Quasi-Stationary Distributions

Alexandre Lecestre,
supervised by Paul Fearnhead,
Lancaster University

15/05/2017-28/07/2016

Contents

Acknowledgements

I want to thank Clément Marteau for helping me in my research of an internship who has been very efficient! I am grateful to my supervisor, Professor Paul Fearnhead, who gave me a really interesting insight of the work of a researcher in statistics. I also want to thank all the staff of STOR-I (Statistics and Operational Research Centre for Doctoral Training in Partnership with Industry) and all the students for their welcome, they helped to make this experience truly enjoyable.

1 Context, Definitions and General Properties

1.1 Introduction

Let us introduce the subjects by recapping basic notions of Markov chains.

Definition 1 Stationary measure
Let p be a Markov kernel on the measurable space (E, Σ).A probability measure μ is a **stationary measure** if:

$$\boxed{\forall A \in \Sigma, \int_A \mu(dx) = \int_E p(x, A)\mu(dx).}$$

If $(X_n)_n$ is a Markov chain on $(\Omega, \mathcal{F}, \mathbb{P})$ with initial probability μ stationary and Markov kernel p, then all the X_n have the same distribution, namely:

$$\forall A \in \Sigma, \forall n, \mathbb{P}_\mu[X_n \in A] = \mu(A).$$

The ergodicity of a Markov chain is very important.
A random process is ergodic if its time average is the same as its average over the probability space.
It allows us to do a lot of things related to the stationary distribution which is related to long-term behavior, such as Markov Chain Monte Carlo algorithms. But there still is a lot of situations where ergodicity does not apply or where the stationary distributions is not of interest

One example of this is the Galton-Watson process, also known as a branching process, which is a stochastic process X_n which evolves according to the recurrence formula $X_0 = 1$ and $X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)}$ where $\{\xi_j^{(n)} : n, j \in \mathbb{N}\}$ is a set of i.i.d. natural number-valued random variables. X_n can be thought of as the number of descendants in the nth generation, and $\xi_j^{(n)}$ can be thought of as the number of children of the jth of these descendants. Then, the recurrence relation states that the number of descendants in the $n+1$st generation is the sum, over all nth generation descendants, of the number of children of that descendant.

One can notice that $X_n = 0 \Rightarrow \forall m \geq n, X_m = 0$. Here 0 is called an absorbing point and therefore δ_0 is a stationary distribution. We say that the process is dead, or extinct, when it arrives at 0. There is no point in studying the process when it is dead. That's why we want to study it with the condition of not being extinct. Even with extinction in finite time almost surely, we can have some results on the process. In several cases we can observe a "stabilization" of the process before it goes to extinction. This is why the notion of Quasi-Stationary Distribution (QSD) has been introduced (work of Yaglom on sub-critical Galton-Watson processes).

The first three sections are widely inspired by [1] which gives a good introduction to the subject of quasi-stationary distributions in the intuitive context of birth and death processes and populations dynamics.

1.2 Birth and Death Processes: Framework and Notations

We will mostly introduce the notions of quasi-stationary distributions within the context of birth and death processes, processes that model the dynamics of the size of a population. We denote this population process by $Z = Z_t$ and assume:

- $Z = (Z_t)_{t\geq 0}$ is an homogeneous Markov process in a subset E of $\mathbb{N}$ or $\mathbb{R}_+$, in a discrete or continuous-time setting.
- No immigration, hence the state 0 (describing the extinction) is an absorbing point.
- If we define $T_0 = \inf\{t > 0/Z_t = 0\}$, which is often called the extinction time, then we assume that for all z in E, $\mathbb{P}_z(T_0 < \infty) = 1$.

If A is the set of the absorbing points, we define $E^* = E\backslash A$. We can now give some notation.

- $\mathbb{P}_\mu$ is the probability associated to Z with initial distribution μ (probability measure on E^*). For all x in E^*, $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.
- $(P_t)_{t\geq 0}$ is the semi-group of the killed process Z, i.e.: $\forall z \in E^*, \forall f \in L_\infty(E^*), P_t f(z) = \mathbb{E}_z\left[f(Z_t)\mathbb{1}_{t<T_0}\right]$, where $L_\infty(E^*)$ is the set of the measurable and bounded functions on E^*
- For μ probability measure and f bounded and measurable, $\mu(f) = \int\limits_{E^*} f(x)\mu(dx)$ and $\mu P_t(f) = \mu(P_t f) = \mathbb{E}_\mu\left[f(Z_t)\mathbb{1}_{t<T_0}\right]$

When it is written $\forall A \subset E^*$ it implies that A is measurable.

1.3 QSD, QLD and Yaglom Limit

We can now introduce the definition of a QSD but before obtaining some results on QSDs, we also need to have the notions of Yaglom limit and Quasi-Limiting Distribution.

Definition 2 Quasi-Stationary Distribution (QSD)
A probability measure α on E^* is a **quasi-stationary distribution** if:

$$\boxed{\forall t \geq 0, \forall A \subset E^*, \alpha(A) = \mathbb{P}_\alpha\left(Z_t \in A | T_0 > t\right)} \tag{1}$$

A quasi-stationary distribution for $(Z_t)_t$ evolving in the state space E is a stationary distribution for the "same" process evolving only in E^*.

Definition 3 Quasi-Limiting Distribution (QLD)
A probability measure α on E^* is a **quasi-limiting distribution** if:

$$\boxed{\exists \nu \text{ probability measure on } E^*, \forall A \subset E^*, \lim_{t\to\infty} \mathbb{P}_\nu\left(Z_t \in A | T_0 > t\right) = \alpha(A).} \tag{2}$$

This means, the distribution of the process Z_t starting with initial distribution ν reaches quasi-stationarity and converges to α, which is quasi-stationary, when t tends to ∞.

Definition 4 Yaglom limit
Z has a **Yaglom limit** if:

$$\boxed{\exists \alpha, \forall x \in E^*, \forall A \subset E^*, \lim_{t\to\infty} \mathbb{P}_x\left(Z_t \in A | T_0 > t\right) = \alpha(A).} \tag{3}$$

One should notice that if the Yaglom limit exists, it is the quasi-limiting distribution for any initial distribution on E^*. *The existence of the Yaglom limit implies the existence and the uniqueness of the QSD.*

Proposition 1 Equivalence QLD-QSD
Let α be a probability measure on E^*. Then, α is QLD for Z if and only if it is a QSD for Z.

Proof.

QLD⇒ QSD This is obvious, as we choose α to be the probability measure μ within the definition of a QLD.

Reciprocity Let α be a QLD for Z. Let ν be a probability measure on E^* such that $\forall A \subset E^*, \lim_{t\to\infty} \mathbb{P}_\nu\left(Z_t \in A | T_0 > t\right) = \alpha(A)$. Let f be a measurable and bounded function on E^*. Then we have:

$$\alpha(f) = \lim_{t\to\infty} \mathbb{E}_\nu\left[f(Z_t)|T_0 > t\right] = \lim_{t\to\infty} \frac{\mathbb{E}_\nu\left[f(Z_t), T_0 > t\right]}{\mathbb{P}_\nu\left(T_0 > t\right)}$$

$f_s : z \mapsto \mathbb{P}_z(T_0 > s)$ is a measurable and bounded function. Then, by the Markov property

$$\mathbb{P}_\alpha(T_0 > s) = \lim_{t\to\infty} \frac{\mathbb{P}_\nu\left(T_0 > s + t\right)}{\mathbb{P}_\nu\left(T_0 > t\right)}.$$

If A is a measurable set we can do the same with $g_{s,A} : z \mapsto \mathbb{P}_z\left(Z_s \in A, T_0 > s\right)$ which gives us

$$\begin{aligned}
\mathbb{P}_\alpha\left(Z_s \in A, T_0 > s\right) &= \lim_{t\to\infty} \frac{\mathbb{P}_\nu\left(Z_{t+s}, T_0 > t + s\right)}{\mathbb{P}_\nu\left(T_0 > t\right)} \\
&= \lim_{t\to\infty} \frac{\mathbb{P}_\nu\left(Z_{t+s}, T_0 > t + s\right)}{\mathbb{P}_\nu(T_0 > t + s)} \frac{\mathbb{P}_\nu(T_0 > t + s)}{\mathbb{P}_\nu\left(T_0 > t\right)} \\
&= \alpha(A)\mathbb{P}_\alpha(T_0 > s)
\end{aligned}$$

Finally, the Bayes formula gives $\alpha(A) = \mathbb{P}_\alpha\left(Z_s \in A | T_0 > s\right)$. Having it for any measurable subset A, we have proved that α is a QSD for Z.

□

Remark: When it exists, the Yaglom limit of Z is unique. There are some processes with an infinity of QSDs, for example a Markov chain with several communicating classes can have an infinity of stationary distributions. Then, the killed process associated has an infinity of QSD. Therefore we know that there are QSDs which are not a Yaglom limit. This can be summarized by:

$$\text{Yaglom limit} \Rightarrow QSD \Leftrightarrow QLD.$$

2 Basic Theory

This section exposes some of the most general and basic results that are essential in the study of the quasi-stationary distributions.

Proposition 2 Exponential extinction rate
Let Z be a Markov process with an absorbing point 0 with extinction almost surely, $\mathbb{P}_z(T_0 < \infty) = 1$. Assume that α is a QSD for the process. Then there exists $\theta(\alpha) > 0$ such that

$$\forall t \geq 0, \mathbb{P}_\alpha(T_0 > t) = e^{-\theta(\alpha)t}. \tag{4}$$

Proof.

By the Markov property, $\begin{aligned}\mathbb{P}_\alpha(T_0 > t + s) &= \mathbb{E}_\alpha\left[\mathbb{P}_{Z_t}(T_0 > s)\mathbb{1}_{T_0>t}\right] \\ &= \mathbb{P}_\alpha[T_0 > t]\mathbb{E}_\alpha[\mathbb{P}_{Z_t}\left(T_0 > s\right)|T_0 > s]\end{aligned}$

Moreover, α is a QSD, then $\mathbb{E}_\alpha\left[\mathbb{P}_{Z_t}(T_0 > s)|T_0 > t\right] = \mathbb{P}_\alpha(T_0 > s)$. The last two equations give us

$$\mathbb{P}_\alpha(T_0 > t + s) = \mathbb{P}_\alpha(T_0 > t)\mathbb{P}_\alpha(T_0 > s).$$

Let us consider $g : t \mapsto \mathbb{P}_\alpha(T_0 > t)$.
Then, we have $g(0) = 1, g(t) \xrightarrow[t\to\infty]{} 0$ and $\forall t, s > 0, g(t+s) = g(s)g(t)$. It can now be easily shown that there is constant $\theta(\alpha) > 0$ such that $\forall t \geq 0, g(t) = e^{-\theta(\alpha)t}$. □

Remark: One can obtain $\theta(\alpha)$ with the formula $\theta(\alpha) = -\frac{\ln[\mathbb{P}_\alpha(T_0>t)]}{t}$ for all $t > 0$.

Corollary 1 Exponential moments
If α is a QSD, we have that

$$\boxed{\forall 0 < \gamma < \theta(\alpha), \mathbb{E}_\alpha\left[e^{\gamma T_0}\right] < \infty \text{ and } \exists z > 0, \mathbb{E}_z\left[e^{\gamma T_0}\right] < \infty.}$$

Proof.

The first inequality follows immediately as we know that T_0 is exponentially or geometrically distributed depending on the time setting. In both time settings we have

$$\mathbb{E}_\alpha\left[e^{\gamma T_0}\right] = \int_{E^*} \mathbb{E}_z\left(e^{\gamma T_0}\right)\alpha(dz) \Rightarrow \exists z \in E^*, \mathbb{E}_z\left[e^{\gamma T_0}\right] < \infty$$

□

2.1 A spectral point of view

We define L to be the infinitesimal generator of the sub-Markovian semi-group (P_t) defined by

$$P_t f(z) = \mathbb{E}_z\left[f(Z_t)\mathbb{1}_{t<T_0}\right].$$

In this subsection we only consider continuous-time processes but we can have a similar intuition for discrete-time settings. This can be observed for the calculation of QSDs in section 4.

Definition 5 Infinitesimal generator
For all function f on E^* and for all x in E^*,

$$Lf(x) = \lim_{t\to 0} \frac{\mathbb{E}_x[f(Z_t)\mathbb{1}_{T_0>t}] - f(x)}{t} \tag{5}$$

Proposition 3
Let α be a probability measure on E^*. We make the following assumption: $\exists D \subset \mathcal{D}(L), \forall A \subset E^*, \exists (f_n)_n$ uniformly bounded in D and converging point-wisely to $\mathbb{1}_A$. Then,

$$\boxed{\alpha \text{ is a QSD for } Z \text{ if and only if } \exists\theta(\alpha) > 0, \forall f \in D, \alpha(Lf) = -\theta(\alpha)\alpha(f).}$$

Remark: Such a D subset always exists when E^* is discrete.

Let us introduce this property by an example before proving it. We consider a "biased" random walk on the state space $\{0, 1, 2, 3\}$ with uniform killing rate, modeling catastrophic events, and without immigration. It has the following transition rate matrix

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 \\ d & -(\lambda+d) & \lambda & 0 \\ d & \mu & -(\mu+2\lambda+d) & 2\lambda \\ d & 0 & \mu & -(\mu+d) \end{pmatrix}$$

One can immediately notice that 0 is an absorbing point due to the absence of immigration. If the state space is discrete we simply have $L = Q|_{E^*}$. We chose a situation in which we know how to calculate the QSD (see section 4). The QSD is $\alpha \propto \begin{pmatrix} 1 & \frac{\lambda}{\mu} & \frac{2\lambda^2}{\mu^2} \end{pmatrix}$ and obviously $\theta(\alpha) = d$.

$$\begin{aligned}
\alpha(Lf) &= \alpha_1[\lambda f_2 - (\lambda+d)f_1] + \alpha_2[2\lambda f_3 - (\mu+2\lambda+d)f_2 + \mu f_1] + \alpha_3[\mu f_2 - (\mu+d)f_3] \\
&= f_1[\mu\alpha_2 - (d+\lambda)\alpha_1] + f_2[\lambda\alpha_1 + \mu\alpha_3 - (\mu+2\lambda+d)\alpha_2] + f_3[2\lambda\alpha_2 - (\mu+d)\alpha_3] \\
&= f_1\alpha_1[\lambda - (\lambda+d)] + f_2\alpha_2[\mu+2\lambda - (\mu+2\lambda+d)] + f_3\alpha_3[\mu - (\mu+d)] \\
&= -d(f_1\alpha_1 + f_2\alpha_2 + f_3\alpha_3)
\end{aligned}$$

Finally, we have $\alpha(Lf) = -\theta(\alpha)\alpha(f)$ for any function f on $\{1,2,3\}$. In this case L is a matrix, f a column vector, α a row vector and $\theta(\alpha)$ a constant. This is true for all f; so $\alpha L = -\theta(\alpha)\alpha$ and α is a left-eigenvector associated to the eigenvalue $-\theta(\alpha)$. Those ideas are expressed clearer in section 4. The idea is that $\mathbb{E}_\alpha[f(Z_t)\mathbb{1}_{T_0>t}]$ is exponentially decreasing over time with the same parameter $\theta(\alpha)$ of the probability of extinction. Therefore, conditioned on non-absorption, this expectation is constant.
Let us now prove the property under the general assumptions.

Proof.

Let α be a QSD for Z.

Let A be a measurable subset. From $\alpha(A) = \frac{\alpha P_t(\mathbb{1}_A)}{\alpha P_t(\mathbb{1}_{E^*})}$ and $\alpha P_t(\mathbb{1}_{E^*}) = e^{-\theta(\alpha)t}$, we deduce that $\alpha P_t(\mathbb{1}_A) = e^{-\theta(\alpha)t}\alpha(A)$. This holds for any subset A, then have $\alpha P_t = e^{-t\theta(\alpha)}\alpha$. The assumption on D and the Kolmogorov's forward equation gives $\forall f \in D, \left|\frac{\partial P_t f}{\partial t}(x)\right| = |P_t L f(x)| \leq \|Lf\|_\infty$. The differentiation of $\alpha P_t f = \int\limits_{E^*} P_t f(x)\alpha(dx)$ under the integral sign is justified and gives

$$\forall f \in D, \alpha(Lf) = -\theta(\alpha)\alpha(f).$$

We assume now that $\forall f \in D, \alpha(Lf) = -\theta(\alpha)\alpha(f)$.

Kolmogorov's backward equation and the differentiation under the integral sign, justified for the same reason, give $\frac{\partial}{\partial t}\alpha(P_t f) = \alpha(LP_t f) = -\theta(\alpha)\alpha P_t(f)$. We deduce from this differential equation that $\forall f \in D, \alpha P_t(f) = e^{-t\theta(\alpha)}\alpha(f)$. The assumption on the sequence $(f_n)_n$ and the dominated convergence theorem gives finally $\alpha P_t(\mathbb{1}_A) = e^{-\theta(\alpha)t}\alpha(\mathbb{1}_A)$. By a simple computation we have $\mathbb{E}_\alpha\left[Z_t \in A | t < T_0\right] = \frac{\alpha P_t(\mathbb{1}_A)}{e^{-t\theta(\alpha)}} = \alpha(A)$. This holds for any measurable subset A therefore we proved that α is a QSD.

□

2.2 Extinction rate

The extinction rate is a value that characterizes the "life expectancy" of the process.

Definition 6 Extinction Rate
The **extinction rate** associated to the initial distribution μ is defined as follows:

Discrete-time setting $\quad r_\mu(t) = \mathbb{P}_\mu\left(T_0 = t+1 | T_0 > t\right)$
Continuous-time setting $\quad r_\mu(t) = -\frac{\frac{\partial}{\partial t}\mathbb{P}_\mu(T_0>t)}{\mathbb{P}_\mu(T_0>t)}$ when it exists and is integrable with respect to μ.

If α is a QSD, then by a previous proposition we have that the extinction rate is constant, given by

$$r_\alpha(t) = \begin{cases} 1 - e^{-\theta(\alpha)}, \text{ in the discrete-time setting} \\ \theta(\alpha), \text{ in the continuous-time setting} \end{cases}$$

Proposition 4 QLD-Extinction Rate
Let α be a QLD for Z with an initial distribution μ on E^*. In the continuous-time setting, we assume moreover that there is a positive constant $h > 0$ such that $L\left(P_h \mathbb{1}_{E^*}\right)$ is well defined and bounded. In both time settings we have that $\lim\limits_{t\to\infty} r_\mu(t) = r_\alpha(0)$.

Proof.

Discrete-time setting :

The semi-group property gives $r_\mu(t) = 1 - \frac{\mu P_t(P_1 \mathbb{1}_{E^*})}{\mu P_t(\mathbb{1}_{E^*})} \xrightarrow[t\to\infty]{} 1 - \alpha\left(P_1 \mathbb{1}_{E^*}\right) = r_\alpha(0)$.

Continuous-time setting :

We apply the Kolmogorov's backward equation $\frac{\partial}{\partial t} P_{t+h} \mathbb{1}_{E^*}(x) = P_t L(P_h \mathbb{1}_{E^*})(x)$. By the definition of the QLD and Proposition 4:

$$\frac{\partial}{\partial t} \mu P_{t+h}(\mathbb{1}_{E^*}) = \mu P_t L(P_h \mathbb{1}_{E^*})$$
$$\frac{\frac{\partial}{\partial t} \mu P_{t+h}(\mathbb{1}_{E^*})}{\mu P_t(\mathbb{1}_{E^*})} = \frac{\mu P_t L(P_h \mathbb{1}_{E^*})}{\mu P_t(\mathbb{1}_{E^*})} \xrightarrow[t\to\infty]{} \alpha(L P_h \mathbb{1}_{E^*}) = -\theta(\alpha)\alpha(P_h \mathbb{1}_{E^*})$$
$$\frac{\mu(P_{t+h} \mathbb{1}_{E^*})}{\mu(P_t \mathbb{1}_{E^*})} \xrightarrow[t\to\infty]{\alpha} (P_h \mathbb{1}_{E^*})$$

With the last two equations we have: $-\frac{\frac{\partial}{\partial t} \mu(P_{t+h} \mathbb{1}_{E^*})}{\mu(P_{t+h} \mathbb{1}_{E^*})} = r_\mu(t+h) \xrightarrow[t\to\infty]{} \theta(\alpha) = r_\alpha(0)$

□

3 Example 1: the Galton-Watson Process

The Galton-Watson process is a good example where it is easy to understand the different cases in which there is or there is not a QSD, and when we can observe the quasi-limiting regime.

3.1 Notations and Definitions

We consider the process Z defined by $Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n)}$ where the random variables $\left(\xi_i^{(n)}\right)_{i,n}$ are independent and identically distributed following the probability measure μ on $\mathbb{N}$. We use the following notation:

- generating function of μ: $g(t) = \mathop{\mathbb{E}}_{X\sim\mu}\left[t^X\right] = \sum_{k=0}^{\infty} \mu(k) t^k$
- $m := \mathbb{E}\left[\xi_1^{(0)}\right]$ average number of children by individual

We have the branching property; the process Z is equal to the sum of Z_0 independent Galton-Watson processes issued from a single individual.

Definition 7 Probability of Extinction
By the branching property we can compute the **probability of extinction**

$$\boxed{\mathbb{P}_1(\exists n \in \mathbb{N}, Z_n = 0) = \lim_{n\to\infty} \mathbb{E}_1\left[0^{Z_n}\right] = \lim_{n\to\infty} g \circ g \circ ... \circ g(0) \text{ (n times).}}$$

The different values of this probability of extinction give three distinct situations:

Sub-critical case (m<1):
Extinction happens in finite time almost surely and $\mathbb{E}_1[T_0] < \infty$.

Critical case (m=1):
$\mathbb{P}_1(\exists n \in \mathbb{N}, Z_n = 0) = 1$ and $\mathbb{E}_1[T_0] = \infty$.

Super-critical case (m>1):
$\mathbb{E}_1[T_0] = \infty$ and $\mathbb{P}_1(\exists n \in \mathbb{N}, Z_n = 0) = 1 - \mathbb{P}_1(\forall n \in \mathbb{N}, Z_n > 0) < 1$.

The following theorem gives us necessary conditions for the existence of QSDs.

3.2 Existence and Uniqueness of the QSD

Theorem 1

- There is no QSD in the critical and the super-critical case.
- In the sub-critical case, the Yaglom limit exits and is the unique QSD of Z.

Proof.

Critical and super-critical case

$\mathbb{E}_1[T_0] = \infty$ implies $\mathbb{E}_\alpha[T_0] = \infty$ for all α probability measure on $\mathbb{N}^*$ by the branching property. We deduce from the exponential moments proposition that there is no QSD.

Sub-critical case

We consider the process with initial distribution $\nu = \delta_1$. For $n \geq 0$, let g_n denote the generating function of Z_n, i.e. for $s \in [0,1], g_n(s) = \mathbb{E}_\nu\left[s^{Z_n}\right]$. One can notice $g_{n+1} = g_n \circ g = g_0(g^{n+1})$ with a simple calculation using the branching property.

$$\begin{aligned}
g_{n+1}(s) = \mathbb{E}_\nu\left[s^{Z_{n+1}}\right] &= \mathbb{E}_\nu\left[\prod_{i=1}^{Z_n} s^{\xi_i^{(n)}}\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_\nu\left[\prod_{i=1}^{Z_n} s^{\xi_i^{(n)}} | Z_n = k\right] \mathbb{P}_\nu(Z_n = k) \\
&= \sum_{k=0}^{\infty} \left(\prod_{i=1}^{k} \mathbb{E}_\mu\left[s^{\xi_i^{(n)}}\right]\right) \mathbb{P}_\nu(Z_n = k) \\
&= \sum_{k=0}^{\infty} g(s)^k \mathbb{P}_\nu(Z_n = k) \\
&= g_n(g(s)) = g_0(g^{n+1}(s))
\end{aligned}$$

Since $\nu = \delta_1$, $g_0(s) = s$ and therefore $\boxed{g_{n+1} = g^{n+1} = g \circ g_n.}$

The same way we note $\hat{g}_n$ the generating function of Z_n conditioned to $\{Z_n > 0\} = \{T_0 > n\}$. For $s$in $[0,1]$ we have:

$$\begin{aligned}
\hat{g}_n(s) = \mathbb{E}_\nu\left[s^{Z_n} | Z_n > 0\right] &= \frac{\mathbb{E}_\nu\left[s^{Z_n}\mathbb{1}_{Z_n>0}\right]}{\mathbb{P}_\nu[Z_n > 0]} \\
&= \frac{\mathbb{E}_\nu\left[s^{Z_n}\right] - \mathbb{P}_\nu(Z_n = 0)}{1 - \mathbb{P}_\nu(Z_n = 0)} = \frac{g_n(s) - g_n(0)}{1 - g_n(0)} \\
&= 1 - \frac{1 - g_n(s)}{1 - g_n(0)} \in [0,1]
\end{aligned}$$

Let us now consider the function $\Gamma : s \mapsto \Gamma(s) = \frac{1-g(s)}{1-s}$ defined on $[0,1[$. The function g being convex implies that Γ is non-decreasing. Then, we have the induction

$$\forall n \geq 0, 1 - \hat{g}_{n+1}(s) = \tfrac{\Gamma(g_n(s))}{\Gamma(g_n(0))}(1 - \hat{g}_n(s)).$$

We know $m = g'(1) < 1, g(1) = 1$ and g is convex. It implies that $\forall x \in [0,1], g(x) \geq x$ and by induction $\forall n \geq 1, g_n(x) \geq x$. In addition, we know that $(g_n(s))$ is non-decreasing in n. We have established that g_n and Γ are non-decreasing, therefore:

$$\begin{aligned}
\Gamma(g_n(s)) \geq \Gamma(g_n(0)) \Rightarrow \frac{\Gamma(g_n(s))}{\Gamma(g_n(0))} &\geq 1 \\
\Rightarrow 1 - \hat{g}_{n+1}(s) &\geq 1 - \hat{g}_n(s)
\end{aligned}$$

We have now that $1-\hat{g}_n(s)$ is non-decreasing in n. In particular, we know that $\hat{g}_n(s)$ converges and we note $\hat{g}(s) = \lim_{n\to\infty} \hat{g}_n(s)$. We denote by α the finite measure associated to $\hat{g}$.

Let us prove that α is a probability measure. That would mean α is a quasi-limiting distribution for the initial distribution δ_1. We have

$$\Gamma(g_n(0))(1-\hat{g}_{n+1}(s)) = \frac{1-g_{n+1}(0)}{1-g_n(0)}\frac{1-g_{n+1}(s)}{1-g_{n+1}(0)} = \frac{1-g_n(g(s))}{1-g_n(0)} = 1-\hat{g}_n(g(s))$$

and $\lim_{n\to\infty} \Gamma(g_n(s)) = \lim_{n\to\infty} \frac{g(1)-g(g_n(s))}{1-g_n(s)} = g'(1) = m$. Therefore, when n tends to ∞ we have

$$m(1-\hat{g}(s)) = 1-\hat{g}(g(s)).$$

Finally we have $\hat{g}(g(s)) = m\hat{g}(s) + 1 - m$ that gives $\hat{g}(1) = 1$ meaning α is a probability measure. Then, its definition implies directly that α is a QLD for Z with initial distribution δ_1

We have to show that there is a QLD for any initial distribution and that the QLD actually does not depend on the initial distribution. First, we prove this for any initial distribution ν such that $\nu = \delta_k$. Let k be an integer greater than or equal to 2. We note $h_n(s) = \mathbb{E}_k\left[s^{Z_n}\right]$ and $\hat{h}_n(s) = \mathbb{E}_k\left[s^{Z_n}|Z_n>0\right]$ the generating functions. We want to prove that $\hat{h}_n$ converges and that its limit $\hat{h}$ is $\hat{g}$.

Like for $\hat{g}_n$ and g_n we have $\hat{h}_n(s) = 1 - \frac{1-h_n(s)}{1-h_n(0)}$. The branching property gives

$$h_n(s) = \mathbb{E}_k\left[s^{Z_n}\right] = \prod_{i=1}^{k}\left(\mathbb{E}_1\left[s^{Z_n^{(i)}}\right]\right) = (g_n(s))^k$$

Therefore

$$\begin{aligned}
1-\hat{h}_n(s) &= \frac{1-g_n(s)^k}{1-g_n(0)^k} \\
&= \frac{1-g_n(s)}{1-g_n(0)}\frac{1+g_n(s)+\cdots+g_n(s)^{k-1}}{1+g_n(0)+\cdots+g_n(0)^{k-1}} \\
&= (1-\hat{g}_n(s))\frac{1+g_n(s)+\cdots+g_n(s)^{k-1}}{1+g_n(0)+\cdots+g_n(0)^{k-1}}
\end{aligned}$$

We already know that $(g_n(s))_n$ converges to 1 for all s in $[0,1]$. As a consequence, $\hat{h}_n(s)$ is also convergent
$1-\hat{h}_n(s) \xrightarrow[n\to\infty]{} (1-\hat{g}(s))\,\frac{k}{k} = 1-\hat{g}(s)$.

We note $\hat{h}$ its limit and we have $\hat{h}(s) = \hat{g}(s)$ for all s in $[0,1]$ which means $\hat{h} = \hat{g}$. This proves that α is the QLD for both initial distributions δ_1 and δ_k. That being true for all $k \geq 2$ implies that α is the Yaglom limit of the process and therefore, its unique QSD.

□

4 Calculate a/the QSD

Theorems give the existence and sometimes the uniqueness of QSDs but in practice we essentially want to have the value of the QSD. Let's find out some situations where we can analytically calculate the QSD. There are two ways to tackle this problem. You can try to find a stationary distribution for the process conditioned to non-extinction using the generator for this "new" process, but then this is not specific to quasi-stationary distributions anymore. We want to know what means being a QSD with respect to the generator without the condition of non-extinction. In this section, we consider processes with a discrete state space and a unique absorbing point 0.

4.1 Continuous-time setting

Let α be a QSD for the process $(Z_t)_t$ with transition rate matrix $P = (p_{ij})_{i,j\in E}$. Let P^* denote the sub-stochastic matrix $P^* = (p_{ij})_{i,j\in E^*}$ and since 0 is the only absorbing state we have $P = (p_{ij})_{i,j\geq 0}$ and $P^* = (p_{ij})_{i,j>0}$. For all $j \geq 1$ and for t close to 0 we have:

$$\begin{aligned}\alpha_j = \mathbb{P}_\alpha(Z_t = j|T_0 > t) &= \frac{\mathbb{P}_\alpha(Z_t = j)}{1 - \mathbb{P}_\alpha(T_0 \leq t)} \\ &= \frac{\sum\limits_{i>0} \alpha_i \mathbb{P}_i(Z_t = j)}{1 - \sum\limits_{i>0} \alpha_i \mathbb{P}_i(T_0 \leq t)} = \frac{\alpha_j + \left(\sum\limits_{i>0} \alpha_i p_{ij}\right) t + o(t)}{1 - \left(\sum\limits_{i>0} p_{i0}\right) t + o(t)} \\ &= \alpha_j + t\left(\sum_{i>0} \alpha_i p_{ij} + \alpha_j \sum_{i>0} p_{i0}\right) + o(t)\end{aligned}$$

This implies $\sum\limits_{i>0} \alpha_i p_{ij} + \alpha_j \sum\limits_{i>0} \alpha_i p_{i0} = 0$, therefore $(\alpha P)_j = -\left(\sum\limits_{i>0} \alpha_i p_{i0}\right) \alpha_j$ and finally $\alpha P^* = -\theta(\alpha)\alpha$ where $\theta(\alpha) = \sum\limits_{i>0} \alpha_i p_{i0}$ is the parameter in the probability of death at quasi-stationarity $\mathbb{P}_\alpha(T_0 > t) = e^{-\theta(\alpha)t}$ (see section 2). Finally, α is a left-eigenvector of P^* associated to the negative eigenvalue $-\theta(\alpha)$. There is a second way to prove that in the continuous-time setting:

$$\alpha \text{ is a QSD} \Rightarrow \alpha P^* = -\theta(\alpha)\alpha$$

One can find in section 2 of [3] that $-\theta(\alpha)$ is the principal eigenvalue, the eigenvalue with the largest magnitude. To calculate α, we need to obtain the set of the eigenvalues of L and find a left-eigenvector v associated to the principal eigenvalue $-\theta(\alpha)$ with non-negative entries. The QSD α is then the normalized vector $v/\|v\|$.
Obtaining eigenvectors and eigenvalues in the finite case is not complicated, it is just solving a set of linear equations, but it becomes computationally expensive when the number of states is very large. Moreover, solving this problem is very hard to do for infinite or continuous state-spaces.

4.2 Discrete-time setting

Let α be a probability measure on E^*. Let $P = (\mathbb{P}_i(Z_1 = j))_{i,j\in E} = (p_{ij})_{i,j\geq 0}$ be the Markov kernel of our process. We denote P^* the sub-stochastic matrix defined on $E^* = E\backslash\{0\}$ that you get by removing the rows and columns from P that relate to the absorbing state 0. If α is a QSD, for all $j > 0$:

$$\alpha_j = \mathbb{P}_\alpha(Z_1 = j|T_0 > 1) = \frac{\mathbb{P}_\alpha(Z_1=j)}{\mathbb{P}_\alpha(T_0>t)} = \frac{\sum\limits_{i>0} \alpha_i p_{ij}}{1-\sum\limits_{i>0} \alpha_i p_{i0}}$$

Finally,

$$\forall j > 0, \sum_{i>0} \alpha_i p_{ij} = \left(1 - \sum_{i>0} \alpha_i p_{i0}\right) \alpha_j$$

We have $\alpha P^* = \lambda\alpha$ where $\lambda = 1 - \sum\limits_{i>0} \alpha_i p_{i0} = 1 - \mathbb{P}_\alpha(T_0 = 1) = \mathbb{P}_\alpha(T_0 > 1) = e^{-\theta(\alpha)}$ is the probability of survival to the next step at quasi-stationarity. Besides, this shows hox to obtain $\theta(\alpha)$ from α. Hence,

$$\alpha P^* = e^{-\theta(\alpha)}\alpha$$

Here again, paper [3] indicates that $e^{-\theta(\alpha)}$ is the principal eigenvalue of P^*, and therefore α is a solution to principal eigenvector problem.
To obtain α, we do the same thing we do for the continuous-time setting. First, calculate the set of eigenvalues of P^* and find a left-eigenvector associated to the principal eigenvalue $1 - e^{-\theta(\alpha)}$ with positive entries. Then, normalize it and you have the QSD α.

We now know the equations we have to resolve to find the QSD of a process. As in the continuous-time setting, solving this is only simple in the finite state space and a reasonably small number of states. That is why we focus on a specific case that allows to calculate analytically the quasi-stationary distribution of a process.

4.3 Birth and death processes with uniform extinction rate

Let us consider birth and death processes with uniform extinction rate, i.e.

$$\forall y, z \in E^*, \forall t > 0, \mathbb{P}_z(T_0 > t) = \mathbb{P}_y(T_0 > t).$$

4.3.1 Discrete-time setting

Let $P = (\mathbb{P}_i(Z_1 = j))_{i,j\geq 0}$ denote the transition probability matrix of the process. We make the following assumptions:

1. $p_{00} = 1$ (i.e. 0 is an absorbing point)
2. $\forall i > 0, 1 = p_{i0} + p_{i,i-1} + p_{ii} + p_{i,i+1}$ with $p_{i,i-1} > 0$ and $p_{i,i+1} > 0$ when they exist.
3. $\exists d > 0, \forall i > 0, p_{i0} = d$

The uniform extinction probability allows us to calculate $\theta(\alpha) = -\ln(1-d)$ and this simplifies our problem. We want now to resolve:

$$\begin{cases} \alpha P^* = (1-d)\alpha & 1. \\ \sum\limits_{i>0} \alpha_i = 1 & 2. \\ \forall i > 0, \alpha_i \geq 0 & 3. \end{cases}$$

Equation 1. can be written $\forall j > 0, \sum\limits_{i>0} \alpha_i p_{ij} = (1-d)\alpha_j$. For an infinite state space, the assumptions we made implies

$$\begin{cases} \alpha_1 p_{11} + \alpha_2 p_{21} = (1-d)\alpha_1 \\ \alpha_{j-1} p_{j-1,j} + \alpha_j p_{jj} + \alpha_{j+1} p_{j+1,j} = (1-d)\alpha_j \text{ for all } j > 1 \end{cases}$$

The first equation and assumption 2 give $1 - d - p_{11} = p_{12}$ and then $\alpha_2 = \frac{p_{12}}{p_{21}}\alpha_1$. Let us assume that $\alpha_j = \frac{p_{j-1,j}}{p_{j,j-1}}\alpha_{j-1} = \left(\prod\limits_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}\right)\alpha_1$. We can now obtain the value of α_{j+1}.

$$\begin{aligned} \alpha_{j+1} &= \frac{1}{p_{j+1,j}}\left(\alpha_j(1-d-pjj) - \alpha_{j-1}p_{j-1,j}\right) \\ &= \frac{\alpha_{j-1}}{p_{j+1,j}}\left(\frac{p_{j-1,j}}{p_{j,j-1}}(p_{j,j-1} + p_{j,j+1}) - p_{j-1,j}\right) \\ &= \frac{\alpha_{j-1}}{p_{j+1,j}}\frac{p_{j-1,j}}{p_{j,j-1}}p_{j,j+1} \\ &= \left(\prod_{i>1}^{j+1} \frac{p_{i-1,i}}{p_{i,i-1}}\right)\alpha_1 \end{aligned}$$

If α is a QSD, we have by induction $\forall j \geq 1, \alpha_j = \left(\prod\limits_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}\right)\alpha_1$. On the other hand, if α is a probability measure on E^* and satisfies $\forall j \geq 1, \alpha_j = \left(\prod\limits_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}\right)\alpha_1$, it is a QSD. We can now state, under the assumptions made earlier, that

$$\boxed{\text{there is a QSD for } (Z_t)_t \text{ if and only if } \sum_{j\geq 1}\left(\prod_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}\right) < \infty.}$$

In this case, there is a unique QSD α and $\forall j > 0, \alpha_j = \dfrac{\prod\limits_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}}{\sum\limits_{j\geq 1}\left(\prod\limits_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}\right)}$.

If the state space is finite, i.e. $E^* = \{1, \ldots, N\}$, there is always a unique QSD α and for all j in $\{1, \ldots, N\}, \alpha_j = \dfrac{\prod\limits_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}}{\sum\limits_{j\geq 1}^{N}\left(\prod\limits_{i>1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}}\right)}$.

4.3.2 Continuous-time setting

Let $P = (p_{ij})_{i,j\geq 0}$ denote the transition rate matrix of the process $(Z_t)_t$. We make the following assumptions:

1. $\forall i > 0, p_{ii} = -(d + p_{i,i-1} + p_{i,i+1})$ with $p_{i,i-1} > 0$ and $p_{i,i+1} > 0$ when they exist.
2. $\exists d > 0, \forall i > 0, p_{i0} = d$

Here again, we know the value of $\theta(\alpha) = d$ and this simplifies our problem. We want now to resolve:

1. $\alpha P^* = -d\alpha$
2. $\sum_{i>0} \alpha_i = 1$
3. $\forall i > 0, \alpha_i \geq 0$

We can apply exactly the same induction we used in the discrete-time setting and the result is still very similar. Under our assumptions,

$$\boxed{\text{there is a QSD for } (Z_t)_t \text{ if and only if } \sum_{j\geq 1}\left(\prod_{i>1}^{j}\frac{p_{i-1,i}}{p_{i,i-1}}\right) < \infty.}$$

In this case, there is a unique QSD α and $\forall j > 0$, $\alpha_j = \dfrac{\prod_{i>1}^{j}\frac{p_{i-1,i}}{p_{i,i-1}}}{\sum_{j\geq 1}\left(\prod_{i>1}^{j}\frac{p_{i-1,i}}{p_{i,i-1}}\right)}$.

If the state space is finite, $E^* = \{1, \dots, N\}$, there is always a unique QSD α and for all $j > 0$, $\alpha_j = \dfrac{\prod_{i>1}^{j}\frac{p_{i-1,i}}{p_{i,i-1}}}{\sum_{j\geq 1}^{N}\left(\prod_{i>1}^{j}\frac{p_{i-1,i}}{p_{i,i-1}}\right)}$.

5 Monte-Carlo methods

The previous sections were about cases with existence and uniqueness of the QSD and the cases allowing to obtain analytically the QSD. However, most of the time we cannot obtain the QSD and therefore we want to use the connection with QLD and Yaglom limit to develop algorithms approximating the QSD. The different methods used here are developed in [3] and [4]. One can also find other methods and motivations in those references.

5.1 Naive Method

Let us consider a process $(Z_t)_t$ which has a Yaglom limit α. This means there exists a unique QSD which is α. Then we have:

$$\forall x \in E^*, \forall A \subset E^*, \lim_{t\to\infty} \mathbb{P}_x[Z_t \in A | T_0 > t] = \alpha(A)$$

Let ν be a distribution on E^* and α_t be the distribution defined by $\alpha_t(A) = \mathbb{P}_\nu(Z_t \in A | T_0 > t)$. Then, α_t converges to α when t tends to ∞ by the definition of the Yaglom limit. Moreover, the strong law of large numbers gives that $\hat{\alpha}_t$ converges to α_t when our sample size N tends to ∞. The intuition is that, assuming the process converges faster to quasi-stationarity than goes to extinction, you can approximate the QSD by looking at the empirical distribution of the process simulated until a sufficiently large time T when it is still alive at this time.

We know that processes are dying at each step and dead individuals are useless because they do not appear in the calculation of $\hat{\alpha}_t$. This method is not very efficient because we need to use very large values of N and T to have good approximations of the QSD. Increasing T would bring processes closer to quasi-stationarity but the death rate would also be higher and that implies a larger error. This means

that we always need to increase both time T and sample size N at the same time if we want to reduce the error. Given the result on exponential killing rate one can find in section 2, N will need to increase exponentially with T to make sure we have a reasonable proportion of particles alive at time T. Therefore, we must find different ways to simulate those processes in order to estimate efficiently α.

5.2 Method A

This method is inspired by the Fleming-Viot method, used by the physicists to model interacting particle system, and by urn processes. One can find more information and proofs about this subsection in [3].

5.2.1 Fleming-Viot

Let us consider N particles evolving independently with respect to the same Markov process with absorption. When one particle gets absorbed, it starts again from a uniformly picked positions from the N-1 particles not absorbed. When time and sample size tend to infinity, this would converge to the quasi-stationary distribution of the process. But to obtain good approximations we need to increase both the sample size and the time of the simulation and this is computationally expensive.

5.2.2 Adapted method

We adapt the idea of the Fleming-Viot method with ideas from urn processes. The next algorithm is said to be a generalization of Polya's urn process.

Algorithm 1 Fleming-Viot
Let N be our sample size and K the number of non-absorbing states. We note $(X_t)_t$ the Markov process we are observing.

1. Initialize a vector $\mu = 0$ with K rows.
2. Take i uniformly in $[\![1, K]\!]$ and initialize $X_0 = i$.
3. Simulate the process from X_0 until it gets absorbed.
   Update μ by adding the number of visits to each non-absorbing state to the corresponding row.
4. Simulate X_0 from μ', which is μ normalized.
5. Repeat N times steps 3 and 4 and return μ'.

We then take μ' as an approximation of the QSD α.

Remark: if we want to, we can also have an approximation of the life expectancy of the process. We just have to memorize the life time of the particles and take the average as an approximation. This algorithm can probably be extended to an infinite or even continuous state space, we just need to use a different initial distribution, and change the way to store the occupation measure. For example, it can be a normalized sum of normal distributions with means equal to the visited states if the state space is $\mathbb{R}$.

5.3 Method B

5.3.1 Motivation

The ideas in [4] were developed originally for simulating QSDs on reducible state spaces, case in which we do not have uniqueness for the QSD. In the precedent algorithm we simulated one process after the other but now we start from a population of size N and simulate every individual to the next generation at the same time. However, we need to do things differently from the naive method.
First, we know that the dead processes can only be used to calculate the extinction rate but not for estimating the QSD. We want to have a constant number of alive processes. The basic idea is to restart the dead individuals from new "alive" positions, as it is done in the original Fleming-Viot method. The question now is what distribution should be used to refill our sample?
One can notice with some practice that most of the methods underestimate the QSD in the low-probability regions and thus overestimate it in the high-probability regions. We still want to visit the

transient states with the lowest probabilities. Then we just refill with a uniform distribution over the transient states but we do not want to lose any information and just keeping the uniform distribution will prevent convergence to happen or severely slow it. The compromise between the two points of view and the main idea here is to assign weight to each individuals. Therefore, low-probability states will still be visited but the weight of the processes will decrease as time increase to finally converge to the QSD value of this state.
This method allows to visit the low-density regions without overestimating the density there. To do so with the naive method, we need a very large sample size and this is very expensive in terms of computation. Weighted particles is therefore a good idea to solve this problem.

5.3.2 Description of Combine-Split Algorithms

The things we want to know are how the weights evolve and how the sample is refilled with the dead particles. We can resume this in three simple steps. We consider the weighted sample $\left((X_t^i)_{1\leq i\leq N}, (w_t^i)_{1\leq i\leq N}\right)$ at time t:

$Refill$	All the dead particles, i.e. with weight equal to 0, are redistributed according to the uniform distribution over the set of the transient states with positive weight.
$Combine$	At each transient state s, we calculate $P_t^s = \sum_{i=1}^{N} \mathbb{1}_{X_t^i=s}$ and $W_t^s = \sum_{i=1}^{N} w_t^i \mathbb{1}_{X_t^i=s}$.
$Split$	At each state s, weight is shared equally between all the particles in the state s. Practically, if $X_t^i = s$, the new weight is $w_t^i = \frac{W_t^s}{P_t^s}$.

Remark: the weight at each state remains constant during combine-split refilling. The paper [4] presents also the method with resampling instead of refiling. Only one point survives at each point, all the others are killed and their weights transferred to the survivor particle.Then you apply the same three steps described above.

Algorithm 2 Combine-Split Refilling
Initialization: simulate N weighted particles $\{(X_0^i, w_0^i) : 1 \leq i \leq N\}$ from the uniform distribution on E^* with normalized weight, i.e. $\forall i, w_0^i = \frac{1}{N}$. While time $t < \delta$:

1. Simulate the N particles from time t to $t + \delta$.
2. For each dead particle $X^i = 0$, remove its weight, $w^i = 0$.
3. For each $s \in E^*, W_t^s := \sum_{i=1}^{N} w_t^i \mathbb{1}_{X_i=s}$ and $E_t^* = \{s \in E^* : W_t^s > 0\}$.
4. For each dead particle, draw a new position to restart the process from the uniform distribution over E_t^*.
5. For each $s \in S^*$ let $P_t^s = \sum_{i=1}^{N} \mathbb{1}_{X_i=s}$.
6. For i = 1, . . . ,N set $w_t^i = W_t^s / P_t^s$ where $s = X_t^i$.

Normalize the weights and return $\{(X_t^i, w_t^i) : 1 \leq i \leq M\}$.

One can find more results and proofs about combine-split algorithms in [4]. This paper focus more on resampling than refilling but they seem to be equally efficient.

Remark: In the case of reducible state spaces, this method allows to control the weight or the number of particles within a communicating class and therefore we can choose which QSD we want to approximate when we do not have uniqueness.

6 Simulations and Results

We now compare the two methods and want to see their characteristics, the situations in which they are efficient and the ones they are not. We are also interested in the dependance on the parameters for both methods. We choose to apply these algorithms for a birth an death process with uniform killing, birth and death probabilities, i.e. the transition probability matrix is such as

$$Q = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ d & 1-(\lambda+d) & \lambda & \dots & 0 & 0 \\ d & \mu & 1-(\lambda+\mu+d) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ d & 0 & 0 & \dots & 1-(\lambda+\mu+d) & \lambda \\ d & 0 & 0 & \dots & \mu & 1-(\mu+d) \end{pmatrix}$$

As it is explained in section 4,we know how to calculate the QSD in this case and it makes easier to compare the results. We ran simulations for the simple case $E^* = \{1, \dots, 10\}$. Let us see the sample size influence on combine-split refilling results. We always use the uniform distribution as the initial distribution for the combine-split algorithm. Figure 1 shows the results for 30 combine/split/refilling operations occurring every 3 new generations.

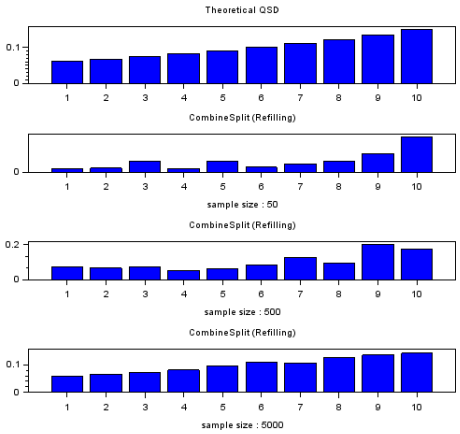
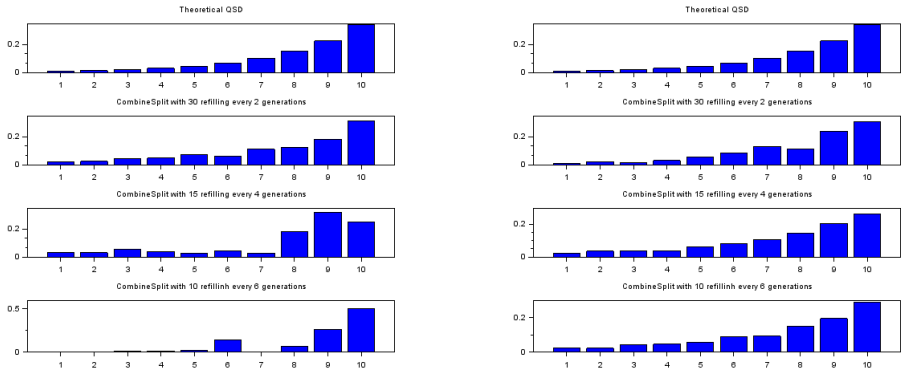


Figure 1: $\lambda = 0.4$, $\mu = 0.33$, $d = 0.2$. Simulation times = 0.2904139,2.0099105,21.928113

Computation time is linear with respect to the sample size. Even with the smallest sample size, we have some weight in the low-density region. This cannot be done with a non-weighted sample. Figure 2 shows that the number of generations between two combine-split refilling is important and its optimization depends on the killing probability d. It is pretty intuitive that when the extinction probability d increases we need to refill more regularly because individuals are dying faster.



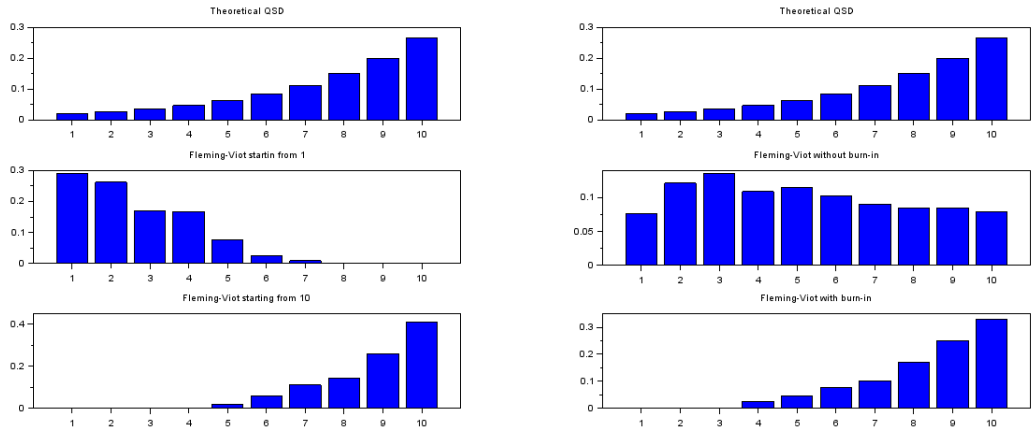
(a) $d = 0.4$ (b) $d = 0.15$

Figure 2: $\lambda = 0.3$, $\mu = 0.2$

The two following graphics show that the results of the Fleming-Viot algorithm can be strongly

dependent on the starting point. Instead of increasing the parameter T, the usual solution is using burn-in.

We can see that starting from 10 is far better than starting from 1, but optimizing the starting point can be done here only because we already know the QSD. Therefore, including burn-in is the best option and figure 3.(b) shows that it is relevant. And most of the time we do not know what the QSD looks like so we cannot figure out what could be a good starting point.

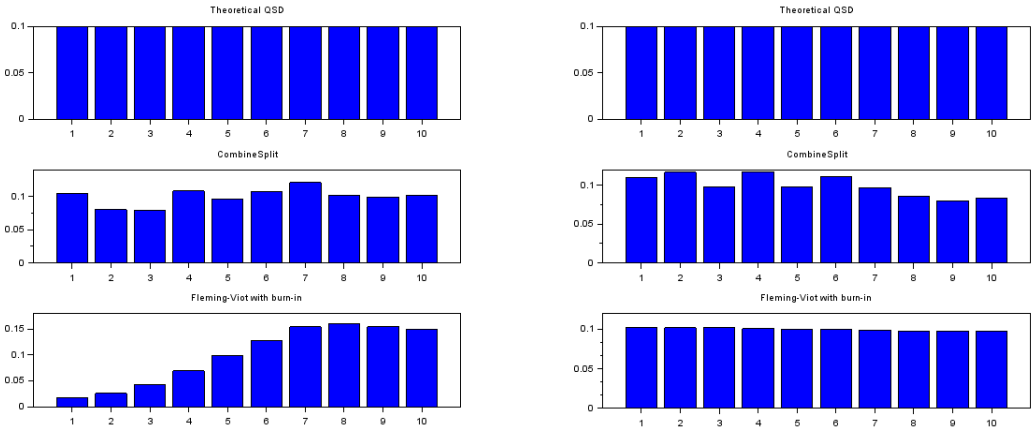


(a) 100 particles (b) 50 particles burnt among 500

Figure 3: $\lambda = 0.3$,$\mu = 0.3$,$d = 0.3$

One can also notice that starting from 10 or adding burn-in improves significantly the results. However, the estimation is still very bad for the low-density region, which is not the case with the combine-split algorithm.

We can look at different situations where one method works significantly better than the other (for the same computational time). Here are two examples with the same (uniform) QSD but different extinction rates and more or less volatility.



(a) d=0.5,lambda=mu=0.2 (b) d=0.1,mu=lambda=0.45

Figure 4: $\lambda = \mu$

7 Application to Extreme Values of GARCH Processes

We now present one application area where quasi-stationary distributions arise as a distribution associated with extremes of a GARCH process. Currently, there are no algorithms for simulating from these extremal distributions, but, by recognising them as QSDs, we might be able to use algorithms from Section 6 to sample from them.

7.1 GARCH Processes

GARCH stands for Generalized Autoregressive Conditionally Heteroskedastic.

Definition 8 GARCH(p,q) Process
$(X_t)_{t\in\mathbb{Z}}$ is a GARCH(p,q) process if for all t, $X_t = \sigma_t Z_t$ where Z_t are i.i.d. random variables that are symmetric with mean 0 and variance 1. And if there are non-negative constants $(\alpha_i)_{0\leq i\leq q}, (\beta_j)_{1\leq j\leq p}$, such that:

$$\boxed{\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2}$$

Remark: GARCH processes are used to model financial time series with time-varying volatility clustering, i.e. periods of swings interspersed with periods of relative calm, such as prices.

Let (A_t, B_t) be an i.i.d. sequence of random $d\times d$ stochastic matrices A_t and d-dimensional vectors B_t. We consider now a stochastic recurrence equation (SRE) for the d-dimensional time series (Y_t)

$$\boxed{Y_t = A_t Y_{t-1} + B_t, t\in\mathbb{Z}}$$

One can notice that GARCH(p,q) processes satisfy this SRE with $d=(p-1)+(q-1)$. The recurrence equation characterizing GARCH processes can be written as follows

$$\sigma_t^2 = \begin{pmatrix}\alpha_1 & \dots & \alpha_q & \beta_1 & \dots & \beta_q\end{pmatrix}\begin{pmatrix}X_{t-1}^2 & \dots & X_{t-q}^2 & \sigma_{t-1}^2 & \dots & \sigma_{t-p}^2\end{pmatrix}^T + \alpha_0.$$

And therefore

$$X_t^2 = \begin{pmatrix}\alpha_1 Z_t^2 & \dots & \alpha_q Z_t^2 & \beta_1 Z_t^2 & \dots & \beta_q Z_t^2\end{pmatrix}\begin{pmatrix}X_{t-1}^2 & \dots & X_{t-q}^2 & \sigma_{t-1}^2 & \dots & \sigma_{t-p}^2\end{pmatrix}^T + \alpha_0 Z_t^2.$$

If we denote $A_t = \begin{pmatrix} Z_t^2 A^{(q-1)} & Z_t^2\alpha_q & Z_t^2 B^{(p-1)} & Z_t^2\beta_p \\ I_{q-1} & 0_{q-1} & 0_{q-1} & 0_{q-1} \\ A^{(q-1)} & \alpha_q & B^{(p-1)} & \beta_p \\ 0_{p-1} & 0_{p-1} & I_{p-1} & 0_{p-1}\end{pmatrix}$ and $B_t = \begin{pmatrix}\alpha_0 Z_t^2 \\ 0_{q-1} \\ \alpha_0 \\ 0_{q-1}\end{pmatrix}$ with
$A^{(q-1)} = \begin{pmatrix}\alpha_1 & \dots & \alpha_{q-1}\end{pmatrix}$ and $B^{(p-1)} = \begin{pmatrix}\beta_1 & \dots & \beta_{p-1}\end{pmatrix}$,
the sequence $Y_t = \begin{pmatrix}X_t^2 & \dots & X_{t-q+1}^2 & \sigma_t^2 & \dots & \sigma_{t-p+1}^2\end{pmatrix}^T$ satisfies the SRE : $Y_t = A_t Y_{t-1} + B_t$, $t\in\mathbb{Z}$.

Remark: there are several different ways to embedded a GARCH process in a SRE.

7.2 Existence of a stationary solution for SRE

One can find more references in [5] about the link between the Lyapunov exponent and stochastic recurrence equations.

Definition 9 Lyapunov exponent
When it is well defined, $\gamma = \inf_{n\geq 1}\left\{\frac{1}{n}\mathbb{E}\left[\ln\|A_1\dots A_n\|\right]\right\}$ is called the Lyapunov exponent for the sequence $(A_i)_i$.

If $\mathbb{E}\left[\ln^+\|A_1\|\right] < \infty$, we have a simpler expression for the Lyapunov exponent; $\gamma = \lim_{n\to\infty}\frac{1}{n}\ln\left(\|A_1\dots A_n\|\right)$ almost surely.
Theorem 2.1 in [6] gives sufficient conditions for the existence and the uniqueness of a stationary solution of the SRE.

Theorem 2 Solution of the SRE
Let us assume $\mathbb{E}\ln^+\|A_1\| < 0$, $\mathbb{E}\ln^+|B_1| < \infty$ and $\gamma < 0$. Then we have that the series defined by $X_t = B_t + \sum_{k=1}^{\infty} A_t A_{t-1}\dots A_{t-k+1}B_{t-k-1}$ converges almost surely and is the unique strictly stationary causal solution of the stochastic recurrence equation.

For $d = 1$, we can even calculate $\gamma = \mathbb{E}[\ln|A_1|]$. Therefore, the conditions become particularly simple, i.e. $-\infty \leq \mathbb{E}[\ln|A_1|] < 0$ and $\mathbb{E}\left[\ln^+|B_1|\right] < \infty$ imply the existence and the uniqueness of the strictly stationary causal solution of the SRE.

7.3 Multivariate Regular Variation

One can find more about multivariate regular variation in [5] and [9].

Definition 10 Regular Variation
The d-dimensional random vector X is said to be *regularly varying* with index $\alpha \geq 0$ if there exists a random vector with values (almost surely) in $\mathbb{S}^{d-1}$, unit sphere in $\mathbb{R}^d$ with respect to the norm $|.|$, such that for all $t > 0$

$$\frac{\mathbb{P}(|X|>tx, X/|X|\in .)}{\mathbb{P}(|X|>x)} \xrightarrow[x\to\infty]{v} t^{-\alpha}\mathbb{P}(\Theta \in .).$$

In this case, the distribution of Θ is called the spectral measure of X.

Remark: Under technical conditions, the stationary distribution of a stochastic recurrence equation is multivariate regularly varying. The section 2.2.2. in [5] describes different situations in which we have the regular variation of solutions of a SRE.

We assume now that the stationary solution Y is regularly varying.

$$\forall t > 0, \frac{\mathbb{P}(|Y|>tx, Y/|Y|\in .)}{\mathbb{P}(|Y|>x)} \xrightarrow[x\to\infty]{v} t^{-\alpha}\mathbb{P}(\Theta \in .).$$

There exists a MCMC method, described in section 5 of [6], to construct a Markov chain whose invariant distribution is the law of the spectral measure of a regularly varying random variable under some conditions. Justifying this method requires the next proposition (5.1 from [6]):

Proposition 5
Assume there exists a stationary solution $(X_t)_{t\in\mathbb{Z}}$ of the stochastic recurrence equation $X_t = A_t X_{t-1} + B_t, t \in \mathbb{Z}$, with $(A_t, B_t), t \in \mathbb{Z}$, an i.i.d. sequence of random $d \times d$ matrices A_t and random d-dimensional vectors B_t such that $\mathbb{E}[\|A_1\|^\gamma]$ is finite for some $\gamma > \alpha$, and $\mathbb{P}(\|B_1\| > x) = o(\mathbb{P}(\|X_0\| > x))$, as $x \to \infty$. If the process $(X_t)_{t\in\mathbb{Z}}$ is also regularly varying of index $\alpha > 0$, then Θ_0 satisfies $\mathbb{E}[\|A_1\Theta_0\|^\alpha] = 1$ and

$$\mathbb{P}(\Theta_0 \in .) = \mathbb{E}\left[\|A_1\Theta_0\|^\alpha; A_1\Theta_0/\|A_1\Theta_0\| \in .\right]. \tag{6}$$

Remark: Those conditions are not sufficient. One should look at the erratum of [6] to find the precise necessary conditions to apply the following algorithm.

Algorithm 3 MCMC for the spectral measure With the additional assumption that there exists $C > 0$ such that $\|A_1\| \leq C$ almost surely, we can construct the Markov kernel we want from a variant of the rejection algorithm. Let us say the current value of the chain is θ_i, we have the following instructions to simulate θ_{i+1}:

1. Sample A from the law of A_1.
2. Set $Y = A\theta_i/\|A\theta_i\|$.
3. With probability $\|A\theta_i\|^\alpha/C^\alpha$ set $\theta_{i+1} = Y$, otherwise reject Y and set $\theta_{i+1} = \theta_i$.

Remark: This method is not valid in the general case (see erratum of [6]). It also requires to know the regularly varying of index $\alpha > 0$ which is not possible for GARCH processes in the general case. In addition, we need the boundedness assumption.

We now want to modify this algorithm to add killing to the Markov process and therefore make the stationary distribution a quasi-stationary distribution. In fact, the distribution defined by (6) is the QSD of a process which evolves by $\theta_i = A\theta_{i-1}/\|A\theta_{i-1}\|$ with death probability $1 - \|A\theta_{i-1}\|^\alpha/C^\alpha$.

As it is described in [5], we will only consider ARCH(1) processes and IGARCH$(1,1)$ processes to work with a simple expression of α, in this case $\alpha = 1$. One can find Monte-Carlo methods, for estimating the index α in the GARCH(1,1) case, in [8].
We need now to choose a distribution for Z_t that satisfies the boundedness assumption and a process where we know the spectral measure. Our idea is to use the new Markov process with killing with the combine-split algorithm. Unfortunately, it has not produced any conclusive results so far, maybe due to a death rate d that is too high in the tested example (almost 80 percents dead at each next generation).

References

[1] Sylvie Méléard and Denis Villemonais, *Quasi-stationary distributuions and population processes*, Probability Surveys, 2012.

[2] Nicolas Champagnat and Denis Villemonais, *Exponential convergence to quasi-stationary distribution and Q-process*, Probab. Theory Relat. Fields, 2016.

[3] Jose Blanchet, Peter Glynn and Shuheng Zheng, *Theoretical analysis of a stochastic approximation approach for computing quasi-stationary distributions.*

[4] A. Griffin, P.A. Jenkins, G.O. Roberts, S.E.F Spencer, *Simulation From Quasi-Stationary Distributions On Reducible State Spaces*, Applied Probability Trust, 2017.

[5] Bojan Basrak, Johan Segers, *Regularly varying multivariate time series*, Stochastic Processes and their Applications, 2009.

[6] Bojan Basrak, Richard A. Davis, Thomas Mikosch, *Regular variation of GARCH processes*, Stochastic Processes and their Applications, 2002.

[7] J. Blanchet, P. Glynn and S. Zheng, *Analysis of a stochastic approximation algorithm for computing quasi-stationary distributions*, 2014.

[8] Fabrizio Laurini, Jonathan A. Tawn, *The extremal index for GARCH(1,1) processes*, 2012.

[9] H. Kesten, *Random difference equations and renewal theory for products of random matrices*, 1973.