye.

1.3. Averaged data

In order to eliminate the random effects which appear in Figure 3, E. J. STEWART *et al.* have averaged the 94 lineages by each unique cell position within the lineage. Figure 4 is the average tree thus produced. It makes

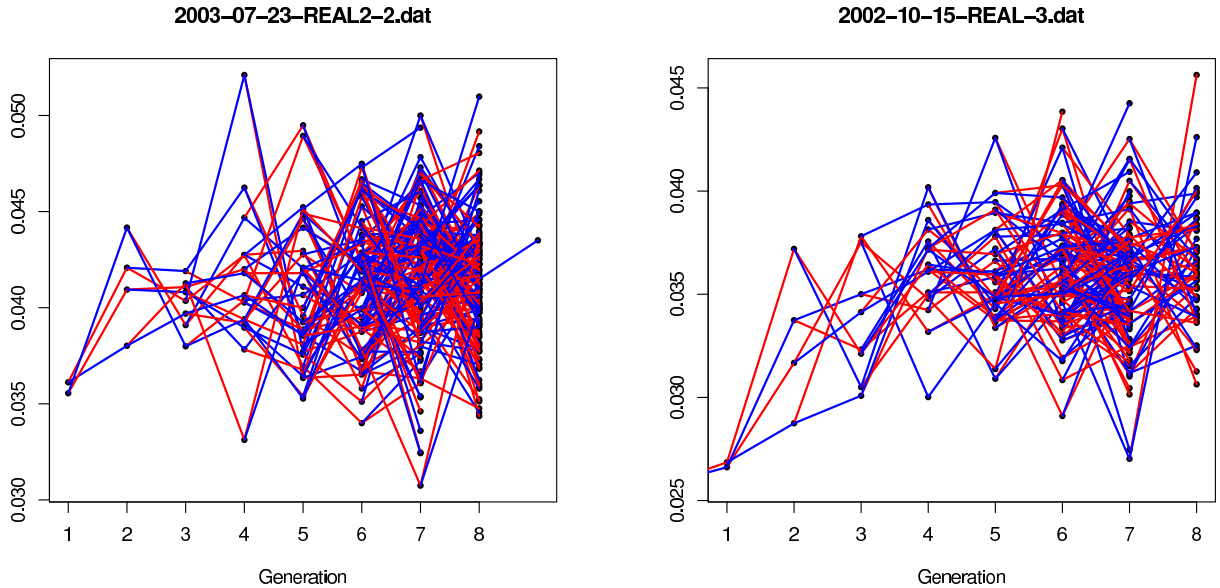


FIGURE 3. Two single-experiment data trees

clearly appear a segregation between the new and old poles. The latter experience cumulatively slowed growth, which is an evidence for aging.

In Figure 5, each branch point is an average cell for that position in the lineage, and the length of the lines connecting cells to their progeny are proportional to the growth rate of the cell. At each division, the cell inheriting the old pole of the progenitor cell is on the right branch, in red. Again there is striking evidence for cumulatively slowed growth in the old poles.

Eventually, E. J. STEWART *et al.* performed a pair-wise comparison of the 7,953 sets of sister cells that were produced at the eighth generation in each of the films, see Figure 6. This analysis shows that the average growth rate of the old pole progeny cell is 2.2% slower than that of the new pole cell.

Such average graphs clearly show a functional asymmetry between the old and new poles. However, averaging between experiments can be problematic, since they are not independent and identically distributed. Indeed, two initial cells giving rise to two different films are actually taken from the same macrocolony, so that there might be a correlation between the experiments. Moreover, as shown in Figure 3, the range of values of the growth rate changes from experiment to experiment, probably due to a slight change in the experimental conditions. In the next section, we propose a statistical model that allows to segregate between the populations of old and new poles experiment-wise. It also has the advantage to take into account the structure of dependency within a lineage.

2. THE MATHEMATICAL MODEL

2.1. Description and objectives

In order to describe the dynamics of the growth rate, we propose the following markovian model with memory 1: $(\rho_2, \rho_3) \in \mathbb{R}_+^2$ and for all $n \geq 2$,

$$\begin{cases} \rho_{2n} &= \alpha_0 \rho_n + \beta_0 + \varepsilon_{2n}, \\ \rho_{2n+1} &= \alpha_1 \rho_n + \beta_1 + \varepsilon_{2n+1}. \end{cases} \quad (1)$$

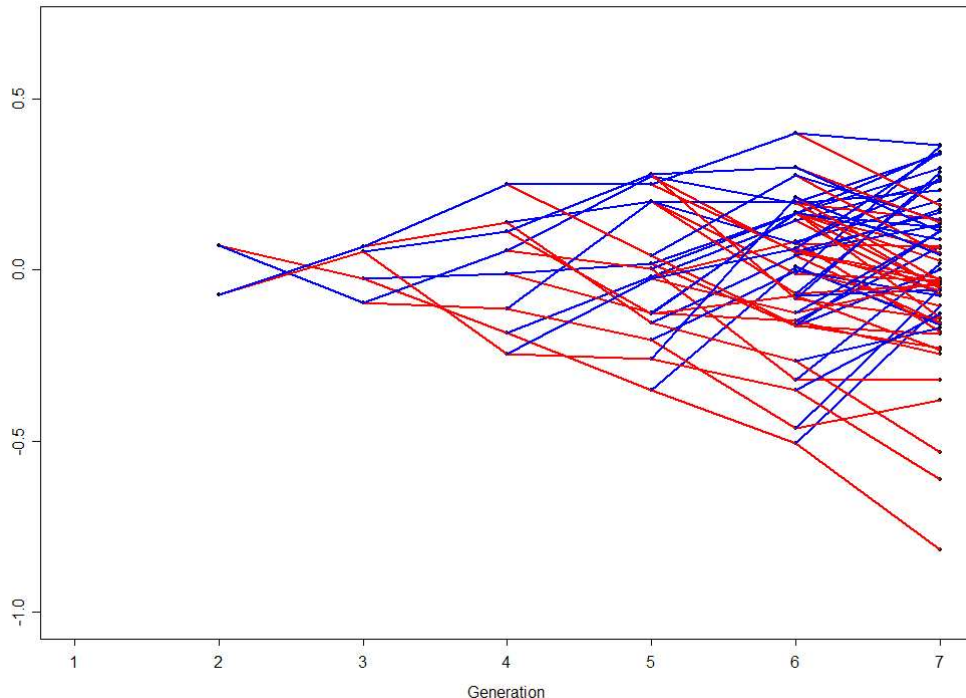


FIGURE 4. The average data tree

In this writing, ρ_i is the growth rate of individual i and n is the mother of $2n$ - the new pole progeny cell - and $2n + 1$ - the old pole progeny cell. The ε_i 's form a sequence of independent and identically distributed (i.i.d.) centered gaussian random variables with common variance σ^2 , independent on the initial condition (ρ_2, ρ_3) . We speak of memory 1 because a cell's growth rate is explained by its mother's. For instance, a markovian model with memory 2 would also take into account the grandmother's growth rate. Since we do not know which of the first two sisters is the new pole progeny cell issued from individual 1, we do not make (ρ_2, ρ_3) part of the dynamics and take it as the initial condition.

On each germinal line, (1) defines a Markov chain on the real axis. This Markov chain *is not* homogeneous since the one-step dynamics depend on whether you are the new pole progeny cell or the old pole one.

The model should also be thought of as a gaussian linear model, since the ε_i 's are i.i.d. gaussian variables and the dependency between a cell's growth rate and its mother's is assumed to be affine: $\mathbf{Y} = M\theta + \varepsilon$ with

$$\mathbf{Y} = \begin{pmatrix} \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \\ \rho_9 \\ \vdots \end{pmatrix}, \quad M = \begin{pmatrix} \rho_2 & 1 & 0 & 0 \\ 0 & 0 & \rho_2 & 1 \\ \rho_3 & 1 & 0 & 0 \\ 0 & 0 & \rho_3 & 1 \\ \rho_4 & 1 & 0 & 0 \\ 0 & 0 & \rho_4 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \quad \theta = \begin{pmatrix} \alpha_0 \\ \beta_0 \\ \alpha_1 \\ \beta_1 \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \vdots \end{pmatrix}. \quad (2)$$

Nevertheless, the explaining factor, the sequence $(\rho_i, i \in \mathbb{N})$ itself, is stochastic, so that one cannot blindly apply the gaussian linear model theory which assumes the explaining factor to be deterministic.

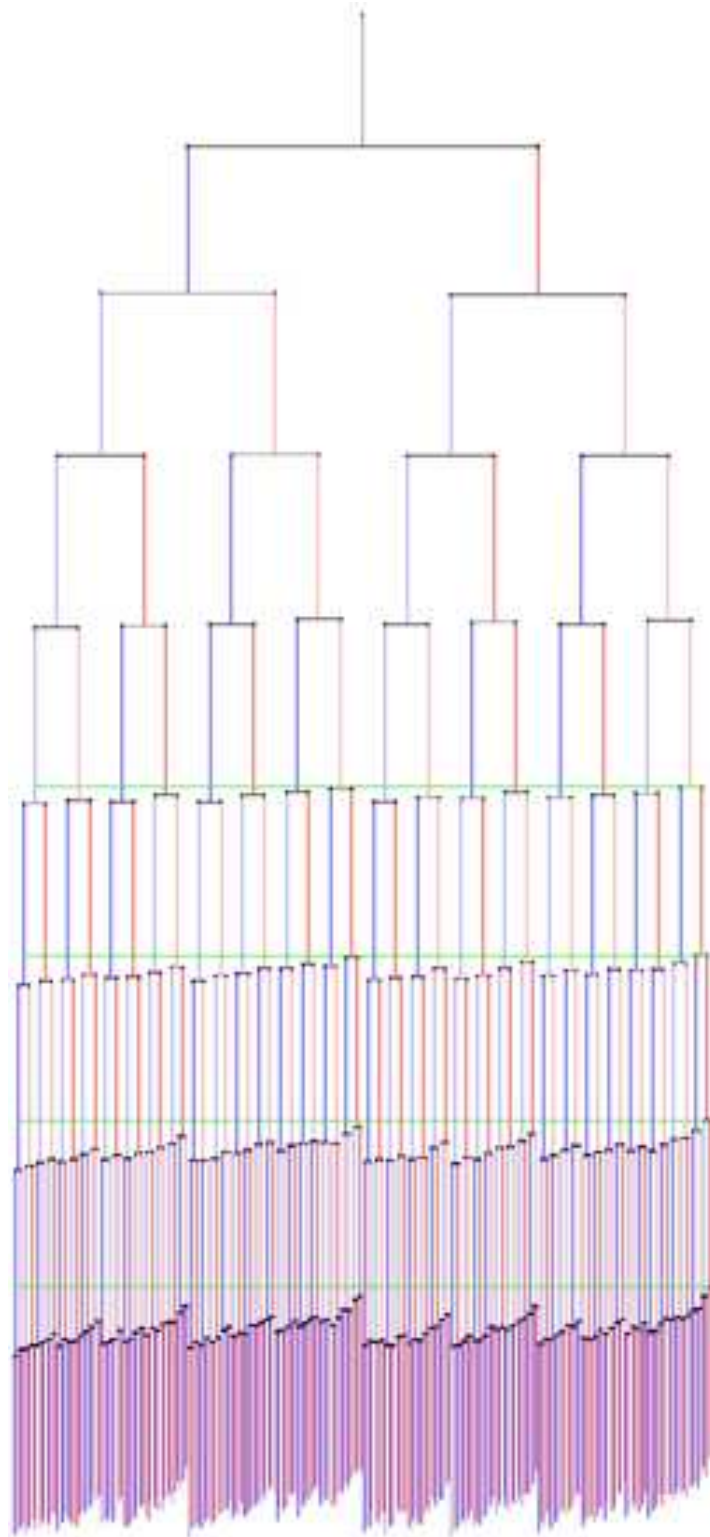


FIGURE 5. Average lineage showing old pole effect on growth rate, from E. J. STEWART *et al.* [1]

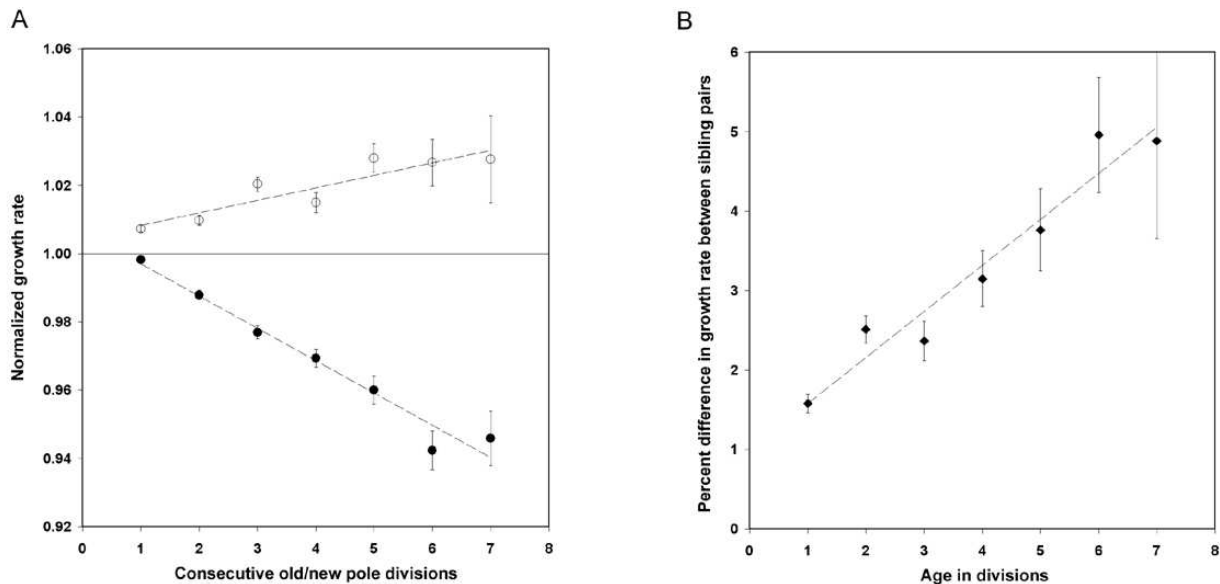


FIGURE 6. The effects of consecutive divisions as an old or new pole on growth rate, from E. J. STEWART *et al.* [1]

We aim at

- (i) estimating the 4-dimensional parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ and σ^2 ,
- (ii) testing the null hypothesis $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ against its alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$,
- (iii) testing the validity of the model.

In view of the biological question addressed here, (ii) is crucial: rejecting H_0 comes down to accepting that the old pole progeny cell experiences slowed growth rate and hence should be considered an aging parent repeatedly producing rejuvenated offspring.

2.2. A parameter estimator

We seek to estimate the parameters of the model. Each of the 94 experiments gives rise to a genealogical binary tree I which is not necessarily complete. I should be seen as a set of individuals, each of them given by its number through the rule: i is the mother of $2i$ - the new pole progeny cell - and $2i + 1$ - the old pole progeny cell. Incompleteness means that we can have $I \cap \{2^n, \dots, 2^{n+1} - 1\} \neq \{2^n, \dots, 2^{n+1} - 1\}$, see for instance Figure 3. However, since we cannot observe a cell without observing its mother, for $i \geq 1$ we have that $2i \in I \Rightarrow i \in I$ and $2i + 1 \in I \Rightarrow i \in I$. We denote by I_0 the set of new pole progeny cells of I , *i.e.* $I_0 = I \cap (2\mathbb{N})$, and by I_1 the set of old pole progeny cells of I , *i.e.* $I_1 = I \cap (2\mathbb{N} + 1)$. We also denote $I_0^\uparrow = \{i \in I | 2i \in I_0\}$ and $I_1^\uparrow = \{i \in I | 2i + 1 \in I_1\}$ the sets of new poles' (resp. old poles') mothers. Since the ε_i 's are i.i.d. gaussian variables, the maximum likelihood estimator $\hat{\theta} = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\alpha}_1, \hat{\beta}_1)$ of θ is also the least squares one:

$$\begin{cases} \hat{\alpha}_0 &= \text{Cov}(\rho_{I_0}, \rho_{I_0^\uparrow}) / \text{Var}(\rho_{I_0^\uparrow}), \\ \hat{\beta}_0 &= \bar{\rho}_{I_0} - \hat{\alpha}_0 \bar{\rho}_{I_0^\uparrow}, \\ \hat{\alpha}_1 &= \text{Cov}(\rho_{I_1}, \rho_{I_1^\uparrow}) / \text{Var}(\rho_{I_1^\uparrow}), \\ \hat{\beta}_1 &= \bar{\rho}_{I_1} - \hat{\alpha}_1 \bar{\rho}_{I_1^\uparrow}. \end{cases} \quad (3)$$

parameter	$\hat{\alpha}_0$	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\sigma}^2$
experiment 1	-0.1089	0.0460	0.0977	0.0370	$1.1 \cdot 10^{-5}$
experiment 2	0.0261	0.0354	-0.1247	0.0399	$7.8 \cdot 10^{-6}$

TABLE 1. Estimation of the parameters of the model

In this writing, $\#$ denoting the cardinality,

$$\begin{aligned}\bar{\rho}_{I_0} &= \frac{1}{\#I_0} \sum_{i \in I_0} \rho_i, \\ \bar{\rho}_{I_0^\dagger} &= \frac{1}{\#I_0} \sum_{2i \in I_0} \rho_i, \\ \text{Cov}(\rho_{I_0}, \rho_{I_0^\dagger}) &= \frac{1}{\#I_0} \sum_{2i \in I_0} \rho_{2i} \rho_i - \bar{\rho}_{I_0} \bar{\rho}_{I_0^\dagger}, \\ \text{Var}(\rho_{I_0^\dagger}) &= \frac{1}{\#I_0} \sum_{2i \in I_0} \rho_i^2 - \bar{\rho}_{I_0^\dagger}^2\end{aligned}$$

respectively stand for the new pole progeny cell's empirical growth rate mean, their mother's empirical growth rate mean, the empirical covariance between mothers' and their new pole progeny cell's growth rates and the empirical variance of the mothers' growth rates. Similar definitions hold for the old poles population.

Note that, like in the linear gaussian model theory, (3) can be written as

$$\hat{\theta} = ({}^t M M)^{-1} {}^t M \mathbf{Y},$$

where M and \mathbf{Y} are respectively the $\#I \times 4$ matrix and the $\#I$ dimensional column vector defined in (2). We then estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{\#I - 4} \left\| \mathbf{Y} - M \hat{\theta} \right\|^2,$$

where $\|\cdot\|$ denotes the euclidian norm.

In Table 1, we give the estimated parameters for the 2 experiments of Figure 3.

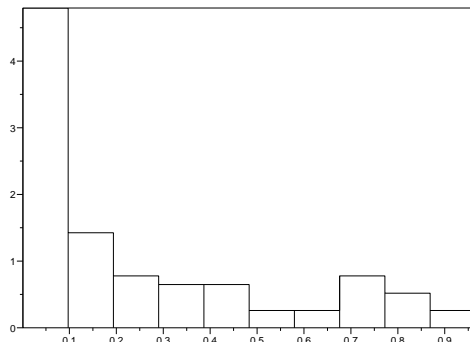
2.3. Two approached tests

Rejecting the null hypothesis $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ boils down to accepting that *E. Coli* experiences aging related to the inheritance of the old pole. To do so, we perform two different approached tests.

2.3.1. Wilcoxon's test

To start with, we perform Wilcoxon's test applied to the growth rates of sister cells. It is a non parametric test which allows to segregate between new and old poles in a more general background than the gaussian one we just presented. The null hypothesis is that growth rates distributions are independant of the age of poles in the cell, and thus if the data was independent old poles and new poles cells' growth rate would follow the same distribution. Two sister cells form a pair. For each pair, we compute the difference between the growth rates. We order these differences according to ascending absolute values. We are then able to attribute a rank k to each pair according to this new order, and call $D(k)$ the difference of growth rate between the sister cells of the k^{th} pair. For each set of data, we compute

$$T^+ = \sum_k k \mathbf{1}_{\{D(k) > 0\}}.$$

FIGURE 7. Histogram of the p -values for Wilcoxon's test

parameter	$\#I$	F	p -value
experiment 1	382	3.1475	0.0441
experiment 2	294	4.3159	0.0142

TABLE 2. Testing $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ against its alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$

Thus, T^+ is the sum of the ranks for pairs showing a positive difference. Let us call p the number of pairs of sister cells. If the data was independent, under the null hypothesis, the expectancy and the variance of T^+ are given by $\mathbb{E}[T^+] = p(p+1)/4$ and $\text{Var}(T^+) = p(p+1)(2p+1)/24$ and $Z = \frac{T^+ - p(p+1)/4}{\sqrt{p(p+1)(2p+1)/24}}$ converges in law towards a gaussian law of parameters (0,1) when p tends to infinity. For each set of data, we compute Z and the associated p -values. Figure 7 shows an histogram of them. A uniform distribution of the p -values would indicate that the old and new poles are indistinguishable. Such is not the case. Values produced are in great majority inferior to 0.1. This first test hence indicates that there is a difference between the new and old poles in terms of growth rate. It is inaccurate since the data is not independent: a cell's growth rate is linked to its mother's, so that there is a dependance between cells throughout the tree, both horizontally and vertically. It is nevertheless useful as a first impression and all the more convincing as it shows a clear answer.

2.3.2. Fisher's test

We then proceed to a Fisher's test. As in the linear gaussian model, we compute the least squares estimator $\hat{\theta}_0$ of θ under H_0 and estimate \mathbf{Y} under H_0 by $\hat{\mathbf{Y}}_0 = M\hat{\theta}_0$ and under the general model by $\hat{\mathbf{Y}} = M\hat{\theta}$. Then we compute Fisher's statistic

$$F = \frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_0\|^2}{2\hat{\sigma}^2}. \quad (4)$$

Even though the explaining factor M is stochastic, which prevents one to state that F follows Fisher's law $\mathcal{F}_{2, \#I-4}$ with 2 and $\#I-4$ degrees of freedom, one expects F to take "small" values under H_0 and to tend to infinity as $\#I$ increases under the alternative hypothesis H_1 . See Section 3 for simulations of the random variable F under both hypothesis.

Making as if F followed $\mathcal{F}_{2, \#I-4}$, we compute its p -value for each of the 94 films. For the two experiments of Figure 3, the results are given in Table 2.

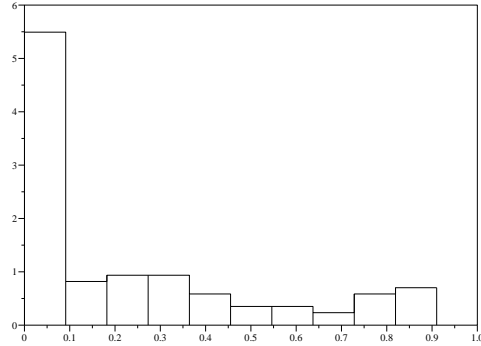
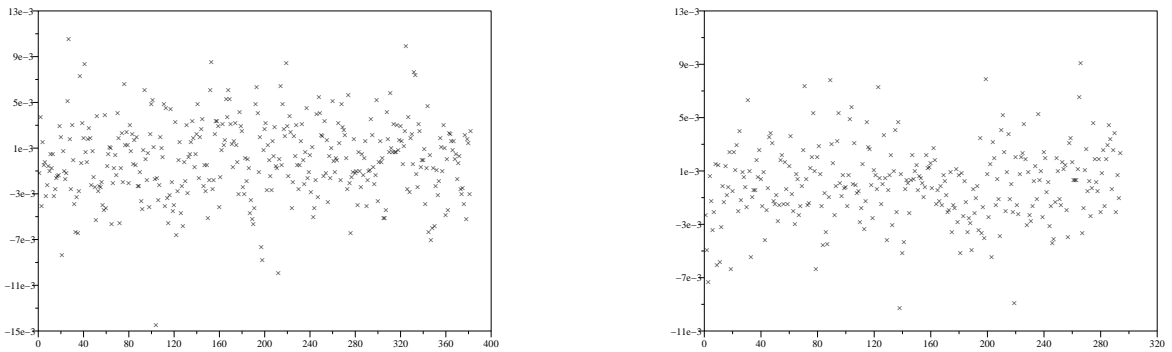
FIGURE 8. Histogram of the p -values produced by Fisher's test

FIGURE 9. Two single-experiment residuals

Figure 8 shows the histogram of the 94 p -values thus produced. Even if F is only an approximated Fisher's random variable, we now have a strong experiment-wise indication for a segregation between new pole progeny cells and old pole ones in terms of growth rate, since under H_0 one would have expected a uniformly distributed histogram of the p -values.

To conclude, Wilcoxon's and Fisher's tests convince us that the experimental data show a difference between the new and old poles.

2.4. Validity of the model

In order to check the validity of the model, we perform two tests.

The first one consists in computing the residual $\hat{\varepsilon} = \mathbf{Y} - M\hat{\theta}$. If (1) is a correct description of the dynamics of the growth rates, one expects $\hat{\varepsilon}$ to be "close" to a realization of a sequence of i.i.d. gaussian variables. For the two experiments of Figure 3, we graph $i \mapsto \hat{\varepsilon}_i$ and the corresponding histograms, see Figures 9 and 10.

Assuming that each such time series are independent draws from a given law, we want to check whether the latter is a centered gaussian law or not. To this end, we evaluate the variance $\hat{\sigma}^2$, divide \mathbb{R} into 10

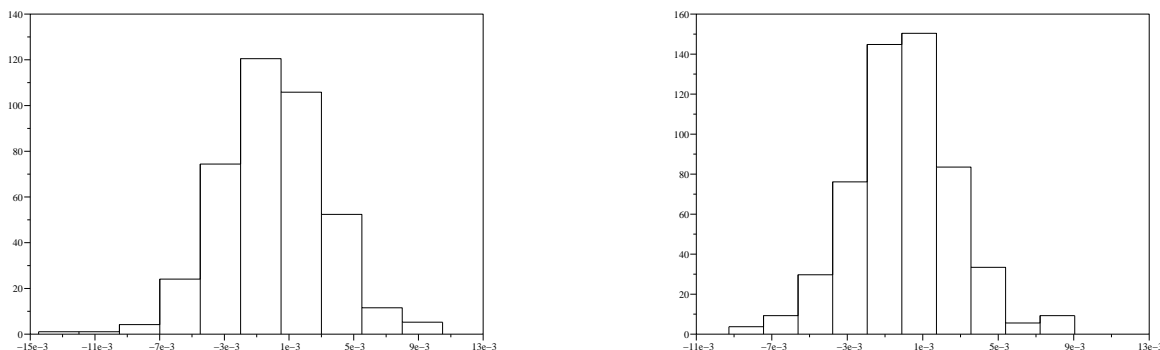


FIGURE 10. Histogram of two single-experiment residuals

parameter	#I	χ^2	p-value	adequation
experiment 1	382	5.8534	0.6636	true
experiment 2	294	12.3946	0.1344	true

TABLE 3. Adequation of the distribution of the ε_i 's to a centered gaussian law

parameter	R^2
experiment 1	0.0164
experiment 2	0.0318

TABLE 4. Percentage of variance explained by the model

$\mathcal{N}(0, \hat{\sigma}^2)$ equiprobable intervals and perform a 8 degrees of freedom χ^2 -test. We come to the conclusion that the distribution of the ε_i 's agrees with a centered gaussian law, see Table 3.

The second test consists in computing the percentage of variance explained by the model, *i.e.* $R^2 = \text{Var}(\hat{\mathbf{Y}})/\text{Var}(\mathbf{Y})$. If (1) correctly captures the aggregated stochastic effects, one expects R^2 to be close to 1. For the two experiments of Figure 3, the results are given in Table 4.

This result is unsatisfying and leads us to envisage more complex models. Such models would take into account a more complex structure of dependency. For example, one can make a cell's growth rate depend on its grandmother's, or on its sister's - this could be done by introducing an unknown correlation between ε_{2n} and ε_{2n+1} .

2.5. Open questions

Many interesting mathematical questions remain open:

- What is the law of $\hat{\theta}$? of $\hat{\sigma}^2$? Do we have an almost sure convergence to the true parameters as the size of the tree tends to infinity ? a central limit theorem ? with which speed ?
- What is the law of F ? How to build an exact test for H_0 ?
- What is the law of $\hat{\varepsilon}$?

parameter	estim	CI	fixed value
α_0	0.4075	0.4051 – 0.4099	0.41
β_0	0.0221	0.0220 – 0.0222	0.022
α_1	0.4269	0.4244 – 0.4294	0.43
β_1	0.0201	0.0200 – 0.0202	0.020
σ^2	$9 \cdot 10^{-6}$	$9 \cdot 10^{-6} - 9 \cdot 10^{-6}$	$9 \cdot 10^{-6}$

TABLE 5. Estimation of parameters of the model

	mean	variance	χ^2	degree	0.95-quantile	adequation
$\hat{\alpha}_0$	0.4075257	0.0015103	6.36	7	14.07	true
$\hat{\alpha}_1$	0.4269006	0.0016162	6.3	7	14.07	true
$\hat{\beta}_0$	0.0221010	0.0000020	3.3	7	14.07	true
$\hat{\beta}_1$	0.0201138	0.0000021	10.08	7	14.07	true
$\hat{\sigma}^2$	0.0000090	$1.5111 \cdot 10^{-13}$	8.36	7	14.07	true
$\hat{\alpha}_0 - \hat{\alpha}_1$	-0.0193746	0.0030630	117.9	8	15.51	false
$\hat{\beta}_0 - \hat{\beta}_1$	0.0019872	0.0000041	1154.9	8	15.51	false

TABLE 6. Adequation of estimated parameters to gaussian laws

As a first step towards answering these questions, we now perform simulations of model (1). See [2] for strong law of large numbers and central limit theorem in bifurcating autoregressive models.

3. SIMULATIONS

Following the analysis of the experimental data, let us run simulations of model (1). We choose a tree depth of 10, and fix ρ_4, ρ_5, ρ_6 and ρ_7 . Thus, we study 1016 cells ($\#I = 1016$).

3.1. Estimation of the parameters

We proceed to 1000 simulations to produce Monte Carlo data. For each simulation, we estimate the parameters as in (3). We can then compute the empiric mean and variance and the corresponding 95%-confidence interval for each parameter. Table 5 shows one set of results. The parameters' estimators seem to converge towards the true values.

We then test the adequation of $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}^2$, to gaussian laws. The parameters used for the gaussian law are the empirical means and variances. We use the same set of simulations (1000), thus the fixed parameters are $\alpha_0 = 0.41, \alpha_1 = 0.43, \beta_0 = 0.22, \beta_1 = 0.20$ and $\sigma^2 = 0.000009$. We find a good adequation to normal distributions (see Table 6), which suggests that a central limit theorem holds.

We proceed to the same type of adequation tests to compare the distributions of $\hat{\alpha}_0 - \hat{\alpha}_1$ and $\hat{\beta}_0 - \hat{\beta}_1$ to *centered* gaussian laws. We still use the empiric variance. We find high values of χ^2 , rejecting the adequation at the 5% significance level: this means that $\hat{\alpha}_0$ seems to be different from $\hat{\alpha}_1$ and $\hat{\beta}_0$ from $\hat{\beta}_1$: $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$ seems to be true (see Table 6). Some of the histograms produced are shown in Figure 11.

3.2. Fisher's test

We also perform Fisher's test as described for the experimental data (see Subsection 2.3.2). The aim is to check the behaviour of the test according to the initial value of the parameters, as, in fact, F does not follow a Fisher's law and we do not know its law. We want to check that F takes "low" values under H_0 and "high"

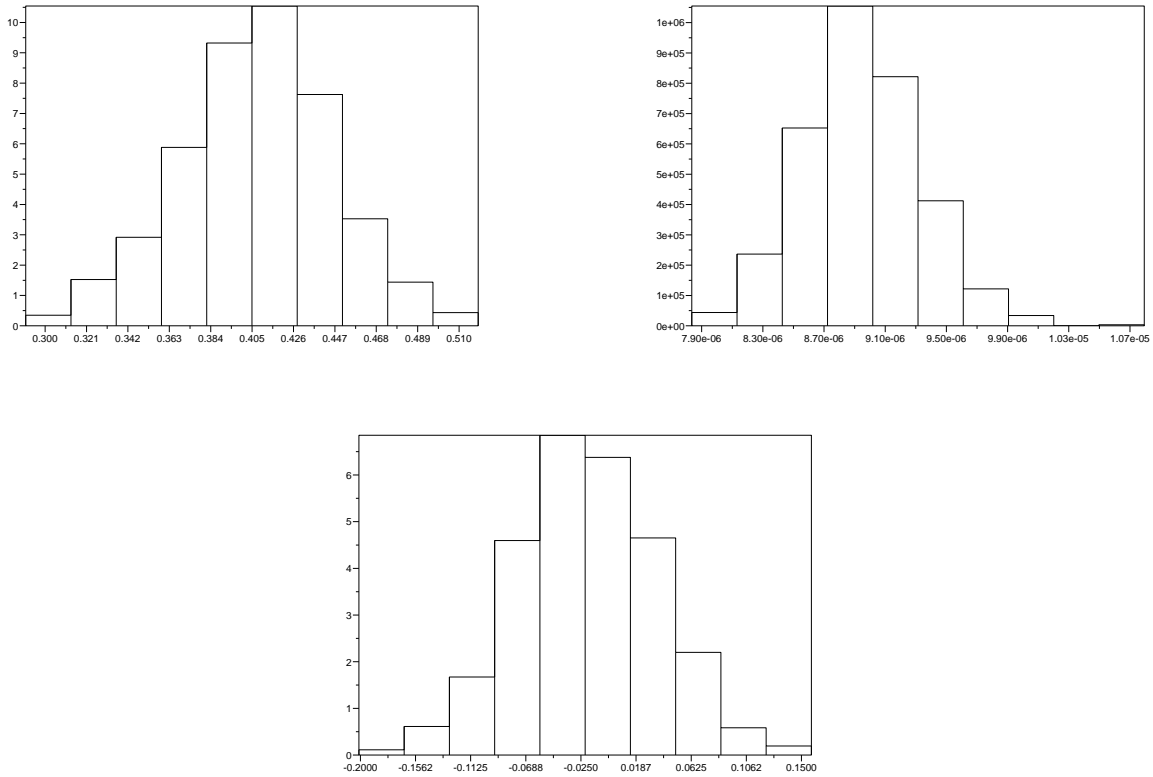


FIGURE 11. Histograms of estimated parameters - Left: $\hat{\alpha}_0$ - Right: $\hat{\sigma}^2$ - Bottom: $\hat{\alpha}_0 - \hat{\alpha}_1$

values under H_1 . In all cases, $\sigma = 0.003$. We first choose $(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)$, with $\alpha_0 = 0.41$, $\beta_0 = 0.022$, $\alpha_1 = 0.43$ and $\beta_1 = 0.020$. We find test values centered around 25. This value of 25 should be compared to 3.00, the quantile of order 0.95 for the Fisher function of parameters (2, 1012). Then, we use $\alpha_0 = \alpha_1 = 0.41$, $\beta_0 = \beta_1 = 0.022$. We find values in majority inferior to 1. Figure 12 shows the 2 histograms of the test values. As a conclusion, the test has the expected behaviour and should allow us to discriminate equality of the parameters.

3.3. The residues

We also study the residues. They are computed using the estimated parameters. We first examine the correlation between the residues of two sister cells, and between the residues of mother and daughter cells. A rank n (60) is arbitrarily chosen. We compute the correlations between $\hat{\varepsilon}_n$ and $\hat{\varepsilon}_{2n}$, $\hat{\varepsilon}_n$ and $\hat{\varepsilon}_{2n+1}$, $\hat{\varepsilon}_{2n}$ and $\hat{\varepsilon}_{2n+1}$, on the basis of 100 simulations. And this is repeated a hundred times in order to get an empirical mean and an empirical variance for the different correlations. We find them very low. The former is about 10^{-8} and the latter is of the order of 10^{-12} . Figure 13 shows the histogram for these correlation values. Such low values should indicate that the $\hat{\varepsilon}_n$ have little correlation.

We then observe adequation of the $\hat{\varepsilon}_n$ to a gaussian law. We still use the arbitrary rank 60 but proceed to 1000 simulations. We calculate the mean m and the variance v of the $\hat{\varepsilon}_{60}$ thus obtained and we find $m = 0.0000368$ and $v = 0.000090$. Following this, we test the adequation of the residues to a gaussian law of parameters (m, v) .

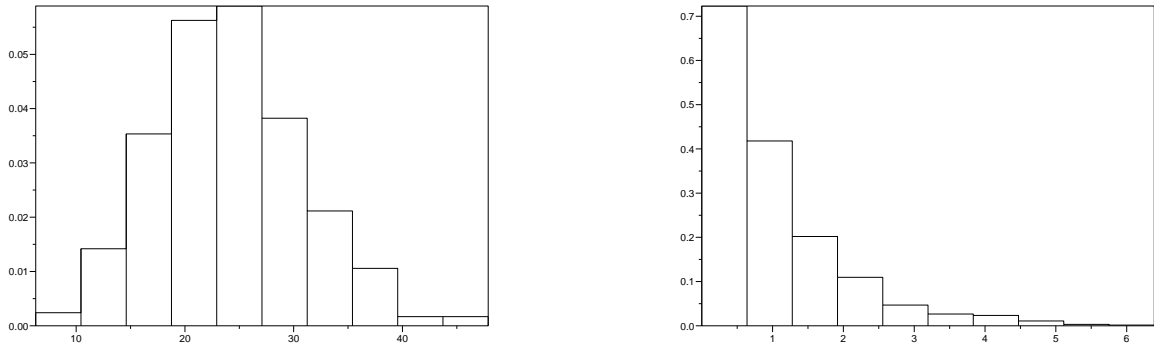


FIGURE 12. Histogram of the values of Fisher's statistic on simulated data - Left: $(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)$ - Right: $(\alpha_0, \beta_0) = (\alpha_1, \beta_1)$

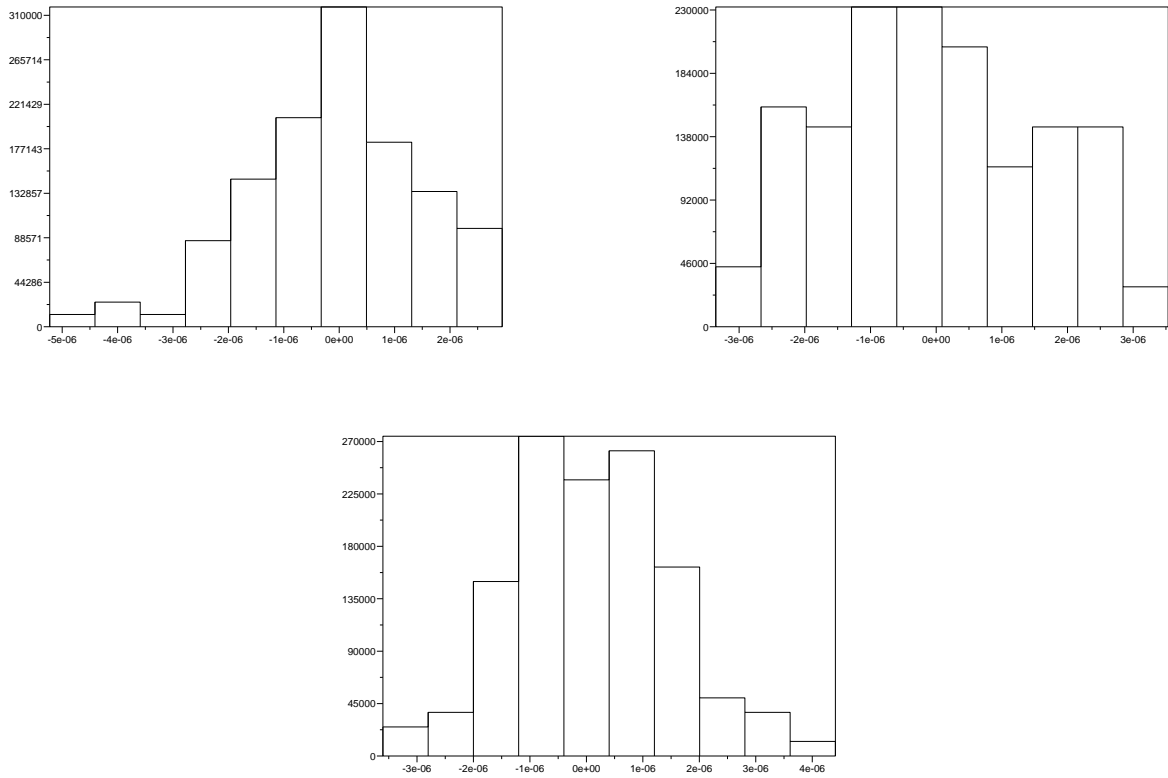


FIGURE 13. Histograms of the empirical correlations of $\hat{\epsilon}_{2n}$ and $\hat{\epsilon}_{2n+1}$, $\hat{\epsilon}_n$ and $\hat{\epsilon}_{2n}$, $\hat{\epsilon}_n$ and $\hat{\epsilon}_{2n+1}$, $n = 60$

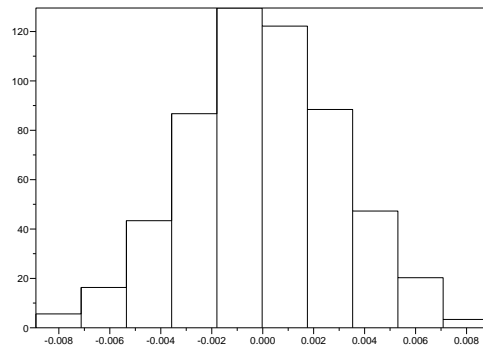


FIGURE 14. Histogram of the $\hat{\epsilon}_n$, $n = 60$, 1000 simulations

In this purpose, we use a χ^2 test with 7 degrees of freedom. The value obtained is 6.76, to compare with 14.07, the 0.95-quantile. This means that $\hat{\epsilon}_{60}$ fits to a gaussian law, suggesting that all the $\hat{\epsilon}_i$'s do so too. Figure 14 shows the distribution of the $\hat{\epsilon}_{60}$.

REFERENCES

- [1] E. J. STEWART, R. MADDEN, G. PAUL and F. TADDÉI. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol*, 3(2): e45, 2005.
- [2] R. M. HUGGINS and I. V. BASAWA. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.*, 36(4):1225–1233, 1999.