Time scales for large populations birth and death processes - Quasi stationary distributions and resilience

Sylvie Méléard, Ecole Polytechnique, France

Les probabilités de demain, Paris, 2019

With J.R. Chazottes, P. Collet and S. Martinez for the last part.
The dynamics of $d$ interacting species is often modeled by a vector field $B(x) - D(x)$ on $\mathbb{R}^d_+$:

$$\frac{dx}{dt} = B(x) - D(x),$$

$x_j$ representing the concentration of species $j$, ($1 \leq j \leq d$).

- $B(x)$ and $D(x)$ have non-negative components.
- $B(0) = D(0) = 0$: absence of spontaneous generation (no immigration).
- Link with individuals: a parameter $K > 0$ (large) - the order of magnitude of the number of individuals compatible with the available resources.
- If $n_j$ is the number of individuals of species $j$, then

$$x_j = \frac{n_j}{K}.$$
Birth and Death Process

The microscopic time evolution is given by the birth and death process \((N^K_t, t \geq 0)\) on \(\mathbb{Z}_d^+\) with transition rates \(K B(n/K)\) and \(K D(n/K)\).

\[
P(N^K_j(t+dt) = n_j+1, N^K_q(t+dt) = n_q, \forall q \neq j \mid N(t) = n) = K B_j(n/K) \, dt ;
\]
\[
P(N^K_j(t+dt) = n_j-1, N^K_q(t+dt) = n_q, \forall q \neq j \mid N(t) = n) = K D_j(n/K) \, dt .
\]

\[
P(N^K(t + dt) = n \mid N(t) = n) = 1 - K \sum_{j=1}^{d} (B_j(n/K) + D_j(n/K)) \, dt.
\]

- For example if \(B_j(x) = \lambda_j x_j\), we have

\[
P(N^K_j(t + dt) = n_j + 1 \mid N(t) = n) = \lambda_j n_j dt.
\]
Theorem (Kurtz ’71) For any $T > 0$, for any $x_0 \in \mathbb{R}^d$, if
\[
\lim_K \frac{N^K(0)}{K} = x_0, \text{ then for any } \varepsilon > 0, \\
\lim_K \mathbb{P}\left( \sup_{t \leq T} \left| \frac{N^K(t)}{K} - x(t) \right| > \varepsilon \right) = 0,
\]
with $dx/dt = B(x) - D(x)$ and $X(0) = x_0$.

On the finite time interval $[0, T]$, the trajectory of the process $N^K/K$ stays close to the trajectory of the diffusion process $Z$ given by
\[
dZ_j(t) = (B_j(Z) - D_j(Z))dt + \frac{\sqrt{B_j(Z) + D_j(Z)}}{\sqrt{K}} dW_t,
\]
where $W$ is a Brownian motion.

This is the standard stochastic fluctuation.
Assumptions on $B$ and $D$

- $B(0) = D(0) = 0$: 0 is an absorbing point.
- $B$ and $D$ are smooth.
- "Descent from infinity":
  \[
  \lim_{x \to \infty} \frac{\sup_j B_j(x)}{\inf_j D_j(x)} = 0.
  \]
- 0 is a repeller
- There exists a unique positive fixed point $x_*$ for $B - D$, lying in $\text{Int}(\mathbb{R}_+^d)$, linearly stable and globally attracting.

Standard assumptions in ecology: logistic birth-and-death process

$B(x) = bx$; $D(x) = x(d + cx)$ and $\frac{dx}{dt} = x(b - d - cx)$. 
What about very large time scale?

Under our assumptions, the process \((N^K_t, t \geq 0)\) attains 0 almost surely in finite time.

Let \(T_0 = \inf \{ t > 0; N^K_t = 0 \}\) be the extinction time.

\[
\forall n \in \mathbb{N}^d \setminus \{0\}, \quad P_n(T_0 < \infty) = 1.
\]

• From Kurtz's Theorem, \(N^K_t\) should be close to \([x_\star K]\) for large \(t\).

• Then the limits in \(t\) and \(K\) cannot be interchanged.

What happens in a larger time scale?

How long does it take for the process to reach 0?

What is the time scale of \(T_0\)?
Trajectories of $N^K_t$

$d = 1, K = 100.$
Theorem (Van Doorn ’91)
For fixed $K$, there exists a unique probability measure $\nu^K$ on $\mathbb{N}^d \setminus \{0\}$ such that

$$
\mathbb{P}_{\nu^K}(N^K_t \in A \mid T_0 > t) = \nu^K(A) \quad \forall t > 0, A \subset \mathbb{N}^d \setminus \{0\}.
$$

Moreover, for all $n \in \mathbb{N}^d \setminus \{0\}$, we have

$$
\lim_{t \to \infty} \mathbb{P}_n(N^K_t \in A \mid T_0 > t) = \nu^K(A).
$$

$\nu^K$ is called a quasi-stationary distribution (QSD).

Large literature on the topics, in particular Cattiaux et al. ’09, M.-Villemonais ’12, Collet-Martinez-San Martin ’13, Champagnat-Villemonais ’16.
One can show that there exists $\rho_0(K) > 0$, extinction rate from the QSD $\nu^K$, such that for all $t > 0$

$$\mathbb{P}_{\nu^K}(T_0 > t) = e^{-\rho_0(K)t}.$$ 

In particular,

$$\mathbb{E}_{\nu^K}(T_0) = \frac{1}{\rho_0(K)}.$$ 

Can we obtain the exact dependence of $\rho_0$ as function of $K$, for large $K$? 

How do the trajectories behave, for large $K$? 

Can we see the QSD? 

Which information can we deduce from the observation of the process?
• If the time it takes for the process to reach the QSD is significantly less than $1/\rho_0(K)$, we can see the QSD.

• We will prove that there exists another time scale $\frac{1}{\rho_1(K)} \ll \frac{1}{\rho_0(K)}$ which describes the time it takes to reach the QSD.
The problem is generically not self-adjoint (except in dimension 1).

A necessary and sufficient condition for the existence and uniqueness of a QSD together with the convergence in total variation is proved by N. Champagnat and D. Villemonais, 2016.

They provide in particular an estimate for the rate of convergence (spectral gap).

$$\sup_{n \in \mathbb{N}^d \setminus \{0\}} \left\| \mathbb{P}_n \left( N_t^K \in \cdot \mid T_0 > t \right) - \nu^K \right\|_{TV} \leq 2 \left( 1 - b_1 b_2 \right) \frac{t}{t_0}.$$ 

They require two conditions.

**Condition A1:** There exist two positive numbers $b_1$ and $t_0$ and a probability measure $\theta_K$ on $\mathbb{N}^d \setminus \{0\}$ such that for any subset $A$ of $\mathbb{N}^d \setminus \{0\}$

$$\inf_{n \in \mathbb{N}^d \setminus \{0\}} \mathbb{P}_n \left( N_{t_0}^K \in A \mid T_0 > t_0 \right) \geq b_1 \theta_K(A).$$

Note that in general $\theta_K$ is not the QSD.

In our case, we choose the uniform distribution on $\mathcal{B}(Kx^*, \sqrt{K})$. 
Condition A2: There exists a positive number $b_2$ such that

$$\mathbb{P}_{\theta_K}(T_0 > t) \geq b_2 \sup_{n \in \mathbb{N}^d \setminus \{0\}} \mathbb{P}_n(T_0 > t).$$

We have proven that for large $K$ the constants $b_1$ and $b_2$ can be chosen independent of $K$ while

$$t_0 = \mathcal{O}(1)_d \log K.$$ 

The proof relies on descent from infinity, Lyapounov function and lower bounds on transition probabilities (no symmetry, no Harnack inequality available, no Gaussian bound known).
We obtain that
\[
\frac{1}{\rho_0(K)} = e^{O(1)K},
\]
with a very precise estimate for \( d = 1 \).

For the convergence rate, we get for some \( a > 0 \) independent of \( K \), for all \( n \),
\[
\| \mathbb{P}_n(N^K_t \in \cdot) - \nu^K(\cdot) \|_{TV} \leq 2 e^{-a t/\log K} + \mathbb{P}_n(T_0 \leq t).
\]

We can prove that for some \( b > 0, c > 0, f > 0 \) and \( D > 0 \),
\[
\mathbb{P}_n(T_0 \leq t) \leq e^{-b(\|n\|_1 \wedge (cK))} + t D e^{-fK}.
\]

Therefore
\[
\| \mathbb{P}_n(N^K_t \in \cdot) - \nu^K(\cdot) \|_{TV} \leq 2 e^{-a t/\log K} + e^{-b(\|n\|_1 \wedge (cK))} + t D e^{-fK}.
\]
This error estimate \( 2 \, e^{-a \frac{t}{\log K}} + e^{-b(\|n\|_1 \wedge (cK))} + t \, D \, e^{-f \, K} \) reflects what we saw in the simulations.

- If the starting point \( n \) is of order one, the error is not small and the population can disappear in a time of order one.
- If the starting point \( n \) is of order \( K \), the error decreases with time at an exponential rate of order \( 1 / \log K \) and becomes small (for large \( K \)).
- If \( t \approx e^{f \, K} \), the error becomes large again.
- Hence if \( \log K \ll t \ll e^{f \, K} \), the distribution of \( N^K(t) \) is very near to \( \nu^K \) (for a starting point of order \( K \)).

Note the huge difference of time scales between \( \log K \) (rate of convergence to \( \nu^K \)), and \( e^{f \, K} \) (lower bound on the time scale of extinction), if \( K \) is large.
Key Properties

Let $S^K_t$ be the semigroup of $N^K$. Then there exists $C$ independent of $K$ s.t.

$$\sup_{n \in \mathbb{N}^d} S^K_1 (e^{\| \cdot \|})(n) \leq e^{C^K}.$$ 

In particular, $S^K_1$ maps polynomially growing functions to bounded functions and is a compact operator in such Banach spaces.

For the QSD, we have

- Exponential moments:
  $$\nu^K (e^{\| n \|}) \leq e^{O(1) K}. $$ 

- $\nu^K(n) = K x_* + O(1)$.
- For $\ell \in \mathbb{N}$, there exist $C_\ell > 0$ and $C' > 0$ such that for all $K \geq 1$,
  $$\nu^K (\| n - \nu^K(n) \|^{2\ell}) \leq C_\ell K^\ell ; \quad \nu^K (\| n - \nu^K(n) \|^2) \geq C' K,$$

- There is a Gaussian approximation of $\nu^K$ near $\nu^K(n)$ with variance of order $K$. 

Properties of the QSD for $d = 1$

For $K$ large enough,

$$\rho_0(K) = \left( a + O\left( \frac{(\log K)^3}{\sqrt{K}} \right) \right) \sqrt{K} \ e^{-bK},$$

where

$$a = \frac{1}{\sqrt{2\pi}} \left( \sqrt{\frac{B'(0)}{D'(0)}} - \sqrt{\frac{D'(0)}{B'(0)}} \right) \sqrt{\frac{D'(x_*)}{D(x_*)} - \frac{B'(x_*)}{B(x_*)}} \ x_* \ B(x_*),$$

and

$$b = \int_0^{x_*} \frac{B(x)}{D(x)} \ dx.$$

We have

$$\sup_{n \in \mathbb{N}^d \setminus \{0\}} \left\| P_nN^K_t \in \cdot \right\| - \alpha_n(K) \nu^K + (1 - \alpha_n(K)) \delta_0 \right\|_{TV} \leq O(1) \times \left( \sqrt{K} \log K \ e^{-cK} + (1 - e^{-\rho_0(K)t}) + Ke^{-d \ t/4} + K^{3/4} e^{\ell K} e^{-\rho_1(K)t} \right)$$

for $c, d, \ell$ positive constants independent of $K$ and

$$\alpha_n(K) = 1 - \left( \frac{D'(0)}{B'(0)} \right)^n + \frac{O(1)}{K}, \quad \rho_1(K) \geq \frac{O(1)}{\log K}.$$
Resilience

Back to the dynamical system. Let

\[ M = J(B - D)(x_*) \].

The engineering resilience is defined by

\[ R = -\sup_{z \in Sp(M)} \Re(z) > 0. \]

Engineering resilience is useful for at least two major purposes:

1) It gives the exponential rate of relaxation to the equilibrium after a (small) perturbation. Large resilience means more stability.

2) It gives an estimation of the change of the equilibrium after a (small) perturbation of the system.
How to determine the resilience?

Can one measure the resilience just by observing and recording the time dynamics of the system?

We prove the relation

$$M \Sigma^K + \Sigma^K M^t + 2D^K = \mathcal{O}(\sqrt{K}),$$

where $D^K$ is the diagonal matrix with entries

$$D^K_{ii} = KD_i(x_\ast) = KB_i(x_\ast)$$

and $\Sigma^K$ is the covariance matrix

$$\Sigma^K_{i,j} = \int (n_i - \mu^K_i)(n_j - \mu^K_j) \nu^K(dn) \ ; \ \mu^K = \int n \nu^K(dn) = \nu^K(n).$$

Given a trajectory $(N^K(t), t \leq T)$, one can estimate $\Sigma^K$ and $D^K$.

$d = 1$:

$$\mathcal{R} = \frac{D^K}{\Sigma^K} \text{ up to } \frac{1}{\sqrt{K}}.$$
Case $d > 1$

The equation $M \Sigma^K + \Sigma^K M^t + 2D^K = 0$ has many solutions for $M$ (which generically is not symmetric).

One uses the time correlations. Define for $\tau > 0$

$$\Sigma^K_{i,j}(\tau) = \mathbb{E}_{\nu^K}((N^K_i(\tau) - \mu^K_i)(N^K_j(0) - \mu^K_j)).$$

Note that $\Sigma^K(0) = \Sigma^K$.

One can prove that

$$e^{\tau M} = \Sigma^K(\tau) \Sigma^K(0)^{-1} + \mathcal{O}(1)(1/\sqrt{K}).$$

The matrices $\Sigma^K(\tau)$ and $\Sigma^K$ can be estimated from the data $(N^K(t), 0 \leq t \leq T)$, and choosing for example $\tau = 1$, one can estimate the matrix $M$ and hence the resilience.
Statistics

One can introduce statistics to estimate the various quantities of interest from the data. For $T > 0$, let

$$S_i^\mu(T, K) = \frac{1}{T} \int_0^T N_i^K(s)ds,$$

$$S_{i,j}^\Sigma(T, K) = \frac{1}{T} \int_0^T (N_i^K(s) - S_i^\mu(T, K))(N_j^K(s) - S_j^\mu(T, K))ds,$$

$$S_i^D(T, K) = \frac{1}{T} \# \{\text{birth of species } i \text{ for } t \in [0, T]\}$$

$$S_{i,j}^C(T, \tau, K) = \frac{1}{T - \tau} \int_0^{T-\tau} (N_i^K(s + \tau) - S_i^\mu(T, K))(N_j^K(s) - S_j^\mu(T, K))ds$$
Rates of convergence of the statistics

The errors in the inferences depend on $T$ and on the starting point.

We have estimates for the $L^2$-distance between each of the above statistics and the quantities to infer, starting from an initial condition or in the QSD.

For example there exist $C > 0$, $a, b, c, d > 0$ such that for all $K > 2$, for all $n$,

$$\mathbb{E}_n(\| S^\mu(T, K) - \nu^K(n) \|^2) \leq C(K^2 + \| n \|^2) \left( \frac{\log K}{T} + T e^{-bK} + e^{-c(\| n \|^\wedge (dK))} \right)$$

$$\mathbb{E}_{\nu^K}(\| S^\mu(T, K) - \nu^K(n) \|^2) \leq C \left( K^2 \frac{\log K}{T} + K^2 (1 + T) e^{-aK} \right)$$

The last inequality has an interest only if $K^2 \log K \ll T \ll e^{aK}$. 
Thank you for your attention!