Quasi stationary distributions and applications to population dynamics.

P. Collet
Centre de Physique Théorique
CNRS UMR 7644
Ecole Polytechnique
F-91128 Palaiseau Cedex (France)
e-mail: collet@cpht.polytechnique.fr

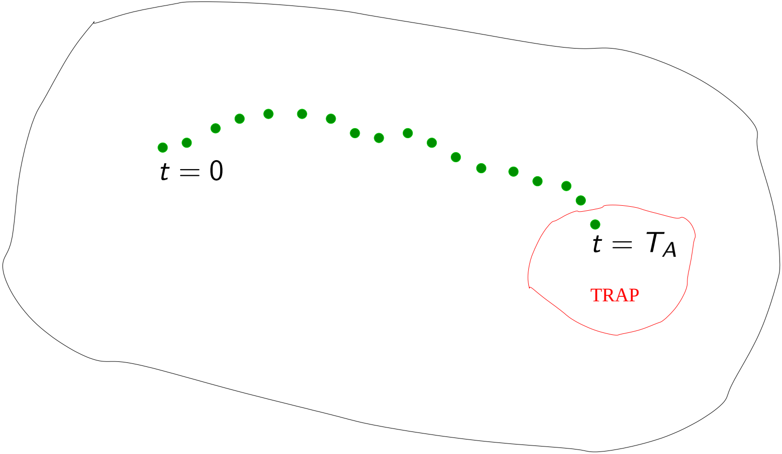
It has been observed that in populations of endangered spices the statistics of some biological traits seem to stabilise.

Renaut, Ferrière and Porter have proposed that qsd (quasi stationary distributions) are responsible for this behaviour, in particular for the Arizona ridge-nose rattlesnake (Crotalus Willardi Willardi). More generally, Steinsaltz and Evans have discussed possible applications of qsd to population dynamics.

Even though qsd may not be the final explanation of this stabilisation phenomena, it is worth looking at them in this context.

Before doing so I will recall some general definitions and results.

We consider a stochastic process (Markov) (X_t) with discrete or continuous time and with values in a metric space $\mathcal{X}$ (metric d). Let A be a subset of $\mathcal{X}$ which is imagined as a trap, namely if a trajectory arrives in A, it disappears (killed, stick etc).



For example in population dynamics, the number n of individuals of a specie is an integer, the phase space is $\mathcal{X} = \mathbb{N}^* = \{0, 1, 2, \ldots\}$. Individuals can die, reproduce, for example in a birth and death process. The number of individuals vary with time $n(t)$. However if there is no spontaneous generation, the state $n = 0$ is a trap, the specie has disappeared. If the system has reached that state it stays there forever.
There are several natural questions associated to this situation.

Question I

Given an initial distribution μ on $\mathcal{X}$, what is the probability that a trajectory has survived up to time $t > 0$? (For example if μ is the Dirac measure on one point).

In other words, if we denote by T_A the first time the particle enters in A (a function of the initial condition and of the randomness of the evolution), what is the behaviour of

$$\mathbb{P}_\mu\big(T_A > t\big) \ .$$

Often one can say something only for large t.

Question II

Assume a trajectory initially distributed with μ has survived up to time $t > 0$, what is its distribution at time t?

In other words, can we say something about

$$\mathbb{P}_\mu\big(X_t \in B \big| T_A > t\big) = \frac{\mathbb{P}_\mu\big(X_t \in B\,,\ T_A > t\big)}{\mathbb{P}_\mu\big(T_A > t\big)}\ ,$$

B a measurable subset of $\mathcal{X}$.

The Yaglom limit of the measure μ is defined as the limit of this measure when t tends to infinity (if the limit exists). It may depend on the initial distribution μ.

One says that ν is the Yaglom limit (no reference to a measure), if for any initial point x, and any Borel set B

$$\lim_{t\to\infty} \frac{\mathbb{P}_x\big(X_t \in B\,,\ T_A > t\big)}{\mathbb{P}_x\big(T_A > t\big)}\ .$$

A related object is a quasi-stationary measure (qsd).

We say that ν (a measure on $\mathcal{X}$) is quasi stationary if for any $t \geq 0$

$$\nu(B) = \mathbb{P}_\nu\big(X_t \in B \big| T_A > t\big) = \nu(B)$$

for any measurable set B. If there is no trap, this is a stationary measure.

Question III

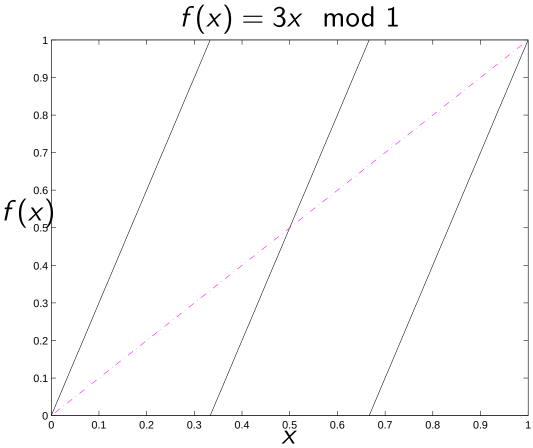
Are there trajectories which never reach the trap A? ($T_A = \infty$).

If so, how are they distributed?

How is this related to Question II?

A simple example from dynamical systems.

Consider the map f of the unit interval $[0, 1]$ given by $f(x) = 3x \pmod 1$.



It is easy to verify that the Lebesgue measure λ is invariant, namely for any Borel set B

$$\lambda\big(f^{-1}(B)\big) = \lambda\big(B\big) \ ,$$

where $f^{-1}(B)$ is the preimage set of B, namely

$$f^{-1}(B) = \big\{x \in [0,1] \,\big|\, f(x) \in B\big\} \ .$$

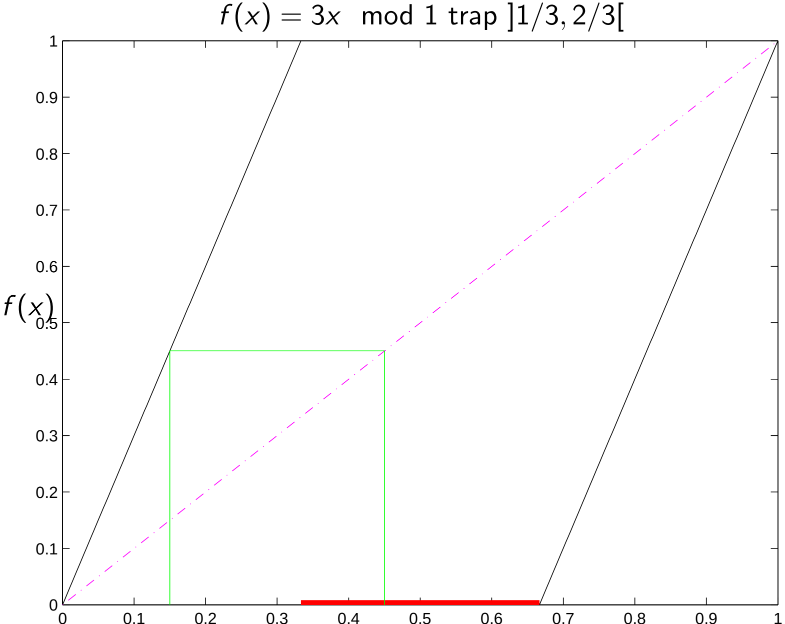
Given an initial distribution μ_0 on $[0, 1$, and a map f of $[0, 1$, we can define a discrete time stochastic process on $[0, 1]$.

The probability space is $[0, 1]$ and the process is defined recursively by $X_{n+1} = f(X_n)$, X_0 being distributed according to μ_0.

The time evolution is deterministic, but the initial condition is chosen at random. This randomness is propagated by the time evolution.

We have a measure on the set of trajectories of the process. In an abstract way any stochastic process can be thought as being a dynamical system.

We use the interval $A =]1/3, 2/3[$ as a trap.



The Lebesgue measure on $A^c = [0, 1/3] \cup [1/3, 1]$ is quasi stationary. The proof is an easy computation from the definition in the discrete time case.

From the Markov property it is enough to verify that for any measurable $B \subset A^c$

$$\mathbb{P}_{\mathrm{Leb}}\big(X_1 \in B \,\big|\, T_A > 1\big) = \mathrm{Leb}(B)\ .$$

Since $B \subset A^c$

$$\mathbb{P}_{\mathrm{Leb}}\big(X_1 \in B \,\big|\, T_A > 1\big) = \frac{\mathbb{P}_{\mathrm{Leb}}\big(X_1 \in B\,,\ T_A > 1\big)}{\mathbb{P}_{\mathrm{Leb}}\big(T_A > 1\big)} = \frac{\mathbb{P}_{\mathrm{Leb}}\big(X_1 \in B\big)}{\mathbb{P}_{\mathrm{Leb}}\big(X_1 \in A^c\big)}$$

$$= \frac{\mathrm{Leb}\big(\big\{x \,\big|\, f(x) \in B\big\}\big)}{\mathrm{Leb}\big(\big\{x \,\big|\, f(x) \in A^c\big\}\big)} = \frac{\mathrm{Leb}\big(f^{-1}(B)\big)}{\mathrm{Leb}\big(f^{-1}(A^c)\big)}\ .$$

Finally, it is enough to verify that this is equal to $\mathrm{Leb}(B)$ when B is a finite union of intervals, and this follows from a direct computation (or Thales theorem).

Consider the sequence of sets

$$\mathscr{S}_n = \{ T_A > n \} \ .$$

It is easy to check that $\mathscr{S}_0 = [0, 1/3] \cup [1/3, 1]$,

$$\mathscr{S}_1 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$$

and so on. Therefore

$$\text{Lebesgue}\big(\mathscr{S}_n\big) = \left(\frac{2}{3}\right)^{n+1} \ .$$

Namely we have the answer to question I: probability of surviving up to time n starting from the Lebesgue measure.

The set of initial conditions which never die is

$$K = \bigcap_n \mathscr{S}_n \ .$$

This is the Cantor set.

K is of zero Lebesgue measure.

Moreover, during the recursive construction, if we start with the Lebesgue measure, all the intervals of $\mathscr{S}_n$ have the same length and hence the same weight. We get at the end the Cantor measure which is very singular (absolutely singular with respect to the Lebesgue measure). So there is little connection with the nice quasi invariant measure.

Intuitively, a qsd (or a Yaglom limit) describes the distribution of trajectories which have survived for a large time but most of them are on the verge of falling in the trap.
Question III deals with trajectories that will never see the trap, this is very different. In the case of dynamical systems, these trajectories concentrate on a very small set which is invariant and disjoint from the trap.
Question II (distribution of survivors at large time) and Question III (eternal life) have very different answers.

There are few general results about qsd.

A first result concerns the entrance time T_A in the trap A for a Markov process $\big(X_t\big)$.

Theorem 1

Let ν be a qsd distribution for $\big(X_t\big)$, then T_A has an exponential law, namely there is a number $\theta > 0$ such that

$$\mathbb{P}_\nu\big(T_A > t\big) = e^{-\theta t} \ .$$

This result follows from the Markov property and the definition of qsd.

A second general result is a relation between the Yaglom limit and qsd.

A qsd is of course a Yaglom limit (of itself) but the converse is not so clear.

Theorem 2

In the case of Markov chains on a countable state space, if there exists a Yaglom limit for some initial Dirac distribution, then there exists a qsd.

The problem of existence of qsd is in a sense similar to the problem of existence of invariant measures.

There is however a supplementary difficulty: the rate of decay θ is unknown ($\theta = 0$ for invariant measures).

Up to now, no general existence theorem of qsd is known, for example there is nothing similar to the Krylov-Bogoliubov theorem for stationary(invariant) measures.

It is easy to construct examples where there is no qsd.

For example, for a dynamical system given by a map f on a space $\mathcal{X}$, if we assume that the trap A satisfies $f(\mathcal{X}\backslash A) \subset A$, there cannot be any qsd.

Indeed, here we have $T_A = 1$ (surely!), but we have seen before that if there is a qsd, this random variable should be exponential.

Another example is given by the Brownian motion (W_t) in dimension one with the trap A equal to the negative real line. There is no qsd but there is renormalised Yaglom limit. One can show easily that (for $x > 0$)

$$\mathbb{P}_x\big(W_t \in [y, y+dy]\,,\ T_0 > t\big) = \frac{1}{\sqrt{2\pi t}}\left(e^{(x-y)^2/2t} - e^{(x+y)^2/2t}\right) dy\ .$$

It follows by a short computation that

$$\lim_{t\to\infty} \mathbb{P}_x\big(W_t/\sqrt{t} \in [y, y+dy] \,\big|\, T_0 > t\big) = ye^{-y^2/2} dy\ .$$

Some existence results have been established using spectral theory. For finite state Markov chains, one can use the standard spectral theory techniques (Perron Frobenius) to prove existence.
For example in continuous time Markov chains, if $\mathscr{L}$ is the generator, a qsd ν should satisfy (in a suitable sense)

$$\mathscr{L}^* \nu = -\theta \nu \ ,$$

with some boundary condition.
For an invariant measures $\theta = 0$, for a qsd θ is unknown (spectral problem).

These spectral ideas work well for diffusions in compact domains, the trap being the boundary since in that case the generator with Dirichlet boundary conditions has often a compact resolvent. Similar spectral methods have been used for some dynamical systems (subshift of finite type, expanding maps of the interval). The Markovian semigroup is replaced in this context by a Ruelle Perron Frobenius operator.

For a continuous time Markov chain on a countable state space there is a general result by Ferrari, Kesten Martínez and Picco. Let (X_t) be a continuous time Markov chain on $\mathbb{N}^*$. Assume 0 is absorbing, and the process restricted to $\mathbb{N}$ is irreducible.

Theorem 3

Assume that for any $t > 0$

$$\lim_{n\to\infty} \mathbb{P}_n\big(T_0 < t\big) = 0 \ ,$$

and $\mathbb{P}_n\big(T_0 < \infty\big) = 1$ for one (and hence all) n. Then a necessary and sufficient condition for the existence of a qsd is that

$$\mathbb{E}_n\left(e^{\lambda T_0}\right) < \infty$$

for some $\lambda > 0$ and for some (and hence all) n.

There are several results for diffusions. I will only mention some of them.

The Ornstein-Uhlenbeck process on the half line $\mathbb{R}^+$ has been investigated in details by Lladser and San Martín.

Consider the diffusion on $\mathbb{R}^+$

$$dX = -Xdt + dW$$

and absorption at the origin (W_t is Brownian motion).

Theorem 4

There exist absolutely continuous qsd (probabilities with density) with any decay rate $\theta \in (0, 1]$.

Denote by ν_θ these qsd.

The qsd ν_θ has a density u_θ satisfying the equation

$$\partial_x \left(\frac{1}{2}\partial_x u_\theta + x u_\theta\right) = -\theta u_\theta \ .$$

The solution should be non negative, and integrable on $\mathbb{R}^+$.

These solutions can be expressed in terms of parabolic cylindric functions (relations with the Harmonic oscillator).

In particular (the "extreme" one, corresponding to $\theta = 1$)

$$u_1(x) = 2xe^{-x^2}$$

is the ground state.

In other words we have a continuum of qsd with decay rates between zero and the ground state eigenvalue (there is no qsd with $\theta > 1$).

Lladser and San Martín have also obtained a result about the domain of attraction of the Yaglom limits.

Theorem 5

Assume a probability density g on $\mathbb{R}^+$ is regularly varying at infinity with exponent $1+\eta$ ($\eta \in (0,1)$), namely for any $c > 0$

$$\lim_{u\to\infty} \frac{g(cu)}{g(u)} = c^{-1-\eta} \, .$$

Then

$$\lim_{t\to\infty} \mathbb{P}_{g(x)\,dx}\big(X_t \in A \,\big|\, T_0 > t\big) = \nu_\eta(A)$$

for any Borel subset of $\mathbb{R}^+$.

In other words, the "abnormal" distributions which decay like a power law at infinity appear only if the initial distribution has a similar behaviour.

With P.Cattiaux, A.Lambert, S.Martínez, S.Méléard, and J.San Martín we investigated another class of diffusions coming from population dynamics.

An example is

$$dX = \big(aX - bX^2\big)dt + \sqrt{X}dW \ ,$$

where $b > 0$ (dW/dt is the white noise). The process is on the half line $\mathbb{R}^+$ with absorption at the origin.

This equation is obtained by starting from a microscopic model and taking a limit of large population. X is a “renormalised” number of individuals (size of population).

In $dX = \left(aX - bX^2\right)dt + \sqrt{X}dW$,

the quadratic term reflects the competition between individuals (for food for example). It is not so important at low population. The coefficient a can be interpreted as the difference between a birth rate and a death rate (at low population).

The square root term is a remnant of the fluctuations in large populations (so called demographic stochasticity, see Champagnat, Ferrière, Méléard, Theoret. Pop. Biology **69**, 297-321 (2006)).

A similar term appears in models of mathematical finance (Cox Ingersoll Ross model) although it seems somewhat ad-hoc in this context.

More generally we have studied the stochastic differential equations

$$dX = h(X)dt + \sqrt{X}dW \ .$$

For the existence of a qsd we have the following result.

Theorem 6

Assume h is C^1

$$\lim_{z\to\infty} \frac{h(z)}{\sqrt{z}} = -\infty \quad \textit{and} \quad \lim_{z\to\infty} \frac{zh'(z)}{h(z)^2} = 0 \ .$$

Then there is a qsd ν which moreover is the Yaglom limit for any initial distribution with compact support (in particular Dirac masses).

Theorem 7

Assume h is strictly negative on $[A, \infty)$ ($A > 0$) and

$$\int_A^\infty \frac{dz}{h(z)} > -\infty \ .$$

Then the qsd is unique.

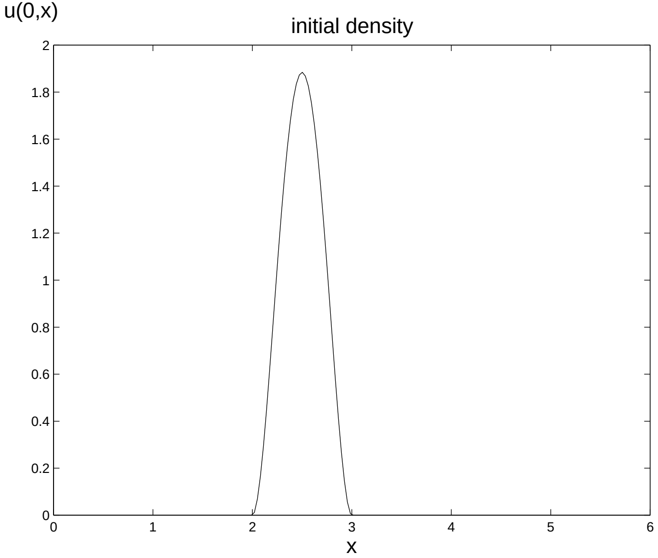
This is very different from the Ornstein-Uhlenbeck process.

In the present case, the process "comes down from infinity". For $y > 0$ we denote by T_y the first time the trajectory hits y. Then for any $A > 0$ there is a $y_A > 0$ such that

$$\sup_{x > y_A} \mathbb{E}_x \left(e^{A T_{y_A}} \right) < \infty \ .$$

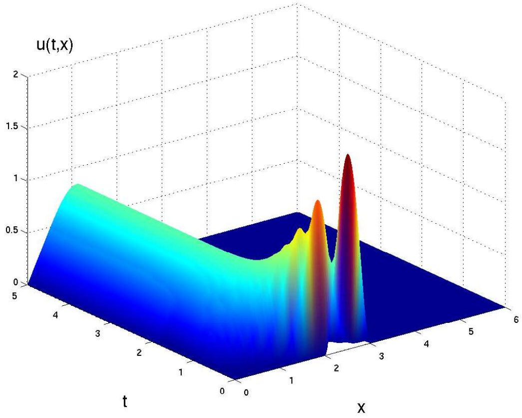
The drift is so strong that even if the process is started at a large point it rapidly goes down to a compact. In other words, it cannot feel much of the tail of an initial distribution.

A simulation with the initial distribution

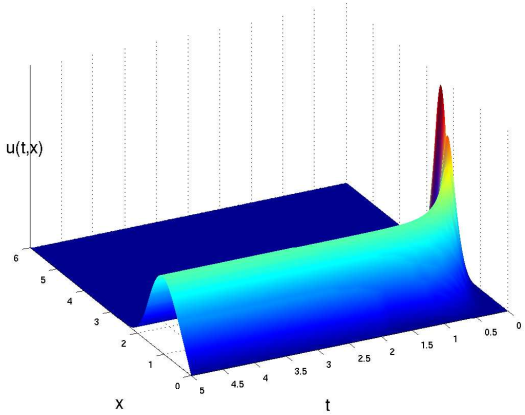


and absorption at the origin

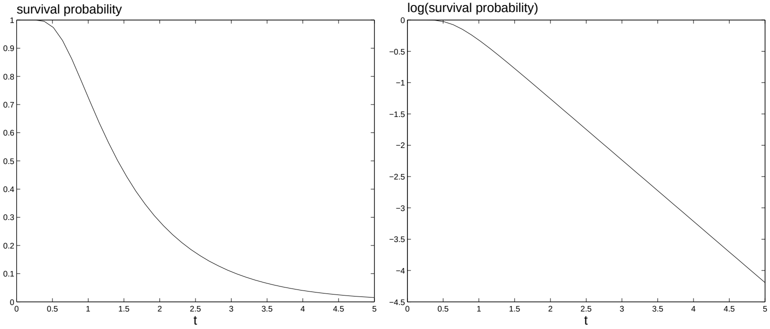
The normalised density (Yaglom limit) as a function of space and time



A view from the side



Decay of the total mass



Remark

The Kolmogorov equation for the case $h(z) = -z^2$ is

$$\partial_t u = \frac{1}{2}\partial_x^2(xu) + \partial_x\left(x^2 u\right) \, .$$

There is a nice stationary solution

$$u_0(x) = \frac{1}{x} e^{-x^2/2} \int_0^x e^{-s^2/2} ds \, .$$

It is positive and integrable.

However the flux at the origin is non zero (equal to $1/2$).

This is compensated by a non zero flux at infinity. In a sense the continuum of qsd of the Orsntein-Uhlenbeck process have disappeared because of the same plague.

A word on the proofs.

We consider the process $Y = \sqrt{X}$. It satisfies a sde

$$dY = \frac{4h(Y^2/4) - 1}{2Y} dt + dW = -q(Y) dt + dW \ .$$

We exchanged a non constant variance to a singular drift.

We can now study the Kolmogorov equation by spectral techniques (Schrödinger operator, Dirichlet form). In particular use some compactness arguments. In a weighted L^2 space, the semigroup of the diffusion process is compact.

We have also proved the existence of a Q process, namely a process where the trajectories never die. Let e denote the density of the qsd of the process Y.

Theorem 8

For any $x > 0$, and any $B_s \in \mathcal{F}_s$

$$\lim_{t \to \infty} \mathbb{P}_x\big(Y \in B_s \,\big|\, T_0 > t\big) = Q_x\big(B_s\big) \ .$$

Moreover Q_x has the transition probabilities

$$q(s, x, y) = e^{\theta s} \frac{e(y)}{e(x)} r(s, x, y) e^{-Q(y)}$$

where r is the transition kernel of the process Y and $Q' = 2q$.

See `arXiv:math/0703781` for details and references.

These results have been extended to the 2d case by S.Méléard and P.Cattiaux for the case of gradient drift, see [`hal-00336156`.

We have recently obtained some results in the general (non gradient) case (with S.Méléard and P.Cattiaux).

Recently with S.Martínez, S.Méléard and J.San Martín we have studied a population dynamics with birth, death and mutations. We have considered the following model.

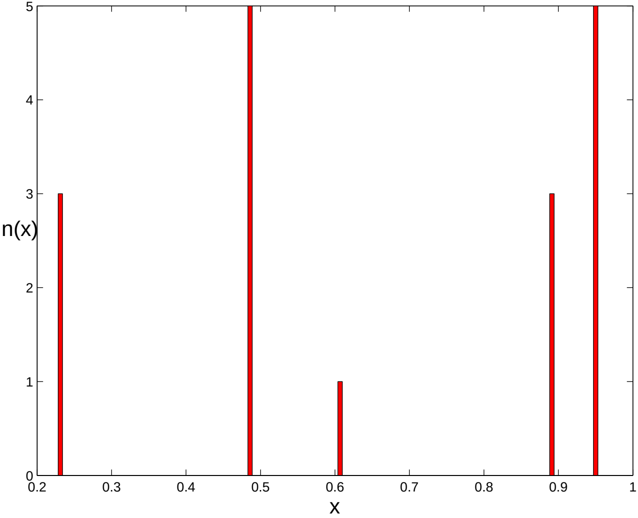
Each individual has a set of biological traits with real values (think of size, color etc.). We assume that the set of traits of an individual belongs to a compact subset $\mathcal{T}$ of $\mathbb{R}^n$ (or a manifold). To simplify I will assume $\mathcal{T}$ is the unit interval.

The configuration of a finite population is given by a set of points $x \in \mathcal{T}$, and for each x by the number of individuals having these traits.

In other words, a configuration is an atomic measure on $\mathcal{T}$ (with integer masses).

We denote by $\mathscr{A}$ the set of all such atomic measures.

A configuration (with $\mathcal{T} = [0, 1]$)



A population of 17 individuals with 5 different trait values.

In the model we considered, an individual with trait x can

1) give birth to an individual with the same trait at rate $b(x, \eta) > 0$ (for $\eta \neq 0$),
2) die at rate $\lambda(x, \eta) > 0$ (for $\eta \neq 0$),
3) give birth to a mutant, at rate $m(x, \eta) > 0$ (for $\eta \neq 0$), where the trait $y \in \mathcal{T}$ of the mutant is randomly chosen according to a transition distribution $K(x, \eta, dy)$.
4) λ (death rate), b (total birth rate), m (mutation rate) and K are assumed continuous. They depend on the trait of the individual but also on the total population (competition is allowed)

For example, if at time t the configuration of the population is $\eta(t)$, the probability that an individual of trait x dies in the time interval $[t, t + dt]$ is $\lambda(x, \eta(t))dt$.
We assume the state with no individuals is absorbing (no spontaneous generation). This corresponds to the measure $\eta = 0$.
In other words, we assume $b(x, 0) = \lambda(x, 0) = m(x, 0) = 0$.

One can then write a generator acting on the set of functions on $\mathscr{A}$. For f a function on $\mathscr{A}$

$$\mathscr{L} f(\eta) = \sum_{x \in \mathrm{Support}\ \eta} b(x,\eta)\eta(x)\big(f\left(\eta+\delta_x\right) - f(\eta)\big)$$

$$+ \sum_{x \in \mathrm{Support}\ \eta} \lambda(x,\eta)\eta(x)\big(f\left(\eta-\delta_x\right) - f(\eta)\big)$$

$$+ \sum_{x \in \mathrm{Support}\ \eta} m(x,\eta)\eta(x) \int_{\mathcal{T}} K(x,\eta,dy)\ \left(f\left(\eta+\delta_y\right) - f(\eta)\right) .$$

This generator is associated to a process with cádlág trajectories (Champagnat, Ferrière, Méléard).

Theorem 9

Assume that

$$\sup_{\eta} \sup_{x \in Support(\eta)} \lambda(x,\eta) < \infty$$

and

$$\inf_{\eta} \inf_{x \in Support(\eta)} \lambda(x,\eta) > \sup_{\eta} \sup_{x \in Support(\eta)} \big(b(x,\eta) + m(x,\eta)\big)$$

Then there exists at least one qsd.

The first condition is convenient to ensure that the process does not immediately extinguishes (the process is well defined).

The second condition imposes that the population is in decline (effective death rate bigger than effective birth rate).

The proof is based on a general abstract Lemma.

Lemma 10

Let $\mathcal{X}$ be a metric space (locally compact). Assume there exists two functions $\varphi_1 \geq 1$ and $\varphi_2 > 0$ (on $\mathcal{X}$) such that for any $u > 0$, the

$$\left\{x \,\middle|\, \frac{\varphi_2(x)}{\varphi_1(x)} \leq u\right\}$$

is compact (or empty). Let S be a continuous linear operator on the set of bounded continuous functions mapping positive functions to positive functions. Assume moreover that there exists three (strictly) positive numbers α, β and D such that

$$S\big(\varphi_1\big) \geq \alpha\varphi_1 \quad et \quad S\big(\varphi_2\big) \leq \beta\varphi_2 + D\varphi_1 \ .$$

Then if $\beta < \alpha$ there exists a probability measure ν on $\mathcal{X}$ and a number $\zeta > 0$ such that $\nu \circ S = \zeta\nu$.

In our case, one can take $\varphi_1 = 1$ and for φ_2 an exponential of the number of individuals.

The above abstract Lemma is proved using Tychonov's fixed point Theorem and does not tel much about the properties of the qsd (or uniqueness). In a more homogeneous case we can say more.

Theorem 11

Assume λ, b and m are constants (they do not depend on the trait of the individual or of the population). Assume the mutation kernel $K(x, \eta, dy)$ is absolutely continuous. Then there exists a unique qsd with decay rate $\lambda - b - m$ and it is absolutely continuous with respect to an absolutely continuous reference measure.

In this case the process of the "number of individuals" is a Markov process (birth and death), and for such processes there are many known results. Our process can be seen a a skew product above this process.

Uniqueness and absolute continuity come from the mutations.

Some references

P.Cattiaux et al. Quasi-stationary distributions and diffusion models in population dynamics. Ann. Probab. to appear. `arXiv:math/0703781`.

P.Cattiaux, S.Méléard. Competitive or weak cooperative stochastic Lotka-Volterra systems conditioned to non-extinction. `hal-00336156`.

P.Collet et al.Quasi-stationary distributions for structured birth and death processes with mutations. `arXiv:0904.3468`.

P. A. Ferrari, H. Kesten, S. Martínez, P. Picco. Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Probab.*, 23:501–521, 1995.

F. Gosselin. Asymptotic behavior of absorbing Markov chains conditional on nonabsorption for applications in conservation biology. *Ann. Appl. Probab.*, 11:261–284, 2001.

A. Lambert. The branching process with logistic growth. *Ann. Appl. Probab.*, 15:1506–1535, 2005.

M. Lladser and J. San Martín. Domain of attraction of the quasi-stationary distributions for the Ornstein-Uhlenbeck process. *J. Appl. Probab.*, 37:511–520, 2000.

P. K. Pollett. Quasi-stationary distributions: a bibliography. Available at http://www.maths.uq.edu.au/~pkp/papers/qsds/qsds.html, regularly updated.

O. Renault, R. Ferrière, and J. Porter. The quasi-stationary route to extinction. Unpublished.

D. Steinsaltz and S. N. Evans. Markov mortality models: implications of quasistationarity and varying initial distributions. *Theor. Pop. Biol.* 65:319–337, 2004.