Statistical inference in transport-fragmentation equations

M. Hoffmann (joint with M. Doumic, N. Krell, L. Robert)

Université Paris-Dauphine

30 mai 2013
We consider (simple) particle systems \(\approx\) toy models 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 size, growth rate and so on.
- Deterministically, the density of structured state variables evolves according to a fragmentation-transport PDE.
- Stochastically, the particles evolve according to a PDMP that evolves along a branching tree.
Toy model: size-structured populations

- $n(t, x)$: density of cells of size $x$.
- Parameter of interest: Division rate $B(x)$.
- 1 cell of size $x$ gives birth to 2 cells of size $x/2$.
- The growth of the cell size by nutrient uptake is given by a growth rate $g(x) = \tau x$ (for simplicity).
Structured populations (cont.)

- **Transport-fragmentation equation**

\[
\partial_t n(t, x) + \partial_x (\tau x n(t, x)) + B(x) n(t, x) = 4B(2x)n(t, 2x)
\]

with \( n(t, x = 0) = 0 \), \( t > 0 \) and \( n(0, x) = n^{(0)}(x), \ x \geq 0 \).

- obtained by mass conservation law:
  - LHS: density evolution + growth by nutrient + division of cells of size \( x \).
  - RHS: division of cells of size \( 2x \).

- Several extensions...
Objectives

- **Main goal**: estimate non-parametrically \( B \) from genealogical data of a cell population of size \( N \) living on a binary tree.


- **Reconcile** the deterministic approach with a rigorous statistical analysis (relaxing the steady-state approximation).
**Figure**: Evolution of a *E. Coli* population.
**Figure:** Evolution of a *E. Coli* population.
Figure: Evolution of a *E. Coli* population.
Strategy

- Construct a stochastic model accounting for the stochastic dependence structure on a tree for which the empirical measure of $N$ particles solves the fragmentation-transport equation (in a weak sense).
- Develop appropriate statistical tools to estimate $B$.
- Additional difficulty & goal: incorporate growth variability (each cell has a stochastic growth rate inherited from its parent).
Result 1

- We construct a Markov process on a binary tree

\[(X_t, V_t) \in \left( \bigcup_{k \geq 0} [0, \infty)^k \right)^2,\]

where \(X_t=\)size and \(V_t=\)growth rate of living cells at time \(t\) (inherited from their parent according to a kernel \(\rho\)).

- \(n(t, \cdot) := \mathbb{E} \left[ \sum_{i=1}^{\infty} \delta_{X_i(t), V_i(t)} \right]\) is a (weak)-solution of an extension of the transport-fragmentation equation:

\[
\partial_t n(t, x, v) + v \partial_x (x n(t, x, v)) + B(x)n(t, x, v) = 4 \int \rho(v', dv)n(t, 2x, dv').
\]

- The initial framework \(g(x) = \tau x\) is retrieved as soon as \(\rho(dv) = \delta_{\tau}(dv)\)
Result 2

- **Genealogical data**: we observe size+variability 
  \((\xi_u, \tau_u)_{u \in \mathcal{U}_N}\), where \(\mathcal{U}_N\) is a (connected) subset of size \(N\) of the binary tree \(\mathcal{U} = \bigcup_{k \geq 0} \{0, 1\}^k\).

- **Main result**: We can construct an estimator 
  \((\hat{B}_N(x), x > 0)\) of the \(s\)-regular division rate \(B(x)\) s.t.

\[
\mathbb{E} \left[ \| \hat{B}_N - B \|_{L^2_{\text{loc}}}^2 \right]^{1/2} \lesssim (\log N) N^{-s/(2s+1)}
\]
Numerical implementation and effect of variability

Figure: Simulated data.
Comparison with the inverse problem approach

Figure: Exploration on simulated data via the global approach (inverse problem), \( N \approx 3000 \).
Numerical implementation

**Figure**: Exploration on real-data. Sparse tree, $N \approx 3000$. 

Statistical inference in transport-fragmentation equations

M. Hoffmann (joint with M. Doumic, N. Krell, L. Robert)
Statistical estimation

- Given a pair $\xi_{u^-}, \zeta_{u^-}$ and $\xi_u$, we can identify $\tau_{u^-}$ through

$$2 \xi_u = \xi_{u^-} e^{\tau_{u^-}} \zeta_{u^-}.$$ 

- 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 $\zeta_{u^-}$ (lifetime of the parent $u^-$) conditional on $\xi_{u^-} = x$ and $\tau_{u^-} = v$:

$$t \sim B(xe^{vt}) \exp \left( - \int_0^t B(xe^{vs}) ds \right).$$
Toward a Markov kernel

Using $2 \xi_u = \xi_u - \exp (\tau_u - \zeta_u)$, we further infer

$$\mathbb{P} (\xi_u \in dx' \mid \xi_u = x, \tau_u = v) = \frac{B(2x')}{vx'} 1\{x' \geq x/2\} \exp \left( - \int_{x/2}^{x'} \frac{B(2s)}{vs} ds \right) dx'.$$

We obtain a simple an explicit representation for the transition kernel on $S = [0, \infty) \times \mathcal{E}$,

$$\mathcal{P}_B(x, dx') = \mathcal{P}_B((x, v), x', dv') dx'.$$
The explicit transition

- The formula is given by

\[ P_B((x, v), x', dv')dx' = \frac{B(2x')}{vx'} \mathbf{1}_{\{x' \geq x/2\}} \exp \left( - \int_{x/2}^{x'} \frac{B(2s)}{\nu s} ds \right) \rho(v, dv'). \]

- \( \rho(a, da') \) : appropriate Markov kernel on \( E \).

- Under appropriate conditions on \( B \), the Markov chain on \( S = [0, \infty) \times E \) is geometrically ergodic. (It is however not reversible.)
Identifying $B$ through the invariant measure

- Under appropriate assumptions, we have existence (and uniqueness) of an **invariant measure** on $S$

  \[ \nu_B(dx) = \nu_B(x, dv) dx \]

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

- More precisely, we have a **contraction property**

  \[ \sup_{|g| \leq V} \left| \mathcal{P}^k_B g(x) - \int_S g(z) \nu_B(dz) \right| \leq RV(x) \gamma^k \]

  for some $\gamma < 1$ locally uniformly in $B$ for an appropriate Lyapunov function $V$. 
Identifying $B$ through the invariant measure

\[ \nu_B(y, dv') \]
\[ = \int_S \nu_B(x, dv) dx \mathcal{P}_B((x, v), y, dv') \]
\[ = \frac{B(2y)}{y} \int_\mathcal{E} \int_0^{2y} \nu_B(x, dv) dx \exp \left( - \int_{x/2}^y \frac{B(2s)}{vs} ds \right) \frac{\rho(v, dv')}{v}. \]

“Survival analysis trick”

\[ \exp \left( - \int_{x/2}^y \frac{B(2s)}{vs} ds \right) = \int_y^\infty \frac{B(2s)}{vs} \exp \left( - \int_{x/2}^v \frac{B(2s')}{vs'} ds' \right) ds \]

and $\mathcal{P}_B$ is involved in the RHS again...
We obtain

\[ \nu_B(y, dv') = \frac{B(2y)}{y} \int \int_0^{2y} \nu_B(x, dv)dx \]

\[ \int \int_y^\infty \frac{B(2s)}{vs} \exp \left( - \int_{x/2}^s \frac{B(2s')}{vs'} ds' \right) ds \frac{\rho(v, dv')}{v} \]

\[ = \frac{B(2y)}{y} \int \int_0^{\infty} \int_{[0,\infty)} 1\{x \leq 2y, s \geq y\} v^{-1} \nu_B(x, dv)dx \mathcal{P}_B((x, v), s, dv') ds \]

and integrate in \( dv' \).
We obtain the key representation

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

We conclude

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

- Introduce a kernel function
  \[ K : [0, \infty) \rightarrow \mathbb{R}, \quad \int_{[0, \infty)} K(y) \, dy = 1 \]
  and set
  \[ K_h(y) = h^{-1}K(h^{-1}y) \quad \text{for} \quad y \in [0, \infty) \quad \text{and} \quad h > 0. \]

- Final estimator

\[
\hat{B}_n(y) = \frac{y}{2} \frac{n^{-1} \sum_{u \in \mathcal{U}_n} K_h(\xi_u - y/2)}{n^{-1} \sum_{u \in \mathcal{U}_n} \frac{1}{\tau_u} \mathbf{1}\{\xi_u^- \leq y, \xi_u \geq y/2\}} \sqrt{\varpi}
\]

- The estimator \( \hat{B}_n(y) \) is specified by \( K \), the bandwidth \( h \) and the threshold \( \varpi \).
Some references

