A probabilistic representation of the quasispecies distribution

Raphaël Cerf and Joseba Dalmau
DMA, École Normale Supérieure

July 6, 2016

Abstract
We give a probabilistic representation of the stationary solutions of Eigen’s model, when the set of possible genotypes is finite and the mutation matrix is primitive. In the long chain regime, we perform a formal passage to the limit to obtain a probabilistic representation of the quasispecies distribution. We prove rigorously the validity of this representation with the help of previously known exact formulas.

1 Introduction
The first forms of life on Earth must have been short macromolecules. With this consideration in mind, Manfred Eigen developed a simple mathematical model for the evolution of the first populations on Earth [5]. The simple structure of these short macromolecules allowed for a description of the reactions governing the evolution of such a population, based on chemical kinetics. The outcome was a beautiful system of differential equations, which is useful to describe not only the first living populations, but many other populations submitted to selection and mutation. Eigen’s model is one of the simplest biological models exhibiting a phase transition phenomenon. Indeed, there exists a critical mutation rate separating two totally different regimes. For high mutation rates, the population at equilibrium is totally random, and for low mutation rates the population at equilibrium is distributed as a quasispecies, that is, a population in which the most adapted genotype is present in a positive concentration, along with a cloud of mutants. A beautiful account of the quasispecies theory can be found in [2].

Quasispecies distributions have attracted special attention in the last few decades, mostly for their relevance in understanding the evolution of viral
populations [3, 4]. These quasispecies distributions are nothing but the stationary solutions to Eigen’s system of differential equations, and of course, they depend both on the choice of the fitness landscape and the mutation scheme. In the general case, they can be characterised by an eigenvector problem [7], but explicit formulas are only available in certain asymptotic regimes, and after fixing the fitness landscape and the mutation scheme [1].

The purpose of our article is to give probabilistic representations of both the general solution and of the approximations.

2 The quasispecies equation

Eigen’s model is defined through a system of differential equations. We present here the equations satisfied by its stationary solutions.

**Notation.** The space of the possible genotypes is denoted by $E$. The fitness of a genotype $u \in E$ is denoted by $f(u)$. A genotype $u$ mutates into another genotype $v$ with probability $M(u, v)$.

**The system.** The stationary solutions of Eigen’s system satisfy:

$$\forall u \in E \quad x(u) \sum_{v \in E} x(v)f(v) = \sum_{v \in E} x(v)f(v)M(v, u). \quad (S)$$

For $u \in E$, the unknown variable $x(u)$ represents the concentration of individuals having genotype $u$. We are looking for solutions to the above system which satisfy the additional constraint:

$$\forall u \in E \quad x(u) \geq 0, \quad \sum_{u \in E} x(u) = 1. \quad (C)$$

3 Finite space and primitive matrix

Suppose that $(x(u))_{u \in E}$ is a solution to $(S)$ which satisfies $(C)$. If we set

$$\Phi = \sum_{v \in E} x(v)f(v),$$

then we observe that $\Phi$ is an eigenvalue of the matrix $(f(u)M(u, v))_{u,v \in E}$, and that $(x(u))_{u \in E}$ is an associated eigenvector, whose components are non-negative. We introduce next an hypothesis which ensures the existence and the uniqueness of the solution.

**Hypothesis ($\mathcal{H}$).** The space $E$ is finite. The function $f$ is positive. The matrix $M$ is irreducible and aperiodic.

Under the above hypothesis, the matrix $(f(u)M(u, v))_{u,v \in E}$ is primitive and all its coefficients are non-negative. We can therefore apply the famous
Perron–Frobenius theorem \[6\]. Let $\lambda$ be the Perron–Frobenius eigenvalue of the matrix $fM$. The corresponding eigenspace has dimension 1 and it contains an eigenvector associated to $\lambda$ whose components are non–negative. Moreover any eigenvector of $fM$ whose components are all non–negative is associated to the eigenvalue $\lambda$. These considerations yield the following result.

**Proposition 3.1** Suppose that $(H)$ holds. The system $(S)$ admits a unique solution satisfying the constraint $(C)$. This solution is the normalized eigenvector $(x(u))_{u \in E}$ of the matrix $fM$, associated to the Perron–Frobenius eigenvalue $\lambda$, which satisfies in addition

$$
\lambda = \sum_{v \in E} x(v)f(v).
$$

4 Probabilistic representation

Let $w$ be an arbitrary point of $E$. Let $(S_n)_{n \in \mathbb{N}}$ be the Markov chain on $E$ with transition matrix $M$ and starting point $w$. We denote by $\lambda$ the Perron–Frobenius eigenvalue of $fM$ and we define

$$
\tau = \inf \{ n \geq 1 : S_n = w \}.
$$

**Theorem 4.1** Suppose that $(H)$ holds. The unique solution to $(S)$ which satisfies the constraint $(C)$ is given by the formula:

$$
\forall u \in E \quad x(u) = \frac{E\left(\sum_{n=0}^{\tau-1} \left(1_{\{S_n=u\}} \lambda^{-n} \prod_{k=0}^{n-1} f(S_k)\right)\right)}{E\left(\sum_{n=0}^{\tau-1} \left(\lambda^{-n} \prod_{k=0}^{n-1} f(S_k)\right)\right)}.
$$

**Proof.** Let us set, for $u \in E$,

$$
y(u) = E\left(\sum_{n=0}^{\tau-1} \left(1_{\{S_n=u\}} \lambda^{-n} \prod_{k=0}^{n-1} f(S_k)\right)\right).
$$

Obviously, the vector $(y(u))_{u \in E}$ is non null and its components are non negative. Let us compute

$$
\sum_{v \in E} y(v)f(v)M(v, u) =
$$
\[
\sum_{v \in \mathcal{E}} \sum_{n \geq 0} E \left( 1_{\{\tau > n\}} \lambda^{-n} \left( \prod_{k=0}^{n-1} f(S_k) \right) 1_{\{S_n = v\}} f(v) M(v, u) \right)
\]

\[
= \sum_{v \in \mathcal{E}} \sum_{n \geq 0} E \left( 1_{\{\tau > n\}} \lambda^{-n} \left( \prod_{k=0}^{n} f(S_k) \right) 1_{\{S_n = v\}} 1_{\{S_{n+1} = u\}} \right)
\]

\[
= E \left( \sum_{n=0}^{\tau-1} 1_{\{S_{n+1} = u\}} \lambda^{-n} \left( \prod_{k=0}^{n} f(S_k) \right) \right)
\]

\[
= \lambda E \left( \sum_{n=1}^{\tau} 1_{\{S_n = u\}} \lambda^{-n} \left( \prod_{k=0}^{n-1} f(S_k) \right) \right).
\]

Suppose that \( u \neq w \). Then the term in the last sum vanishes for \( n = 0 \) or \( n = \tau \), and we recover the identity

\[
\sum_{v \in \mathcal{E}} y(v) f(v) M(v, u) = \lambda y(u).
\]

For \( u = w \), we obtain

\[
\sum_{v \in \mathcal{E}} y(v) f(v) M(v, w) = \lambda E \left( \lambda^{-\tau} \prod_{k=0}^{\tau-1} f(S_k) \right).
\]

The last expectation can be rewritten as

\[
E \left( \lambda^{-\tau} \prod_{k=0}^{\tau-1} f(S_k) \right) = \sum_{n \geq 1} E \left( 1_{\{\tau = n\}} \lambda^{-n} \prod_{k=0}^{n-1} f(S_k) \right)
\]

\[
= \sum_{n \geq 1, v_1, \ldots, v_{n-1} \neq w} \lambda^{-n} f(w) f(v_1) \cdots f(v_{n-1})
\]

\[
\times P(S_1 = v_1, \ldots, S_{n-1} = v_{n-1}, S_n = w)
\]

\[
= \sum_{n \geq 1, v_1, \ldots, v_{n-1} \neq w} \lambda^{-n} f(w) M(w, v_1) \cdots f(v_{n-1}) M(v_{n-1}, w).
\]

This last sum is equal to 1 by proposition A.1. Noticing that \( y(w) = 1 \), we conclude that

\[
\sum_{v \in \mathcal{E}} y(v) f(v) M(v, w) = \lambda y(w).
\]

Therefore the vector \( (y(u))_{u \in \mathcal{E}} \) is an eigenvector of \( fM \) associated to \( \lambda \).

We normalize it in order to satisfy the constraint \( (C) \) and we obtain the formula stated in the theorem. \( \square \)

Let us denote \( E_u \) the expectation for the Markov chain \( (S_n)_{n \in \mathbb{N}} \) starting from \( u \). By taking \( w = u \) in the formula stated in theorem 4.4, we obtain the following corollary.
Corollary 4.2 Suppose that $(\mathcal{H})$ holds. The unique solution to $(\mathcal{S})$ which satisfies the constraint $(\mathcal{C})$ is given by the formula:

$$\forall u \in E \quad x(u) = 1 / \left( E_u \left( \sum_{n=0}^{\tau_u-1} \left( \sum_{k=0}^{n-1} \lambda^{-n} f(S_k) \right) \right) \right),$$

where $\lambda$ is the Perron–Frobenius eigenvalue of $fM$ and

$$\tau_u = \inf \left\{ n \geq 1 : S_n = u \right\}.$$

This formula is a generalization of the classical formula for the invariant probability measure of a Markov chain. Indeed, in the particular case where $f$ is constant equal to 1, then $\lambda = 1$ as well, and the system $(\mathcal{S})$ reduces to

$$\forall u \in E \quad x(u) = \sum_{v \in E} x(v) M(v, u),$$

while the formula in corollary 4.2 becomes the well–known formula

$$\forall u \in E \quad x(u) = 1 / E_u(\tau_u).$$

This probabilistic representation of the Perron–Frobenius eigenvector can be applied to any primitive matrix of size $n \times n$. It is enough to define a fitness function $f$ by setting $f(i) = A(i, 1) + \cdots + A(i, n)$, and a mutation matrix $M$ by $M(i, j) = A(i, j)/f(i)$.

5 Class–dependent fitness landscapes

In the previous sections we have dealt with the system $(\mathcal{S})$ in its general form. However, we are often interested in measuring the impact of different fitness landscapes on the evolution of a population. To this end, it is customary to fix the space of genotypes and the mutation matrix. Both for practical and historical reasons, a common choice is to take the space of genotypes to be the $\ell$–dimensional hypercube $\{0, 1\}^\ell$, and to assume that mutations happen independently on each site of the chain, with probability $q$. In order to measure the impact of selection in this particular setting, Eigen considered the sharp peak fitness landscape: all genotypes but one, the master sequence, share the same fitness, while the master sequence has a higher fitness. A natural generalisation is to consider class–dependent fitness landscapes: the fitness of an individual depends on its genotype only through the number of point mutations to the master sequence. Precisely,
the Hamming distance on the hypercube measures the number of point mutations between two chains:

$$\forall u, v \in \{0, 1\}^\ell \quad d_H(u, v) = \text{card} \left\{ 1 \leq h \leq \ell : u(h) \neq v(h) \right\}.$$

We may thus group the different genotypes into Hamming classes, by lumping together all the genotypes at the same Hamming distance from the master sequence. It turns out that the mutation mechanism can be factorised through the Hamming classes, allowing us to write the system \((S)\) for the concentrations of the different Hamming classes:

$$\forall k \in \{1, \ldots, \ell\} \quad x(k) \sum_{h=0}^{\ell} x(h) f(h) = \sum_{h=0}^{\ell} x(h) f(h) M(h, k). \quad (S_H)$$

The variable \(x(k)\) represents the concentration of individuals in the Hamming class \(k\), while \(f(k)\) represents the fitness common to the Hamming class \(k\), and \(M(h, k)\) the probability for an individual to mutate from the class \(h\) to the class \(k\). Since the Hamming class 0 corresponds to the master sequence, i.e., the fittest genotype, we introduce the following assumption:

**Hypothesis \(\mathcal{H}'\).** The function \(f\) is positive and satisfies \(f(0) > f(k)\) for all \(k \geq 1\).

Applying corollary 4.2 to the system \((S_H)\), we obtain that the concentration of the master sequence is equal to

$$x(0) = \frac{1}{E_0 \left( \sum_{n=0}^{\tau_0-1} \left( \frac{\lambda^{-n}}{\prod_{k=0}^{n-1} f(S_k)} \right) \right)},$$

where \(\lambda\) is the Perron–Frobenius eigenvalue of \(fM\) and

$$\tau_0 = \inf \{ n \geq 1 : S_n = 0 \}.$$

### 6 Long chain regime

For the joy of the walking mathematician, nature has set things so that even the simplest living creatures have extremely long genomes, meaning that we might as well take it to be infinite. Therefore, we consider the system \((S_H)\) in the following asymptotic regime:

$$\ell \to \infty \quad q \to 0 \quad \ell q \to a.$$
The limiting mutation probabilities are given by:

\[ \lim_{\ell \to \infty, q \to 0} M(h, k) = \begin{cases} e^{-a} \frac{a^{k-h}}{(k-h)!} & \text{if } k \geq h, \\ 0 & \text{if } k < h. \end{cases} \]

We can thus write the limiting system of equations:

\[ x(k) \sum_{h=0}^{\infty} x(h) f(h) = \sum_{h=0}^{k} x(h) f(h) e^{-a} \frac{a^{k-h}}{(k-h)!}, \quad k \geq 0. \quad (S_{\infty}) \]

Heuristically, we expect that in the asymptotic regime introduced above, the Perron–Frobenius eigenvalue \( \lambda \) converges towards the maximum between \( f(0)e^{-a} \) and 1, while the law of the Markov chain \( (S_n)_{n \in \mathbb{N}} \) converges to the law of the random walk \( (S_n^\infty)_{n \in \mathbb{N}} \) defined as follows. Let \( (X_n)_{n \geq 1} \) be a sequence of i.i.d. random variables with distribution the Poisson law \( \mathcal{P}(a) \) of parameter \( a \). We define

\[ \forall n \geq 1 \quad S_n^\infty = X_1 + \cdots + X_n. \]

In the limit, the random time \( \tau_0 \) is equal to 1 with probability \( e^{-a} \) and to \( \infty \) with probability \( 1 - e^{-a} \). If we perform a formal passage to the limit in the above formula, we expect that \( x(0) \) converges towards

\[ \frac{1}{1 + (1 - e^{-a}) \sum_{n \geq 1} \left( \max(f(0)e^{-a}, 1) \right)^{-n} E\left( \prod_{k=0}^{n-1} f(S_k^\infty) \bigg| X_1^\infty > 0 \right)}}. \]

It is a delicate matter to prove rigorously this convergence, already the convergence of the Perron–Frobenius eigenvalue is far from obvious. Therefore we take a different road. Having in mind the above expression, we start from the explicit formula obtained in \( \prod \) and we try to rewrite it as a functional of the random walk \( (S_n^\infty)_{n \in \mathbb{N}} \).

**Definition 6.1** We say that \( (x_k)_{k \geq 0} \) is a quasispecies associated to \( f \) if it is a non-negative solution of \( (S_{\infty}) \) such that \( x(0) > 0 \) and \( \sum_{k \geq 0} x(k) = 1 \).

We denote by \( E^{1 >} \) the expectation for the random walk \( (S_n^\infty)_{n \in \mathbb{N}} \) conditioned on the event that the first step \( X_1 \) is positive, i.e., for any function \( \phi \) of \( S_1^\infty, \ldots, S_n^\infty \), we have

\[ E^{1 >} \left( \phi(S_1^\infty, \ldots, S_n^\infty) \right) = E \left( \phi(S_1^\infty, \ldots, S_n^\infty) \bigg| X_1 > 0 \right). \]
Theorem 6.2 A quasispecies associated to \( f \) exists if and only if
\[
\sum_{n \geq 1} E^1 > \left( \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}} \right) < +\infty.
\]

If this condition holds, then we have
\[
x^\infty(0) = \frac{1}{1 + (1 - e^{-a}) \sum_{n \geq 1} E^1 > \left( \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}} \right)}
\]
and for \( k \geq 1, \)
\[
x^\infty(k) = x^\infty(0) (1 - e^{-a}) \sum_{n \geq 1} E^1 > \left( 1 \{ S_n^\infty = k \} \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}} \right).
\]

Thanks to this theorem, we have now a probabilistic criterion for the existence of a quasispecies on a class-dependent fitness landscape. Moreover we have a probabilistic formula expressing the concentration of each Hamming class as a functional of a Poisson random walk.

7 Sharp peak landscape

Before embarking into the proof, let us examine the formula in the specific case of the sharp peak landscape, i.e., the fitness function \( f \) is equal to \( \sigma > 1 \) at 0 and to 1 elsewhere. We have then
\[
\sum_{k \geq 1} \sum_{n \geq 1} E^1 > \left( 1 \{ S_n^\infty = k \} \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}} \right) = \sum_{n \geq 1} \frac{1}{\sigma^{n-1} e^{-na}} = \frac{\sigma}{\sigma e^{-a} - 1}
\]
whence
\[
x^\infty(0) = \frac{\sigma e^{-a} - 1}{\sigma - 1}.
\]

Let now \( k \geq 1 \). We have
\[
\sum_{n \geq 1} E^1 > \left( 1 \{ S_n^\infty = k \} \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}} \right) = \sum_{n \geq 1} \left( \frac{1}{\sigma^{n-1} e^{-na}} P^1 > (S_n^\infty = k) \right).
\]

We compute separately \( P^1 > (S_n^\infty = k) \) as follows:
\[
P^1 > (S_n^\infty = k) = P(S_n^\infty = k \mid X_1 > 0) = \frac{1}{1 - e^{-a}} P(S_n^\infty = k, X_1 > 0)
\]
\[
= \frac{1}{1 - e^{-a}} \left( P(S_n^\infty = k) - P(S_n^\infty = k, X_1 = 0) \right)
= \frac{1}{1 - e^{-a}} \left( P(S_n^\infty = k) - e^{-a} P(S_{n-1}^\infty = k) \right).
\]

Now we have
\[
\sum_{n \geq 1} \frac{P^{1 >}(S_n^\infty = k)}{\sigma^{n-1}e^{-na}} = \frac{\sigma - 1}{1 - e^{-a}} \sum_{n \geq 1} \frac{P(S_n^\infty = k)}{\sigma^n e^{-na}} = \frac{\sigma - 1}{1 - e^{-a}} \sum_{n \geq 1} \frac{(an)^k}{k! \sigma^n}
\]
and we recover the quasispecies formula
\[
x^\infty(k) = (\sigma e^{-a} - 1) \frac{a^k}{k!} \sum_{n \geq 1} \frac{n^k}{\sigma^n}.
\]

8 Proof of theorem 6.2

We shall prove first an intermediate result, which is interesting on its own. We denote by \(E^>\) the expectation for the random walk \((S_n^\infty)_{n \in \mathbb{N}}\) conditioned on the event that all the steps are positive, i.e., for any function \(\phi\) of \(S_1^\infty, \ldots, S_n^\infty\), we have
\[
E^>\left( \phi(S_1^\infty, \ldots, S_n^\infty) \right) = E\left( \phi(S_1^\infty, \ldots, S_n^\infty) \right| X_1 > 0, \ldots, X_n > 0).
\]

**Proposition 8.1** A quasispecies associated to \(f\) exists if and only if
\[
\sum_{n \geq 0} (e^a - 1)^n E^> \left( \frac{1}{f(S_n^\infty)} \prod_{t=1}^{n} \frac{f(S_t^\infty)}{f(0) - f(S_t^\infty)} \right) < +\infty.
\]
If this condition holds, and if we let
\[
\forall k \geq 1 \quad \tau_k = \inf \left\{ n \geq 1 : S_n^\infty = k \right\},
\]
then we have
\[
\forall k \geq 1 \quad x^\infty(k) = \frac{1}{f(k)} \sum_{n \geq 0} (e^a - 1)^n E^> \left( \frac{1}{f(S_n^\infty)} \prod_{t=1}^{n} \frac{f(S_t^\infty)}{f(0) - f(S_t^\infty)} \right).
\]
Proof. Our starting point is the formula of theorem 2.3 of [1]: for \( k \geq 1 \), we have

\[
y_k = \frac{a^k}{k!} \frac{f(0)}{f(k)} \sum_{i_0 < \cdots < i_h = k} \left( \prod_{t=1}^{h} \frac{f(i_t)}{f(0) - f(i_t)} \right).
\]

Let us denote by \( P^> \) the probability associated to \( E^> \). For \( 1 \leq h \leq k \) and \( 0 = i_0 < \cdots < i_h = k \), we have

\[
a^k \frac{(i_1 - i_0)! \cdots (i_h - i_{h-1})!}{(i_1 - i_0)! \cdots (i_h - i_{h-1})!}
= (e^a - 1)^h P^> (X_1 = i_1 - i_0, \ldots, X_h = i_h - i_{h-1})
= (e^a - 1)^h P^> (S^\infty_1 = i_1, \ldots, S^\infty_h = i_h).
\]

Using this identity, setting \( \tau_k = \inf \{ n \geq 1 : S^\infty_n = k \} \), we can rewrite \( y_k \) as follows:

\[
y_k = \frac{f(0)}{f(k)} \sum_{i_0 < \cdots < i_h = k} (e^a - 1)^h P^> (S^\infty_1 = i_1, \ldots, S^\infty_h = i_h) \prod_{t=1}^{h} \frac{f(i_t)}{f(0) - f(i_t)}
= \frac{f(0)}{f(k)} E^> \left( 1_{\{\tau_k < \infty\}} (e^a - 1)^{\tau_k} \prod_{t=1}^{\tau_k} \frac{f(S^\infty_t)}{f(0) - f(S^\infty_t)} \right).
\]

We compute next the sum of the \( y_k \)'s:

\[
\sum_{k \geq 1} y_k = E^> \left( \sum_{k \geq 1} \frac{f(0)}{f(k)} 1_{\{\tau_k < \infty\}} (e^a - 1)^{\tau_k} \prod_{t=1}^{\tau_k} \frac{f(S^\infty_t)}{f(0) - f(S^\infty_t)} \right)
= E^> \left( \sum_{n \geq 1} \frac{f(0)}{f(S^\infty_n)} (e^a - 1)^n \prod_{t=1}^{n} \frac{f(S^\infty_t)}{f(0) - f(S^\infty_t)} \right).
\]

This last equality holds because the set of the indices \( k \geq 1 \) such that \( \tau_k < \infty \) is exactly the increasing sequence \((S^\infty_n)_{n \geq 1}\). The statement of the proposition follows from this formula and lemma 2.2 of [1]. □

We shall now prove theorem 6.2. We introduce the jumping times of the random walk \((S^\infty_n)_{n \in \mathbb{N}}\). We set \( T_0 = 0 \) and for \( k \geq 1 \), we define iteratively

\[
T_k = \inf \{ n > T_{k-1} : S^\infty_n > S^\infty_{T_{k-1}} \}.
\]

We define also, for \( k \geq 1 \),

\[
N_k = \inf \{ n \geq 1 : S^\infty_n = k \}.
\]
Let us denote by $P^>$ (respectively $P^{1>}$) the probability associated to $E^>$ (respectively $E^{1>}$). With these definitions, for any $k \geq 1$, the joint law of $(S_{n\in N}^\infty, \tau_k)$ under $P^>$ is equal to the joint law of $(S_{n\in N}^{\infty}, N_k)$ under $P^{1>}$. We can thus rewrite the formula obtained in proposition 8.1 with the help of the probability $P^{1>}$. For $k \geq 1$, we have

$$E^>(1_{\tau_k<\infty}(e^a - 1)\tau_k \prod_{t=1}^{\tau_k} f(0) - f(S_t^\infty))$$

$$= E^{1>}(1_{\{N_k<\infty\}}(e^a - 1)^{N_k} \prod_{t=1}^{N_k} f(S_t^\infty) f(0) - f(S_t^\infty))$$

$$= \sum_{n \geq 1} E^{1>}(1_{\{N_k=n\}}(e^a - 1)^n \prod_{t=1}^{n} (f(S_t^\infty) f(0) - f(S_t^\infty)))$$

We remark that, under $P^{1>}$, the time intervals $T_{t+1} - T_t$, $t \geq 1$, are independent of $N_k$, and they are i.i.d. with distribution the geometric law of parameter $1 - e^{-a}$. Therefore we can rewrite the previous sums as

$$\sum_{n \geq 1} E^{1>}(1_{\{N_k=n\}}(e^a - 1)^n \prod_{t=1}^{n} \left( f(S_t^\infty) f(0) - e^a(T_{t+1} - T_t) \right))$$

$$= \sum_{n \geq 1} E^{1>}(1_{\{N_k=n\}} \prod_{t=1}^{n} \left( f(S_t^\infty) f(0) - e^a(T_{t+1} - T_t) \right))$$

$$= E^{1>}(1_{\{N_k<\infty\}} \prod_{t=1}^{T_{\tau_k+1}} f(S_t^\infty) f(0)e^{-a})$$

$$= E^{1>}(1_{\{N_k<\infty\}} \prod_{t=1}^{T_{\tau_k+1}} f(S_t^\infty) f(0)e^{-a})$$

Let us set $\gamma = f(k)/(f(0)e^{-a})$. We compute

$$E(\gamma^T \mid S_0^\infty = k) = \sum_{n \geq 0} \gamma^{n+1} e^{-an} (1 - e^{-a}) = \gamma \frac{1 - e^{-a}}{1 - \gamma e^{-a}}.$$
So we have obtained that

\[ y_k = \frac{f(0)(e^a - 1)}{f(0) - f(k)} E^{1>}(1_{\{N_k < \infty\}} \prod_{t=1}^{T_{N_k}^{-1}} \frac{f(S_t^\infty)}{f(0)e^{-a}}) \, . \]

On the other hand, we have

\[ \sum_{n \geq 0} E^{1>}(1_{\{S_n^\infty = k\}} \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}}) = E^{1>}(1_{\{N_k < \infty\}} \sum_{n=T_{N_k}}^{T_{N_k+1}^{-1}} \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}}) \]

\[ = E^{1>}(1_{\{N_k < \infty\}} \prod_{t=0}^{T_{N_k}^{-1}} \frac{f(S_t^\infty)}{f(0)e^{-a}} \sum_{n=T_{N_k}}^{T_{N_k+1}^{-1}} \frac{f(k)}{f(0)e^{-a}}^{n-T_{N_k}}) \]

\[ = E^{1>}(1_{\{N_k < \infty\}} \prod_{t=0}^{T_{N_k}^{-1}} \frac{f(S_t^\infty)}{f(0)e^{-a}} E^{1>}(T_{1}^{-1}) \sum_{n=0}^{T_{1}^{-1}} \frac{f(k)}{f(0)e^{-a}}^{n} | S_0^\infty = k) \, . \]

Let us set \( \gamma = f(k)/(f(0)e^{-a}) \). We compute

\[ E\left(\sum_{n=0}^{T_{1}^{-1}} \gamma^n \right) = E\left(\frac{1 - \gamma T_{1}}{1 - \gamma} \right) \]

\[ = \frac{1}{1 - \gamma} \left(1 - \sum_{n=0}^{\infty} \gamma^{n+1}e^{-an}(1 - e^{-a})\right) \]

\[ = \frac{1}{1 - \gamma} \left(1 - \gamma \frac{1 - e^{-a}}{1 - \gamma e^{-a}}\right) = \frac{1}{1 - \gamma e^{-a}} \, . \]

Therefore, we have

\[ \sum_{n \geq 0} E^{1>}(1_{\{S_n^\infty = k\}} \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}}) = \]

\[ \frac{f(0)}{e^{-a}(f(0) - f(k))} E^{1>}(1_{\{N_k < \infty\}} \prod_{t=1}^{T_{N_k}^{-1}} \frac{f(S_t^\infty)}{f(0)e^{-a}}) \, . \]

From these computations, we conclude that, for \( k \geq 1 \), we have

\[ y_k = (1 - e^{-a}) \sum_{n \geq 0} E^{1>}(1_{\{S_n^\infty = k\}} \prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}}) \, . \]

Summing over \( k \geq 1 \) yields

\[ \sum_{k \geq 1} y_k = (1 - e^{-a}) \sum_{n \geq 1} E^{1>}(\prod_{t=0}^{n-1} \frac{f(S_t^\infty)}{f(0)e^{-a}}) \, . \]

12
Taking into account that \( y_0 = 1 \), the statements of theorem 6.2 follow from the previous formulas together with lemma 2.2 of [1].

### Appendix

**Proposition A.1** Let \( A \) be a non-negative primitive matrix indexed by a finite set \( E \). Its Perron–Frobenius eigenvalue \( \lambda \) satisfies the following identity: for any \( w \in E \),

\[
1 = \frac{1}{\lambda} A(w, w) + \frac{1}{\lambda^2} \sum_{v_1 \neq w} A(w, v_1)A(v_1, w) + \cdots \\
+ \frac{1}{\lambda^n} \sum_{v_1, \ldots, v_{n-1} \neq w} A(w, v_1)A(v_1, v_2) \cdots A(v_{n-1}, w) + \cdots
\]

**Proof.** Let \( (x(u))_{u \in E} \) be a non-negative eigenvector associated to the Perron–Frobenius eigenvalue \( \lambda \) of \( A \):

\[
\forall u \in E \quad \sum_{v \in E} x(v)A(v, u) = \lambda x(u) .
\]

Since \( A \) is primitive, all the components of \( x \) are positive. Let \( w \in E \) be fixed. We have thus

\[
1 = \frac{1}{\lambda x(w)} \sum_{v \in E} x(v)A(v, w) = \frac{1}{\lambda^2} A(w, w) + \sum_{v \neq w} \frac{x(v)}{\lambda x(w)}A(v, w) .
\]

We replace \( x(v) \) in the last sum and we get

\[
1 = \frac{1}{\lambda} A(w, w) + \sum_{v \neq w} \frac{1}{\lambda^2} A(w, v)A(v, w) + \sum_{v, v' \neq w} \frac{x(v')}{\lambda x(w)}A(v', v)A(v, w) .
\]

Iterating this procedure, we obtain, for \( n \geq 1 \),

\[
1 = \sum_{k=0}^{n-1} \frac{1}{\lambda^{k+1}} \sum_{v_1, \ldots, v_k \neq w} A(w, v_1)A(v_1, v_2) \cdots A(v_k, w) + \\
+ \frac{1}{\lambda^n} \sum_{v_1, \ldots, v_{n-1} \neq w} \frac{x(v_1)}{x(w)}A(v_1, v_2) \cdots A(v_{n-1}, w) .
\]
Let $B$ be the matrix obtained from $A$ by filling with zeroes the line and the column associated to $w$. The last term of the previous identity can be rewritten as

$$\frac{1}{\lambda^n} \sum_{u,v \neq w} \frac{x(u)}{x(w)} B^{n-2}(u,v)A(v,w).$$

Yet it follows from part (e) of theorem 1.1 of [6] that the spectral radius of $B$ is strictly less than $\lambda$, whence

$$\forall u,v \in E \lim_{n \to \infty} \frac{1}{\lambda^n} B(u,v)^{n-2} = 0.$$

Thus the previous sum vanishes as $n$ goes to $\infty$. Passing to the limit, we obtain the desired identity.

□

References

[1] Raphaël Cerf and Joseba Dalmau. Quasispecies on class-dependent fitness landscapes. Bulletin of Mathematical Biology, pages 1–21, 2016.

[2] E. Domingo and P. Schuster. Quasispecies: From Theory to Experimental Systems. Current Topics in Microbiology and Immunology. Springer International Publishing, 2016.

[3] Esteban Domingo. Quasispecies theory in virology. Journal of Virology, 76(1):463–465, 2002.

[4] Esteban Domingo, Christof Biebricher, Manfred Eigen, and John J. Holland. Quasispecies and RNA virus evolution: principles and consequences. Landes Bioscience, Austin, Tex., 2001.

[5] Manfred Eigen. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften, 58(10):465–523, 1971.

[6] E. Seneta. Nonnegative matrices and Markov chains. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.

[7] Colin J. Thompson and John L. McBride. On Eigen’s theory of the self-organization of matter and the evolution of biological macromolecules. Math. Biosci., 21:127–142, 1974.