

# Mutation models: probabilistic study and parameter estimation

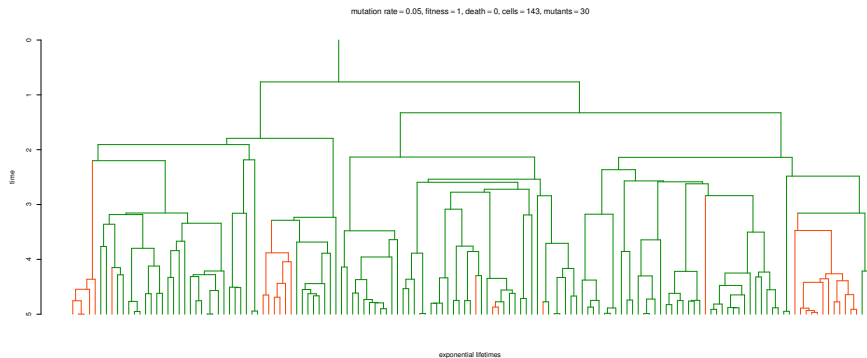
Adrien Mazoyer, supervised by Bernard Ycart

Laboratoire Jean Kuntzmann, UGA GRENoble

JPS 2016



# Example



(source : <http://ljk.imag.fr/membres/Bernard.Ycart/>)

# Motivations

$N_{\text{mut}}$	$N_f$
------------------	-------

2	1.36e9
---	--------

3	1.05e9
---	--------

0	4.28e8
---	--------

0	6.24e8
---	--------

5	7.36e8
---	--------

6	4.90e8
---	--------

110	1.36e9
-----	--------

1	9.56e8
---	--------

0	6.82e8
---	--------

Parameters of interest:

→  $\pi$  : Probability of mutation

→  $\alpha$  : Mean number of mutations

→  $\rho$  : "Fitness"

...

# Mutation model

Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$  }  $\Rightarrow$  Asymptotic model

Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$  }  $\Rightarrow$  Asymptotic model

## 3 ingredients

- Occurrences of random mutations during a growth process.

Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$  }  $\Rightarrow$  Asymptotic model

## 3 ingredients

- Occurrences of random mutations during a growth process.  
 $\Rightarrow$  **Poisson distribution (law of small numbers)**

Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$  }  $\Rightarrow$  Asymptotic model

## 3 ingredients

- Occurrences of random mutations during a growth process.  
 $\Rightarrow$  **Poisson distribution (law of small numbers)**
- Growth of a clone starting from a mutant cell for a random time.

Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$  }  $\Rightarrow$  Asymptotic model

## 3 ingredients

- Occurrences of random mutations during a growth process.  
 $\Rightarrow$  **Poisson distribution (law of small numbers)**
- Growth of a clone starting from a mutant cell for a random time.  
 $\Rightarrow$  **Sequence of independent exponential times for each clone.**



Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$  }  $\Rightarrow$  Asymptotic model

## 3 ingredients

- Occurrences of random mutations during a growth process.  
 $\Rightarrow$  **Poisson distribution (law of small numbers)**
- Growth of a clone starting from a mutant cell for a random time.  
 $\Rightarrow$  **Sequence of independent exponential times for each clone.**
- Number of cells in a mutant clone that develops for a finite time.

Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$  }  $\Rightarrow$  Asymptotic model

## 3 ingredients

- Occurrences of random mutations during a growth process.  
 $\Rightarrow$  **Poisson distribution (law of small numbers)**
- Growth of a clone starting from a mutant cell for a random time.  
 $\Rightarrow$  **Sequence of independent exponential times for each clone.**
- Number of cells in a mutant clone that develops for a finite time.  
 $\Rightarrow$  **Depends on model assumptions.**

# The Luria-Delbrück model ( $LD$ )

## Assumptions

- At time 0 a homogeneous culture of  $n$  normal cells.
- The generation time of any normal cell is a random variable with Malthusian parameter  $\nu$ .
- A splitting normal cell is replaced by :
  - One normal and one mutant cell with probability  $\pi$
  - Two normal cells with probability  $1 - \pi$ .
- The generation time of any mutant cell is exponentially distributed with parameter  $\mu$ .
- A splitting mutant cell is replaced by two mutant cells.
- All random variables and events (division times and mutations) are mutually independent.

# The Luria-Delbrück model ( $LD$ )

## Assumptions

- At time 0 a homogeneous culture of  $n$  normal cells.
- The generation time of any normal cell is a random variable with Malthusian parameter  $\nu$ .
- A splitting normal cell is replaced by :
  - One normal and one mutant cell with probability  $\pi$
  - Two normal cells with probability  $1 - \pi$ .
- The generation time of any mutant cell is exponentially distributed with parameter  $\mu$ .
- A splitting mutant cell is replaced by two mutant cells.
- All random variables and events (division times and mutations) are mutually independent.

# Stastical model of $LD$

Number of divisions tends to  $\infty$   
Mutation probability tends to 0  
Observation at time which tends to  $\infty$

}  $\Rightarrow$  Asymptotic model

## 3 ingredients

- Occurrences of random mutations during a growth process.  
 $\Rightarrow$  **Poisson distribution (law of small numbers)**
- Growth of a clone starting from a mutant cell for a random time.  
 $\Rightarrow$  **Sequence of independent exponential times for each clone.**
- Number of cells in a mutant clone that develops for a finite time.  
 $\Rightarrow$  **Sequence of independent geometric numbers (Yule process).**

## Asymptotic assumptions

Let  $t_n$  et  $\pi_n$  two sequences and  $\alpha > 0$  such that :

$$\lim_{n \rightarrow \infty} \pi_n = 0, \quad \lim_{n \rightarrow \infty} t_n = +\infty, \quad \lim_{n \rightarrow \infty} \pi_n n e^{\nu t_n} = \alpha$$

## Asymptotic assumptions

Let  $t_n$  et  $\pi_n$  two sequences and  $\alpha > 0$  such that :

$$\lim_{n \rightarrow \infty} \pi_n = 0, \quad \lim_{n \rightarrow \infty} t_n = +\infty, \quad \lim_{n \rightarrow \infty} \pi_n n e^{\nu t_n} = \alpha$$

## Initial result

As  $n \rightarrow \infty$ , the final number of mutants at time  $t_n$ , starting with  $n$  normal cells, converges to the distribution with probability generating function

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

where  $h_\rho(z)$  is the probability generating function of the Yule distribution with parameter  $\rho = \nu/\mu$ .

## An explicit asymptotic distribution

- Compound Poisson of an exponential mixture of geometric distributions;
- Two parameters:
  - $\alpha$ : the mean number of mutations;
  - $\rho$ : “fitness” parameter.
- Heavy tail distribution with tail exponent  $\rho$ .



## Estimation methods for $\alpha$ and $\rho$

- Maximum Likelihood estimators (but heavy tail distribution...)
- Compound Poisson  $\Rightarrow$  Generating function estimators ;
- $p_0$  estimators (relies on  $\mathbb{P}[0 \text{ mutants}] = e^{-\alpha}$ ).

## Estimation methods for $\alpha$ and $\rho$

- Maximum Likelihood estimators (but heavy tail distribution...)
- Compound Poisson  $\Rightarrow$  Generating function estimators ;
- $p_0$  estimators (relies on  $\mathbb{P}[0 \text{ mutants}] = e^{-\alpha}$ ).
- Deduce  $\hat{\pi}$  dividing  $\hat{\alpha}$  by the final count of cells.

## Estimation methods for $\alpha$ and $\rho$

- Maximum Likelihood estimators (but heavy tail distribution...)
- Compound Poisson  $\Rightarrow$  Generating function estimators ;
- $p_0$  estimators (relies on  $\mathbb{P}[0 \text{ mutants}] = e^{-\alpha}$ ).
- Deduce  $\hat{\pi}$  dividing  $\hat{\alpha}$  by the final count of cells.

## Bias sources

- Ignoring cells death;
- Fluctuations of the final count of cells;
- Exponentially distributed lifetime;
- Time homogeneity.

# Bias sources: fluctuation of final count $N$

Instead of being constant,  $N$  is a random variable.

## Link between $\alpha$ and $\pi$

- If  $\mathcal{L}[z] = \mathbb{E} [e^{-zN}]$

$$g(z) = \mathcal{L} [\pi(h(z) - 1)] ;$$

# Bias sources: fluctuation of final count $N$

Instead of being constant,  $N$  is a random variable.

## Link between $\alpha$ and $\pi$

- If  $\mathcal{L}[z] = \mathbb{E} [e^{-zN}]$

$$g(z) = \mathcal{L} [\pi(h(z) - 1)] ;$$

- Instead of  $\pi = \alpha/N$ :

$$\pi = \frac{\mathcal{L}^{-1} [\exp (\alpha(h(z) - 1))]}{(h(z) - 1)} ;$$

# Bias sources: fluctuation of final count $N$

Instead of being constant,  $N$  is a random variable.

## Link between $\alpha$ and $\pi$

- If  $\mathcal{L}[z] = \mathbb{E} [e^{-zN}]$

$$g(z) = \mathcal{L} [\pi(h(z) - 1)] ;$$

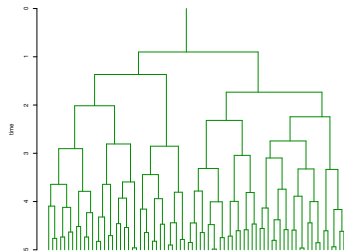
- Instead of  $\pi = \alpha/N$ :

$$\pi = \frac{\mathcal{L}^{-1} [\exp (\alpha(h(z) - 1))]}{(h(z) - 1)} ;$$

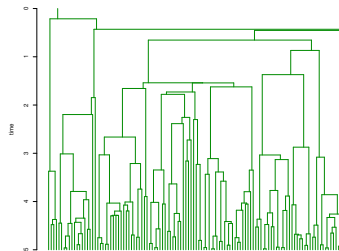
- In practice: only empirical mean and variance of are known;
- Reduce the bias using approximation of  $\mathcal{L}$ .

# Bias sources: lifetime distribution

Lifetime is not exponentially distributed with rate  $\mu$ .

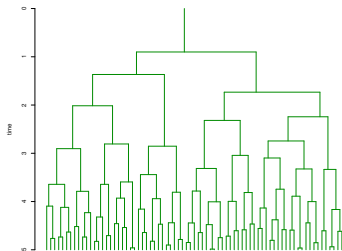


log-normal lifetimes

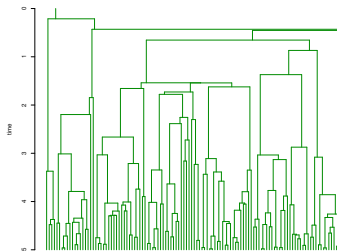


exponential lifetimes

Lifetime is not exponentially distributed with rate  $\mu$ .



log-normal lifetimes



exponential lifetimes

## Features of the new distribution

- $\mu$  is the Malthusian parameter of new lifetime distribution.
- Explicit distribution only for exponential and constant division times.



The lifetime distribution depends on the birth time of the cell.

The lifetime distribution depends on the birth time of the cell.

## New assumptions

The generation time of a normal (resp. mutant) cell born at time  $s \geq 0$  has cumulative distribution function

$$F_\nu(s, t) = 1 - e^{-\nu(s, t)} \quad \left( \text{resp. } F_\mu(s, t) = 1 - e^{-\mu(s, t)} \right)$$

where  $\nu$  (resp.  $\mu$ ), is positive, differentiable and increasing, such that

$$\lim_{t \rightarrow \infty} \nu(s, t) = +\infty \quad \text{and} \quad \forall t \in [0; s], \nu(s, t) = 0.$$

## Main Result (M.A.)

As  $n \rightarrow \infty$ , the final number of mutants at time  $t_n$ , starting with  $n$  normal cells, converges to the distribution with probability generating function

$$g(z) = \exp \{ \alpha (\mathcal{I}(z, +\infty) - 1) \} ,$$

where  $\mathcal{I}(z, t)$  depends on  $\nu$  and  $\mu$ .

## Main Result (M.A.)

As  $n \rightarrow \infty$ , the final number of mutants at time  $t_n$ , starting with  $n$  normal cells, converges to the distribution with probability generating function

$$g(z) = \exp \{ \alpha (\mathcal{I}(z, +\infty) - 1) \},$$

where  $\mathcal{I}(z, t)$  depends on  $\nu$  and  $\mu$ .

Particular case :  $\mu(s, t) = \frac{\nu(s, t)}{\rho}$

As  $n \rightarrow \infty$ , the final number of mutants at time  $t_n$ , starting with  $n$  normal cells, converges to the distribution with probability generating function

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

where  $h_{\rho}(z)$  is the probability generating function of the Yule distribution with parameter  $\rho$ .

## Theoretical work

- Results for non-trivial cases.
- Extend to multitype branching processes.
- Including normal cell deaths.

## Theoretical work






- Results for non-trivial cases.
- Extend to multitype branching processes.
- Including normal cell deaths.

## Practical work

Implementation of the R package `f1an` including

- Simulation ;
- Estimation ;
- Statistical tests

for the different possible approaches.

-  W. ANGERER, *An explicit representation of the Luria-Delbrück distribution*, J. Math. Biol., 42 (2001), pp. 145–174.
-  K. ATHREYA AND P. NEY, *Branching processes*, Springer Berlin Heidelberg, 1972.
-  R. BELLMAN AND T. HARRIS, *On age-dependent binary branching processes*, Ann. Math., 55 (1952), pp. 280–295.
-  A. HAMON AND B. YCART, *Statistics for the Luria-Delbrück distribution*, Elect. J. Statist., 6 (2012), pp. 1251–1272.
-  S. LURIA AND M. DELBRÜCK, *Mutations of bacteria from virus sensitivity to virus resistance*, Genetics, 28 (1943), pp. 491–511.

Thank you for your attention !