Statistical inference for structured populations alimented by transport-fragmentation.

M. Doumic, M. Hoffmann, N. Krell, L. Robert.

INRIA, ENSAE & Paris-Est, Rennes 1, INRA

Cuba, March 7th 2012

Content

Goal

- We consider (simple) particle systems that may serve as a toy model for the evolution of cells or bacteria.
  - Each particle grows by ingesting a common nutrient.
  - After some time, each particle gives rise to two offsprings by cell division.
- We structure the model by state variables like age, size, growth rate and so on.
- The state variables are measured to within a certain accuracy and for specific observation schemes.

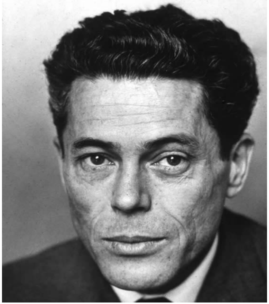
- Our control experiments are data set of the evolution of 88 microcolonies of *E. Coli* bacteria cultures.
- Each colony starts with a single ancestor and is followed up to a few hundred descendants.
- The biological hypotheses refer to (suprisingly old) classical studies that go back to 1942 J. Monod thesis.

JACQUES MONOD

Recherches sur la croissance des cultures bactériennes

DEUXIÈME ÉDITION

THÈSE DE 1942



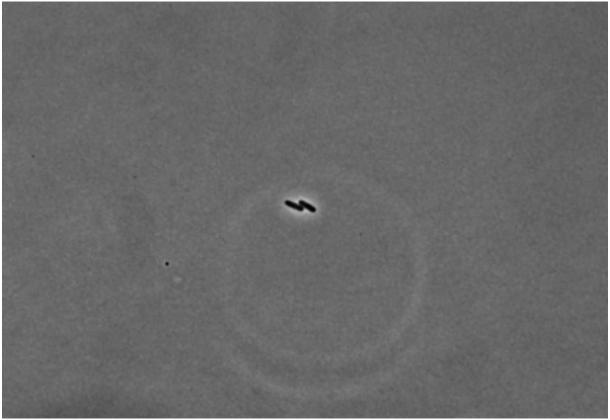


Figure: Evolution of a *E. Coli* population.

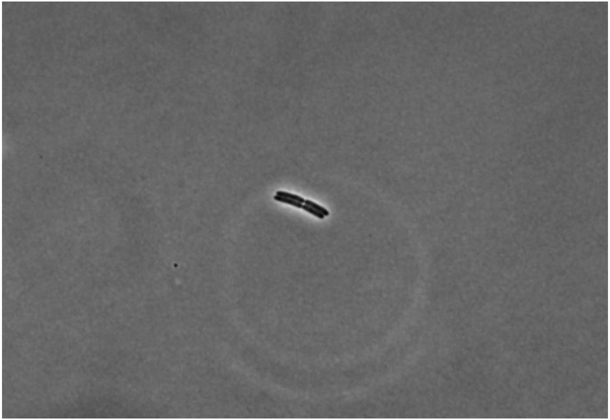


Figure: Evolution of a *E. Coli* population.

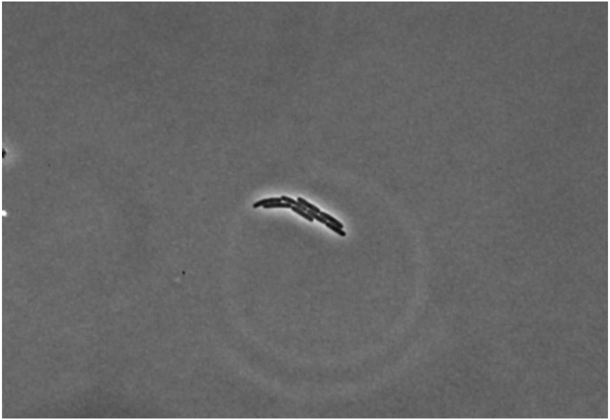


Figure: Evolution of a *E. Coli* population.

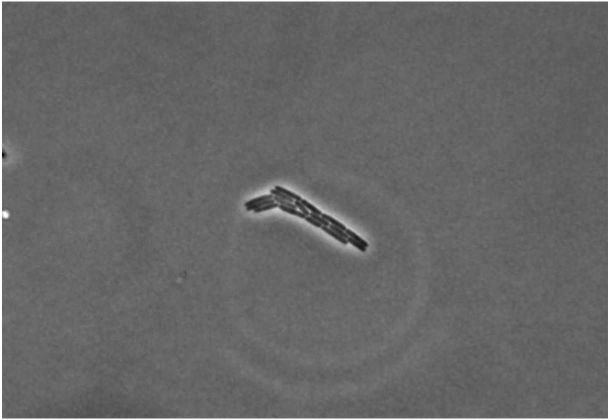


Figure: Evolution of a *E. Coli* population.

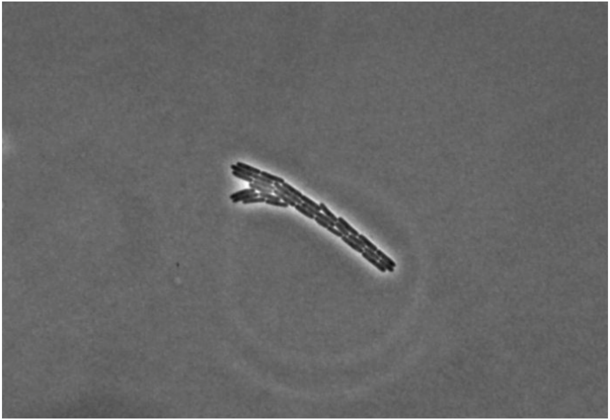


Figure: Evolution of a *E. Coli* population.

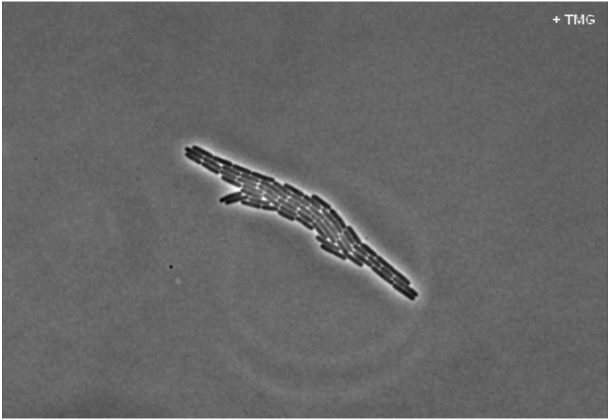


Figure: Evolution of a *E. Coli* population.

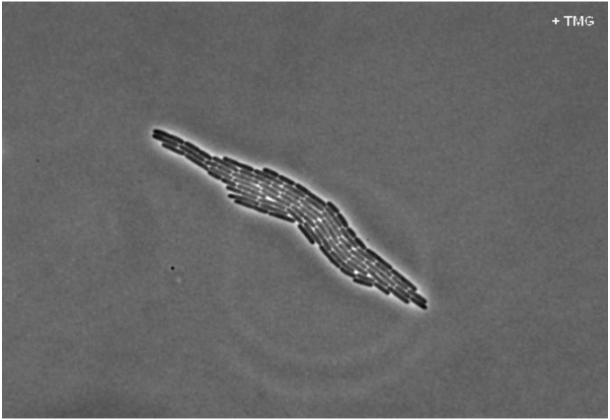


Figure: Evolution of a *E. Coli* population.

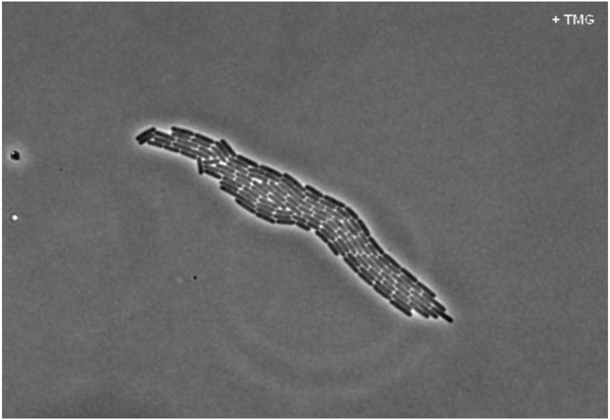


Figure: Evolution of a *E. Coli* population.

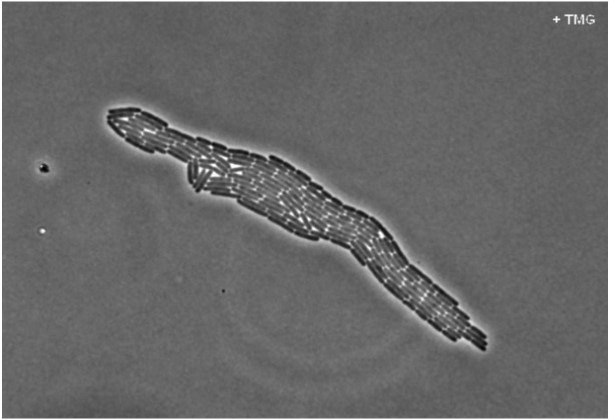


Figure: Evolution of a *E. Coli* population.

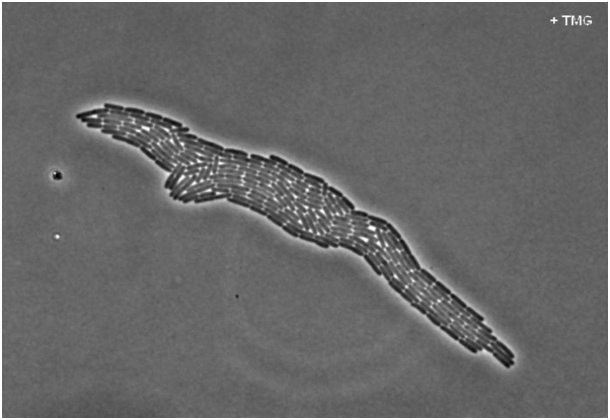


Figure: Evolution of a *E. Coli* population.

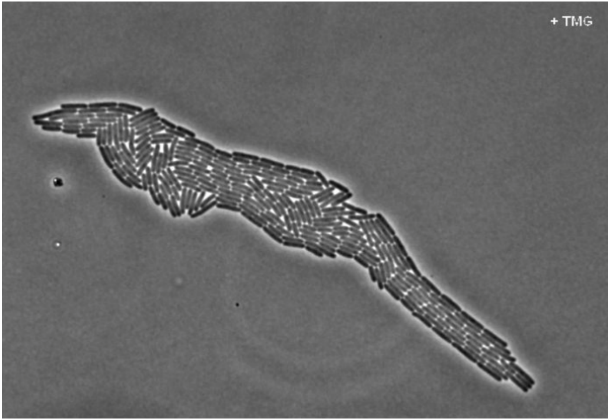


Figure: Evolution of a *E. Coli* population.

- Stochastically, the particles (cells, bacteria) evolve according to a piecewise deterministic Markov processes that evolve along a branching tree.
- Deterministically, the density of structured state variables evolves according to fragmentation-transport PDEs.

- The choice of modelling will usually be governed by underlying observation scheme, which govern in turn the accuracy of estimation of the parameters of the model.
- Considering realistic observation schemes is technically more difficult (both mathematically and experimentally) but leads to statistically more informative models.

About the growth rate

- For bacteria population growth, it is commonly admitted that the assumption $g(x) = \kappa x$ holds for a given cell. This goes back to Monod (1942).

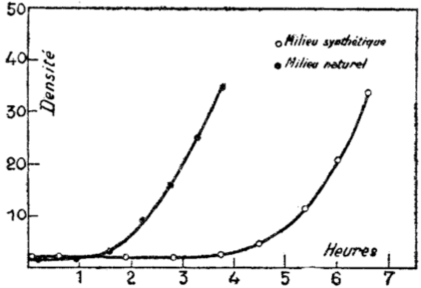


Fig. 2. — Début de la croissance de deux cultures de *B. subtilis*, en milieu synthétique et en bouillon. La phase de latence est beaucoup plus marquée en milieu synthétique.

Figure: Monod's 1942 thesis on *B. Coli* culture cells.

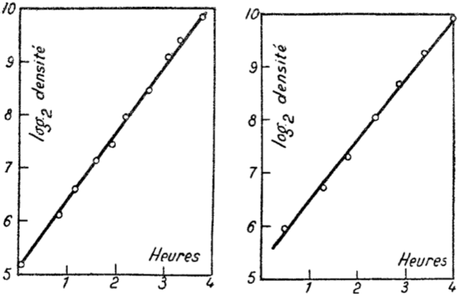


FIG. 10. — Phase exponentielle de la croissance d'une culture de *B. coli* en milieu synthétique, avec 300 mgr. par l. de glucose. Coordonnées semi-logarithmiques.

FIG. 11. — Phase exponentielle de la croissance d'une culture de *B. subtilis* en milieu synthétique, avec 500 mgr. par l. de saccharose. Coordonnées semi-logarithmiques.

Figure: Monod's 1942 thesis on *B. Coli* culture cells.

Variability in the growth rate

- Variability of the growth rate from one cell to another: exogeneous and endogeneous factors.
- In a first approach, we will ignore variability and assume a constant κ for every cell.
- We will discuss experimentally these limitations afterwards and
- subsequently propose an approach that incorporates growth variability.

The microscopic approach

- We start with a singe cell of size x_0. The cell grows exponentially according to a constant rate κ.
- The mother cell gives rize to two offsprings, at a rate $B(x)$ that depend on its size x.
- The two offsprings have initial size $x_1/2$, where x_1 is the size of the mother at division.
- The two offsprings start independent growth according to the rate κ and divide according to the rate $B(x)$.

The microscopic approach (cont.)

- The population evolution is associated with an infinite random marked tree. Let

$$\mathcal{U} = \bigcup_{n=0}^{\infty} \{0,1\}^n \quad \text{with} \quad \{0,1\}^0 := \emptyset.$$

- To each node $u \in \mathcal{U}$, we associate a cell with size at birth given by ξ_u, a lifetime ζ_u and a birth time a_u.
- $u-$ denotes the parent of u. Thus

$$\boxed{2\xi_u = \xi_{u-} \exp\left(\kappa \zeta_{u-}\right).}$$

The microscopic approach (cont.)

- $X(t) = \big(X_1(t), X_2(t), \dots\big)$ process of the sizes of the population at time t.
- We can identify $X(t)$ with a finite point measures on $\mathbb{R}_+ \setminus \{0\}$ thanks to

$$\mathcal{M}_{X(t)} = \sum_{i=1}^{\sharp X(t)} \delta_{X_i(t)}.$$

- Identity between point measures

$$X(t) = \sum_{i=1}^{\infty} \mathbf{1}_{\{X_i(t)>0\}} \delta_{X_i(t)} = \sum_{u \in \mathcal{U}} \delta_{\xi_u e^{\kappa(t-a_u)}} \mathbf{1}_{\{a_u \le t < a_u + \zeta_u\}}.$$

Main probabilistic tools

- $\mathbb{P}_x$ law of X started at $\xi_\emptyset = x$.
- Branching property: conditional on $\mathcal{M}_{X(s)} = \sum_{i=1}^{\sharp X(s)} \delta_{s_i}$, the process $\mathcal{M}_{X(t+s)}$ has the same law as $\sharp X(s)$ independent processes $\mathcal{M}_{X^{(i)}(t)}$, where the $X^{(i)}$ are independent with marginal law $\mathbb{P}_{s_i}$.
- Mass conservation: Let $X_u(t) = \xi_u e^{\kappa(t-a_u)} \mathbf{1}_{\{a_u \leq t < a_u+\zeta_u\}}$. Then

$$\sum_{t\in[a_u,a_u+\zeta_u)} X_u(t) \frac{e^{-\tau_u(t)}}{x} \equiv 1$$

where $\tau_u(t)$ denotes the cumulative growth along the node u (and $\tau_u(t) := 0$ after the time of death of the cell associated to the node u).

The tagged fragment approach

- Inspired from fragmentation processes techniques (Bertoin, Haas, among others).
- Pick a cell at random at each division and follow its size $\chi(t)$ through time.

$$\boxed{\chi(t) = \xi_\emptyset \frac{e^{\tau_t}}{2^{N_t}}}$$

  where
  - N_t is the number of divisions of the tagged fragment up to time t.
  - $\tau_t = \kappa t$ is the cumulative growth of the tagged fragment (very simple when no variability in the population)
- This enables to obtain a many-to-one formula.

A many-to-one formula

- Exists in other contexts for Branching Markov processes in a general setting (*e.g.* Bansaye *et al.*, 2009, Cloez, 2011).
- We have, for every $f \in \mathcal{C}_c\big((0,\infty) \times (0,\infty)\big)$

$$\mathbb{E}\Big[f\big(\chi(t), \tau_t\big)\Big] = \mathbb{E}\Big[\sum_{u \in \mathcal{U}} \xi_u(t) \frac{e^{-\tau_u(t)}}{x} f\big(\xi_u(t), \tau_u(t)\big)\Big]$$

from which we obtain

$$\boxed{\mathbb{E}\Big[\frac{f\big(\chi(t)\big)}{\chi(t)} x e^{\tau_t}\Big] = \mathbb{E}\Big[\sum_{i=1}^{\infty} f\big(X_i(t)\big)\Big].}$$

- Proof: genealogical representation + fragmentation technique.

Transport-fragmentation equation

- Set, for $f \in \mathcal{C}_c\big((0,\infty)\big)$,

$$\langle \mu_t, f\rangle := \mathbb{E}\Big[\sum_{i=1}^{\infty} f\big(X_i(t)\big)\Big].$$

Then, we have (in a weak sense)

$$\boxed{\partial_t \mu_t(x) + \partial_x\big(\kappa x \mu_t(x)\big) + B(x)\mu_t(x) = 4B(2x)\mu_t(2x).}$$

- Therefore the mean empirical distribution of $X(t)$ satisfies the deterministic transport-fragmentation equation.

Statistical reconstruction

- What is a relevant observation scheme ?
- Natural candidates : $\big(X(t), t \in [0, T]\big)$ or $\big((\xi_u, \zeta_u), |u| \leq n\big)$ with asymptotics taken as T or $n \to \infty$.
- What we rather have is a stopping line, compare for instance cell subcultivation for *E. Coli*.
- Denote by $\mathcal{U}_n \subset \mathcal{U}$ a set of nodes of n individuals "before" a stopping line; in particular

$$u \in \mathcal{U}_n \Longrightarrow u- \in \mathcal{U}_n.$$

- Observation scheme

$$\big\{(\xi_u, \zeta_u), \ \ u \in \mathcal{U}_n\big\},$$

asymptotics taken as $n \to \infty$.

Statistical estimation

- Given a pair ξ_{u-}, ζ_{u-} and ξ_u, we can identify κ through $2\xi_u = \xi_{u-} e^{\kappa \zeta_{u-}}$.
- How about the nonparametric estimation of B ?
- We have

$$\mathbb{P}(\zeta_u \in [t, t+dt] \,|\, \zeta_u \geq t, \xi_u = x) = B(xe^{\kappa t})dt$$

from which we obtain the density of the lifetime ζ_u conditional on the size at birth $a_u = x$:

$$\boxed{f(t,x) = B(xe^{\kappa t}) \exp\Big(- \int_0^t B(xe^{\kappa s})ds \Big).}$$

Estimation of B

- Conditional on $\xi_u = x$, the variable $\xi_u e^{\kappa \zeta_u}$ has density

$$p(y,x) = \frac{B(y)}{\kappa y}\mathbf{1}_{\{y \geq x\}} \exp\Big(- \int_0^y \frac{B(s)}{\kappa s}\mathbf{1}_{\{s \geq x\}} ds\Big)$$
$$= \lambda(y,x) \exp\Big(- \int_0^y \lambda(s,x) ds\Big),$$

with

$$\boxed{\lambda(y,x) = \frac{B(y)}{\kappa y}\mathbf{1}_{\{y \geq x\}}.}$$

- Reminiscent of conditional survival function estimation.

Elementary survival analysis

- Let f be a density function on $\mathbb{R}_+$ of the form

$$f(y) = \lambda(y) \exp\big(- \int_0^y \lambda(s)ds \big).$$

Then

$$\lambda(y) = \frac{f(y)}{1 - F(y)} = \frac{f(y)}{\mathbb{P}(X_1 \geq y)}.$$

- We mimic the same scheme: let $K_h(y) = h^{-1}K(h^{-1}y)$ denote a smooth kernel with bandwidth $h > 0$.

Nonparametric estimation of B

- For bandwidths h and $x, y > 0$, let

$$A_n(y)_h := \sum_{u \in \mathcal{U}_n} K_h(\xi_u e^{\kappa \zeta_u} - y).$$

- Then

$$\boxed{A_n(y)_{h_1} \approx \sum_x \color{red}{p(x,y)}\ \mu_n(x)}$$

where $\mu_n(x)$ is the "density" of the ξ_u's.

Nonparametric estimation of B (cont.)

- Define likewise

$$D_n(y) := \sum_{u \in \mathcal{U}_n} \mathbf{1}_{\{\xi_u e^{\kappa \zeta_u} \geq y\}} \mathbf{1}_{\{\xi_u \leq y\}}.$$

- Similarly, we have

$$\boxed{D_n(y) \approx \sum_x \left(1 - F(y,x)\right) \mu_{n(x)}}$$

- Finally

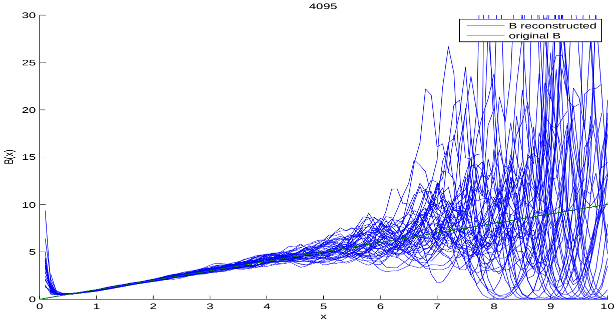
$$\frac{A_n(y)_{h_1}}{D_n(y)} \approx \frac{B(y)}{\kappa y},$$

so that eventually

$$\kappa y \frac{A_n(y)_h}{D_n(y)} \approx B(y).$$

Nonparametric estimation of B (cont.)

- Final estimator : for appropriate bandwidths $h = h(n)$, we set
$$\boxed{\widehat{B}_n(y) := \kappa y \frac{A_n(y)_{h(n)}}{D_n(y)}}.$$
- Error estimates If $B \in H^s$, for appropriate bandwidths + SRC, we have
$$\|\widehat{B}_n - B\|_{L^2(K)} \lesssim_{\mathbb{P}} n^{-s/(2s+1)} \ll n^{-s/(2s+3)}.$$
- This rate is provably optimal and is to be compared with the global approach.


4095
B reconstructed
original B
B(x)
x
0
5
10
15
20
25
30
0
1
2
3
4
5
6
7
8
9
10

About the growth rate (cont.)

- Variability of the growth rate from one cell to another: exogeneous and endogeneous factors.
- The dataset consits of 88 microcolonies followed for a few hours (average time of division is of order 20 minutes):
  - Approximately 5 microcolonies are followed everyday, for 16 days.
  - Variability in growth rate may vary from one day to the next (exogeneous factor).
  - Variability in growth rate may vary within a microcolony if specific factors are transmitted from parents to offsprings. (endogeneous factor).

Variability in growth rate: experimental results

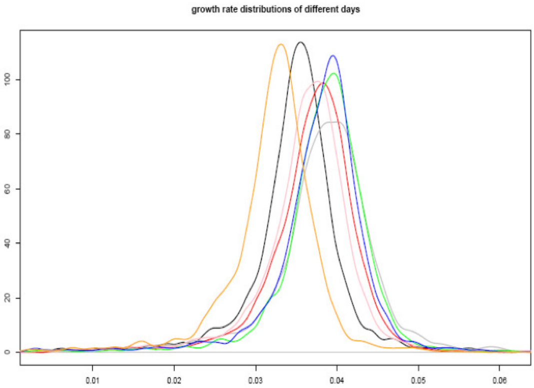


Figure: one curve = 1 day

Variability within microcolonies for given days

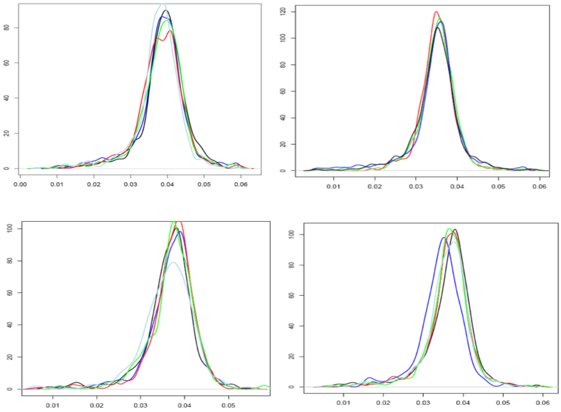


Figure: one curve = 1 microcolony

Variability within microcolonies for given days

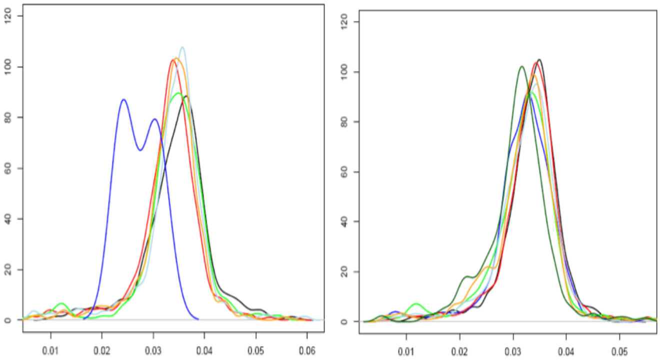


Figure: one curve = 1 microcolony; beware of artefacts!

Incorporating variability

- To each cell labeled by u, we associate a birth time a_u and a random growth rate κ_u.
- Conditional on κ_{u-}, the variability is distributed according to a (nice) Markov kernel

$$\rho(\kappa_{u-}, d\kappa_u).$$

- We now have the identity between point measures

$$X(t) = \sum_{i=1}^{\infty} \mathbf{1}_{\{X_i(t)>0\}} \delta_{X_i(t)} = \sum_{u \in \mathcal{U}} \delta_{\xi_u e^{\kappa_u (t-\zeta_u)}} \mathbf{1}_{\{a_u \leq t < a_u + \zeta_u\}}.$$

The corresponding many-to-one formula

- $\kappa_u(t)$: growth rate associated with the node u at time t.
- $\kappa(t)$: growth rate of the tagged-fragment.
- The many-to-one formula becomes

$$\mathbb{E}\left[\frac{f\left(\chi(t),\kappa_t\right)}{\chi(t)} x e^{\tau_t}\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} f\left(X_i(t),\kappa_i(t)\right)\right].$$

for f in $\mathcal{C}_c^1\big((0,\infty)\times(0,\infty)\big)$.

What of the transport-fragmentation PDE?

- In the context of variability, there is no hope to obtain a transport-fragmentation equation in $n(t,x)$.
- However, if the equation is structured in both size and variability, such a representation is still possible.
- Define, for every $f \in \mathcal{C}^1_c\big((0,\infty)\times(0,\infty)\big)$

$$\langle \mu_t, f(x,\kappa)\rangle := \mathbb{E}\Big[\sum_{i=1}^{\infty} f\big(X_i(t),\kappa_i(t)\big)\Big]$$

(slight abuse of notation).

The corresponding transport-fragmentation equation

- We have (in a weak sense)

$$\partial_t \mu_t(x,\kappa) + \kappa\,\partial_x\big(x\mu_t(x,\kappa)\big) + B(x)\mu_t(x,\kappa)$$

$$= 4\int_{\mathbb{R}_+} \rho(\kappa, d\kappa')\mu_t(2x,\kappa').$$

- What about statistical estimation? We may reasonably assume an observation scheme of the form

$$\big\{(\xi_u, \zeta_u, \kappa_u),\ \ u \in \mathcal{U}_n\big\},$$

and we need to localise further the previous estimates.

Final estimators

- Set

$$A_n(y)_h := \sum_{u \in \mathcal{U}_n} K_h(\xi_u e^{\kappa_u \zeta_u} - y)$$

and

$$D_n(y) := \sum_{u \in \mathcal{U}_n} \mathbf{1}_{\{\xi_u e^{\kappa_u \zeta_u} \geq y\}} \mathbf{1}_{\{\xi_u \leq y\}} \frac{1}{\kappa_u}$$

- With growth variability, the estimator of B becomes

$$\boxed{\widehat{B}_n(y) = \frac{y A_n(y)_{h(n)}}{D_n(y)}.}$$

Conclusion

- For constant growth rate and a global observation scheme, estimation of the division rate is ill-posed.
- Richer observation schemes enable to overcome the ill-posedness.
- Link between stochastic and deterministic modelling via many-to-one formulas for transport-fragmentation processes.
- Variablity encompassed into richer stochastic models, with deterministic counterparts if we enlarge the state space
- Other issues: stationarity of the growth rate, relative size of two offsprings.

About the size ratio between two offsprings

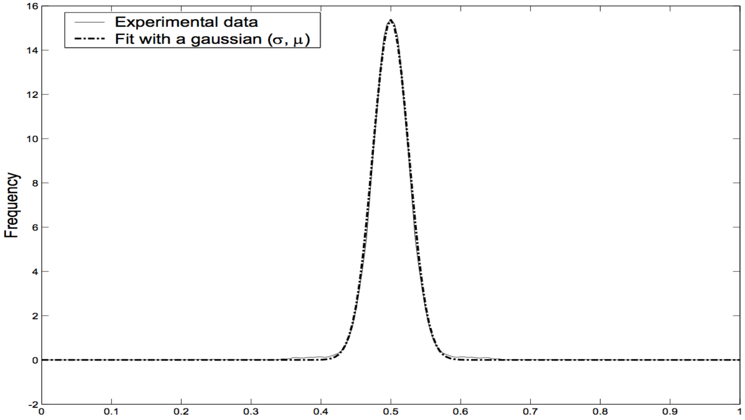


Figure: size ratio between two offsprings

Fragmentation-transport equation when two offsprings have different size

- Assume that at division

$$\xi_{(u-,0)} = \alpha\,\xi_{u-} \;\text{ and }\; \xi_{(u-,1)} = (1-\alpha)\xi_{u-}$$

with $\alpha \sim \nu(\alpha)d\alpha$ (such that $\alpha \overset{(d)}{=} 1-\alpha$).

- We obtain an extension of the fragmentation-transport equation

$$\partial_t \mu(x,\kappa) + \kappa\partial_x\big(x\mu_t(x,\kappa)\big) + B(x)\mu_t(x,\kappa)$$

$$= \int_{\mathbb{R}_+} d\kappa' \int_{[0,1]} \frac{\nu(d\alpha)}{\alpha^2}\rho(\kappa,\kappa')B(x/\alpha)\mu_t(x/\alpha,\kappa').$$

- Subsequently statistical analysis can presumably be carried over in this context.