A distribution weighting a set of laws whose initial states are grouped into classes

Servet Martinez

November 20, 2013

Abstract

Let $I$ be a finite alphabet and $S \subset I$ be a nonempty strict subset. The sequences in $I^\mathbb{Z}$ are organized into connected regions which always start with a symbol in $S$. The regions are labelled by types $C(s)$, thus a region starting at $s' \in C(s)$ has the same type as one starting at $s$. Let $(P_s : s \in S)$ be a family of distributions on $I^\mathbb{N}$ where each $P_s$ charges sequences starting with the symbol $s$. We can define a natural distribution $P$ on $I^\mathbb{N}$, that counts the number of visits to the states from $P_s$, properly weighted. A dynamics of interest is such that at the first occurrence of $s' \in S \setminus C(s)$ the law regenerates with distribution $P_{s'}$. In this case we are able to find simple conditions for $P$ to be stationary. In addition, we study the following more complex model: once a symbol $s' \in S \setminus C(s)$ has been encountered, there is a decision to be made, either a new region of type $C(s')$ governed by $P_{s'}$ starts or the region continues to be a $C(s)$ region. This decision is modeled as random and depends on $s'$. In this setting a similar distribution to $P$ can be constructed and the conditions for stationarity are supplied. These models are inspired by genomic sequences where $I$ is the set of codons, the classes $(C(s) : s \in S)$ group codons defining similar genomic classes, e.g. in bacteria there are two classes corresponding to the start and stop codons, and the random decision to continue a region or to begin a new region of a different class reflects the well-known fact that not every appearance of a start codon marks the beginning of a new coding region.

Keywords: Markov chains; Kac’s Measure; Genomics; Regeneration; Renewal Theorem.

AMS Subject Classification: 60J10, 60J20, 92D10, 92D20.

1 Introduction

Here we give an abstract description of the linear organization of sequences into different types of regions whose beginnings are marked by a distinguished number of symbols. The regions are organized in a sequential way, each one
starts at some prescribed set of symbols and ends at some other fixed set which is also the initial symbols of a region of a different type. In bacterial genomes there are two types of regions: genic and intergenic. Start codons mark the site where translation into a polypeptide sequence begins and stop codons define where the translation ends. So, stop codons define the beginnings of intergenic regions. In our model, we assume that there could be an arbitrary number of types.

Let $I$ be an alphabet of symbols. The infinite sequences of symbols $I^\mathbb{N}$ are assumed to be organized into connected regions labelled by different types. Let $S$ be the subset of symbols marking the beginning of a region and assume it is partitioned into equivalence classes ($C(s) : s \in S$), each class defining a different type of region. In genomics the alphabet $I$ is the set of 64 codons which are triplets of the bases \{A,C,G,T\}. The set $S$ is constituted by the codons \{ATG, GTG, TTG, TAA, TAG, TGG\}, the first three are the starting codons for genic regions and the other three are the stopping codons marking their ends, so \{ATG, GTG, TTG\} and \{TAA, TAG, TGG\} are the two classes.

In Section 6 we supply the main results of our work, but many of the concepts and intermediate results are presented in Sections 2, 3 and 4. In fact, the proofs of the main results employ similar computations to those used in the simpler models introduced in the initial sections.

Our work has as input a class of distributions $(P_s : s \in S)$ on $I^\mathbb{N}$. The law $P_s$ governs a region starting with $s$ and it is said to be of type $C(s)$. In Theorem 2.2 of Section 2 we show that there is a natural distribution $\mathbb{P}$ on $I^\mathbb{N}$ that allows the distributions $(P_s : s \in S)$ to be mixed. This probability measure depends on a vector of positive weights $\pi = (\pi_s : s \in S)$ and it counts the number of visits to the states as in Kac’s construction of the stationary vector in Markov chains (see Chapter I in [1] or Chapter 10 in [9]).

In Section 3 we assume the laws have a regenerative structure. If we start at $s \in S$, the sequence of letters evolves with the distribution $P_s$ until $T^1$ which is the time (or site) when a state $s^1 \in S \setminus C(s)$ is first reached. We assume the law regenerates at $T^1$, that is, at this time the sequence restarts its evolution with law $P_{s^1}$ until time $T^2$ when it first reaches $s^2 \in S \setminus C(s^1)$, and so on. The study of stationarity of $\mathbb{P}$ is made through the chain of states \{s^1, s^2, \ldots\} at times \{T^1, T^2, \ldots\}. In one of our main results, Theorem 3.3 we show that $\mathbb{P}$ is stationary in time if and only if the vector of weights $\pi$ is invariant for this chain.

We use this result to prove in Theorem 4.1 of Section 4 that, when $\pi$ is invariant and \{T^1, T^2, \ldots\} is aperiodic, the probability measure $\mathbb{P}$ is the asymptotic measure of any starting distribution that is a convex combination of $P_s$.

We note that stationarity is not a property totally foreign to genomes, in fact we show in Section 5 that the well-established Chargaff’s second parity rule (CSPR) implies stationarity and when this law is only assumed to be valid for \(k\)-mers then the stationarity holds for cylinders of length $k - 1$. In genomics,
CSPR has been proven to hold in the alphabet of nucleotides for \( k \)-mers of length \( k \approx 10 \). Then stationarity in the alphabet of codons is for length \( k \approx 3 \). CSPR was first observed experimentally in *Bacillus subtilis* \([16]\) and confirmed in sufficiently long sequences for small polymer chains in \([12]\). More recent empirical studies assessing its validity can be found in \([10], [6], [18]\).

In Section 6 we supply a richer model and give the main results of this work. Here, a choice must be made at each site where a region of type \( C \) encounters a symbol \( s' \notin C \): either it starts a new region governed by \( P_{s'} \) or it continues the former region of type \( C \). This decision is modeled by a sequence of independent random variables in the unit interval and the random choice also depends on \( s' \). The conditions for stationarity of this process are given in Theorem 6.5. In Theorem 6.6 it is stated that the law of the process can be also seen as an asymptotic law when starting from an initial weighted distribution.

Some of the most relevant works in the statistical analysis of DNA sequences have been devoted to describing the statistical differences between regions of different types. Thus, in \([19], [20]\) it is discovered that intergenic sequences have long-range correlations while short-range correlations prevail in genic sequences. An important tool constructed in \([19]\) was a map from the nucleotide sequences onto a walk. Then, correlations and other statistical quantities could be computed in walks and translated to DNA sequences. These methods were used in \([3], [4]\) to study stationarity, where a detailed statistical discussion about stationarity or non-stationarity of genic and intergenic regions can be found, together with an examination of power-type decreasing correlation functions. We wish to emphasize that in our model the laws for the regions \( \{P_s : s \in S\} \) may have long- or short-range correlations, or neither. They do not need to satisfy any Markovian condition, there is no need of hidden Markov chains or other kinds of models used in annotation as in \([11, 13, 21]\) or in references therein. Also, these laws do not need to exhibit any kind of stationarity. We take them as an input to study how they can combine and organize together into a unique law that under some conditions turns out to be stationary.

We are aware that the models we introduce and study are far from having the necessary degree of complexity to make it realistic for describing nucleotide or codon DNA sequences in bacterial genomes, but they provide some insights for their analysis. Thus, even if the statistical laws of nucleotide or codon sequences in genomes are claimed not to be stationary, our results imply that non-stationarity will not simply arise due to the existence of two types of regions, genic and intergenic, but from other phenomena that would contradict our hypotheses. This could be the case for the regenerative property which is one of the main ingredients for studying stationarity. One might be tempted to think this condition is too strong, but this is not so clear because a direct consequence of it, that the sequence of symbols marking the beginnings of regions is an homogeneous Markov chain, was shown to hold in annotated bacterial genomes in recent joint work with A. Hart \([15]\).

We point out that there is no intersection, not in any obvious way at least,
between the probabilistic study carried out in this work and the probabilistic studies devoted to genome evolution. Finally, there is a large bibliography on the statistics of codon and nucleotide sequences of bacterial DNA. Here, we have only cited papers that have a direct relationship to the present study. For a more complete view of this body of work, the reader is directed to the references contained in those that we have cited.

2 A law based on visits

Our goal is to supply a global law that mixes, in some natural way, distinct probability distributions starting at symbols belonging to a defined subset. We will do it similarly to the Kac’s construction of invariant probability measures for Markov chains. In genomics this problem corresponds on how the laws of the genic and intergenic regions can be mixed to obtain a global law for the genome.

From now on I denotes a finite alphabet, S is a nonempty subset of I and its elements are called initial symbols of the alphabet. We suppose that S is partitioned into equivalence classes which define regions of the same type. We denote by C(s) the class containing s ∈ S.

Let us introduce some notation and basic concepts. Every countable set L is endowed with the discrete σ-field $\mathcal{S}(L) = \{K : K \subseteq L\}$. We set $\mathbb{N} = \{0, 1, 2, \ldots\}$ and $\mathbb{N}^* = \{1, 2, \ldots\}$.

Define $X_n : I^N \to I, x \to x_n$ to be the n-th coordinate function, so $X_n(x) = x_n$ for $x \in I^N$. For each $n \in \mathbb{N}$,
$$\mathcal{B}_n^X = \sigma(X_0, \ldots, X_n)$$
denotes the σ-field generated by the coordinates $X_0, \ldots, X_n$ and
$$\mathcal{B}_\infty^X = \sigma(X_n : n \in \mathbb{N})$$
denotes the σ-field generated by all the coordinates. The product set $I^N$ is endowed with the σ-field $\mathcal{B}_\infty^X$. For $q \in \mathbb{N}$, the shift map in $q$-steps of time is
$$\Theta_q : I^N \to I^N, (\Theta_q x)_n = x_{n+q} \forall n \in \mathbb{N}.$$ (1)
The random variables are $\mathcal{B}_\infty^X$-measurable functions $W : I^N \to \mathbb{R} \cup \{-\infty, \infty\}$, so $W(x)$ is the value of this variable at $x \in I^N$.

The set $B \circ \Theta_N^{-1}$ has characteristic function $1_B \circ \Theta_N$ because $x \in B \circ \Theta_N^{-1}$ if and only if $\Theta_N(x) \in B$.

If $P$ is a probability measure on $(I^N, \mathcal{B}_\infty^X)$ then the process $X = (X_n : n \in \mathbb{N})$ is said to have distribution $P$. When we want to emphasize the dependence on $P$ we say under (law or distribution) $P$. 

4
Let $I^+ = \bigcup_{n \in \mathbb{N}} I^n$ be the set of non-empty finite words, so $\mathcal{S} \times I^+ = \bigcup_{n \in \mathbb{N}} (\mathcal{S} \times I^n)$ is the set of words with length at least two starting with some symbol in $\mathcal{S}$.

Below we use the usual convention $\inf \emptyset = \infty$.

As said $I$ is endowed with the $\sigma$-field $\mathcal{S}(I)$, and this last class of subsets is endowed with the $\sigma$-field $\mathcal{S}(\mathcal{S}(I))$. Let $\mathcal{J} : I \to \mathcal{S}(I)$, $i \to \mathcal{J}(i)$, be a map. Then, the function $I^N \to \mathcal{S}(I), x \to \mathcal{J}(x_0)$ is $\mathcal{B}_0^X$-measurable. So, $\mathcal{J}(x_0)$ is a random set. Let $T_J$ be the random time to hit $\mathcal{J}$ in the future,

$$T_J = \inf\{n > 0 : X_n \in \mathcal{J}(X_0)\}, \quad \text{so } T_J(x) = \inf\{n > 0 : x_n \in \mathcal{J}(x_0)\}.$$

It defines the sequence of successive returns to $\mathcal{J}$,

$$T^n_J := T_J \text{ and } \forall n \geq 1 : \quad T^{n+1}_J = T^n_J + T_J \circ \Theta_{T^n_J}. \quad (2)$$

Here $T^n_J = \infty$ implies $T^{n'}_J = \infty$ for $n' \geq n$. Sometimes, the dependence on $X_0$ will be written explicitly so we put indistinctly $T_J(X_0)$ or $T_J$.

Let $(\mathbf{P}_s : s \in \mathcal{S})$ be a family of probability distribution on $I^N$. Under $\mathbf{P}_s$ the process $X = (X_n : n \geq 0)$ starts from $s$, so $\mathbf{P}_s(X_0 = s) = 1$. We denote by $E_s$ the expectation defined by $\mathbf{P}_s$. We assume the set $\mathcal{S} \setminus C(s)$ is attained in finite time $\mathbf{P}_s$-a.s.. Hence, when $X_0 \in \mathcal{S}$ we can define the random time

$$T := T_{\mathcal{S}\setminus C(X_0)} = \inf\{n > 0 : X_n \in \mathcal{S} \setminus C(X_0)\},$$

which is $\mathbf{P}_s$-a.s. finite,

$$\forall s \in \mathcal{S} : \quad \mathbf{P}_s(T < \infty) = 1. \quad (3)$$

The sequence of successive returns is,

$$T^1 = T \text{ and } T^{n+1} = T^n + T \circ \Theta_{T^n} \text{ for } n \geq 1.$$

The time $T^n$ is called the $n$-th hitting time of a different class and $(T^n : n \in \mathbb{N}^*)$ is called the sequence of hitting times of different classes. Note that it is not guaranteed that $T^n$ is finite for $n > 1$. By definition we have

$$T^{n+1} < \infty \Rightarrow C(X_{T^{n+1}}) \neq C(X_{T^n}).$$

A family $(\mathbf{P}_s : s \in \mathcal{S})$ does not determine a common law. In the sequel we define a probability measure $\mathbb{P}$ on $I^N$. To avoid trivial situations we assume $I \setminus \mathcal{S} \neq \emptyset$ (as is the case in genomics) because in the contrary it will be sufficient to weight the laws $(\mathbf{P}_s : s \in \mathcal{S})$ in a simple way. The probability that at coordinate 0 the process takes the value $i \in I$, will be obtained by weighting the number of visits done to the state $i$ by the laws $(\mathbf{P}_s : s \in \mathcal{S})$ previous to hit a region of a different type. These visits will be weighted with a strictly positive vector $\pi = (\pi_s : s \in \mathcal{S}), \pi_s$ being the weight given to $\mathbf{P}_s$. Even if the distribution we define depends on $\pi$ we shall not explicit it to avoid overburden notation.
Definition 2.1. For the family $(P_s : s \in S)$ and a strictly positive vector $\pi = (\pi_s : s \in S)$ we define $P$ on $\mathbb{N}$ by:

$$\forall B \in B_\infty \mathcal{X} : \quad P(B) = \sum_{s \in S} \pi_s \left( \sum_{n \geq 0} E_s(1_{T>n} \mathbf{1} \circ \Theta_n) \right)$$

$$= \sum_{s \in S} \pi_s \left( \sum_{n \geq 0} P_s(T > n, B \circ \Theta_n^{-1}) \right). \quad (4)$$

□

(As usual, in the last expression the event $A_1 \cap A_2$ is written $(A_1, A_2)$). Obviously $P$ is a measure on $\mathbb{N}$. Note that for all $(i_l : l = 0, .., m) \in I^{m+1}$ we have

$$P(X_l = i_l, l = 0, .., m) = \sum_{s \in S} \pi_s \sum_{n \geq 0} P_s(T > n, X_{l+n} = i_l, l = 0, .., m)). \quad (5)$$

It is useful to develop (5) in two different cases. For $(i_l : l = 0, .., m) \in S \times I^m$ we have

$$P(X_l = i_l, l = 0, .., m) = \sum_{s \in C(i_0)} \pi_s \sum_{n \geq 0} P_s(T > n+1, X_{l+n+1} = i_l, l = 0, .., m)) \quad (6)$$

For $(i_l : l = 0, .., m) \in (I \setminus S) \times I^m$ we get,

$$P(X_l = i_l, l = 0, .., m) = \sum_{s \in S} \sum_{n \geq 0} P_s(T > n+1, X_{l+n+1} = i_l, l = 0, .., m)). \quad (7)$$

Theorem 2.2. There exists some strictly positive vector $\pi = (\pi_s : s \in S)$ such that the measure $P$ defined by (2) is a probability measure if and only if it is satisfied

$$\forall s \in S : \quad E_s(T) < \infty. \quad (8)$$

In this case, the condition on $\pi$:

$$\sum_{s \in S} \pi_s E_s(T) = 1, \quad (9)$$

is necessary and sufficient in order that $P$ is a probability measure on $\mathbb{N}$.

Proof. We will show that condition (9) is equivalent to $P(X_0 \in S) = 1$. Let $s_0 \in S$. From (3) we have

$$P(X_0 = s_0) = \sum_{s \in C(s_0)} \pi_s \left( \sum_{n \geq 0} P_s(T > n, X_n = s_0) \right) + \pi_{s_0}$$

$$= \sum_{s \in C(s_0)} \pi_s E_s(\sum_{n=1}^T 1_{X_n = s_0}) + \pi_{s_0}. \quad (10)$$

...
Hence
\[ P(X_0 \in S) = \sum_{s \in S} \pi_s \mathbb{E}_s \left( \sum_{n=1}^{T} 1_{\{X_n \in C(s)\}} \right) + \sum_{s \in S} \pi_s. \quad (10) \]

On the other hand from (7) we obtain
\[ P(X_0 \in I \setminus S) = \sum_{s \in S} \pi_s \left( \sum_{n \geq 0} P_s(T > n + 1, X_n+1 \in C(s)) \right) \]
\[ = \sum_{s \in S} \sum_{n \geq 0} \pi_s \left( E_s(T) - P_s(T < \infty) - E_s \left( \sum_{n=1}^{T} 1_{\{X_n \in C(s)\}} \right) \right) \]
\[ = \sum_{s \in S} \pi_s \left( E_s(T) - E_s \left( \sum_{n=1}^{T} 1_{\{X_n \in I \setminus S\}} \right) \right) - \sum_{s \in S} \pi_s. \quad (11) \]

From (10) and (11) we get,
\[ \sum_{i \in I} P(X_0 = i) = \sum_{s \in S} \pi_s \mathbb{E}_s(T). \]

Hence, condition (9) is necessary and sufficient in order that \( P \) is a probability measure on \( I^N \). So (8) is a necessary and sufficient condition in order that there exists such a strictly positive vector \( \pi \). \( \Box \)

Note that relation (9), together with \( \pi > 0 \) and \( E_s(T) \geq 1 \) for \( s \in S \), imply \( \sum_{s \in S} \pi_s \leq 1 \). Moreover \( \sum_{s \in S} \pi_s = 1 \) if and only if \( E_s(T) = 1 \) for all \( s \in S \), which is equivalent to \( P_s(T = 1) = 1 \) for all \( s \in S \). In this case the dynamics we study further will be trivial. So, we can assume \( \pi \) is a strictly positive and strictly substochastic vector.

From now on we assume (8) always hold and that \( \pi \) satisfies (9), so \( P \) is a probability measure on \( I^N \). We denote by \( \mathbb{E} \) its mean expected value.

Remark 2.3. From (4) and by using (9), we get formally
\[ P(T < \infty) = \sum_{s \in S} \pi_s \left( \sum_{n \geq 0} P_s(T > n, T < \infty) \right) = \sum_{s \in S} \pi_s \mathbb{E}_s(T) = 1. \quad (12) \]

This is formal because the definition of \( T \) requires a pointwise construction of \( P \) where the type of region at the initial time is explicitly known. This will be done in Section 3 for a class of laws \( P_s \) that satisfy a regenerative condition and for vectors \( \pi \) that define a stationary law \( P \). \( \Box \)
3 Regeneration and conditions for stationarity

Let us introduce some notation and recall some basic notions. For a probability measure \( P \) on \( I^N \), \( E \) denotes its associated expectation and \( E(\cdot \mid B') \) the mean expected value operator with respect to a sub-\( \sigma \)-field \( B' \subseteq B \) and \( P \). For \( i \in I \) we denote by \( P_i = P(\cdot \mid X_0 = i) \) the conditional distribution to start from \( i \in I \) and by \( E_i \) the expectation associated with \( P_i \).

A random time \( T' \) taking values in \( \mathbb{N} \cup \{\infty\} \) is a stopping time with respect to the filtration \( (B_n^X : n \in \mathbb{N} \cup \{\infty\}) \) when \( \{T' \leq n\} \in B_n^X \) is satisfied for all \( n \in \mathbb{N} \). Its associated \( \sigma \)-field is

\[
B_n^X = \{ B \in B : B \cap \{T' \leq n\} \in B_n^X, \forall n \in \mathbb{N} \}.
\]

It is easy to see that for every \( B(X_0) \) measurable random set \( J = J(X_0) \), the return time \( T_J(X_0) = \inf\{n > 0 : X_n \in J(X_0)\} \) is a stopping time. So, for \( X_0 \in S, T = T_{S\cap C(X_0)} \) is a stopping time. From (3), the random time \( T \) is finite \( P_s \)-a.s. for all \( s \in S \).

Let us define a regenerative time in a larger sense than in [17], Section 3.7 in [15] or Chapter V in [1], where it is required that at such a time the process restarts independently as a replica of the initial one. We only need that at a regenerative time the process starts in an independent way, unique requirement set in [2]. As said in [9] Section 2.4, at a regenerative time the strict past is forgotten.

**Definition 3.1.** Let \( T' \) be a stopping time with respect to the filtration \( (B_n^X : n \in \mathbb{N} \cup \{\infty\}) \). We say that \( P \) regenerates at \( T' \) if for all bounded measurable function \( h : I^N \rightarrow \mathbb{R} \) we have

\[
E(1_{\{T' < \infty\}} h \circ \Theta_{T'} \mid B_{T'}) = 1_{\{T' < \infty\}} E_{X_{T'}}(h) \quad P \text{-a.s.}.
\]

\[
\square
\]

Let us define a new family of probability measures \( (P_s^*: s \in S) \) from \( (P_s : s \in S) \) by regeneration at \( T \). To fix it we introduce some new notation. For a sequence \( \uparrow = (i_l : l = 0, \ldots, m) \in S \times I^+ \) let \( (\tau_n(\uparrow) : n \geq 0) \) be the set of indexes given by

\[
\tau_0(\uparrow) = 0 \quad \text{and} \quad \forall n \geq 1 : \tau_n(\uparrow) = \inf\{l > \tau_{n-1} : i_l \in S \setminus C(i_{\tau_{n-1}(\uparrow)})\}.
\]

Let \( \chi(\uparrow) = \sup\{n \geq 0 : \tau_n(\uparrow) < \infty\} \). From definition,

\[
\forall n \in \{1, \ldots, \chi(\uparrow)\} : C(i_{\tau_n(\uparrow)}) \neq C(i_{\tau_{n-1}(\uparrow)}).
\]

Let us define the laws \( (P_s^* : s \in S) \). Take \( \uparrow = (i_l : l = 0, \ldots, m) \in I^+ \), note the functions \( \tau_k(\uparrow) \) by \( \tau_k \), but in \( \chi(\uparrow) \) keep the dependence on \( \uparrow \). We set,

\[
P_s^*(X_l = i_l, l = 0, \ldots, m) = 1_{\{i_0 = s\}} \prod_{k=0}^{\chi(\uparrow)-1} P_{i_{\tau_k}}(X_l = i_{\tau_k+l}, l = 1, \ldots, \tau_{k+1} - \tau_k)
\]

\[
\times P_{i_{\tau_{\chi(\uparrow)+l}}}(X_l = i_{\tau_{\chi(\uparrow)+l}}, l = 1, \ldots, m - \tau_{\chi(\uparrow)}).
\]
An inductive argument on \( \chi(\mathbb{1}) = 0,...,m \) shows that \( P_s^* \) is well-defined. Note that \( P_s^*(X_0 = s) = 1 \) for all \( s \in S \). From (15) we find \( P_s^*(T < \infty) = 1 \) for all \( s \in S \). Moreover, from definition (14), we can apply Borel-Cantelli lemma to get

\[
\forall s \in S, \forall n \in \mathbb{N}^*: \quad P_s^*(T^n < \infty) = 1.
\] (15)

We denote by \( E^*_s \) the mean expected value associated with \( P_s^* \). Note that \( P_s(B \cap \{T \leq n\}) = P_s^*(B \cap \{T \leq n\}) \) for all \( B \in \mathcal{B}_T \) and \( n \in \mathbb{N} \). In particular \( P_s^*(T > n) = P_s(T > n) \), so \( E^*_s(T) = E_s(T) \).

**Proposition 3.2.** For all probability vector \( \gamma = (\gamma_s : s \in S) \) and all \( n \in \mathbb{N}^* \) the distribution \( P_s^* = \sum_{s \in S} \gamma_s P_s^* \) regenerates at \( T^n \). In particular for all \( s \in S \), \( P_s^* \) regenerates at \( T \).

**Proof.** It suffices to show the statement for \( P_s^* = P_s^* \), that is for an extremal vector \( \gamma \). Also by an inductive argument it suffices to prove the result for \( n = 1 \), that is for \( T^1 = T \). Since \( P_s^*(T < \infty) = 1 \), we must show the following equality holds for \( (j_1,...,j_q) \in I^+ \),

\[
E^*_s(1_{\{X_{k+T} = j_k, k = 1,...,q\}} | B_T^X) = E^*_{X_T}(1_{\{X_{k+T} = j_k, k = 1,...,q\}}) \quad P^*_s \text{ a.s.} \quad (16)
\]

Let \( (i_0,...,i_m) \in I^+ \) be such that \( i_0 = s \), \( i_l \in C(s) \) for \( l = 1,...,m-1 \) and \( i_m \notin C(s) \). Let \( B_m = \{T = m, X_l = i_l, l = 0,...,m\} \). Then, (16) will be shown once we prove the equality

\[
\int_{B_m} 1_{\{X_{k+T} = j_k, k = 1,...,q\}} dP^*_s = \int_{B_m} P^*_{i_m}(X_k = j_k, k = 1,...,q) dP^*_s.
\]

But, this follows straightforwardly from a recurrence argument on property (14).

From now on we define the distribution \( \mathbb{P}^* \) as in Definition (2.1) but for the family of probability measures \( (P_s^*: s \in S) \) instead of \( (P_s: s \in S) \). It suffices to replace \( E_s \) by \( E^*_s \) in (14). Since \( E_s(T) = E^*_s(T) \) for \( s \in S \), the condition (15) is the same and \( \pi \) must satisfy the same condition (9). Thus, \( \mathbb{P}^* \) is a probability measure on \( I^N \) and we denote by \( E^* \) its mean expected value.

Let \( \gamma = (\gamma_s : s \in S) \) be a probability vector and \( \mathbb{P}^* = \sum_{s \in S} \gamma_s P_s^* \) be the associated distribution on \( I^N \), so \( \mathbb{P}^*(X_0 \in S) = 1 \). From relation (14), \( \mathbb{P}^* \) satisfies

\[
\mathbb{P}^*(X_l = i_l, l = 0,...,m, T = m, X_{m+k} = j_k, k = 1,...,t) = \mathbb{P}^*(X_l = i_l, l = 0,...,m, T = m) P^*_{i_m}(X_k = j_k, k = 1,...,t).
\]

Consider the following sequence of variables \( (\Xi_n : n \geq 0) \) taking values on \( S \),

\[
\Xi_0 = X_0 \text{ and } \forall n \geq 1: \quad \Xi_n = X_{T^n}.
\]

By (15) this is a well defined process. Under \( \mathbb{P}^* \), and by using Proposition 3.2 we get that the sequence \( (\Xi_n : n \in \mathbb{N}) \) is a Markov chain taking values in \( S \).
with transition matrix $Q = (q_{ss'} : s, s' \in S)$ given by $q_{ss'} = P_s(X_T = s')$ for $s, s' \in S$. In fact, from Proposition 3.2 we have

$$P^*(\Xi_{k+1} = s_{k+1} | \Xi_k = s_k, ..., \Xi_0 = s_0) = P^*_s(X_T = s_{k+1}).$$

Since $\Xi_{k+1} \in S \setminus C(\Xi_k)$, we get that $q_{ss'} > 0$ implies $s' \notin C(s)$.

Recall that a positive vector $\rho = (\rho_s : s \in S)$ is invariant for $Q$ if it satisfies the set of equalities

$$\forall s \in S : \rho_s = \sum_{s' \in S} \rho_{s'} q_{s's}.$$

There always exists invariant positive vectors, moreover, if $Q$ is irreducible an invariant positive vector is unique up to a multiplicative constant.

On the other hand, $P^*$ is a stationary distribution on $P^N$ if for all $m \in \mathbb{N}$ and all $(i_0, ..., i_m) \in I^{m+1}$ we have

$$\forall t \geq 1 : P^*(X_{k+t} = i_k, k=0, ..., m) = P^*(X_k = i_k, k=0, ..., m). \quad (17)$$

By an inductive argument, (17) is satisfied once it holds for $t = 1$, so stationarity is verified when for all $m \in \mathbb{N}$ and all $(i_0, ..., i_m) \in I^{m+1}$ it holds

$$P^*(X_k = i_k, l=0, ..., m) = \sum_{j \in I} P^*(X_0 = j, X_{k+1} = i_k, k=0, ..., m). \quad (18)$$

**Theorem 3.3.** Assume that the strictly positive vector $\pi = (\pi_s : s \in S)$ satisfies the condition (4). Then, $P^*$ is stationary if and only if $\pi$ is invariant for $Q$, that is it satisfies

$$\forall s \in S : \pi_s = \sum_{s' \in S} \pi_{s'} q_{s's} \text{ where } q_{ss'} = P_s(X_T = s'). \quad (19)$$

**Proof.** From (18) $P^*$ is stationary if for all $m \in \mathbb{N}$ and all $(i_0, ..., i_m) \in I^{m+1}$ it is satisfied

$$P^*(X_l = i_l, l=0, ..., m) = \sum_{j \in I} P^*(X_0 = j, X_{l+1} = i_l, l=0, ..., m). \quad (20)$$

Let,

$$G = \{X_l = i_l, l=0, ..., m\} \text{ and } G \Theta_n^{-1} = \{X_{l+n} = i_l, l=0, ..., m\}$$

be the $n$-shifted set. The stationarity condition is $P^*(G) = P^*(G \Theta_1^{-1})$. From (17) we obtain

$$P^*(X_0 \in I \setminus S, G \Theta_n^{-1}) = \sum_{s \in S} \sum_{n \geq 0} \pi_s P_s^*(T > n+1, X_{l+n+1} \in I \setminus S, G \Theta_{n+2}^{-1})$$

$$= \sum_{s \in S} \sum_{n \geq 0} \pi_s (P_s^*(T > n+1, G \Theta_{n+2}^{-1}) - P_s^*(T > n+1, X_{n+1} \in C(s), G \Theta_{n+2}^{-1})).$$

10
From (6) we get
\[ \sum_{j \in \mathcal{S}} P^*(X_0 = j, G \circ \Theta_1^{-1}) = \sum_{s \in \mathcal{S}} \left( \sum_{n \geq 0} P^*_s(T > n + 1, X_{n+1} \in \mathcal{C}(s), G \circ \Theta_{n+2}^{-1}) \right) \]
\[ + \sum_{s \in \mathcal{S}} \pi_s P^*_s(G \circ \Theta_1^{-1}). \]

From the last two expressions we obtain
\[ \sum_{j \in \mathcal{I}} P^*(X_0 = j, G \circ \Theta_1^{-1}) = \sum_{s \in \mathcal{S}} \pi_s \left( \sum_{n \geq 0} P^*_s(T > n + 1, G \circ \Theta_{n+2}^{-1}) \right) \]
\[ + \sum_{s \in \mathcal{S}} \pi_s P^*_s(G \circ \Theta_1^{-1}) \]
\[ = \sum_{s \in \mathcal{S}} \pi_s \left( \sum_{n \geq 0} P^*_s(T > n + 1, G \circ \Theta_{n+1}^{-1}) \right). \quad (21) \]

For \( i_0 \in \mathcal{I} \setminus \mathcal{S} \), expression (17) implies,
\[ P^*(G) = \sum_{s \in \mathcal{S}} \pi_s \left( \sum_{n \geq 0} P^*_s(T > n + 1, G \circ \Theta_{n+1}^{-1}) \right). \quad (22) \]

Hence, (21) and (22) show that, with no additional hypothesis, the stationary equality (20) is satisfied when \((i_0, \ldots, i_m) \in (I \setminus \mathcal{S}) \times I^+\).

Now, when \( i_0 \in \mathcal{S} \), from (21) we obtain,
\[ \sum_{j \in \mathcal{I}} P^*(X_0 = j, G \circ \Theta_1^{-1}) = \sum_{s \in \mathcal{C}(i_0)} \pi_s \left( \sum_{n \geq 0} P^*_s(T > n + 1, G \circ \Theta_{n+1}^{-1}) \right) \]
\[ + \sum_{s \in \mathcal{S} \setminus \mathcal{C}(i_0)} \pi_s \left( \sum_{n \geq 0} P^*_s(T = n + 1, G \circ \Theta_{n+1}^{-1}) \right). \quad (23) \]

Now we use \( P^*_s(T < \infty) = 1 \) for all \( s \in \mathcal{S} \) as well as the definition done in (14), to get
\[ \sum_{s \in \mathcal{S} \setminus \mathcal{C}(i_0)} \pi_s \left( \sum_{n \geq 0} P^*_s(T = n + 1, G \circ \Theta_{n+1}^{-1}) \right) \]
\[ = \left( \sum_{s \in \mathcal{S} \setminus \mathcal{C}(i_0)} \pi_s \left( \sum_{n \geq 0} P^*_s(T = n + 1, X_T = i_0) \right) \right) \cdot P^*_{i_0}(G) \]
\[ = \left( \sum_{s \in \mathcal{S} \setminus \mathcal{C}(i_0)} \pi_s P^*_s(X_T = i_0) \right) \cdot P^*_{i_0}(G). \quad (24) \]

On the other hand by using \( i_0 \in \mathcal{S} \) formula (6) gives
\[ P^*(G) = \sum_{s \in \mathcal{C}(i_0)} \pi_s \left( \sum_{n \geq 0} P^*_s(T > n + 1, G \circ \Theta_{n+1}^{-1}) \right) + \pi_{i_0} P^*_{i_0}(G). \quad (25) \]

Therefore, from equalities (23), (24) and (25) we deduce that the equality (20) is satisfied if and only if the following relation holds
\[ \forall i_0 \in \mathcal{S} : \pi_{i_0} = \sum_{s \in \mathcal{S} \setminus \mathcal{C}(i_0)} \pi_s P^*_s(X_T = i_0). \]
Since $P^*_s(X_T = s') = 0$ when $C(s') = C(s)$, we have proven that $P^*$ is stationary if and only if the following condition is satisfied
\[
\forall s' \in S : \pi_{s'} = \sum_{s \in S} \pi_s P^*_s(X_T = s') = \sum_{s \in S} \pi_s q_{ss'} .
\]
This shows the theorem. \qed

When $P^*$ is stationary we can extend it to the set of bi-infinite sequences $I^\mathbb{Z}$ by putting
\[
P^*(X_{l+k} = i_k, k = 0, \ldots, m) = P^*(X_k = i_k, k = 0, \ldots, m)
\]
for all $l \in \mathbb{Z}$, $m \geq 0$ and $(i_k : k = 0, \ldots, m) \in I^+$. Note that this equality obviously holds for $l \in \mathbb{N}$ because $P^*$ is stationary.

4 A renewal property of the law

Define the probability vector
\[
\hat{\pi} = (\hat{\pi}_s : s \in S) \text{ with } \hat{\pi}_s = \pi_s \left( \sum_{s' \in S} \pi_{s'} \right)^{-1}.
\]
Consider the distribution $P^- = \sum_{s \in S} \hat{\pi}_s P_s$ on $B^\infty_X$ and let $E^- = \sum_{s \in S} \pi_s E_s$ be its mean expected value. From (15) we have $P^- (T^n < \infty) = 1$ for all $n \in \mathbb{N}^*$, being $T = T_{S \setminus C(s)}$. By condition (9) we also find
\[
E^- (T) = \sum_{s \in S} \pi_s \left( \sum_{s \in S} \pi_s E_s (T) \right) = \left( \sum_{s \in S} \pi_s \right)^{-1}.
\]
Let $P^*_\hat{\pi}$ be given by $P^*_\hat{\pi} = \sum_{s \in S} \hat{\pi}_s P^*_s$ on $B^\infty_X$ and $E^*_\hat{\pi}$ be its mean expected value. By previous relations
\[
\forall n \in \mathbb{N}^* \quad P^*_\hat{\pi} (T^n < \infty) = 1 \quad \text{and} \quad E^*_\hat{\pi} (T)^{-1} = \sum_{s \in S} \pi_s .
\]

We will extend the probability measure spaces $(I^\mathbb{N}, B^\infty_X)$ where the probabilities $P^n_s$, $P^*_s$, $P^*$ are defined, to include a countable number of independent copies of $X$. Since this is a simple extension, the probability distributions on this space will be continue to be noted $P^n_s$, $P^*_s$, $P^*$, respectively.

Consider the distribution of $T$: $P^*_\hat{\pi} (T = l), l \in \mathbb{N}^*$. It is aperiodic if the greatest common divisor of its support satisfies
\[
g.c.d. \{ l > 0 : P^*_\hat{\pi} (T = l) > 0 \} = 1.
\]
Theorem 4.1. Assume \( \pi \) satisfies (22), (24) and the distribution of \( T \) satisfies (25). Then,
\[
\forall B \in \mathcal{B}_\infty^X: \quad \mathbb{P}^*(B) = \lim_{N \to \infty} \mathbb{P}^*_\hat{\pi}(B \circ \Theta_N^{-1}). \tag{29}
\]
If in addition the matrix \( Q \) is aperiodic then for all probability vector \( \gamma = (\gamma_s : s \in S) \) the probability measure \( \mathbb{P}_\gamma^* = \sum_{s \in S} \gamma_s \mathbb{P}_s^* \) satisfies
\[
\forall B \in \mathcal{B}_\infty^X: \quad \mathbb{P}^*_\gamma(B) = \lim_{N \to \infty} \mathbb{P}^*_\gamma(B \circ \Theta_N^{-1}). \tag{30}
\]

Proof. Let us first prove the statement (29). It is sufficient to show the equality for \( B = (X_k = i_k, k = 0, \ldots, m) \) with \( (i_k: k = 0, \ldots, m) \in I^+ \).

Since \( \hat{\pi} \) is invariant for the stochastic matrix \( Q \),
\[
\forall s \in S: \quad \hat{\pi}_s = \mathbb{P}^*_\hat{\pi}(X_T = s). \tag{31}
\]

From Proposition 3.2 the probability distribution \( \mathbb{P}_s^* \) regenerates at times \( (T^n : n \in \mathbb{N}^*) \). An inductive argument on \( \mathbb{N}^+ \) gives \( \hat{\pi}_s = \mathbb{P}_s^*(X_{T^n} = s) \), so \( X_{T^n} \) is distributed as \( \hat{\pi} \). Then, by using (13), for all bounded and measurable \( h : I^N \to \mathbb{R} \) we have
\[
\mathbb{E}_s^*(h \circ \Theta_{T^n} | \mathcal{B}_{T^n}) = \mathbb{E}_s^*(h) \quad \mathbb{P}_s^* - \text{a.s.}. \tag{32}
\]

Under \( \mathbb{P}_s^* \), the increments \( (T^1, T^{n+1} - T^n : n \in \mathbb{N}^*) \) are independent equally distributed, each increment having the same distribution as \( T \). For proving (29) it is useful to give a renewal construction of \( \mathbb{P}_s^* \).

Consider a sequence \( (X^{(m)} : m \in \mathbb{N}^*) \) of independent copies of the process \( (X_n : n \leq T) \) with distribution \( \mathbb{P}_s^* \), so \( X^{(m)} = (X^{(m)}_n : n \leq T^{[m]}) \) where \( (T^{[m]} : m \in \mathbb{N}^*) \) is a sequence of independent copies of \( T \). Define the process \( \hat{X} = (\hat{X}_n : n \geq 0) \) by
\[
\hat{X}_n = X^{(m)}_{n'} \quad \text{if} \quad n = \sum_{l=1}^{m-1} T^{[l]} + n' \quad \text{and} \quad n' \leq T^{[m]}.
\]

Since \( \mathbb{P}_s^* \) regenerate at \( (T^n : n \in \mathbb{N}) \) and \( X_{T^n} \) has distribution \( \hat{\pi} \), we get that \( \hat{X} \) and \( X \) are equally distributed with distribution \( \mathbb{P}_s^* \).

Let \( \hat{T}^n = \sum_{k=1}^{n} T^{[m]} \) and \( \hat{T} = \hat{T}^1 \). Then, \( (\hat{T}^n : n \in \mathbb{N}^*) \) is the sequence of hitting times of different classes of \( \hat{X} \) and it has the same distribution as \( (T^n : n \in \mathbb{N}^*) \). By construction \( (T^{[m]} : m \in \mathbb{N}^*) \) are independent identically distributed random variables with common distribution \( \mathbb{P}^*_\pi(T^{[m]} = l) = \mathbb{P}^*_\hat{\pi}(\hat{T} = l) \) for \( l \in \mathbb{N}^* \). Since these variables have finite mean \( \mathbb{E}^*_\pi(\hat{T}) = (\sum_{s \in S} \pi_s)^{-1} \), we can apply the renewal theorem, see Chapter II in [8]. Define,
\[
\forall N > 0: \quad \beta_N = \sup \{ \hat{T}^n : \hat{T}^n \leq N, n \in \mathbb{N}^* \},
\]
where we put $\beta_N = 0$ if $\hat{T}^n > N$ for all $n \in \mathbb{N}^*$. Note that the distribution of $\beta_N$ only depends on the sequence $(\hat{T}^n : n \geq 1)$. By aperiodicity of the distribution of $\hat{T}$ the renewal theorem gives:

$$\forall l \geq 0: \lim_{N \to \infty} P^*_\pi(\beta_N = N - l) = E^{\pi}(T)^{-1}P^*_\pi(\hat{T} > l) = (\sum_{s \in S}\pi_s)P^*_\pi(\hat{T} > l) \quad (33)$$

and,

$$\lim_{N \to \infty} P^*_\pi(N \in \{\hat{T}^n : n \in \mathbb{N}^*\}) = E^{\pi}(T)^{-1} = \sum_{s \in S}\pi_s. \quad (34)$$

We have

$$P^*_\pi(\hat{X}_{N+k} = i_k, k = 0, \ldots, m) = \sum_{l=0}^N P^*_\pi(\hat{X}_{N+k} = i_k, k = 0, \ldots, m; \beta_N = N - l).$$

Let $\epsilon > 0$ and $r > 0$. From (33) we obtain

$$\exists N' = N'(\epsilon, r) \text{ such that } \forall N \geq N': P^*_\pi(\beta_N < N - r) < \epsilon. \quad (35)$$

Hence

$$|P^*_\pi(\hat{X}_{N+k} = i_k, k = 0, \ldots, m) - \sum_{l=0}^r P^*_\pi(\hat{X}_{N+k} = i_k, k = 0, \ldots, m; \beta_N = N - l)| < \epsilon.$$

By regeneration at times \{\hat{T}^n : n \in \mathbb{N}^*\} (see (32)) we get

$$P^*_\pi(\hat{X}_{N+k} = i_k, k = 0, \ldots, m; \beta_N = N - l) = P^*_\pi(\hat{X}_{N+k} = i_k, k = 0, \ldots, m; \beta_N = N - l \in \{\hat{T}^n : n \in \mathbb{N}^*\}),$$

$$(N - l, N] \cap \{\hat{T}^n : n \in \mathbb{N}^*\} = \emptyset$$

$$= P^*_\pi(\hat{X}_{l+k} = i_k, k = 0, \ldots, m, \hat{T} > l)P^*_\pi(N - l \in \{\hat{T}^n : n \in \mathbb{N}^*\}). \quad (36)$$

From (34) we get the existence of $N''(r, \epsilon) > N'$ such that for all $N > N''$ we have

$$\forall l \in \{0, \ldots, r\}: |P^*_\pi(N - l \in \{\hat{T}^n : n \in \mathbb{N}^*\}) - (\sum_{s \in S}\pi_s)| < \frac{\epsilon}{r}. \quad (37)$$

Therefore

$$|P^*_\pi(\hat{X}_{N+k} = i_k, k = 0, \ldots, m) - \sum_{l=0}^r P^*_\pi(\hat{X}_{l+k} = i_k, k = 0, \ldots, m; T > l)(\sum_{s \in S}\pi_s)| < 2\epsilon. \quad (38)$$

Since $X$ and $\hat{X}$ are equally distributed we get

$$\lim_{N \to \infty} P^*_\pi(X_{N+k} = i_k, k = 0, \ldots, m)$$

$$= \sum_{l=0}^\infty P^*_\pi(T > l, X_{l+k} = i_k, k = 0, \ldots, m)(\sum_{s \in S}\pi_s)$$

$$= \sum_{l=0}^\infty \sum_{s \in S}\pi_s P^*_s(T > l, X_{l+k} = i_k, k = 0, \ldots, m)$$

$$= P^*_s(X_k = i_k, k = 0, \ldots, m). \quad (39)$$
Then, the proof of (29) for $B = \{X_k = i_k, k = 0, \ldots, m\}$ is finished.

Let us now show (30). Since $Q$ is aperiodic, $\hat{\pi}$ is the unique invariant probability measure for $Q$ and every probability vector $\nu = (\nu_s : s \in S)$ satisfies

$$\lim_{N \to \infty} \nu'Q^N = \hat{\pi}'.$$  \hspace{1cm} (40)

where we note by $\nu'$ the row vector, transpose of $\nu$. From Proposition 3.2 the probability distribution $P_s^\ast$ regenerates at times $(T^n : n \in \mathbb{N})$ (they are finite $P^\ast_s$-a.s.). Denote by $\gamma(n)$ the distribution of $X_{T^n}$ on $S$ when we start from $P^\ast_s$. Since $X_{T^{n-1}}$ is distributed as $\gamma(n-1)$ and $P^\ast_{\gamma(n-1)}(X_T = s') = (\gamma(n')Q)_{s'}$ we get $\gamma(n') = \gamma'Q^n$. So, (40) gives

$$\lim_{N \to \infty} \gamma(n) = \hat{\pi}.\hspace{1cm} (41)$$

Note that for all event $D \in \mathcal{B}^N_\infty$ and all probability vector $\nu$ we have

$$|P^\ast_s(D) - P^\ast_{\hat{\pi}}(D)| = |\sum_{s \in S} (\nu_s - \hat{\pi}_s)P^\ast_s(D)| \leq \sum_{s \in S} |\nu_s - \hat{\pi}_s|.\hspace{1cm} (42)$$

Let us fix $\epsilon > 0$. Since $P_s(D) = \sum_{s \in S} P_s(D)$, when $P_{\hat{\pi}}(D) < \epsilon$ we get $P_s(D) < \epsilon/\hat{\pi}_s$. Since $\hat{\pi} > 0$, from (35) we obtain that for all $\epsilon > 0$ and $r > 0$,

$$\exists N(\epsilon, r) \text{ such that } \forall s \in S, \forall N \geq N(\epsilon, r) : P^\ast_s(\beta_N < N - r) < \epsilon.\hspace{1cm} (43)$$

Define the sequence of random variables $(\eta_N : n \in \mathbb{N}^*)$ by

$$\eta_N = \sup\{n \in \mathbb{N}^* : T^n \leq N\},$$

where we put $\eta_N = 0$ if $T^n > N$ for all $n \in \mathbb{N}^*$. The sequence $(\eta_N : N \in \mathbb{N}^*)$ is increasing and

$$\forall s \in S : \ P^\ast_s(\lim_{N \to \infty} \eta_N = \infty) = 1.$$

Then,

$$\forall \tilde{r} \in \mathbb{N}^* \exists N'(\tilde{r}, \epsilon) \forall N \geq N'(\tilde{r}, \epsilon) \forall s \in S : P^\ast_s(\eta_N \leq \tilde{r}) < \epsilon.$$\hspace{1cm} (44)

Hence, for all $N \geq N'(\tilde{r}, \epsilon)$ we have $P^\ast_s(\eta_N \leq \tilde{r}) < \epsilon$. This last relation and (41) implies the existence of $N''(\epsilon)$ that satisfies

$$\forall N \geq N''(\epsilon) : \ E^\ast_s(\sum_{s \in S} |\gamma(\eta_N)_s - \hat{\pi}_s|) < \epsilon.$$\hspace{1cm} (45)

Then, above relation and (42) implies that for all event $D \in \mathcal{B}^N_\infty$ and $N \geq N''(\epsilon)$ it is satisfied

$$|E^\ast_s(P^\ast_{\gamma(\eta_N)}(D)) - P^\ast_{\hat{\pi}}(D)| = |E^\ast_s(P^\ast_{\gamma(\eta_N)}(D) - P^\ast_{\hat{\pi}}(D))| \leq E^\ast_s(\sum_{s \in S} |\gamma(\eta_N)_s - \hat{\pi}_s|) < \epsilon.\hspace{1cm} (46)$$
From (43) we get that for all $N \geq N(\epsilon, r)$ it holds

$$|P^*_{\gamma}(X_{N+k} = i_k, k = 0, m) - \sum_{l=0}^{r} P^*_{\gamma}(X_{N+k} = i_k, k = 0, m, \beta_N = N-l)| < \epsilon.$$ 

By regeneration at times $\{T^n : n \in \mathbb{N}^*\}$, see (43), and since the law of $X_{T^n}$ is $\gamma(n)$ we obtain for all $l = 0, \ldots, r$:

$$P^*_{\gamma}(X_{N+k} = i_k, k = 0, m; \beta_N = N-l) = \sum_{l=0}^{r} P^*_{\gamma}(N-l \in \{T^n : n \in \mathbb{N}^*\})P^*_{\gamma}(N-l \in \{T^n : n \in \mathbb{N}^*\})\{N-l, N\} \cap \{T^n : n \in \mathbb{N}^*\} = \emptyset.$$

From (44) we get for all $l = 0, \ldots, r$:

$$P^*_{\gamma}(X_{N+k} = i_k, k = 0, m; \beta_N = N-l) = E^*_\gamma(P^*_{\gamma(n+\gamma)}(X_{l+k} = i_k, k = 0, m; T > l) - P^*_{\gamma}(X_{l+k} = i_k, k = 0, m; T > l) < \epsilon.$$

From (39), (37), (38) and (39), we get that the proof will be complete once we show

$$\lim_{N \to \infty} P^*_{\gamma}(N \in \{T^n : n \in \mathbb{N}^*\}) = \lim_{N \to \infty} P^*_{\pi}(N \in \{T^n : n \in \mathbb{N}^*\}). \quad (45)$$

From (43) we get for all $N \geq N(\epsilon, r)$

$$P^*_{\pi}(\beta_N < N-r) + P^*_{\pi}(\beta_N < N-r) < 2\epsilon.$$

Then, for all $k > 0$ we obtain

$$|P^*_{\pi}(N+k \in \{T^n : n \in \mathbb{N}^*\}) - P^*_{\pi}(N+k \in \{T^n : n \in \mathbb{N}^*\})| \leq |P^*_{\pi}(N+k \in \{T^n : n \in \mathbb{N}^*\}, \beta_N \geq N-r) - P^*_{\pi}(N+k \in \{T^n : n \in \mathbb{N}^*\}, \beta_N \geq N-r)| + 2\epsilon.$$

We have

$$P^*_{\pi}(N+k \in \{T^n : n \in \mathbb{N}^*\}, \beta_N \geq N-r) \leq \sum_{l=0}^{r} (P^*_{\pi}(\beta_N = N-l)P^*_{\pi}(k+l \in \{T^n : n \in \mathbb{N}^*\})$$

$$-P^*_{\gamma}(\beta_N = N-l)E^*_\gamma(P^*_{\gamma(n+\gamma)}(k+l \in \{T^n : n \in \mathbb{N}^*\}).$$

For all $N \geq N(\epsilon, r)$ we have

$$|E^*_\gamma(P^*_{\gamma(n+\gamma)}(k+l \in \{T^n : n \in \mathbb{N}^*\}) - P^*_{\pi}(k+l \in \{T^n : n \in \mathbb{N}^*\})| < \epsilon.$$

We use $P^*_{\pi}(k+l \in \{T^n : n \in \mathbb{N}^*\}) = (\sum_{s \in S} \pi_s)^{-1}$ and put all these relations together to conclude,

$$|P^*_{\pi}(N+k \in \{\tilde{T}^n : n \in \mathbb{N}^*\}, \beta_N \geq N-r) - P^*_{\gamma}(N+k \in \{\tilde{T}^n : n \in \mathbb{N}^*\}, \beta_N \geq N-r)|$$

$$\leq \left| \sum_{l=0}^{r} (P^*_{\pi}(\beta_N = N-l) - P^*_{\pi}(\beta_N = N-l)) \right| (\sum_{s \in S} \pi_s)^{-1} + \epsilon$$

$$\leq 2\epsilon(\sum_{s \in S} \pi_s)^{-1}.$$
Hence (45) is shown. Therefore, relation (30) is proven. □

**Remark 4.2.** Note that in the proof of property (30) we require a starting measure of the type $\sum_{s \in S} \gamma_s P_s^*$ because, on the one hand the definition of $T$ needs that the starting state in $S$ is defined and, on the other hand we use the regenerative equality of this measure as stated in (14). □

Let $P^*$ be the law defined on $\mathbb{I}^\mathbb{Z}$ by (26). The same proof showing property (29) in Theorem 4.1 allows us to prove that for all $l \in \mathbb{Z}$ and all $(i_k : k = 0,..,m) \in I^+$ we have

$$P^*(X_{l+k} = i_k, k = 0,..,m) = \lim_{N \to \infty} P_{\hat{\pi}}^*(X_{l+k+N} = i_k, k = 0,..,m). \quad (46)$$

From (43) we deduce that the random variable

$$T^0 = \sup \{T^n : T \leq 0\}$$

which takes value in $\{l \in \mathbb{Z} : l \leq 0\}$, has a proper distribution. That is $T^0$ is finite $P^*$-a.s.. Since there is regeneration at $T_0$ the random variable $X_{T_0}$ has distribution $\hat{\pi}$. Then,

$$P^*(X_0 = i_k, k = 0,..,m) = \sum_{n \in \mathbb{N}} P^*(T^0 = -n, X_0 = i_k, k = 0,..,m)$$

$$= \sum_{n \in \mathbb{N}} P^*(X_0 = i_k, k = 0,..,m \mid T^0 = -n) P^*(T^0 = -n). \quad (47)$$

Since $P^*$ regenerates at each $T^n$ with law $\hat{\pi}$, we have

$$P^*(X_0 = i_k, k = 0,..,m \mid T^0 = -n) = P_{\hat{\pi}}^*(X_n = i_{k+n}, k = 0,..,m \mid T > n). \quad (48)$$

On the other hand from (33) we get,

$$P_{\hat{\pi}}^*(T^0 = -n) = (\sum_{s \in S} \pi_s) P_{\hat{\pi}}^*(T > n). \quad (49)$$

Then, we retrieve the definition done in (41),

$$P^*(X_0 = i_k, k = 0,..,m) = \sum_{s \in S} \sum_{n \in \mathbb{N}} \pi_s P_s^*(X_n = i_{k+n}, k = 0,..,m; T > n).$$

Hence (47), (48) and (49) give a probabilistic insight to definition $P^*$ and allow us to have a good definition of $T$ under law $P^*$, as claimed in Remark 2.3.

5 Stationarity and Chargaff second parity rule

Let $L$ be an alphabet and $Y_n : L^\mathbb{N} \to L$ be the $n$-th coordinate function: $Y_n(y) = y_n$ for $y \in L^\mathbb{N}$. Let $\varphi : L \to L$ be a convolution, this means $\varphi$ is one-to-one and $\varphi^{-1} = \varphi$. Since $\varphi$ is a bijection we have $L = \{\varphi(h) : h \in L\}$.  


Let $P$ be a probability measure on $L^N$. We say that $P$ satisfies the Chargaff second parity rule (CSPR) with respect to $\varphi$ if for all $m \in \mathbb{N}$, all $(l_0, \ldots, l_m) \in L^{m+1}$ and all $t \in \mathbb{N}$ it is satisfied:

$$P(Y_{k+t} = l_k, k = 0, \ldots, m) = P(Y_{k+t} = \varphi(l_{m-k}), k = 0, \ldots, m).$$  \hspace{1cm} (50)

We claim that (50) is satisfied if it holds for $t = 0$. That is, if for all $m \in \mathbb{N}$ and all $(l_0, \ldots, l_m) \in L^{m+1}$,

$$P(Y_k = l_k, k = 0, \ldots, m) = P(Y_k = \varphi(l_{m-k}), k = 0, \ldots, m).$$  \hspace{1cm} (51)

In fact, from (51) we get,

$$P(Y_k = h_k, k = 0, \ldots, t-1; Y_{t+k} = l_k, k = 0, \ldots, m; Y_{t+m+k} = c_k, k = 0, \ldots, t-1)$$

$$= P(Y_k = \varphi(c_{t-k}), k = 0, \ldots, t-1; Y_{k+t} = \varphi(l_{m-k}), k = 0, \ldots, m;$$

$$Y_{k+t+k} = \varphi(h_{t-1-k}), k = 0, \ldots, t-1).$$

Hence, by summing on $(h_0, \ldots, h_{t-1}) \in L^t$ and $(c_0, \ldots, c_{t-1}) \in L^t$ we get (50).

**Proposition 5.1.** If $P$ verifies the CSPR then it is stationary.

**Proof.** Assume $P$ satisfies the CSPR. For all $m \in \mathbb{N}$ we have

$$\sum_{h \in L} P(Y_0 = h, Y_{k+1} = l_k, k = 0, \ldots, m)$$

$$= \sum_{h \in L} P(Y_{m+1} = \varphi(h), Y_{m-k} = \varphi(l_k), k = 0, \ldots, m)$$

$$= P(Y_{m+1} \in L, Y_{m-k} = \varphi(l_k), k = 0, \ldots, m)$$

$$= P(Y_{m-k} = \varphi(l_k), k = 0, \ldots, m) = P(Y_k = l_k, k = 0, \ldots, m).$$

Then, the result follows.

Let us fix $d \in \mathbb{N}^*$ and consider $I := L^d$ as a new alphabet. Take the following transformation $\zeta : L^N \to I^N, y \to \zeta y$ with $x_n = (\zeta y)_n = (y_{dn}, y_{d(n+1)-1})$. Let $P \circ \zeta^{-1}$ be the induced law on $I^N$. We claim that if $P$ is stationary, then also $P \circ \zeta^{-1}$ is stationary. Let $X_n : I^N \to I$ be the $n$-th coordinate function, we must prove that for all $m \in \mathbb{N}$ and $(l_{dk}, \ldots, l_{d(k+1)-1}) : k = 0, \ldots, m) \in I^{m+1}$ we have

$$P \circ \zeta^{-1}(X_k = (l_{dk}, \ldots, l_{d(k+1)-1}); k = 0, \ldots, m)$$

$$= \sum_{(e_0, \ldots, c_k) \in L^d} P \circ \zeta^{-1}(X_0 = (e_0, \ldots, c_{d-1}), X_{k+1} = (l_{dk}, \ldots, l_{d(k+1)-1}, k = 0, \ldots, m).$$

This relation is equivalent to,

$$P(Y_t = l_t, t = 0, \ldots, dm - 1)$$

$$= \sum_{(e_0, \ldots, c_{d-1}) \in L^d} P(Y_0 = e_0, \ldots, Y_{d-1} = c_{d-1}; Y_{t+d} = l_t, t = 0, \ldots, dm - 1),$$

18
which is equivalent to the equality
\[ P(Y_t = l_t, t = 0, \ldots, dm - 1) = P(Y_{t+d} = l_t, t = 0, \ldots, dm - 1). \]
This last relation follows straightforward from the stationarity of \( P \), proving that \( P \circ \psi^{-1} \) is stationary.

For \( I = L^d \), let \( \psi : I \to A \) be an onto-function and consider the function \( \Psi : L^N \to A^N, \ y \to \Psi y \) by \( (\Psi y)_n = \psi((\zeta y)_n) \). We claim that if \( P \) is stationary, then also \( P \circ \Psi^{-1} \) is stationary. Denote by \( Z_n : A^N \to A \) the \( n \)-th coordinate function, we must show that
\[ P \circ \Psi^{-1}(Z_k = a_k, k = 0, \ldots, m) = \sum_{b \in I} P \circ \Psi^{-1}(X_0 = b, X_{k+1} = a_k, k = 0, \ldots, m). \]
From the equality,
\[ P \circ \Psi^{-1}(Z_k = a_k, k = 0, \ldots, m) = \sum_{(l_d, \ldots, l_{d(k+1)-1}) \in \psi^{-1}(a_k), k = 0, \ldots, m} P \circ \zeta^{-1}(X_k = (l_{dk}, \ldots, l_{d(k+1)-1}), k = 0, \ldots, m), \]
the stationarity of \( P \circ \Psi^{-1} \) is retrieved from the stationarity of \( P \circ \zeta^{-1} \).

Let us state that a weaker condition of CSPR implies a weaker stationary property. Assume that the CSPR is verified only for words of length smaller or equal to \( t \). This means that for all for all \( m < t \), all \((l_0, \ldots, l_m) \in L^{m+1}\) and all \( u \in \mathbb{N} \) it is satisfied:
\[ P(Y_{k+u} = l_k, k = 0, \ldots, m) = P(Y_{k+u} = \varphi(l_{m-k}), k = 0, \ldots, m). \]
Let us prove that in this case the stationarity only holds for cylinders of length strictly smaller than \( t \).

**Proposition 5.2.** Let \( t \geq 2 \). Assume \( P \) verifies the CSPR for cylinders defined by words of length smaller or equal to \( t \), then for all \( m < t-1 \) and all \((l_0, \ldots, l_m) \in L^{m+1}\) we have
\[ \forall u \geq 1 : P(Y_{k+u} = l_k, k = 0, \ldots, m) = P(Y_k = l_k, k = 0, \ldots, m). \]

**Proof.** We prove it by induction on \( u \geq 1 \). For \( u = 1 \) the proof is the same as the one done in Proposition 5.1. Assume it has been shown up to \( u \), let us prove it for \( u + 1 \). Since \( m + 2 \leq t \) we get,
\[ P(Y_{u+1+k} = l_k, k = 0, \ldots, m) = \sum_{h \in L} P(Y_u = h, Y_{u+1+k} = l_k, k = 0, \ldots, m) \]
\[ = \sum_{h \in L} P(Y_{u+m-k} = \varphi(l_k), k = 0, \ldots, m; Y_{u+1+m} = \varphi(h)) \]
\[ = P(Y_{u+m-k} = \varphi(l_k), k = 0, \ldots, m) = P(Y_{u+k} = l_k, k = 0, \ldots, m). \]
Then, an inductive argument is applied to get \( P(Y_{u+1+k} = l_k, k = 0, \ldots, m) = P(Y_k = l_k, k = 0, \ldots, m) \). Hence, the result follows. □
In a genomic framework $L = \{A, C, G, T\}$ and $\varphi : \{A, C, G, T\} \to \{A, C, G, T\}$ is the involution given by $\varphi(A) = T$, $\varphi(C) = G$. The CSPR is satisfied for the empirical probability measure on the DNA nucleotide sequence of bacterial genome. So, $d = 3$, $I = L^3$ is the list of codons and $P = P \circ \zeta^{-1}$ on $I^N$. The alphabet $A$ is the list of aminoacids and $\psi : L^3 \to A$ is the genetic code. Hence, a consequence of CSPR is that the probability distribution on the nucleotide sequences, and so on codon sequences, is stationary. If one accepts that CSPR is only valid for small $t$–mers of nucleotides with $t \approx 10$, then the weak stationarity property stated in Proposition 5.2 implies stationarity for 9–mers in the nucleotide sequence, which in the alphabet of codons means stationarity for triplets of codons.

For a fundamental explanation of CSPR, it is argued in [7] that it would be a probabilistic consequence of the reverse complementarity between paired strands, because symmetry of chemical energy implies Gibbs distribution is invariant by reverse complementarity which is exactly CSPR.

6 Random model

We will modify the model studied in Sections 2, 3 and 4 to approach some of the phenomena occurring in codon sequences of bacteria genome. Up to now a region of a new type starts when a state of a different class is hit. Nevertheless, it is known in genome annotation that when an intergenic region hits a start codon of a genic region only a small proportion of these start codons mark the beginning of a genic region. Some signals must be present in the neighborhood of the codon to trigger a beginning. Nowadays, there is an active research on this domain, either on the list of motifs and on the localization they must be with respect to the starting codons. A recent discussion on this topic can be found in [14].

So, at each time a site containing a state of a different class is hit a decision must be made: either a new region starts, or this beginning is postponed and continues to be governed by the symbol of the former region. We will model this decision by a random choice, in this purpose we use a sequence of independent random variables uniformly distributed in the unit interval. Our model admits that the decision depends on the hit state.

From now on we assume that each symbol $s \in S$ has a probability $\epsilon(s) \in (0, 1]$ of start governing a new region when it is hit by a region of type different from $C(s)$. Note that $\epsilon(s) > 0$ is a natural constraint, in fact in the contrary we could delete $s$ from $S$. The case $\epsilon(s) = 1$ means that it is sure that when a region of type $C$, with $C \neq C(s)$, encounters a site containing a symbol $s$ then a region governed by $P_s$ starts. The sample region for the random choice is the unit interval $R := [0, 1]$ which is endowed with the Borel $\sigma$–field $B(R)$ and the Lebesgue measure denoted by $|\cdot|$. So $|R|$ is the Lebesgue measure of a Borel set $R \subseteq R$. The product spaces
$R^N$ and $R^{N^*}$ are respectively endowed with the Borel product $\sigma$–fields noted respectively by $\mathcal{B}(R^N)$ and $\mathcal{B}(R^{N^*})$. By $e$ me denote a random variable uniformly distributed in $R$ and $P^e$ denotes this distribution. Let $\tilde{e} = (e_m : n \in N^*)$ be a sequence of independent random variables uniformly distributed in $R$. So, the distribution of $\tilde{e}$ in $R^{N^*}$ is the product measure, $P^\tilde{e} := P^e \otimes N^*$. Consider the projection $Z_n : R^N \to R$, $z \in R^N \to z_n \in R$ for $n \in N$. For $R'_1, \ldots, R'_m \in \mathcal{B}(R)$ we have

$$P^\tilde{e}(Z_k \in R'_k, k = 1, \ldots, m) = \prod_{k=1}^m |R'_k|.$$ 

Define the space $\mathcal{K} := I \times R$, whose points are pairs $(i, r) \in I \times R$.

To each $s \in S$ we associate a fixed interval $R_s \subseteq R$ with size $|R_s| = \epsilon(s)$.

We consider the following dynamics: if a region governed by $s$ encounters a pair $(s', r')$ then a new region starts if and only if $C(s') \neq C(s)$ and $r' \in R_{s'}$. In this new setting, the set of starting states is

$$\mathcal{V} = \bigcup_{s \in S} \{s\} \times R_s.$$ 

Hence $(S \times R) \setminus \mathcal{V} = \bigcup_{s \in S} \{s\} \times (R \setminus R_s)$ are the states having a starting symbol but with a value in the sample region that prevent it to start governing a new region.

The class $C(v)$ associated with $v = (r, s) \in \mathcal{V}$ is defined to be

$$C(v) = \bigcup_{s' \in C(s)} \{s'\} \times R_{s'}.$$ 

Hence $(C(s, r) = C(s', r')) \Leftrightarrow (C(s) = C(s'), r \in R_s, r' \in R_{s'})$.

For all $s \in S$ we define the conditional law $P^{\epsilon(s)}(\cdot \mid s)$ to be uniformly distributed on $R_s$, that is

$$P^{\epsilon(s)}(R' \mid s) = |R' \cap R_s|/|R_s|, \ R' \in \mathcal{B}(R).$$

Consider the product space $\mathcal{K}^N = (I \times R)^N = I^N \times R^N$. We set $X_n : \mathcal{K}^N \to \mathcal{K}$, $w \in \mathcal{K}^N \to X'(w) = w_n \in \mathcal{K}$ the projection onto the $n$–th component. Let $w_n = (x_n, r_n)$, we denote $X_n : \mathcal{K}^N \to I$, $w \to x_n$ and $Z_n : \mathcal{K}^N \to R$, $w \to r_n$. So, we can write $X_n = (X_n, Z_n)$. This is an abuse of notation with $X_n$ and $Z_n$, in fact we will also continue writing $X_n : I^N \to I$, $x \in I^N \to x_n \in I$ and $Z_n : R^N \to R$, $z \in R^N \to z_n \in R$. We keep the same notation for the shift $\Theta_q : \mathcal{K}^N \to \mathcal{K}^N$, $(\Theta_q w)_n = w_{n+q}$, as the one introduced in $[\mathbb{1}]$ for $I^N$ and use the same notation $\Theta_q$ for the shift in $R^N$.

As before we endow $\mathcal{K}^N$ with the $\sigma$–field $\mathcal{B}_X^\infty = \sigma(X_n : n \in N)$ and we denote $\mathcal{B}_X^N = \sigma(X_0, \ldots, X_n)$. Let $P$ be a probability measure on $(\mathcal{K}^N, \mathcal{B}_X^\infty)$. A random time $T'_\cdot : \mathcal{K}^N \to N \cup \{\infty\}$ is a stopping time with respect to the filtration
(\(B_n^X : n \in \mathbb{N}\)) when \(\{T' \leq n\} \in B_n^X\) is satisfied for all \(n \in \mathbb{N}\). The \(\sigma\)-field associated to a stopping time \(T'\) was already defined and denoted by \(B_n^T\).

Assume \(X_0 \in \mathcal{V}\). Then, the random time
\[
T := T_{\mathcal{V}\setminus C(X_0)} = \inf\{n > 0 : C(X_n) \neq C(X_0)\},
\]
is well-defined (it can take the value \(\infty\)) and it is a stopping time. As already done for a random time in (2), we define the sequence of times
\[
T^1 = T \text{ and for } n \in \mathbb{N}^*: \ T^{n+1} = T^n + \Theta T^n,
\]
which are also stopping times. We have \(T^1 = T\) and \(T^{n+1}\) finite implies \(C(X_{T^{n+1}}) \neq C(X_T)\).

Let \((P_s : s \in S)\) be a family of probability distribution on \(\mathcal{I}^S\) such that for all \(s \in S\), \(P_s(X_0 = s) = 1\) and satisfies condition (3). Each \(P_s\) defines the following probability measure \(P^1_s\) on \(\mathcal{K}^S\): for all \(m \in \mathbb{N}\), \((i_0, \ldots, i_m) \in \mathcal{I}^{m+1}\) and \(R'_0, \ldots, R'_m \in B(R)\),
\[
P^1_s(X_k = i_k, Z_k \in R'_k, k = 0, \ldots, m)
= 1_{\{i_0 = s\}} P^s_{R'_0}(s) P_s(X_k = i_k, k = 1, \ldots, m) P_s(Z_k \in R'_k, k = 1, \ldots, m).
\]

Then, the initial distribution of \(P^1_s\) is the uniform one on \(\{s\} \times R_s\). Note that
\[
\forall R' \in B(R), R' \supseteq R_s : \ P^1_s(X_0 = s, Z_0 \in R') = 1, \quad (52)
\]
in particular \(P^1_s(X_0 = s, Z_0 \in R) = P^1_s(X_0 = s) = 1\).

Since condition (3) ensures \(S \setminus C(s)\) is attained in finite time \(P_s\)-a.s., we apply the Borel-Cantelli Lemma to the independent random variables \((Z_n : n \in \mathbb{N}^+)\) to get
\[
\forall s \in S : \ P^1_s(\mathcal{T} < \infty) = 1. \quad (53)
\]
Let \(E^1_s\) be the expected value defined by \(P^1_s\).

The following definition will depend on a strictly positive vector vector \(\pi^1 = (\pi^1_s : s \in S)\).

**Definition 6.1.** For the family \((P^1_s : s \in S)\) and \(\pi^1 = (\pi^1_s : s \in S) > 0\) we define \(P^1\) on \(\mathcal{K}^N\) by:
\[
\forall B \in B_\infty^X : \ P^1(B) = \sum_{s \in S} \pi^1_s \left( \sum_{n \geq 0} E^1_s(1_{\mathcal{T} > n} 1_B \circ \Theta_n) \right) = \sum_{s \in S} \pi^1_s \left( \sum_{n \geq 0} P^1_s(\mathcal{T} > n, B \circ \Theta^{-1}_n) \right), \quad (54)
\]
where \(\Theta\) is the shift operator on \(\mathcal{K}^N\).
Obviously $\mathbb{P}^\dagger$ is a measure. For all $m \in \mathbb{N}^*$, $(i_0, \ldots, i_m) \in I^{m+1}$ and $R'_{i_0}, \ldots, R'_{i_m} \in \mathcal{B}(R)$ we have
\[
\mathbb{P}^\dagger(X_k = i_k, Z_k \in R'_{i_k}, k = 0, \ldots, m) = \sum_{s \in S} \pi^\dagger_s \left( \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n, X_{k+n} = i_l, Z_{k+n} \in R'_{i_k}, k = 0, \ldots, m) \right). \tag{55}
\]

**Theorem 6.2.** There exists some strictly positive vector $\pi^\dagger = (\pi^\dagger_s : s \in S)$ such that $\mathbb{P}^\dagger$ defined by (55) is a probability measure if and only if it is satisfied
\[
\forall s \in S : \ E_s^\dagger(T) < \infty. \tag{56}
\]
In this case, the condition on $\pi^\dagger$
\[
\sum_{s \in S} \pi^\dagger_s E_s^\dagger(T) = 1, \tag{57}
\]
is necessary and sufficient in order that $\mathbb{P}^\dagger$ is a probability measure on $I^\mathbb{N}$.

**Proof.** We must show that condition (57) is equivalent to $\mathbb{P}^\dagger(X_0 \in S, Z_0 \in R) = 1$. Let $s_0 \in S$. From (52) we get
\[
\mathbb{P}^\dagger(X_0 = s_0, Z_0 \in R) = \sum_{s \in S} \pi^\dagger_s \left( \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n + 1, X_{n+1} = s_0, Z_0 \in R) \right) + \pi^\dagger_{s_0}
\]
\[
= \sum_{s \in S} \pi^\dagger_s \left( \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n + 1, X_{n+1} = s_0) \right) + \pi^\dagger_{s_0}
\]
\[
= \sum_{s \in S} \pi^\dagger_s \left( \sum_{n=1}^{T} 1_{\{X_n = s_0\}} \right) + \pi^\dagger_{s_0}.
\]
Hence
\[
\mathbb{P}^\dagger(X_0 \in S, Z_0 \in R) = \sum_{s \in S} \pi^\dagger_s \left( \sum_{n=1}^{T} 1_{\{X_n \in s\}} \right) + \sum_{s \in S} \pi^\dagger_s. \tag{58}
\]
On the other hand
\[
\mathbb{P}^\dagger(X_0 \in \bar{S}, Z_0 \in R) = \sum_{s \in S} \pi^\dagger_s \left( \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n + 1, X_{n+1} \in \bar{S}, Z_{n+1} \in R) \right)
\]
\[
= \sum_{s \in S} \pi^\dagger_s \left( \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n + 1, X_{n+1} \in \bar{S}) \right).
\]
We have
\[
\mathbb{P}^\dagger_s(T > n + 1, X_{n+1} \in \bar{S}) \in \mathcal{B}(\mathbb{S} \times \mathbb{R}) \setminus \mathcal{V}
\]
an\d\ and
\[
\mathbb{P}^\dagger_s(T > n + 1, X_{n+1} \in (\mathcal{S} \times \mathcal{R}) \setminus \mathcal{V}) = \mathbb{E}_s^\dagger(1_{\{X_{n+1} \in (\mathcal{S} \times \mathcal{R}) \setminus \mathcal{V}, T > n + 1\}})
\]
\[
= \mathbb{E}_s^\dagger(1_{\{X_{n+1} \in \mathcal{S}, T > n + 1\}}).
\]

23
Then
\[ P^\dagger(\pi \in I \setminus S, Z_0 \in R) = \sum_{\tau \in S} \pi^\dagger(\sum_{n \geq 0} P^\dagger_s(\pi \in I_{X_n \in S} | s > n + 1, s < T > n + 1)) \]
\[ = \sum_{\tau \in S} \pi^\dagger(\sum_{n \geq 1} P^\dagger_s(\tau < \infty, \sum_{n=1}^{\tau} 1_{\{X_n \in S\}})). \]

We conclude
\[ P^\dagger(\pi \in I \setminus S, Z_0 \in R) = \sum_{\tau \in S} \pi^\dagger(\sum_{n \geq 0} P^\dagger_s(\tau < \infty, \sum_{n=1}^{\tau} 1_{\{X_n \in S\}})). \]

From (58) and (59) we find, from (61)
\[ \sum_{\tau \in S} \pi^\dagger(\sum_{n \geq 1} P^\dagger_s(\tau < \infty, \sum_{n=1}^{\tau} 1_{\{X_n \in S\}})) = \sum_{\tau \in S} \pi^\dagger. \]

Hence, condition (57) is necessary and sufficient in order that \( P^\dagger \) is a probability measure on \( \mathcal{K}^N \). So (59) is a necessary and sufficient condition in order that there exists some strictly positive vector \( \pi^\dagger \) fulfilling (57).

From now on we assume (56) always hold and that \( \pi^\dagger \) satisfies (57), so \( P^\dagger \) is a probability measure on \( \mathcal{K}^N \). Let \( \mathbb{E}^\dagger \) be its associated mean expected value.

Remark 6.3. From (67) and by using (57), we get formally
\[ P^\dagger(\tau < \infty) = \sum_{\tau \in S} \pi^\dagger(\sum_{n \geq 1} P^\dagger_s(\tau < \infty, \sum_{n=1}^{\tau} 1_{\{X_n \in S\}})) = \sum_{\tau \in S} \pi^\dagger(\tau < \infty, \sum_{n=1}^{\tau} 1_{\{X_n \in S\}}) = 1. \]

Similar comments to those of Remark 2.3 can be made. □

Let us define the family of probability measures \( \{ P_{s}^\dagger : s \in S \} \) on \( \mathcal{K}^N \) by regeneration at \( \tau = T_{\mathcal{K} \setminus \mathcal{C}(\mathbb{k})} \). In this purpose for \( \mathbb{k} = (\kappa_l : l = 0, \ldots, m) \in \mathcal{V} \times \mathcal{K}^m \), define the sequence of indexes \( (\tau_n(\mathbb{k}) : n \geq 0) \) by
\[ \tau_0(\mathbb{k}) = 0 \quad \text{and} \quad \forall n \geq 1 : \tau_n(\mathbb{k}) = \inf\{ l > \tau_{n-1} : \kappa_l \in \mathcal{V} \setminus \mathcal{C}(\kappa_{\tau_{n-1}(\mathbb{k})}) \} \].

Let \( \chi(\mathbb{k}) = \sup\{ k \geq 0 : \kappa_k(\mathbf{x}) < \infty \} \). From definition,
\[ \forall n \in \{ 1, \ldots, \chi(\mathbb{k}) \} : \mathcal{C}(\kappa_{\tau_n(\mathbb{k})}) \neq \mathcal{C}(\kappa_{\tau_{n-1}(\mathbb{k})}) \].

Let us simply note \( \tau_{\chi(\mathbb{k})}(\mathbf{x}) \) by \( \tau_{\chi(\mathbb{k})} \), in \( \chi(\mathbb{k}) \) we keep the dependence on \( \mathbb{k} \). Let us define \( P_{s}^\dagger \). For \( m \in \mathbb{N}^*, (i_0, \ldots, i_m) \in I^{m+1} \) and \( R_0', \ldots, R_m' \in \mathcal{B}(R) \) we put
\[ P_{s}^\dagger(X_i = i_l, Z_l \in R'_l, l = 0, \ldots, m) \]
\[ = 1_{\{i_0 = s\}} P^\dagger (R_0' | s) \prod_{k=0}^{\chi(\mathbb{k})-1} P^\dagger (X_i = i_{\tau_k+l}, Z_l \in R'_{\tau_k+l}, l = 1, \ldots, \tau_k+1 - \tau_k) \]
\[ \times P_{i_{\chi(\mathbb{k})}}^\dagger (X_i = i_{\tau_{\chi(\mathbb{k})}+l}, Z_l \in R'_{\tau_{\chi(\mathbb{k})}+l}, l = 1, \ldots, m - \tau_{\chi(\mathbb{k})}). \]
An inductive argument on \( \chi(s) \) shows that \( P_s^\dagger \) is well-defined by (61). From definition, \( P_s^\dagger(X_0 = s) = 1 \) for all \( s \in S \).

Let us fix \( T = T_{Y \cap C(X_0)} \). From definition \( (61) \), we can apply Borel-Cantelli lemma to get
\[
\forall s \in S, \forall n \in \mathbb{N}^* : \quad P_s^\dagger(T^n < \infty) = 1.
\]

(62)

We denote by \( E_s^\dagger \) the mean expected value associated with \( P_s^\dagger \). Note that \( P_s^\dagger(B \cap \{T \leq n\}) = P_s^\dagger(B \cap \{T < n\}) \) for all \( B \in \mathcal{B}_T \) and \( n \in \mathbb{N} \). In particular \( P_s^\dagger(T > n) = P_s^\dagger(T > n) \), so \( E_s^\dagger(T) = E_s^\dagger \).

Similarly to Proposition 3.2 we can state the regeneration property.

**Proposition 6.4.** For all probability vector \( \gamma = (\gamma_s : s \in S) \) and all \( n \in \mathbb{N}^* \) the distribution \( P_s^\dagger = \sum_{s \in S} \gamma_s P_s^\dagger \) regenerates at \( T^n \). In particular for all \( s \in S \), \( P_s^\dagger \) regenerates at \( T \).

**Proof.** It suffices to show the statement for \( P_s^\dagger = P_s^\dagger \), that is for an extremal vector \( \gamma \). Also by an inductive argument it suffices to prove the result for \( n = 1 \), that is for \( T^1 = T \). Since \( P_s^\dagger(T < \infty) = 1 \), we must show the following equality for \( j_k \in I, R_k \in \mathcal{B}(R), k = 1, \ldots, q \):
\[
E_s^\dagger(1_{\{X_k+T = j_k, Z_k+T \in R_k^\dagger; k = 1, \ldots, q\}} \mid B_T^X) = E_{X_T^\dagger}(1_{\{X_k = j_k, Z_k \in R_k^\dagger; k = 1, \ldots, q\}}) P_s^\dagger - a.s.
\]
(63)

Let \( i_l, j_k \in I, R_k^\dagger \in \mathcal{B}(R), l = 0, \ldots, m, \) be such that \( i_0 = s, i_l \in C(s) \) for \( l = 1, \ldots, m - 1 \) and \( i_m \notin C(s) \); and \( R_0^\dagger \subseteq R_s, R_m^\dagger \subseteq R_m \). Let \( B_m = \{T = m, X_l = i_l, Z_l \in R^\dagger_l, l = 0, \ldots, m\} \). Then, (63) will be shown once we prove the equality
\[
\int_{B_m} 1_{\{X_k+T = j_k, Z_k+T \in R_k^\dagger; k = 1, \ldots, q\}} dP_s^\dagger = \int_{B_m} P_{i_m}^\dagger(X_k = j_k, Z_k \in R_k^\dagger; k = 1, \ldots, q) dP_s^\dagger.
\]

This follows from a recurrence argument on (61). \( \square \)

We define \( P^\dagger \) for the family \( (P_s^\dagger : s \in S) \) simply by putting \( E_s^\dagger \) instead of \( E_s^\dagger \) in Definition 6.1. Since \( E_s^\dagger(T) = E_s^\dagger \) the condition (62) supplied by Theorem 6.2 in order that \( P^\dagger \) is a probability measure is the same as for \( P^\dagger \), that is the vector \( \pi^\dagger \) must satisfy (57). Let \( E^\dagger \) be the mean expected value associated with \( P^\dagger \).

Assume \( X_0 \in \mathcal{V} \). Let \( \gamma = (\gamma_s : s \in S) \) be a probability vector on \( S \) and let \( S^n \) be endowed with the distribution \( P^\dagger = \sum_{s \in S} \gamma_s P_s^\dagger \), so \( P^\dagger(X_0 \in S) = 1 \). By (62) the times \( \{T^n : n \in \mathbb{N}^*\} \) are finite \( P^\dagger \)-a.s.. Define the sequence \( (\Xi_n : n \geq 0) \) by \( \Xi_0 = X_0 \) and \( \Xi_n = X_{T^n} \) for \( n \geq 1 \). Proposition 6.4 implies that the sequence \( (\Xi_n : n \in \mathbb{N}) \) is a Markov chain. The transition matrix \( Q = (q_{ss'} : s, s' \in S) \) of this chain is given by
\[
\forall s, s' \in S : \quad q_{ss'}^\dagger = P_s^\dagger(X_T = s').
\]
By definition of \( T \) we have \( C(\Xi_k+1) \neq C(\Xi_k) \), so \( q_{ss'}^\dagger > 0 \) implies \( C(s') \neq C(s) \).

A positive vector \( (\rho = (\rho_s : s \in S) \) is invariant for \( Q^\dagger \) if it verifies the set of equalities
\[
\forall s \in S : \quad \rho_s = \sum_{s' \in S} \rho_{s'} q_{s's}^\dagger.
\]

There always exist invariant positive vectors. Moreover, if \( Q^\dagger \) is irreducible the invariant positive vectors are unique up to a multiplicative constant. In a similar way as we did in Theorem 3.3 we can state the following condition for stationarity of \( \mathbb{P}^\dagger \).

**Theorem 6.5.** Assume that the strictly positive vector \( \pi^\dagger = (\pi_s^\dagger : s \in S) \) satisfies the condition \( \text{(57)} \). Then, \( \mathbb{P}^\dagger \) is stationary if and only if \( \pi^\dagger \) is an invariant vector for \( Q^\dagger \), that is it satisfies
\[
\forall s \in S : \quad \pi_s^\dagger = \sum_{s' \in S} \pi_s q_{s's}^\dagger \quad \text{where} \quad q_{s's}^\dagger = \mathbb{P}^\dagger_s(X_T = s'). \tag{64}
\]

**Proof.** We have that \( \mathbb{P}^\dagger \) is stationary if for all \( m \in \mathbb{N} \) and all \( i \in I, R'_i \in \mathcal{B}(R), l = 0, \ldots, m \), it is satisfied
\[
\mathbb{P}^\dagger(X_i = i_t, Z_l \in R'_l, l = 0, \ldots, m) = \mathbb{P}^\dagger(X_{i+1} = i_t, Z_{l+1} \in R'_l, l = 0, \ldots, m). \tag{65}
\]

From now on we denote
\[
G = \{X_i = i_t, Z_l \in R'_l, l = 0, \ldots, m\}
\]
and the \( n \)–shifted set
\[
G \circ \Theta_n^{-1} = \{X_{i+n} = i_t, Z_{l+n} \in R'_l, l = 0, \ldots, m\}.
\]

So, the relation \( \text{(65)} \) that we want to show is
\[
\mathbb{P}^\dagger(G) = \mathbb{P}^\dagger(G \circ \Theta_1^{-1}).
\]

We have
\[
\mathbb{P}^\dagger(G \circ \Theta_1^{-1}) = \mathbb{P}^\dagger(X_0 \in I \setminus S, G \circ \Theta_1^{-1}) + \mathbb{P}^\dagger(X_0 \in S, Z_0 \not\in R_{X_0}, G \circ \Theta_1^{-1}) + \mathbb{P}^\dagger(X_0 \in S, Z_0 \in R_{X_0}, G \circ \Theta_1^{-1}).
\]

Now
\[
\mathbb{P}^\dagger(X_0 \in I \setminus S, G \circ \Theta_1^{-1}) = \sum_{s \in S} \pi_s^\dagger \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n+1, X_{n+1} \in I \setminus S, G \circ \Theta_n^{-1})
\]
and
\[
\mathbb{P}^\dagger(X_0 \in S, Z_0 \not\in R_{X_0}, G \circ \Theta_1^{-1}) = \sum_{s \in S} \pi_s^\dagger \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n+1, X_{n+1} \in S, Z_{n+1} \not\in R_{X_{n+1}}, G \circ \Theta_n^{-1})
\]
\[
= \sum_{s \in S} \pi_s^\dagger \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n+1, X_{n+1} \in C(s), Z_{n+1} \not\in R_{X_{n+1}}, G \circ \Theta_n^{-1})
\]
\[
+ \sum_{s \in S} \pi_s^\dagger \sum_{n \geq 0} \mathbb{P}^\dagger_s(T > n+1, X_{n+1} \in S \setminus C(s), Z_{n+1} \not\in R_{X_{n+1}}, G \circ \Theta_n^{-1}).
\]

26
Now, by using $P^\dagger_s(X_0 = s, Z_0 \in R_X) = 1$ we find
\[
P^\dagger(X_0 \in S, Z_0 \in R_X, G \circ \Theta_1^{-1})
= \sum_{s \in S} \pi^\dagger_s (\sum_{n \geq 0} \sum_{m \geq 0} P^\dagger_s(T > n + 1, X_{n+1} \in C(s), Z_{n+1} \in R_{X_{n+1}}, G \circ \Theta_{n+2}^{-1}))
+ \sum_{s \in S} \pi^\dagger_s (\sum_{n \geq 0} P^\dagger_s(T = n + 1, X_{n+1} \in S \setminus C(s), Z_{n+1} \in R_{X_{n+1}}, G \circ \Theta_{n+2}^{-1}))
+ \sum_{s \in S} \pi^\dagger_s (G \circ \Theta_1^{-1}).
\]

On the other hand we have
\[
P^\dagger_s(T > n + 1, G \circ \Theta_{n+2}^{-1})
= P^\dagger_s(T > n + 1, X_{n+1} \in I \setminus S, G \circ \Theta_{n+2}^{-1}) + P^\dagger_s(T > n + 1, X_{n+1} \in C(s), G \circ \Theta_{n+2}^{-1})
+ P^\dagger_s(T > n + 1, X_{n+1} \in S \setminus C(s), Z_{n+1} \notin R_{X_{n+1}}, G \circ \Theta_{n+2}^{-1}).
\]

We put the previous elements together to get
\[
P^\dagger(G \circ \Theta_1^{-1}) = \sum_{s \in S} \pi^\dagger_s (\sum_{n \geq 0} P^\dagger_s(T > n + 1, G \circ \Theta_{n+2}^{-1}))
+ \sum_{s \in S} \pi^\dagger_s (\sum_{n \geq 0} P^\dagger_s(T = n + 1, X_{n+1} \in S \setminus C(s), Z_{n+1} \in R_{X_{n+1}}, G \circ \Theta_{n+2}^{-1})).
\]

Hence
\[
P^\dagger(G \circ \Theta_1^{-1}) = \sum_{s \in S} \pi^\dagger_s (\sum_{n \geq 0} P^\dagger_s(T > n + 1, G \circ \Theta_{n+1}^{-1}))
+ \sum_{s \in S} \pi^\dagger_s (\sum_{n \geq 0} P^\dagger_s(T = n + 1, G \circ \Theta_{n+1}^{-1})). \tag{66}
\]

Recall $\Theta_1$, $P^\dagger(G) = \sum_{s \in S} \pi^\dagger_s (\sum_{n \geq 0} P^\dagger_s(T \geq n, G \circ \Theta_n^{-1}))$,

and $G = \{X_l = i_l, Z_l \in R'_l, l = 0, ..., m\}$. In both cases: $i_0 \in I \setminus S$, or $i_0 \in S$ and $R'_0 \subseteq R \setminus R_{i_0}$; we get $P^\dagger(G) = 0$ and $P^\dagger(T = n + 1, G \circ \Theta_{n+1}^{-1}) = 0$ for all $n \geq 0$.

Then, in these cases, the stationarity property $P^\dagger(G) = P^\dagger(G \circ \Theta_1^{-1})$, is a straightforward consequence of formulae (66) and (71).

We are left to study the case $i_0 \in S$ and $R'_0 \subseteq R_{i_0}$. In this case
\[
\sum_{s \in S} \pi^\dagger_s (G) = \pi^\dagger_{i_0} \frac{|R'_0|}{|R_{i_0}|} P^\dagger(X_l = i_l, Z_l \in R'_l, l = 1, ..., m).
\]
On the other hand, since $P_{s}^{*\dagger}$ is defined by using the regeneration property, we get

$$
\sum_{s \in S} \left( \sum_{n \geq 0} P_{s}^{*\dagger} \left( T = n+1, G \circ \Theta_{n+1}^{-1} \right) \right) \\
= \left( \sum_{n \geq 0} P_{s}^{*\dagger} \left( T = n+1, X_T = i_0, R_T \in R_0 \right) \right) \cdot P_{i_0}^{*\dagger} \left( X_{i} = i_l, Z_i \in R_i, l = 1, \ldots, m \right).
$$

We have

$$
\sum_{s \in S} \pi_s^{\dagger} \left( \sum_{n \geq 0} P_{s}^{*\dagger} \left( T = n+1, X_T = i_0, R_T \in R_0 \right) \right) = \frac{|R'_0|}{|R_{i_0}|} \left( \sum_{s \in S} \pi_s^{\dagger} P_{s}^{*\dagger} \left( X_T = i_0 \right) \right).
$$

Therefore, from (66) and (54) we obtain the equivalence

$$
\left( \mathbb{P}^{*}(G) = \mathbb{P}^{*}(G \circ \Theta_{n}^{-1}) \right) \iff \left( \forall i_0 \in S : \pi_{i_0}^{\dagger} = \sum_{s \in S} \pi_s^{\dagger} P_{s}^{*\dagger} \left( X_T = i_0 \right) \right).
$$

We have proven that $\mathbb{P}^*$ is stationary if and only if the following condition is satisfied

$$
\forall s' \in S : \pi_{s'}^{\dagger} = \sum_{s \in S} \pi_s^{\dagger} P_{s}^{*\dagger} \left( X_T = s' \right) = \sum_{s \in S} \pi_s^{\dagger} q_{ss'}.
$$

This shows the theorem.

Similarly as in we did in (20), when $\mathbb{P}^*$ is stationary we can extend it to the set $\mathbb{K}_s^Z$ by putting

$$
\mathbb{P}^*(X_{i+k} = i_k, Z_{i+k} \in R'_{i_k}, k = 0, \ldots, m) = \mathbb{P}^*(X_k = i_k, Z_k \in R'_{i_k}, k = 0, \ldots, m)
$$

for all $l \in \mathbb{Z}$, $m \geq 0$; $i_k \in I$, $R'_{i_k} \in \mathcal{B}(R)$ for $k = 0, \ldots, m$.

Now we state the equivalent of Theorem 4.1 in Section 4. Define the probability vector

$$
\hat{\pi}^{\dagger} = (\hat{\pi}_s^{\dagger} : s \in S) \text{ with } \hat{\pi}_s^{\dagger} = \pi_s^{\dagger} \left( \sum_{s' \in S} \pi_{s'}^{\dagger} \right)^{-1}.
$$

Consider the distribution $P_{\hat{\pi}^{\dagger}} = \sum_{s \in S} \hat{\pi}_s^{\dagger} P_{s}^{*\dagger}$ on $\mathbb{B}_s^X$ and let $E_{\hat{\pi}^{\dagger}}$ be its mean expected value. From (62) we have $P_{\hat{\pi}^{\dagger}}(T^n < \infty) = 1$ for all $n \in \mathbb{N}^*$, where $T = T_{V \setminus \mathcal{C}(X_0)}$. By condition (61) we also find

$$
E_{\hat{\pi}^{\dagger}} \left( T \right) = \left( \sum_{s \in S} \pi_s^{\dagger} \right)^{-1} \left( \sum_{s \in S} \pi_s^{\dagger} E_{s}^{\dagger} \left( T \right) \right) = \left( \sum_{s \in S} \pi_s^{\dagger} \right)^{-1}.
$$

Let $P_{\hat{\pi}^{\dagger}}$ be given by $P_{\hat{\pi}^{\dagger}} = \sum_{s \in S} \hat{\pi}_s P_{s}^{*\dagger}$ on $\mathbb{B}_s^X$ and let $E_{\hat{\pi}^{\dagger}}$ be its mean expected value. By previous relations,

$$
\forall n \in \mathbb{N}^* \quad P_{\hat{\pi}^{\dagger}}(T < \infty) = 1 \quad \text{and} \quad E_{\hat{\pi}^{\dagger}}(T)^{-1} = \sum_{s \in S} \pi_s^{\dagger}.
$$

28
The following result is proven in a similar way as we did for Theorem 4.1. In fact, it is a corollary of Theorem 6.5 because this last result allows us to construct the process $\mathcal{X}$ with distribution $\mathbb{P}_{\hat{\pi}}$ with independent copies between the sequence of hitting times of different classes. Since the increments of this sequence of times are independent identically distributed variables and its distribution has a finite mean the renewal theorem can be applied as in the proof of Theorem 4.1 and also the other arguments in this proof work in an entirely analogous way. Therefore we can state:

**Theorem 6.6.** Assume that $\pi^*$ satisfies (67) and (64) and that the distribution of $\mathcal{T}$ under $\mathbb{P}_{\hat{\pi}}^*$ is aperiodic. Then,

$$\forall B \in \mathcal{B}_\infty^X : \quad \mathbb{P}_{\hat{\pi}}^*(B) = \lim_{N \to \infty} \mathbb{P}_{\pi^*}^*(B) \circ \Theta_N^{-1}.$$  

Moreover, if in addition the matrix $Q^*$ is aperiodic then for all probability vector $\gamma = (\gamma_s : s \in \mathcal{S})$ the probability measure $\mathbb{P}_{\gamma}^* = \sum_{s \in \mathcal{S}} \gamma_s \mathbb{P}_{s}^*$ satisfies

$$\forall B \in \mathcal{B}_\infty^X : \quad \mathbb{P}_{\gamma}^*(B) = \lim_{N \to \infty} \mathbb{P}_{\gamma}^*(B) \circ \Theta_N^{-1}.$$  

□

Let $\mathbb{P}_{\gamma}^*$ the law defined on $\mathcal{K}_\mathbb{Z}$ in (67). We can also show that for all $k \geq 0$, $i_k \in I$, $R_k^* \in \mathcal{B}(R)$, $k = 0, \ldots, m$, we have

$$\mathbb{P}_{\gamma}^*(X_i + k = i_k, Z_{i+k} \in R_k^*, k = 0, \ldots, m) = \lim_{N \to \infty} \mathbb{P}_{\gamma}^*(X_i + k + N = i_k, Z_i + k + N \in R_k^*, k = 0, \ldots, m).$$

Therefore, $\mathcal{T}_0 = \sup \{ \mathcal{T}^n : \mathcal{T} \leq 0 \}$ is finite $\mathbb{P}_{\gamma}^*$-a.s. and $X_{\mathcal{T}_0}$ has distribution $\hat{\pi}$. Then,

$$\mathbb{P}_{\gamma}^*(X_i = i_k, Z_k \in R_k^*, k = 0, \ldots, m) = \sum_{n \in \mathbb{N}} \mathbb{P}_{\gamma}^*(\mathcal{T}^0 = -n, X_0 = i_k, k = 0, \ldots, m)$$

$$= \sum_{n \in \mathbb{N}} \mathbb{P}_{\gamma}^*(X_0 = i_k, Z_k \in R_k^*, k = 0, \ldots, m | \mathcal{T}^0 = -n) \mathbb{P}_{\gamma}^*(\mathcal{T}^0 = -n).$$

(68)

Since $\mathbb{P}_{\gamma}^*$ regenerates at each $\mathcal{T}^n$ with law $\hat{\pi}$, we have

$$\mathbb{P}_{\gamma}^*(X_0 = i_k, k = 0, \ldots, m | \mathcal{T}^0 = -n) = \mathbb{P}_{\gamma}^*(X_0 = i_k, k = 0, \ldots, m | \mathcal{T} > n).$$

Similarly to (49) we have $\mathbb{P}_{\gamma}^*(\mathcal{T}^0 = -n) = (\sum_{s \in \mathcal{S}} \pi_s) \mathbb{P}_{\gamma}^*(\mathcal{T} > n)$. Thus, we retrieve the definition in (49),

$$\mathbb{P}_{\gamma}^*(X_0 = i_k, k = 0, \ldots, m) = \sum_{s \in \mathcal{S}} \sum_{n \in \mathbb{N}} \pi_{\gamma} \mathbb{P}_{s}^*(X_i = i_k + n, k = 0, \ldots, m; \mathcal{T} > n).$$

Hence, from (68) we get a probabilistic insight to definition $\mathbb{P}_{\gamma}^*$ and a good definition of $\mathcal{T}$ under law $\mathbb{P}_{\gamma}^*$, as claimed in Remark 6.3.

**Remark 6.7.** We have found conditions in order that $\mathbb{P}^*$ or $\mathbb{P}_{\gamma}^*$ are stationary laws. The ergodic description of these measures is part of an on-going study of the author.
Acknowledgments

The author thanks the Center for Mathematical Modeling (CMM) Basal CONICYT Program PFB 03 and INRIA-CHILE program CIRIC for supporting this work. He is indebted to Andrew Hart for fruitful discussions.

References

[1] S. Asmussen. Applied probability and queues. Second Edition. Collection Application of Mathematics Springer-Verlag, New York (2003), 438 pp.

[2] S. Asmussen, R. Biard. Ruin probabilities for a regenerative Poisson gap generated risk process. Eur. Actuar. J. (2011), 1, 1, 3-22.

[3] N. Bouaynaya, D. Schonfeld. Non-stationary Analysis of Coding and Non-coding Regions in Nucleotide Sequences. IEEE J. of Selected Topics in Signal Processing (2008), 2, 3, 357-364.

[4] N. Bouaynaya, D. Schonfeld. Emergence of New Structure from Non-Stationary Analysis of Genomic Sequences. IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS’08), Phoenix, June 2008, 1-4.

[5] A. Hart, S. Martínez. Markovianness and Conditional Independence in Annotated Bacterial DNA (2013). arXiv:1311.4411 [q-bio.QM].

[6] A. Hart, S. Martínez. Statistical testing of Chargaff’s second parity rule in bacterial genome sequences. Stoch. Models 27 (2011), 2, 272-317.

[7] A. Hart, S. Martínez, F. Olmos. A Gibbs approach to Chargaff’s second parity rule. Journal Statistical Physics 146 (2012), 2, 408-422.

[8] T. Lindvall. Lectures on the coupling method. Wiley, New-York (1992), 257 pp.

[9] S.P. Meyn, R.L. Tweedie. Markov chains and stochastic stability. Springer-Verlag, New-York (1993), 550 pp.

[10] D. Mitchell, R. Bridge. A test of Chargaff’s second rule. Biochem. Biophys. Res. Commun. (2006), 340, 1, 90-94.

[11] P. Nicolas, L. Bize, F. Muri, M. Hoebeke, F. Rodolphe, S. D. Ehrlich, B. Prum, P. Bessires. Mining Bacillus subtilis chromosome heterogeneities using Hidden Markov Models. Nucleic Acids Res. (2002), 30, 14181426.
[12] V.V. Prabhu. Symmetry observations in long nucleotide sequences. Nucleic Acids Res. (1993), 21, 12, 2797-2800.

[13] B. Prum. Markov models and hidden Markov models in genome analysis. Short course 6th Brazilian School of Probability (2002).

[14] E. Richardson, M. Watson. The automatic annotation of bacterial genomes. Briefings in bioinformatics (2013) 14, 1, p. 1-12.

[15] S. Ross. Stochastic processes. John Wiley & Sons (1983), 309 pp.

[16] R. Rudner, J.D. Karkas, E. Chargaff. Separation of B. subtilis DNA into complementary strands. III. Direct analysis. Proc. Natl. Acad. Sci. USA (1968), 60, 921-922.

[17] W.L. Smith. Regenerative Stochastic Processes. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences (1955), 232, 1188, 631.

[18] S.H. Zhang, Y.Z. Huang. Limited contribution of stem-loop potential to symmetry of single-stranded genomic dna. Bioinformatics (2010) 26, 4, 478-485.

[19] C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H. E. Stanley. Long-range correlations in nucleotide sequences. Nature (1992) 356, 6365, 168170.

[20] W. Li, K. Kaneko. Long-range correlation and partial 1/f spectrum in a noncoding DNA sequence. Europhysics Letters (1992) 17, pp. 655.

[21] B-J. Yoon. Hidden Markov Models and their Applications in Biological Sequence Analysis. Current Genomics (2009) 10, 6, 402415.

SERVET MARTÍNEZ
Departamento Ingeniería Matemática and Centro Modelamiento Matemático, Universidad de Chile, UMI 2807 CNRS, Casilla 170-3, Correo 3, Santiago, Chile. e-mail: smartine@dim.uchile.cl