Branching processes and bacterial growth

B. Cloez, B. De Saporta, M. Doumic, M. Hoffmann, N. K., L. Robert , T. Roget

Montpellier, INRIA, Dauphine, Rennes 2, INRA

ISCPS, 27th July, 2025

Cell division models: Presentation of the model

The microscopic approach structured by size

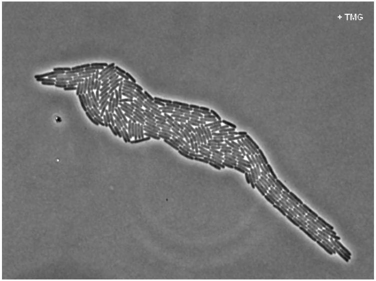
- without the type
- with two types

Statistical estimation in the microscopic model

- without the type
- with two types.

Observations of population dynamics

The data set of Stewart (2005) is the evolution of 88 microcolonies of *E. Coli* bacteria cultures.



Observations of population dynamics

Different cell characteristics may be observed. In the previous movie:

- age distribution
- size distribution
- size of the 2 daughters cells
- growth rate distribution
- age-at-division distribution
- size-at-division distribution
- genealogical influence (inheritance of some traits)...

Question: Can we deduce laws from our observations?

Direct observations

The exponential growth for Bacteria is now, after much debate, commonly admitted:

$$x_t = x_0 e^{\tau t}.$$

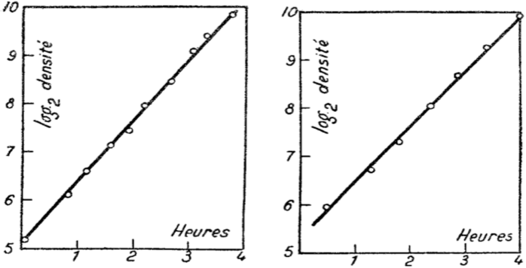


Fig. 10. — Phase exponentielle de la croissance d'une culture de *B. coli* en milieu synthétique, avec 300 mgr. par l. de glucose. Coordonnées semi-logarithmiques.

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.

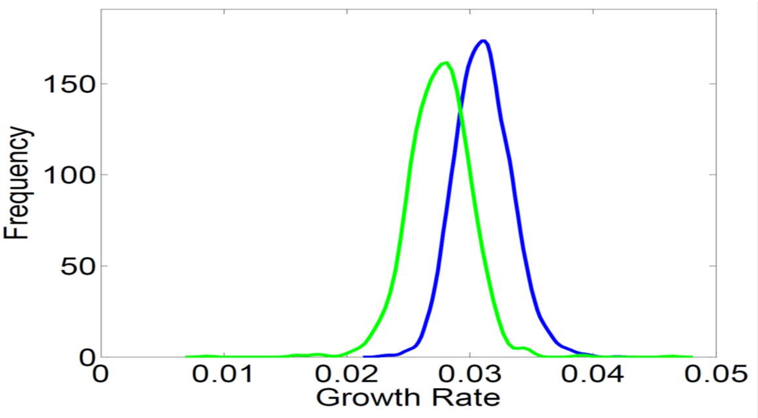
Models assumption:

The division rate B depends on

- size
- age (A. Olivier and M. Hoffmann. SPA. 2016.)
- nothing
- the increment of size (Adder model)
- and/or previous elements and/or something else...

✏Doumic, Hoffmann, K., Robert, Aymerich et Robert. BMC Biology. 2014.

Variability among exponential growth rates



In a first approach, we ignore variability and assume a constant τ for all cells.

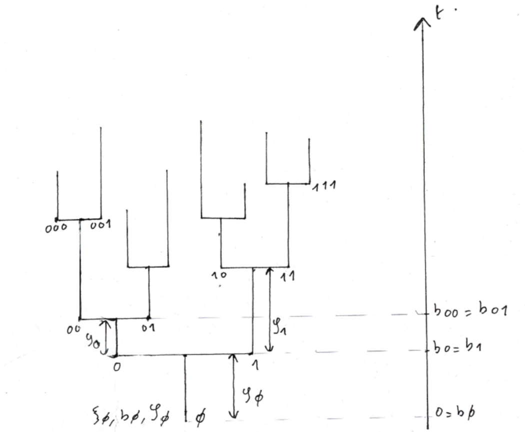
✏Delyon, de Saporta, K., Robert. CSBIGS. 2018.

Table of contents

The microscopic approach

- Initially a singe cell of size x_0.
- Exponential growth $x_t = x_0 e^{\tau t}$.
- Two offsprings, at a rate $B(x_t)$. Division occurs at time T.
- The two offsprings have initial size $x_T/2$
- And so on...

Figure: Random marked tree

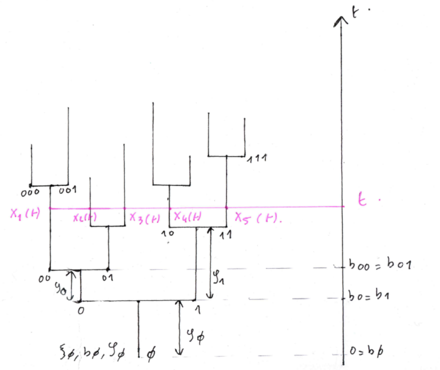


ξ_u the birth size b_u the birth time ζ_u the life time

$$u \in \{\emptyset, 0, 1, 00, 01, 10, 11, 000, \ldots\} .$$

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.



$X_1(t) = \xi_{00} e^{\tau(t-b_{00})}$ ... $X_5(t) = \xi_{11} e^{\tau(t-b_{11})}$ ξ_u the birth size b_u the birth time ζ_u the life time

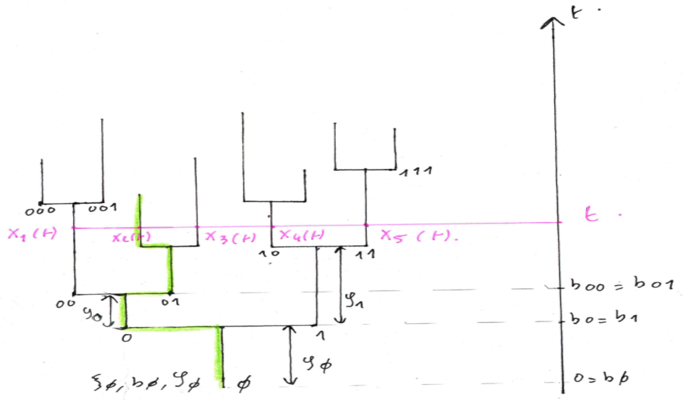
Main probabilistic tools

- Branching property
- Mass (size) conservation:

$$\sum_i X_i(t) = x_0 e^{\tau t}$$

The tagged bacterium approach

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



The tagged bacterium approach

- $$\chi(t) = x_0 \frac{e^{\tau t}}{2^{N_t}}$$
  where N_t is the number of divisions of the tagged bacterium up to time t.
- **$\chi(t)$ is a PDMP**
- This makes it possible 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,

$$\mathbb{E}\Big[f\big(\chi(t)\big)\Big] = \mathbb{E}\Big[\sum_i X_i(t)\frac{e^{-\tau t}}{x_0} f\big(X_i(t)\big)\Big]$$

from which we obtain

$$\mathbb{E}\Big[\frac{f\big(\chi(t)\big)}{\chi(t)} x_0 e^{\tau t}\Big] = \mathbb{E}\Big[\sum_i f\big(X_i(t)\big)\Big].$$

Transport-fragmentation equation

The mean empirical distribution

$$\boxed{\partial_t n_t(x) + \partial_x\big(\tau x n_t(x)\big) + B(x) n_t(x) = 4B(2x) n_t(2x)}$$

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

Table of contents

Two types.

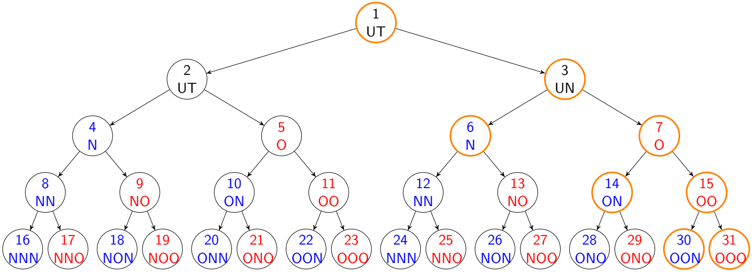
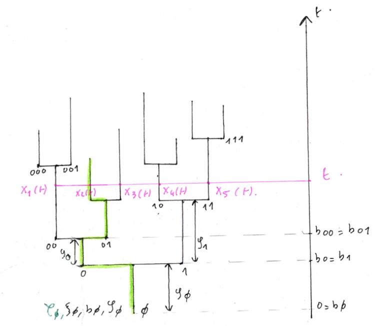


Figure: Cell division binary tree with the type of each cell

Incorporating the type and variability

- To each cell labeled by u, we associate a random growth rate
$$\tau_u \in [e_{\min}, e_{\max}].$$
- Conditional on τ_{u-}, the variability is distributed according to a (nice) Markov kernel
$$\rho_i(\tau_{u-}, d\tau_u)$$
for the type i.
- The proportion of the mother cell inherited at birth is θ_0 for old cells and θ_1 for young cells.

Figure: Random marked tree



ξ_u the size $\quad$ b_u the birth time $\quad$ ζ_u the life time $\quad$ τ_u the growth rate $\quad$ θ_u the proportion of the mother size

$$X_1(t) = \xi_{00} e^{\tau_{00}(t-b_{00})} \qquad Z_1(t) = \tau_0 \qquad \xi_{00} = \theta_0 \xi_0 e^{\tau_0 \zeta_0}$$

The corresponding many-to-one formula

- The many-to-one formula becomes
$\mathbb{E}\left[\frac{f\left(\chi(t),\tilde{Z}(t),P(t)\right)}{\chi(t)}x_0 e^{\overline{\mathcal{V}}(t)}\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} f\left(X_i(t), Z_i(t), P_i(t)\right)\right].$
where $\tilde{Z}(t)$ is the instantaneous growth rate, $\overline{\mathcal{V}}(t)$ accumulated growth rate and $P(t)$ type of the tagged bacterium.
- **$(\chi(t), \tilde{Z}(t), P(t))$ is a PDMP**
- We then have the representation

$$\chi(t) = xe^{\overline{\mathcal{V}}(t)}\theta_0^{C_t^o}\theta_1^{C_t^1} \tag{1}$$

with C_t^o the number of divisions resulting in a bacterium with a old pole and C_t^1 the one with a new pole.

What of the transport-fragmentation PDE?

The mean empirical distribution

$$\begin{cases} \partial_t n(t,x,v,i) + v\,\partial_x\big(xn(t,x,v,i)\big) + B(x)n(t,x,v,i) \\ = \int_{\mathcal{E}} \frac{\phi(x,v',0)}{\theta_0^2}\rho_0(v,dv')B(x/\theta_0)n(t,x/\theta_0,dv',i) \\ + \int_{\mathcal{E}} \frac{\phi(x,v',1)}{\theta_1^2}\rho_1(v,dv')B(x/\theta_1)n(t,x/\theta_1,dv',i), \\ n(0,x,v,i) = n^{(0)}(x,v,i), x \geq 0. \end{cases} \qquad (2)$$

with

$$\langle n(t,\cdot),\phi\rangle = \mathbb{E}_\mu\Big[\sum_{i=1}^{\infty}\phi\big(X_i(t),Z_i(t),P_i(t)\big)\Big] \text{ for every } \phi \in \mathcal{C}_0^1(\mathcal{S})$$

and $n_i(t,x,v)$ the density of $n_i(t,dx,dv)$.

✏K., Proceedings of IWBPA24

Table of contents

Statistical estimation without the type

- The dynamic is determined by the division rate $B(x)$ and the variability kernel $\rho(a, da')$ (and an initial condition $(\xi_\emptyset, \tau_\emptyset)$.)
- Observation scheme

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

with

$$\boxed{\sharp\, \mathcal{U}_n = n}$$

- Asymptotics taken as $n \to \infty$.
- We want to estimate nonparametrically $x \rightsquigarrow B(x)$.

Statistical estimation without the type

- We have

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

from which we obtain the density of the lifetime ζ_{u-} conditional on $\xi_{u-} = x$ and $\tau_{u-} = a$:

$$\boxed{t \rightsquigarrow B(xe^{at}) \exp\Big(-\int_0^t B(xe^{av})dv\Big).}$$

- Using $2\,\xi_u = \xi_{u^-} \exp\big(\tau_{u^-}\zeta_{u^-}\big)$, we further infer

$$\begin{aligned}&\mathcal{P}_B\big((x,a), x', da'\big)dx' \\ =&\frac{B(2x')}{ax'}\mathbf{1}_{\{x' \geq x/2\}} \exp\big(-\int_{x/2}^{x'} \tfrac{B(2v)}{av}dv\big)\rho(a, da').\end{aligned}$$

Identifying B through the invariant measure

- Under some assumptions, we have existence (and uniqueness) of an invariant measure on $\mathcal{S}$

$$\nu_B(d\boldsymbol{x}) = \nu_B(x, da)dx$$

i.e. such that $\nu_B \mathcal{P}_B = \nu_B$.

- More precisely, we have a contraction property

$$\sup_{|g| \leq V} \left| \mathcal{P}_B^k g(\boldsymbol{x}) - \int_{\mathcal{S}} g(\boldsymbol{z}) \nu_B(d\boldsymbol{z}) \right| \leq RV(\boldsymbol{x})\gamma^k$$

uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c})$, $\rho \in \mathcal{M}(\rho_{\min})$, for an appropriate Lyapunov function V.

Identifying B through the invariant measure

$$\nu_B(y, da') = \int_{\mathcal{S}} \nu_B(x, da) dx \, \mathcal{P}_B\big((x,a), y, da'\big)$$

$$= \frac{B(2y)}{y} \int_{\mathcal{E}} \int_0^{2y} \nu_B(x, da) dx \exp\Big(- \int_{x/2}^{y} \frac{B(2v)}{av} dv\Big) \frac{\rho(a, da')}{a}.$$

"Survival analysis trick"

$$\exp\Big(- \int_{x/2}^{y} \frac{B(2v)}{av} dv\Big) = \int_y^{\infty} \frac{B(2v)}{av} \exp\Big(- \int_{x/2}^{v} \frac{B(2v')}{av'} dv'\Big) dv$$

and $\mathcal{P}_B$ appears on the RHS again...

Identifying B through the invariant measure

- We obtain

$$\nu_B(y, da') = \frac{B(2y)}{y} \int_{\mathcal{E}} \int_0^{2y} \nu_B(x, da) dx$$
$$\int_y^{\infty} \tfrac{B(2v)}{av} \exp\Big(- \int_{x/2}^{v} \tfrac{B(2v')}{av'} dv' \Big) dv \, \frac{\rho(a, da')}{a}$$
$$= \frac{B(2y)}{y} \int_{\mathcal{S}} \int_{[0,\infty)} \mathbf{1}_{\{x \leq 2y,\, v \geq y\}} a^{-1}$$
$$\nu_B(x, da) dx \, \mathcal{P}_B\big((x, a), v, da'\big) dv.$$

- Integrating (in da') yields the key representation

$$\nu_B(y) = \frac{B(2y)}{y} \mathbb{E}_{\nu_B}\Big[\frac{1}{\tau_{u^-}} \mathbf{1}_{\{\xi_u^- \leq 2y,\, \xi_u \geq y\}} \Big]$$

Key representation

- We conclude

$$\boxed{B(y) = \frac{y}{2}\frac{\nu_B(y/2)}{\mathbb{E}_{\nu_B}\left[\frac{1}{\tau_{u^-}}\mathbf{1}_{\{\xi_u^- \leq y,\ \xi_u \geq y/2\}}\right]}.$$

- Final estimator

$$\widehat{B}_n(y) = \frac{y}{2}\frac{n^{-1}\sum_{u\in\mathcal{U}_n} K_h(\xi_u - y/2)}{n^{-1}\sum_{u\in\mathcal{U}_n}\frac{1}{\tau_{u^-}}\mathbf{1}_{\{\xi_{u^-} \leq y, \xi_u \geq y/2\}} \bigvee \varpi},$$

is specified a kernel function K, the bandwidth h and the threshold ϖ.

Proposition

Work under the previous assumptions. Specify

$$h_n = c_0 n^{1/(2s+1)}, \;\; \varpi_n = (\ln(n))^{-1}.$$

We have

$$\boxed{\mathbb{E}_\mu\big[\|\widehat{B}_n - B\|^2_{L^2(\mathcal{D})}\big]^{1/2} \lesssim (\ln(n))n^{-s/(1+2s)}}$$

uniformly in $B \in \mathcal{F} \cap \mathcal{H}^s(\mathcal{D})$.

✏Doumic, Hoffmann, K., Robert. Bernoulli. 2015.

Numerical implementation

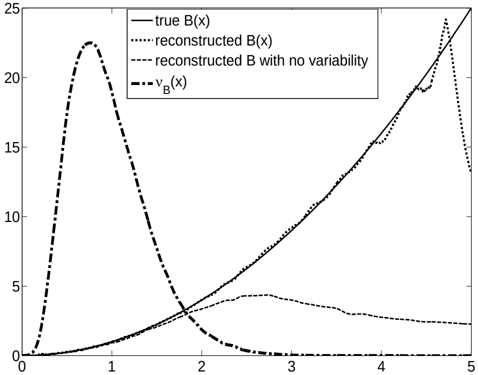


Figure: 100 Monte-Carlo estimations, dense tree case. Target function $B(x) = x$, $\tau = 1$. Reconstruction for $n = 2^{17}$ and $\varphi = n^{1/2}$.

Table of contents

The goal is to generalize what precedes to stick even more to the reality

- take into account the difference between "young" and "old" poles
- allow that division does not give 2 bacteria of the same size

This is a work in progress with Bertrand Cloez, Benoîte de Saporta and Tristan Roget.

We obtain an adaptive estimator $\hat{\nu}_{\hat{m}(y)}$ of the invariant probability. To estimate the jump rate, we construct a quotient estimator. Let us consider the estimator

$$\hat{\lambda}_n(y) = y \frac{\hat{\nu}_{\hat{m}(y)}}{\hat{\mathbf{D}}_n(y)} \mathbb{1}_{\left\{\hat{\nu}_{\hat{m}(y)} \geq 0\right\}} \mathbb{1}_{\left\{\hat{\mathbf{D}}_n(y) \geq \frac{1}{\ln(n)}\right\}} \tag{3}$$

where

$$\hat{\mathbf{D}}_n(y) := \frac{1}{n} \sum_{u \in \mathcal{U}} \frac{1}{\tau_j} \mathbb{1}\left\{\theta_j d_{u^-} \leq y \leq d_u,\ p_u = j\right\}.$$

Numerical implementation

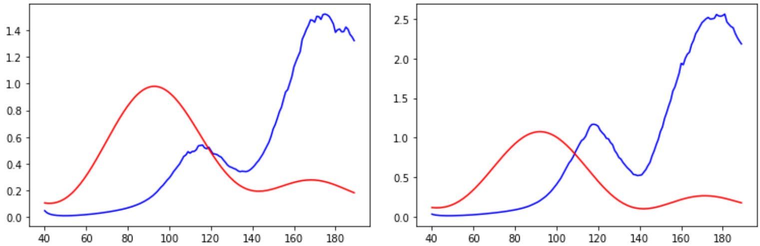


Figure: In red the estimated invariant probability and in blue the estimated division rate. On the left for the old cell and on the right for the young.

- Doumic, M., Hoffmann.,M., K., N. and Robert, L. (2015) *Statistical inference across scales for size-strutured models under growth variability. Bernoulli, 21, 1760–1799.*
- Doumic, M., Hoffmann.,M., K., N., Robert, L., Aymerich S. and Robert J. (2014) *Division Control in Escherichia coli is Based on a Size-sensing rather than Timing Mechanism. BMC Biology 12:17, 2014.*
- B. Delyon, B. de Saporta, N. K.and L. Robert. *Investigation of asymmetry in E. coli growth rate Case Studies in Business, Industry and Government Statistics, Société Française de Statistique, 2018, 7 (1), pp.1-13*
- K., N., and Schmisser E. (2021) *Nonparametric estimation of jump rates for a specific class of piecewise deterministic Markov processes. Bernoulli, 27(4):2362–2388,*
- B. Cloez, B. de Saporta, N. K. and T. Roget. *Model estimation and selection for cell division data.* Work in progress.
- N. K. *Branching processes and bacterial growth* To appear at Proceedings of IWBPA24.