SSD draft: Selection and Agafonov's theorem

HLMS

January 24, 2019

Abstract

This note is an expository discussion of Agafonov's theorem.

1 Equivalent Definitions of Normality

We first recall the notion of normality due to Borel [3].

Let Σ be a finite alphabet with $|\Sigma| \geq 2$. Let α be a probability measure on Σ. We can extend α to a probability measure α_1 on product space Σ^* by by $\alpha_1(u) := \prod_{i=0}^{|u|-1} \alpha(u[i])$. Now consider infinite sequences in Σ^∞. We denote $C_u = \{z \in \Sigma^\infty \mid$ where u is a prefix of $z\}$ and called it a *cylinder*. Again we extend α to a set function from cylinders to $[0, 1]$, defined by

$$\overline{\alpha}_2(C_u) := \alpha_1(u)$$

and then further extend $\overline{\alpha}_2$ to the *Borel field* generated by all cylinders, and to the *augmented field* when consider extra sets with measure zero. For more details about product (probability) measures, one can refer to Billingsley [2].

For convenience' sake, we abuse the notation α to denote α, α_1, α_2 and their further extensions.

Let $A \subseteq \Sigma^*$ and $B \subseteq \Sigma^* \cup \Sigma^\infty$, and $\chi_A : B \to \{0, 1\}$ be the characteristic function (indicator function) of $A \subseteq B$. For $u \in \Sigma^*$, the notation $\Sigma^* u$ means the set of string ends with u. Specially, when $b \in \Sigma$, we use $\Sigma^* b$ to denote the set of string ends with a symbol b. We then have

Definition. (*Simply Normal*) An infinite sequence $z \in \Sigma^\infty$ is simply normal if for all $b \in \Sigma$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_{\Sigma^* b}(z \upharpoonright i) = \alpha(b). \tag{1}$$

Definition. (*Borel Normality: sliding definition*) We say $z \in \Sigma^\infty$ is α-normal if for all $u \in \Sigma^*$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_{\Sigma^* u}(z \upharpoonright i) = \alpha(u). \tag{2}$$

Definition. (*Borel Normality: blocking definition*) A sequence $z \in \Sigma^\infty$ is α-normal if for all $u \in \Sigma^*$ such that $|u| = k$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_{\Sigma^* u}(z \upharpoonright i \cdot k) = \alpha(u). \tag{3}$$

Note that in this definition, a prefix of z of with length nk is divided into blocks of length k. Since these blocks are not overlapping to each other, it is more convenient to count positions that satisfy some desired property by using blocks. We'll see that in the proof of the main lemma (Lemma 3.4).

Another way to see Definition 1 is to treat u as a single symbol in the new alphabet Σ^k. Definition 1 then says a sequence is normal if it is simply normal to bases $|\Sigma|$, $|\Sigma|^2$, $|\Sigma|^3 \cdots$. These two definitions of normal numbers are equivalent.

Theorem 1.1. *(*Niven [7]*, Theorem 8.13) The above sliding definition and blocking definition of normal numbers are equivalent. Or formally, a sequence is normal as in Definition 1 if and only if it is simply normal to bases $|\Sigma|$, $|\Sigma|^2$, $|\Sigma|^3 \cdots$. That is, $z \in \Sigma^\infty$ is normal if for all $k \in \mathbb{N} \setminus \{0\}$ and for all $u \in \Sigma^k$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_{\Sigma^* u}(z \upharpoonright i \cdot k) = \alpha(u).$$

Proof. Todo: adapt Niven's proof to make it work for non-uniform distributions. His proof is just counting arguments. Should be straight forward to get an version for non-uniform distributions. □

2 Strategies, Selection Rules and Collectives

In the early stage of the development of probability theory, Richard von Mises attempts to build it on the notion of a *collective* (see also in [10, 1]). His idea goes like the following.

We can view the alphabet Σ as outcomes of a trial of some experiment. We call an infinite sequence of trials of this experiment a *game of chance*. That is, our sample space is Σ^∞. A player bets on the outcome of the trials can have his own *system* or *strategy* to select a subsequence of trials he'd like to bet. Mises considers a infinite sequence $z \in \Sigma^\infty$ a *collective* if

I. The asymptotic frequency of z occurrences of any $b \in \Sigma$ is $\alpha(b)$.

II. Property I persists for any subsequence of out comes derived from the collective by place selection rule.

However, Church in [4] showed collectives do not exist; collectives can exist if we only allow *admissible* selecting strategies.

2.1 Strategies

First, we need to define strategies formally. We shall restrict ourselves to non-anticipatory strategies, since, after all, no one really has a crystal ball.

Definition. (*Strategy [1, 5]*) A strategy is a function $\tau : \Sigma^* \to \{0, 1\}$. That is, a strategy is a *predicate* defined on all finite string.

Notation. Given a strategy τ and $x = x_0x_1x_2 \cdots \in \Sigma^* \cup \Sigma^\infty$, we denote

$$\tau_0(u) = \tau(\lambda),\ \tau_1(x) = \tau(x_0),\ \tau_2(x) = \tau(x_0x_1),\ \cdots, \tau_n = \tau(x \upharpoonright n), \cdots$$

We can view the function τ_i as an indicator to select i-th bit of x or not.

We denote $\theta(n, x)$ be the n-th value of j for which $\tau_j(x_0, \cdots, x_{j-1}) = 1$. That is, $\theta(n, x)$ is the index of the n-th digit being selected by strategy τ.

Definition. (Transformation) Let τ be a strategy and $z = z_0z_1 \cdots \in \Sigma^\infty$. Denote z' the sequence with $z'_n = z_{\theta(n,z)}$ for all $n \in \mathbb{N}$. That is, z' is the subsequence selected by τ out of z. We say $T : \Sigma^\infty \to \Sigma^\infty$ defined by $T(z) = z'$ the *transformation* induced by τ. In the case $z' \notin \Sigma^\infty$, we say $T(z)$ is *undefined*.

The following theorem says that non-anticipatory strategies are measure-preserving. It plays an important role in proving Agafonov's theorem. We will revisit it later.

Theorem 2.1. *(Doob [5]) Let τ be a strategy and T the transformation induced by τ. If T is defined almost everywhere on Σ^∞, then T is measure-preserving. That is, for all measurable set $M \subseteq \Sigma^\infty$,*

$$\alpha(T^{-1}(M)) = \alpha(M).$$

Proof. Todo: maybe rewrite Doob's proof in our notation? Note that Doob's proof already takes care of non-uniform distributions. □

2.2 Selection Rules

In this subsection, we will discuss a class of strategies "generated" by the so-called selection rules.

Definition. (Selection Rules) A selection rule S is a set of finite strings, i.e., $S \in \Sigma^*$. A strategy τ is say to be *generated* by a selection rule S, if for all $u = u_0 \cdots u_{n-1} \in \Sigma^*$

$$\tau_n(u) = 1 \Leftrightarrow u \in S, \text{ where } \tau_n \text{ is defined as in Notation 2.1.}$$

We can use the notation τ_S to make it explicit that it is a strategy generated by S. Similarly, we use T_S for the translations induced by τ_S.

We now focus on two classes of specific selection rules:

1. Strings end with a fixed string/block $\lambda \neq u \in \Sigma^*$, i.e.,

$$S_u = \Sigma^* u,$$

and we denote the set of strategies

$$\mathcal{F}_b := \{\text{all strategies generate by } S_u, \text{ with } u \text{ ranging over } \Sigma^* \setminus \{\lambda\}\}.$$

2. Regular languages and strategies generated by them. That is,

$$\mathcal{F}_r := \{\text{all strategies generated by regular languages.}\}$$

We will also say $\mathcal{F}_r$ are strategies *generated by finite automata*.

It clear that the set $\Sigma^* u$ is a regular language. Therefore,

Observation 2.2. $\mathcal{F}_b \subseteq \mathcal{F}_r$.

2.3 Collectives

We care about collectives that are related to some concrete strategies, especially strategies in $\mathcal{F}_b$ and $\mathcal{F}_r$. We first define this notion formally.

Definition. (Collectives relative to a strategy) Let α be a probability measure on Σ, τ be a strategy, and $z \in \Sigma^\infty$, we say z is a *collective relative to* τ if

I. The asymptotic frequency of z occurrences of any $b \in \Sigma$ is $\alpha(b)$. In short, z is simply normal.

II. Property I persists for $T(z)$, where T is the transformation induced by τ. That is, $T(z)$ is also simply normal.

We will also say z is a collective relative to a *set* of strategies $\mathcal{F}$, if z is a collective relative to every $\tau \in \mathcal{F}$. We now consider selecting a subsequence out of a normal number, based on the preceding n symbols, for some fix n. We have

Theorem 2.3. *(Postnikova [8]) A number is normal if and only if it is a collective relative to $\mathcal{F}_b$.*

By Observation 2.2, the following implication is immediately.

Corollary 2.4. *If a number is a collective relative to $\mathcal{F}_r$, then it is normal.*

This is the easy direction of Agafonov's theorem. To proof the converse, we need to develop a helper lemma in the next section.

3 A Main Lemma

3.1 Ergodic theorem

Let I be a countable set denoted as *the space state*. For each $n \in \mathbb{N}$ let X_n be a random variable with values in I, intuitively X_n is a marker of what state our system is in at step n.

Definition. $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with initial distribution λ and transition matrix $P = (p_{ij} \,|\, i, j \in I)$, shortened to $Markov(\lambda, P)$, if

- λ is the probability distribution on X_0;
- given that $X_n = i$, $(p_{ij} \,|\, j \in I)$ is the probability distribution on X_{n+1} and is independent of X_k for $0 \leq k < n$, i.e. $P(X_{n+1} = j \,|\, X_n = i) = p_{ij}$.

For each $k \geq 1$, let $(p_{ij}^{(k)}) = P^k$, that is, the kth matrix multiplication power.

Definition. A Markov chain is *irreducible* if for every $i, j \in I$,

$$P(X_n = j \text{ for some } n \in \mathbb{N} \,|\, X_0 = i) > 0.$$

Let $i \in I$, we define T_i to be *the first passage time*, that is

$$T_i = \inf\{n \geq 1 \,|\, X_n = i\}.$$

Let $m_i = E(T_i \,|\, X_0 = i)$, i.e. *the expected return time* to state i.

Let $i \in I$, $n \in \mathbb{N}$, *the number of visits to i before step n* is

$$V_i(n) = \sum_{k=0}^{n-1} [\![X_k = i]\!].$$

Theorem 3.1. *(Ergodic Theorem) Let $(X_n)_{n\in\mathbb{N}}$ be $Markov(\lambda, P)$ and irreducible. Then for every $j \in I$*

$$P\left(\lim_n \frac{V_j(n)}{n} = \frac{1}{m_j}\right) = 1.$$

Definition. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain. We say that a state i is recurrent if $P(X_n = i \text{ for infinitely many } n \,|\, X_0 = i) = 1$.

Property 3.2. *If $(X_n)_{n\in\mathbb{N}}$ is irreducible and it has a recurrent state then all states are recurrent.*

Theorem 3.3. *(Ergodic Theorem 2) Let $(X_n)_{n\in\mathbb{N}}$ be $Markov(\lambda, P)$, irreducible and with a recurrent state. Then for every $j \in I$*

$$P\left(\lim_n \frac{V_j(n)}{n} = \pi_j\right) = 1,$$

for $\pi_j = \lim_n \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)}$, where π_j is a value independent of i and $\pi_j \in \Delta(I)$.

Remark: I think the only part that requires recurrent states is $\pi_j \in \Delta(I)$. (See Corollary 1.6 (page 50) of Sericola book for the rest).

3.2 From strongly connected automaton to Markov chain

Let Σ be a finite alphabet. Let $\alpha \in \Delta(\Sigma)$, that is, $\alpha : \Sigma \to [0,1]$ with $\sum_{a\in\Sigma} \alpha(a) = 1$.

Let α be the extension of α to Σ^∞.

Definition. Let $\mathcal{U} = (\Sigma, S, \delta, s_0)$ be a finite automaton, S the state set, $\delta : S \times \Sigma \to S$, δ^* denoted as δ. Then $\mathcal{U}$ is *strongly connected* if for every $s, s' \in S$ there is a $w \in \Sigma^*$ with $\delta(s, w) = s'$.

Let $\mathcal{U} = (\Sigma, S, \delta, 1)$ be a strongly connected finite automaton, $S = \{1, \ldots, m\}$ the state set.

Fix $q \in S$ as the initial state, that is, the initial distribution is $p_q = 1$, $p_i = 0$ for $i \neq q$.

For $n \in \mathbb{N}$ we define the random variables $Z_n : \Sigma^\infty \to S$ as

$$Z_n(z) = \delta(q, z \restriction n).$$

Then $(Z_n)_{n\in\mathbb{N}}$ is a Markov chain with transition matrix

$$p_{ij} = \sum_{\delta(i,a)=j} \alpha(a).$$

The n-step transition matrix is $(p_{ij}^{(n)}) = (p_{ij})^n$ and

$$p_{ij}^{(n)} = \sum_{|u|=n} [\![\delta(i,u) = j]\!]\alpha(u).$$

Since $\mathcal{U}$ is strongly connected and S is finite, $(Z_n)_{n\in\mathbb{N}}$ is irreducible and all states are recurrent. Therefore by the ergodic theorem (Theorem 3.3),

$$P\left(\lim_n \frac{V_j(n)}{n} = \pi_j\right) = 1,$$

for $\pi_j = \lim_n \frac{1}{n}\sum_{k=1}^n p_{ij}^{(k)}$ and $V_j(n) = \sum_{k=0}^{n-1} [\![Z_k = j]\!]$.

Since for each $z \in \Sigma^\infty$,

$$V_j(z)(n) = \sum_{k=0}^{n-1} [\![Z_k(z) = j]\!] = \sum_{k=0}^{n-1} [\![\delta(q, z \restriction k) = j]\!],$$

$$\forall q, j \in S \qquad \alpha\left(\left\{ z \in \Sigma^\infty \;\middle|\; \lim_n \frac{1}{n}|\{k \leq n \mid \delta(q, z \restriction k) = j\}| = \pi_j \right\}\right) = 1 \tag{1}$$

For $n \in \mathbb{N}$ we define the random variables $\overline{Z}_n : \Sigma^\infty \to S \times \Sigma$ as

$$\overline{Z}_n(z) = (\delta(q, z \restriction n), z_{n+1}).$$

Then $(\overline{Z}_n)_{n\in\mathbb{N}}$ is a Markov chain with transition matrix

$$\overline{p}_{(i,a)(j,b)} = [\![\delta(i,a) = j]\!]\alpha(b).$$

The n-step transition matrix is $(\overline{p}^{(n)}_{(i,a)(j,b)}) = (\overline{p}_{(i,a)(j,b)})^n$ and

$$\overline{p}^{(n)}_{(i,a)(j,b)} = \sum_{|u|=n, u_1=a} [\![\delta(i,u) = j]\!]\alpha(u)\alpha(b)/\alpha(a).$$

Therefore

$$\overline{p}^{(n)}_{(i,a)(j,b)} = p^{(n-1)}_{\delta(i,a)j}\alpha(b).$$

Since $\mathcal{U}$ is strongly connected and S is finite, $(\overline{Z}_n)_{n\in\mathbb{N}}$ is irreducible and all states are recurrent. Therefore by the ergodic theorem (Theorem 3.3),

$$P\left(\lim_n \frac{\overline{V}_{(j,b)}(n)}{n} = \overline{\pi}_{(j,b)}\right) = 1,$$

for $\overline{\pi}_{(j,b)} = \lim_n \frac{1}{n}\sum_{k=1}^n \overline{p}^{(k)}_{(i,a)(j,b)}$ and $\overline{V}_{(j,b)} = \sum_{k=0}^{n-1}[\![\overline{Z}_k = (j,b)]\!]$.

For each $z \in \Sigma^\infty$,

$$\overline{V}_{(j,b)}(z)(n) = \sum_{k=0}^{n-1}[\![\overline{Z}_k(z) = (j,b)]\!] = \sum_{k=0}^{n-1}[\![\delta(q, z\upharpoonright k) = j \wedge z_{k+1} = b]\!].$$

$$\overline{\pi}_{(j,b)} = \lim_n \frac{1}{n}\sum_{k=1}^n p^{(k-1)}_{\delta(i,a)j}\alpha(b) = \pi_j\alpha(b)$$

Therefore we have

$\forall q, j \in S\, \forall b \in \Sigma$

$$\alpha\left(\left\{z \in \Sigma^\infty \;\middle|\; \lim_n \frac{1}{n}|\{k \le n \mid \delta(q, z\upharpoonright k) = j \wedge z_{k+1} = b\}| = \pi_j\alpha(b)\right\}\right) = 1 \tag{2}$$

3.3 Proof of the Main Lemma

Notation: We use $\forall^\infty k$ to denote "for almost every k"

Definition. $z \in \Sigma^\infty$ is *α-normal* if for any $u \in \Sigma^*$, if $k = |u|$, $N(u, z\upharpoonright(nk)) = \sum_{w=1}^n [\![z[(w-1)k+1..wk] = u]\!]$

$$\lim_n \frac{N(u, z\upharpoonright(nk))}{n} = \alpha(u).$$

In other words, when considering disjoint length-k blocks of z, the experimental frequency of u converges to $\alpha(u)$.

Lemma 3.4. *Let $\mathcal{U} = (\Sigma, S, \delta, 1)$ be a strongly connected finite automaton with $S = \{1, \ldots, m\}$ its state set. Then there exist $\pi_j > 0$ $(j = 1, \ldots, m)$ such that for every α-normal sequence z, $i \in S$, $b \in \Sigma$,*

$$\lim_n \frac{1}{n} |\{r \leq n \,|\, \delta(q, z \upharpoonright r) = j \wedge z_{r+1} = b\}| = \pi_j \alpha(b).$$

Proof. Let π_j $(j = 1, \ldots, m)$ be as in (2). Due to (2) and the fact that there are only finitely many $i, j \in S$ and $b \in \Sigma$, for each $\epsilon, \epsilon' > 0$ we have that

$$\begin{aligned} \alpha(\{z \in \Sigma^\infty & \\ & |\, \forall^\infty k \forall q, j \in S \, \forall b \in \Sigma \\ & \frac{1}{k} |\{t < k \,|\, \delta(q, z \upharpoonright t) = j \wedge z_{t+1} = b\}| - \pi_j \alpha(b)| < \epsilon \pi_j \alpha(b)\}) = 1 \end{aligned}$$

and therefore

$$\forall^\infty k \quad \sum_{u \in \Sigma^k} \alpha(u) [\![\forall q, j \in S \, \forall b \in \Sigma \, |\{t < k \,|\, \delta(q, u \upharpoonright t) = j \wedge u_{t+1} = b\}| > k \pi_j \alpha(b)(1 - \epsilon)]\!] > 1 - \epsilon' \tag{3}$$

For every α-normal sequence z, every $k \in \mathbb{N}$ and every $\epsilon'' > 0$ we have that

$$\forall^\infty n \, \forall u \in \Sigma^k \quad \frac{N(u, z \upharpoonright (nk))}{n} > \alpha(u)(1 - \epsilon'') \tag{4}$$

Let $k \in \mathbb{N}$ be such that (3) holds. Let $u \in \Sigma^k$ be such that

$$\forall q, j \in S \, \forall b \in \Sigma \, |\{t < k \,|\, \delta(q, u \upharpoonright t) = j \wedge u_{t+1} = b\}| > k \pi_j \alpha(b)(1 - \epsilon). \tag{5}$$

We will refer to a u with property (5) as a "good u".

Fix $n \in \mathbb{N}$ be large enough such that (4) holds. By the equivalent definition of normal numbers, i.e., Theorem 1.1, we have

$$N(u, z \upharpoonright (nk)) > n \alpha(u)(1 - \epsilon'')$$

Therefore whenever $z[(w-1)k+1..wk] = u$ we have at least $k \pi_j \alpha(b)(1-\epsilon)$ positions r for which $\delta(q, z \upharpoonright r) = j \wedge z_{r+1} = b$. Since there are at least $n\alpha(u)(1 - \epsilon'')$ such w with $w \leq n$ we have that the number of positions for which $\delta(q, z \upharpoonright r) = j \wedge z_{r+1} = b$ corresponding to occurrences of the same u are at least

$$k \pi_j \alpha(b)(1 - \epsilon) n \alpha(u)(1 - \epsilon'')$$

From equation (3) we know that if we add $\alpha(u)$ for all good u is

$$\sum_{u \text{ is good}} \alpha(u) > 1 - \epsilon'.$$

The number of positions for which $\delta(q, z \upharpoonright r) = j \wedge z_{r+1} = b$ $(r \leq nk)$ corresponding to all good u is at least

$$\sum_{u \text{ is good}} k\pi_j \alpha(b)(1-\epsilon) n\alpha(u)(1-\epsilon'') > kn\pi_j \alpha(b)(1-\epsilon)(1-\epsilon')(1-\epsilon'').$$

Therefore

$$\lim_n \frac{1}{n} |\{r \leq n \mid \delta(q, z \upharpoonright r) = j \wedge z_{r+1} = b\}| \geq \pi_j \alpha(b).$$

Since π_j $(j = 1, \ldots, m)$ in (2) is a probability distribution on S we have the equality. □

4 Agafonov's Theorem

We can now prove Agafonov's Theorem, the proof is taken from [1, 9].

Theorem 4.1. *(a) If a number $z \in \Sigma^\infty$ is collective relative to $\mathcal{F}_r$, then z is normal.*

(b) If a number $z \in \Sigma^\infty$ is normal, for any $\tau \in \mathcal{F}_r$, let T be the transformation induce by τ, then $T(z)$ is either not defined or is a collective relative to τ.

Proof. Statement (a) is Corollary 2.4. We only need to show case (b).

Denote the set of normal numbers as $\mathcal{N}$. We have $\alpha(\mathcal{N}) = 1$.

Let $z \in \mathcal{N}$, $\tau \in \mathcal{F}_r$, $\mathcal{U}$ be the finite automata generate τ, T be the transformation induced by τ and $T(z)$ be defined, i.e. is an infinite sequence.

Let's first assume $\mathcal{U}$ be a strongly connected automata, S be its state set and F be the set of accepting states.

Consider the first n symbols of $T(z)$, Let m be length of the shortest prefix of z that $\mathcal{U}$'s selecting result be $T(z) \upharpoonright n$. For a fixed $b \in \Sigma$, we have

$$\frac{\#_n(b)}{n} = \sum_{s \in F} \frac{\mathrm{Occ}_{s,m}(b)}{m} = \sum_{s \in F} \frac{\mathrm{Occ}_m(s)}{m} \frac{\mathrm{Occ}_{s,m}(b)}{\mathrm{Occ}_m(s)},$$

where $\#_n(b)$ counts the number of occurrences of b in $T(z) \upharpoonright n$, $\mathrm{Occ}_m(s)$ the number of times we visit s during the selection of $T(z) \upharpoonright n$, and $\mathrm{Occ}_{s,m}(b)$ the number of times we leave state s by consuming the symbol b in the same process.

Note that as $n \to \infty$, m goes to ∞. Then by Lemma 3.4, both of the following limit

$$\lim_{n \to \infty} \frac{\mathrm{Occ}_{s,m}(b)}{\mathrm{Occ}_m(s)}, \quad \lim_{m \to \infty} \frac{\mathrm{Occ}_m(s)}{m}$$

exist as n goes to ∞, and the limit is not dependent on the sequence z but only determined by the finite automata $\mathcal{U}$.

Hence the limit

$$L := \lim_{n\to\infty} \frac{\#_n(b)}{n}$$

exist and the same for every normal sequence z. Suppose $L \neq \alpha(b)$.

Let

$$M = \{z \mid \text{asymptotic frequency of } b \in \Sigma = L \neq \alpha(b)\}.$$

By the above discussion, we can see T maps every normal number to some element of M. That is, $\mathcal{N} \subseteq T^{-1}(M)$.

Therefore $\alpha(T^{-1}(M)) = \alpha(\mathcal{N}) = 1$. However, since elements of M do not satisfy the strong law of large number, the set M must have measure zero. A contradiction to Theorem 2.1.

For a finite automata that is not necessarily strong connective, we just need to first apply Proposition 2.5 of [9], to first find a prefix u of z that will lead us to a ergodic state. The repeat the measure preserving argument on the set $u\mathcal{N}$, i.e. the set of concatenations of u and normal sequences. □

Corollary 4.2. *Let $z \in \Sigma^\infty$ be a normal sequence, T be as in Theorem 4.1, then $T(z)$ is also normal whenever $T(z)$ is defined.*

Proof. It suffices to show $T(z)$ is a collective relative to $\mathcal{F}_r$. By Theorem 4.1, since $T(z)$ is a resulting sequence after selection of an finite automata, it is simply normal. Only need to show the resulting sequence of $T(z)$ after selection of an finite automata is again simply normal.

Let T be the transformation induced by $\mathcal{U} = (\Sigma, Q_1, \delta_1, F_1, I_1)$, and let $\mathcal{U}' = (\Sigma, Q_2, \delta_2, F_2, I_2)$ be an arbitrary finite state machine, T' be its induced transformation. We want to show $T'(T(z))$ is simply normal.

Consider the following finite automata, $\mathcal{V} = (Q_1 \times Q_2, \delta, F_1 \times F_2, I_1 \times I_2)$, where

$$\delta((q_1, q_2), b) = \begin{cases} \Big(\delta_1(q_1, b), q_2\Big) & \text{if } q_1 \notin F_1 \\ \Big(\delta_1(q_1, b), \delta_2(q_2, b)\Big) & \text{if } q_1 \in F_1 \end{cases}$$

The intuition is to treat $\mathcal{V}$ a pipeline of $\mathcal{U}_1$ and $\mathcal{U}_2$: $\mathcal{U}_2$ only treat as input the output of $\mathcal{U}_1$. Let $T_\mathcal{V}$ be the transformation induced by $\mathcal{V}$. It is obvious that

$$T_\mathcal{V} = T' \circ T.$$

Then by Theorem 4.1 again, we can see $T'(T(z)) = T_\mathcal{V}(z)$ is simply normal, when every it is defined.

Therefore, $T(z)$ is a collective relative to finite automata and hence is normal. □

5 Extension to more general selection rules

(This section is not finished, I have also some handwritten notes I am not including)

Then they consider the equivalence relation on finite strings defined $u = v$ if $\{w|uw \in S\} = \{w|vw \in S\}$

Then they consider the case where $S/=$ is finite (the original Agafonov paper would be that case $\{0,1\}^*/=$ being finite).

Notice that $S/=$ being finite is equivalent to considering automata with countably many states and finitely many Accepting states.

The ergodic theorem works for countably many states. Our lemma 3.1 giving the frequency of reaching state j with next symbol b as $\pi_j * \alpha(b)$ works for countably many states.

Then the final result of this selection rules preserving normality should work because we only need to consider a finite number of Accepting states.

6 Dichotomy

In this section, we will first revisit Schnorr-Stimm's dichotomy theorem and the quantify it by use

7 Generalization of Niven's Proof of Equivalence of Definition of Normality

We generate Niven's proof of equivalence of definitions of normality for non-uniform distribution. Our proof directly follows the structure of Niven's proof as in [7].

7.1 Notations and Definitions

Notation. In what follows, we use an alphabet Σ of size r and use $b \in \Sigma$ be a symbol in Σ. Let $x = x_1x_2x_3\cdots$ be an infinite decimal to base r, and let X_n denote the block of digits $x_1x_2\cdots x_n$. Let $N(b, X_n)$ denote the number of occurrences of b in the block X_n.

The following notation is very useful in denoting counts of block after a slide.

Notation. Let $B_k = b_1b_2\cdots b_k$ and $X_n = x_1x_2\cdots x_n$. We use $N_s(B_k, X_n)$ to denote the number of occurrences of B_k in X_n, say $b_1 = x_j$, $b_2 = x_{j+1}$, etc., which satisfy the condition $j \equiv s(\text{mod } k)$.

One way to understand the notation $N_s(B_k, X_n)$ is to view it as counting the occurrence of B_k only starting from $s + m \cdot k$ position, where $m \in \mathbb{N}$. Since nevertheless any position falls in some residue class of k, we have

Observation 7.1.

$$N(B_k, X_n) = \sum_{1}^{k} N_s(B_k, X_n). \tag{6}$$

We can now rewrite the definition of normality in the above notation. Recalled that we use α for a distribution and its extension on Σ, Σ^* and Σ^∞.

Definition. (Sliding) A number x is normal to base r if

$$\lim_{n\to\infty} \frac{1}{n} N(B_k, X_n) = \alpha(B_k) \tag{7}$$
$$\text{for all } B_k,\ k = 1, 2, 3, \cdots .$$

E. Borel originally defined a number x is normal to base r if each of x, rx, r^2x,$\cdots$ is simple normal to all of the bases r, r^2, r^3, $\cdots$. That is saying

Definition. (Borel) A number x is normal to base r if

$$\lim_{n\to\infty} \frac{1}{n} N_s(B_k, X_m) = \frac{\alpha(B_k)}{k} \quad \text{for all } k,\ s,\ B_k. \tag{8}$$

We can write the blocking definition by
Definition. (Blocking) A number x is normal to base r if

$$\lim_{n\to\infty} \frac{1}{n} N_1(B_k, X_n) = \frac{\alpha(B_k)}{k}. \tag{9}$$

It is clear that by applying Observation 7.1, Borel's definition (8) implies the sliding definition (7). It is also obvious that Borel's definition (8) implies blocking definition (9).

To show the equivalence of the above three definitions, it suffices to further show sliding definition (7) implies Borel's definition (8) and blocking definition (9) implies sliding definition (7).

We will come back to this. Before that we need to develop some counting lemmas.

7.2 Counting and Measure

In this subsection, we will develop some counting lemmas that lead us to Strong Law of Large Number and the measure of normal numbers.

For a fixed digit $b \in \Sigma$ we define $p(n, j)$ be the measure of all blocks of digits $X_n = x_1x_2\cdots x_n$ having exactly j occurrences of the digit b, that is, having $N(b, X_n) = j$. Taking $0 \le j \le n$, we see that

$$p(n, j) = \binom{n}{j} \alpha(b)^j (1 - \alpha(b))^{n-j} = \frac{n!}{j!(n-j)!} \alpha(b)^j (1 - \alpha(b))^{n-j}. \tag{10}$$

Note that we can define $p(n,j)=0$ in case $j<0$ or $j>n$. We also write $r_b=\frac{1}{\alpha(b)}>1$ for convenient.

We will use the following inequality frequently.

Lemma 7.2. *If $0<\alpha<1$, then $1-\alpha<\exp(-\alpha)$.*

Proof.

$$\exp(-\alpha)=1-\alpha+\frac{\alpha^2}{2!}-\frac{\alpha^3}{3!}+\cdots>1-\alpha.$$

□

We establish the next two lemmas to give a bound on $p(n,j)$. Specifically, the fixed symbol $b\in\Sigma$ with probability $\alpha(b)$. The expected occurrences of b in a length n block should be closed to $n\alpha(b)$. That is to say, in a length $n/\alpha(b)=nr_b$ block, there should be n occurrences of b. Since nr_b is not a integer in usual, so we have to talk about $\lfloor nr_b\rfloor$ instead.

Lemma 7.3. *For $j\geq 2$ we have*

$$p(\lfloor nr_b\rfloor,n+j)<\exp(-j^2/4nr_b)$$

Proof. From formula (10) and Lemma 7.2 we see that

$$\begin{aligned}
\frac{p(\lfloor nr_b\rfloor,n+j)}{p(\lfloor nr_b\rfloor,n)}&=\frac{\binom{\lfloor nr_b\rfloor}{n+j}}{\binom{\lfloor nr_b\rfloor}{n}}\cdot\left(\frac{\alpha(b)}{1-\alpha(b)}\right)^j\\
&=\frac{(\lfloor nr_b\rfloor-n)(\lfloor nr_b\rfloor-n-1)\cdots(\lfloor nr_b\rfloor-n-j+1)}{(n+1)(n+2)\cdots(n+j)}\cdot\frac{1}{(r_b-1)^j}\\
&\leq\frac{(nr_b-n)(nr_b-n-1)\cdots(nr_b-n-j+1)}{n^j(1+1/n)(1+2/2)\cdots(1+j/n)}\cdot\frac{1}{(r_b-1)^j}\\
&<\frac{(r_b-1)(r_b-1-\frac{1}{n})\cdots(r_b-1-\frac{j-1}{n})}{(r_b-1)^j}\\
&=\left(1-\frac{1}{n(r_b-1)}\right)\left(1-\frac{2}{n(r_b-1)}\right)\cdots\left(1-\frac{j-1}{n(r_b-1)}\right)\\
&<\exp\left(-\frac{1}{n(r_b-1}-\frac{2}{n(r_b-1)}\cdots-\frac{j-1}{n(r_b-1)}\right)\\
&=\exp\left(-\frac{j(j-1)}{2n(r_b-1)}\right)<\exp\left(-\frac{j^2}{4nr_b}\right).
\end{aligned}$$

Also rather obviously we have

$$p(\lfloor nr_b\rfloor,n)<1.$$

So the lemma follows from these inequalities. □

Lemma 7.4. *For $j \geq 2$ we have*

$$p(\lfloor nr_b \rfloor, n-j)\exp(-j^2/4nr_b)$$

Proof. Lemma 7.2 and formula (10) imply that

$$\begin{aligned}
\frac{p(\lfloor nr_b \rfloor, n-j)}{p(\lfloor nr_b \rfloor, n)} &= \frac{\binom{\lfloor nr_b \rfloor}{n-j}}{\binom{\lfloor nr_b \rfloor}{n}} \cdot \left(\frac{(1-\alpha(b))}{\alpha(b)}\right)^j \\
&= \frac{n(n-1)\cdots(n-j+1)(r_b-1)^j}{(\lfloor nr_b \rfloor - n+1)(\lfloor nr_b \rfloor - n+2)\cdots(\lfloor nr_b \rfloor - n+j)} \\
&\leq \frac{n^j \cdot 1 \cdot (1-\frac{1}{n})\cdots(n-\frac{j-1}{n})(r_b-1)^j}{(nr_b-1-n+1)(nr_b-1-n+2)\cdots(nr_b-1-n+j)} \\
&< \frac{(r_b-1)^j}{(r_b-1)(r_b-1+1/n)\cdots(r_b-1+\frac{j-1}{n})} \times (1-\frac{1}{n})\cdots(n-\frac{j-1}{n}) \\
&< (1-\frac{1}{n})\cdots(n-\frac{j-1}{n}) \\
&< \exp\left(-\frac{1}{n}-\frac{2}{n}-\cdots-\frac{j-1}{n}\right) \\
&= \exp\Big(-j(j-1)/2n\Big) < \exp\Big(-j^2/4nr_b\Big) < \exp(-j^2/4nr_b)
\end{aligned}$$

The last two inequalities hold because $j \geq 2$ and $r_b > 1$. The lemma follows from the above and $p(\lfloor nr_b \rfloor, n) < 1$. □

Lemma 7.5. *Let $\epsilon > 0$ be given. For all sufficiently large n the measure of blocks $X_{\lfloor nr_b \rfloor}$ such that*

$$|N(b, X_{\lfloor nr_b \rfloor} - n)| > n\epsilon \tag{11}$$

is less than $\lfloor nr_b \rfloor (1+c_1)^n$, where c_1 is a positive constant depending on ϵ but not on n.

Proof. The measure of blocks $X_{\lfloor nr_b \rfloor}$ of $\lfloor nr_b \rfloor$ digits satisfying Formula (11) is simply

$$\sum_{k>n+n\epsilon} p(\lfloor nr_b \rfloor, k) + \sum_{k<n-n\epsilon} p(nr, k) = \sum_{|j|>n\epsilon} p(\lfloor nr_b \rfloor, n+j).$$

We estimate this sum as follows. Taking n sufficiently large so that $n\epsilon > 1$, we note that the sum has fewer than $\lfloor nr_b \rfloor$ non-zero terms. By the preceding two lemmas we have

$$p(\lfloor nr_b \rfloor, n+j) < \exp(-j^2/4nr_b) < \exp(-(n\epsilon)^2/4nr_b) = \exp(-n\epsilon^2/4r_b)$$

Thus the measure of the blocks $X_{\lfloor nr_b \rfloor}$ satisfying Formula (11) is at most

$$\lfloor nr_b \rfloor \exp(-n\epsilon^2/4r_b) = \lfloor nr_b \rfloor (1+c_1)^{-n},$$

where c_1 is defined by the relation $1 + c_1 = \exp(\epsilon^2/4r_b)$. □

Lemma 7.6. *Let $\epsilon > 0$ be given. For all sufficiently large m the measure of blocks X_m satisfying*

$$|N(b, X_m) - m\alpha(b)| > m\epsilon \tag{12}$$

is less than $m(1+c)^{-m}$, where c is a positive constant not dependent on m.

Proof. The case $m = \lfloor nr_b \rfloor$ is discuss in Lemma 7.5. We only need to extend Lemma 7.5 for $m = \lfloor nr_b \rfloor + d$, where n is the largest n such that $m \geq \lfloor nr_b \rfloor$ and $d \in \mathbb{N}$.

Note that if $r_b \in \mathbb{N}$, then $m = \lfloor nr_b \rfloor + d = nr_b + d$, we have $0 \leq d < r_b$. We now let $r_b \notin \mathbb{N}$, by the way we define n, we have

$$\begin{aligned}
\lfloor nr_b \rfloor \leq m < \lfloor (n+1)r_b \rfloor &\implies (n+1)r_b \geq m+1 \\
&\implies (nr_b - \lfloor nr_b \rfloor) + r_b \geq m + 1 - \lfloor nr_b \rfloor, \quad \text{by subtracting } \lfloor nr_b \rfloor \\
&\implies r_b + 1 \geq d + 1 \implies r_b \geq d.
\end{aligned}$$

Since $r_b \notin \mathbb{N}$ and $d \in N$, we have $r_b > d$. Therefore, in either case, we have $0 \leq d < r_b$.

We now show that, for any m that is sufficiently large, a block X_m satisfies inequality (12), then the first $\lfloor nr_b \rfloor$ digits of this block, $X_{\lfloor nr_b \rfloor}$, must satisfy inequality (11). This induce a upper bound of the desired measure.

To see this, we assume that a certain X_m satisfies (12). Then by triangle inequality we have

$$\begin{aligned}
m\epsilon < |N(b, X_m) - m\alpha(b)| &\leq |N(b, X_m) - N(b, X_{\lfloor nr_b \rfloor})| + \\
&\quad |N(b, X_{\lfloor nr_b \rfloor}) - n| + |n - \frac{m}{r_b}| \\
&< d + |N(b, X_{\lfloor nr_b \rfloor}) - n| + 2.
\end{aligned}$$

The last inequality holds since

$$|n - \frac{m}{r_b}| = |\frac{nr_b - n}{r_b}| \leq |\frac{nr_b - \lfloor nr_b \rfloor}{r_b}| + |\frac{\lfloor nr_b \rfloor - m}{r_b}| = \frac{1}{r_b} + \frac{d}{r_b} < 2.$$

Therefore we have, for sufficiently large m (and hence n),

$$\begin{aligned}
|N(b, X_{\lfloor nr_b \rfloor}) - n| &> m\epsilon - d - 2 \\
&\geq \lfloor nr_b \rfloor \epsilon - d - 2 \geq nr_b\epsilon - d - 2 > n\epsilon
\end{aligned}$$

and this is inequality (11). The last inequality holds since $r_b > 1$ and when n sufficiently large $n(r_b - 1)\epsilon - d - 2 > 0$, we have

$$nr_b\epsilon - d - 2 = n\epsilon + (n(r_b - 1)\epsilon - d - 2) > n\epsilon.$$

Hence the blocks of $X_{\lfloor nr_b \rfloor}$ that satisfy inequality (11) form a cover of the blocks of X_m that satisfy inequality (12). By Lemma 7.5, this gives an upper bound of the measure of later as

$$\lfloor nr_b \rfloor (1+c_1)^{-n} \leq m\left((1+c_1)^{\frac{n}{m}}\right)^{-m} < m\left((1+c_1)^{1/2r_b}\right)^{-m} = m(1+c)^{-m},$$

where we let $c = (1+c_1)^{\frac{n}{m}} - 1$. □

With this lemma in hand, it is now easy to establish the following results.

Theorem 7.7. *(Strong Law of Large Number) The set S of infinite decimal $x = .x_1x_2x_3\cdots$ which are not α-simply normal to base r has measure zero.*

Proof. (Sketched.) By preceding lemma and BorelCantelli lemma. □

Theorem 7.8. *Almost all real numbers are absolutely α-normal.*

Proof. (Sketched.) By the preceding theorem and the same process as in Theorem 8.11 of [7]. □

7.3 Equivalence of Definitions

We return to the proof of equivalence of the definitions of normality.

Lemma 7.9. *Let r, k, and s be given positive integers with $r \geq 2$, $1 \leq s \leq k$. Let B_k be a fixed block of k digits to base r and ϵ any positive real number. Then for all t sufficiently large the inequality*

$$N_s(B_k, A_t) > \frac{t\alpha(B_k)}{k} - \frac{2\epsilon t}{k}$$

is satisfied by all blocks A_t of t digits apart from a set of exceptional blocks with a measure of at most ϵ.

Proof. First we reformulate Lemma 7.6 with the base r replaced by r^k. Thus the measure of blocks X_m of m digits to base r^k satisfying

$$|N(b, X_m) - m\alpha(b)| > m\epsilon$$

is less than $m(1+c)^{-m}$, where the positive constant c in independent of m. Note again here b is a single digit to base r^k.

As a result, we can say that the inequality

$$N(b, X_m) \geq m\alpha(b) - m\epsilon \tag{13}$$

holds for all X_m apart form a set of blocks with a measure of at most $m(1+c)^{-m}$ blocks.

We now interpret this in base r, so that b becomes a block of B_k of k digits and X_m becomes say X_{mk} with mk digits (to base k). Also $N(b, X_m)$ becomes $N_1(B_k, X_{mk})$, the subscript 1 occurring now because we can view X_m to base r^k a mk digits block X_{mk} to base r, and we divide the later into blocks of length k, starting form the first digit. We can extend this to get

$$N_1(B_k, X_{mk}) = N_s(B_k, A_t) \tag{14}$$

by regarding A_t as the block X_{mk} with $s-1$ digits attached at the left end, and $t-mk-(s-1)$ digits attached at the right end. We can be certain that equation (14) holds if the number of digits attached at the right end is fewer than k. This we do by regarding the integers t, s and k as given, and from them we determine m as the quotient when we apply the division algorithm to $t-s+1$ and k,

$$t-s+1=mk+u, \quad 0\leq \mathrm{u} < \mathrm{k}.$$

This implies that

$$mk \leq t < mk+2k, \quad \text{and} \quad \frac{t}{k} \geq m > \frac{t}{k}-2. \tag{15}$$

Thus we have obtain equality (14) by identifying most of the digits of A_t with the block X_{mk},

$$\begin{aligned} A_t &= a_1 a_2 \cdots a_t \\ &= a_1 a_2 \cdots a_{s-1} x_1 x_2 \cdots x_{mk} a_{t-u+1} \cdots a_t, \end{aligned}$$

with fewer than k digits both prior to and following X_{mk} because $s-1<k$ and $u<k$. Corresponding to a fixed set of values for X_{mk} there are r^{t-mk} blocks A_t with the same measure in total as the fixed X_{mk}.

Thus we can rewrite inequality (13) in view of equality (14) and other discussion, and conclude that for all sufficiently large t the inequality

$$N_s(B_k, A_t) \geq m\alpha(B_k) - m\epsilon \tag{16}$$

holds for all block A_t apart from a set of exceptional blocks with a measure of at most

$$m(1+c)^{-m}.$$

As t increase indefinitely, so does m, and so $m(1+c)^{-m}$ tends to zero. Thus (16) holds for all A_t apart from exceptional blocks with a measure of at most ϵ, provided t is sufficiently large.

Also we note that $\epsilon t/k > 2\alpha(B_k)$ for t sufficiently large, and so by inequality (15)

$$\begin{aligned} m\alpha(B_k) - m\epsilon &> \big(\frac{t}{k}-2\big)\alpha(B_k) - \frac{t}{k}\epsilon \\ &= \frac{t\alpha(B_k)}{k} - 2\alpha(B_k) - \frac{t\epsilon}{k} > \frac{t\alpha(B_k)}{k} - \frac{2t\epsilon}{k}. \end{aligned}$$

In view of this, we see that the inequality (7.9) is satisfied by any block A_t which satisfies inequality (16), and so the lemma is proved. □

We will also need to use the following lemma in the rest of our proof, which give us some convenience by consider just one-side of the limiting behavior.

Lemma 7.10. *Let $f_1, f_2, \cdots, f_m$ be real functions of the real variable x satisfying the conditions*

$$\lim_{x\to\infty}\left\{f_1(s)+f_2(x)+\cdots+f_m(x)\right\}=l,$$

and

$$\liminf_{x\to\infty} f_i(x)\geq l_i \quad \textit{for} \quad i=1,2,\cdots,m,$$

where $\sum_{i=1}^{m} l_i = l$. Then

$$\lim_{x\to\infty} f_i(x)=l_i \quad \textit{for} \quad i=1,2,\cdots,m.$$

Proof. See Lemma 8.2 of [7]. □

Now we are ready to show the equivalence of definitions of normality. The following theorem says that the sliding definition implies Borel's original definition.

Theorem 7.11. *If the infinite string $x = x_1x_2\cdots$ satisfies (7), then x satisfies (8).*

Proof. We begin by observing that (7) implies that for any block of t digits A_t and any positive ϵ

$$N(A_t, X_n) > n\alpha(A_t) - \epsilon n\alpha(A_t) = n(1-\epsilon)\alpha(A_t) \tag{17}$$

for sufficiently large n.

Recall that $N_s(B_k, X_n)$ counts the number of occurrences of B_k in X_n of a certain type, namely $b_1 = x_m,\ b_2 = x_{m+1},\ \cdots,\ b_k = x_{m+k-1}$ with $m \equiv s(\text{mod } k)$. We establish that, for $t > k$,

$$(t-k+1)N_s(B_k, X_n) \geq \sum_{j=1}^{n-t+1} N_s(B_k, x_jx_{j+1}\cdots x_{j+t-1}) \tag{18}$$

where the notation N_s in the sum on the right is to be interpreted as counting occurrences of B_k subject to the same condition on the subscripts: namely, occurrences of the form $b_1 = x_m, \cdots,\ b_k = x_{m+k-1}$ with $m \equiv s(\text{mod } k)$. Any such occurrence of B_k in X_n is counted exactly $t-k+1$ times by the expression on the left side of equation (18), and at most $t-k+1$ times by the expression on the right as j ranges over its assigned values. Having established equation (18), we now approximate the sum on the right side of this inequality. As j ranges over the values 1, 2, $\cdots$, $n-t+1$, the block $x_jx_{j+1}\cdots x_{j+t-1}$ is identical with any specified block A_t of t digits at least $n\alpha(A_t)(1-\epsilon)$ times, by inequality (17). Hence we can change the sum in inequality (18) to a sum over all the r^t possible values of A_t, thus

$$(t-k+1)N_s(B_k, X_n) > \sum_{A_t} n\alpha(A_t)(1-\epsilon)N_\sigma(B_k, A_t).$$

This sum ranges over all possible blocks A_t, and the indeterminate subscript σ can remain entirely unspecified because we now apply Lemma 7.9. Thus we apply the

inequality (7.9) to N_σ in all cases of nonexceptional blocks A_t, or "good" A_t's, which we know have a measure of at least $1-\epsilon$, since the exceptional ones have a measure of no more than ϵ. Disregarding the exceptional blocks, we obtain

$$\begin{aligned}
(t-k+1)N_s(B_k,X_n) &> \sum_{A_t} n\alpha(A_t)(1-\epsilon)N_\sigma(B_k,A_t) \\
&= n(1-\epsilon)\sum_{A_t}\alpha(A_t)N_\sigma(B_k,A_t) \\
&> n(1-\epsilon)\sum_{A_t \text{ good}}\alpha(A_t)\left(\frac{t\alpha(B_k)}{k}-\frac{2\epsilon t}{k}\right) \\
&= n(1-\epsilon)\left(\frac{t\alpha(B_k)}{k}-\frac{2\epsilon t}{k}\right)\sum_{A_t \text{ good}}\alpha(A_t) \\
&\geq nt\left(\frac{\alpha(B_k)}{k}-\frac{2\epsilon}{k}\right)(1-\epsilon)^2 \\
&> nt\left(\frac{\alpha(B_k)}{k}-2\epsilon\right)(1-2\epsilon) \\
&> nt\left(\frac{\alpha(B_k)}{k}-4\epsilon\right)
\end{aligned}$$

for n and t sufficiently large. (The steps in this chain of inequalities are incorrect in case ϵ is large, for example, $\epsilon = 1$, but the final result is obviously correct for such values of ϵ.) Hence we have, for any positive ϵ,

$$\frac{N_s(B_k,X_n)}{n} > \frac{t}{t-k+1}\left(\frac{\alpha(B_k)}{k}-4\epsilon\right) > \frac{\alpha(B_k)}{k}-4\epsilon,$$

and therefore

$$\liminf_{n\to\infty}\frac{N_s(B_k,X_n)}{n} \geq \frac{\alpha(B_k)}{k}.$$

This result holds for the r^k possible blocks B_k and the k possible values of s. Hence we can apply Lemma 7.10 to obtain (8), and the proof of the theorem is complete. □

Now since Borel's definition clearly implies the blocking definition, the only missing piece of the puzzle is to prove blocking definition implies sliding definition, which will be complete by the following theorem.

Theorem 7.12. *Formula (9) implies Formula (7).*

Proof. By Formula (9) we see that, for any block of digits A_t and any positive ϵ,

$$N_1(A_t,X_n) > \frac{n\alpha(A_t)}{t}-\frac{\epsilon n\alpha(A_t)}{t} \tag{19}$$

for all n sufficiently large.

Next we established that, with $k < t < n$,

$$N(B_k, X_n) \geq \sum_{A_t} N_1(A_t, X_n) > \left(\frac{n}{t} - \frac{\epsilon n}{t}\right) \sum_{A_t} \alpha(A_t) N(B_k, A_t), \tag{20}$$

the second inequality arising from (19). This is seen by partitioning X_n into consecutive blocks of t digits and then counting $N_1(A_t, X_n)$. Then we count $N(B_k, A_t)$, the number of occurrences of B_k in A_t, perform the obvious multiplication, and sum over all possible A_t. This sum may not yield exactly $N(B_k, X_n)$, because we lose any B_k that straddles blocks of length t in the partitioning of X_n.

The expression $N(B_k, A_t)$ on the right side of (20) can be written

$$N(B_k, A_t) = \sum_{s=1}^{k} N_s(B_k, A_t),$$

and Lemma 7.9 is applied to these terms on the nonexceptional A_t's, or the "good" A_t's. Thus (20) implies

$$\begin{aligned}
N(B_k,\ X_n) &> \left(\frac{n}{t} - \frac{\epsilon n}{t}\right) \sum_{A_t} \alpha(A_t) N(B_k, A_t) \\
&> \left(\frac{n}{t} - \frac{\epsilon n}{t}\right) k \left(\frac{t\alpha(B_k)}{k} - \frac{2\epsilon t}{k}\right) \sum_{A_t \text{ good}} \alpha(A_t) \\
&> n(\alpha(B_k) - 2\epsilon)(1-\epsilon)^2 \qquad \text{since the measure of good } A_t\text{'s} > 1-\epsilon \\
&> n(\alpha(B_k) - 2\epsilon)(1-2\epsilon) \\
&> n(\alpha(B_k) - 4\epsilon),
\end{aligned}$$

for all n sufficiently large. (As in the proof of the previous theorem, the steps in this chain of inequalities are incorrect if ϵ is large, but the final result holds nevertheless.)

Thus we have

$$\liminf_{n\to\infty} \frac{N(B_k,\ X_n)}{n} \geq \alpha(B_k)$$

for any block B_k of k digits. Application of Lemma 7.10 leads to the conclusion (7), which proves the theorem. □

8 Discussion

1. Measure preservation or measure bounded is not sufficient to preserve normality. Agafonov [1] concluded his proof by only measure selection rules defined by finite automata are measure bounded (preserved). Actually, Doob [5] proved that all non-predicted strategies are measure-preserved. Yet, in Merkle and Reimann [6] a non-predicted strategy fails to preserve normality. The trouble is that we

can have a subset of normal numbers with measure zero, that be mapped into non-normal numbers. This is enough to break normality preservation but not measure preservation.

2. The main lemma save us from this trouble and is the real engine here in this proof. It is a statement about all normal numbers. Under this lemma, the selection result of normal number will either all be normal or all not normal. Hence we can apply the measure-preservation rule. That is, if the selection result in mapping some normal numbers to non-normal ones, then **all** normal numbers will be mapped to non-normal ones. That is, mapping a measure 1 set to a measure 0 set, which breaks the measure bounded/preserved property.

3. More our-main-lemma-like theorem or "Kolmogorov program" style theorem on normal numbers will be useful in our future research.

References

[1] Valerii Nikolaevich Agafonov. Normal sequences and finite automata. In *Doklady Akademii Nauk*, volume 179, pages 255–256. Russian Academy of Sciences, 1968.

[2] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[3] M Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 27(1):247–271, 1909.

[4] Alonzo Church. On the concept of a random sequence. *Bulletin of the American Mathematical Society*, 46(2):130–135, 1940.

[5] JL Doob. Note on probability. *Annals of Mathematics*, pages 363–367, 1936.

[6] Wolfgang Merkle and Jan Reimann. Selection functions that do not preserve normality. *Theory of computing systems*, 39(5):685–697, 2006.

[7] Ivan Niven. *Irrational numbers*. Number 11. Cambridge University Press, 2005.

[8] LP Postnikova. On the connection between the concepts of collectives of mises-church and normal bernoulli sequences of symbols. *Theory of Probability & Its Applications*, 6(2):211–213, 1961.

[9] Claus-Peter Schnorr and Hermann Stimm. Endliche automaten und zufallsfolgen. *Acta Informatica*, 1(4):345–359, 1972.

[10] Benjamin Weiss. *Single orbit dynamics*. Number 95. American Mathematical Soc., 2000.