SOME REMARKS ON OUTPUT MEASURES

Tomasz Downarowicz and R. Daniel Mauldin

March 17, 2004

Abstract. An *output measure* is an image of a uniform Bernoulli measure on finitely many states. We discuss "generalized Bernoulli" measures of [F-J], and answer several questions posed in that paper, in particular we establish a condition when such a measure is a finitary output measure. Answering D. Goldstein's question, we characterize output measures via block codes in terms of walks on labeled graphs with special adjacency matrices.

1. Introduction

By a measure-preserving transformation one understands a system (X, Σ, μ, T), where (X, Σ, μ) is a standard probability space and $T : X \to X$ is a measurable invertible map preserving the measure μ. Two such systems (X, Σ, μ, T) and (X', Σ', μ', T') are *isomorphic* if there is a measurable and invertible map $\pi : X \to X'$ sending μ to μ' and *equivariant*, *i.e.,* such that $\pi \circ T = T' \circ \pi$ almost everywhere.

A finite partition $\mathcal{A}$ of X is a *generator* (or *alphabet*) if it distinguishes orbits of almost all points. In such case the system (X, Σ, μ, T) is isomorphic to the shift map on the set of all (bi-infinite) $\mathcal{A}$-names of points equipped with an appropriate measure. A measure-preserving transformation together with a distinguished generating partition and represented in the above mentioned symbolic form is often called a *process*.

Throughout this note, by a *dynamically Bernoulli process* (sometimes also called a *B-process*) we shall mean a finite state stationary process (*i.e.,* shift-invariant measure on sequences over a finite alphabet), which is isomorphic to an independent identically distributed (i.i.d.) process *i.e.,* to a product measure of the form $\sigma^{\mathbb{Z}}$ on $\mathcal{A}^{\mathbb{Z}}$, where σ is some distribution on a finite alphabet $\mathcal{A}$.

1991 *Mathematics Subject Classification.* Primary 37A35; Secondary 37B10.
Key words and phrases. finite block code, uniform Bernoulli measure, walks on labeled graphs.
The research of the first named author supported by KBN grant 2 P03 A 04622,
The research of the second named author supported by NSF DMS 0100078

There are several known characterizations of dynamically Bernoulli processes. In Ornstein's monograph [O3] they are described as *finitely determined* processes. This condition is hardly effectively verifiable as it requires comparing the weak convergence and the $\bar{d}$-convergence of other processes to the given one. A more effective characterization is the intrinsic *very weak Bernoulli* property. Ornstein [O2] and Ornstein and Weiss [O-W] proved its equivalence to the finitely determined property. Another fairly checkable equivalent condition is the *almost block independent* property of Shields' [Sh1]. We mainly follow the terminology of Shields' book [Sh2] in this note.

By a fundamental result of Ornstein [O2], the class of B-processes is closed under measure-theoretic (nontrivial) factors, and every such process has factors with arbitrary smaller positive entropy. A celebrated Ornstein's theorem [O1] asserts that two such processes with equal entropy are isomorphic. Thus, every dynamically Bernoulli measure can be thought of as an *output measure* defined as a measure-theoretic factor of a uniform Bernoulli process, *i.e.,* of the uniform product measure $\lambda_d = \{\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\}^{\mathbb{Z}}$ on d (finitely many) symbols. On the other hand, by definition, they are all isomorphic to finite state i.i.d. processes.

The notion of a "process", however, incorporates also a distinguished generating partition (or alphabet), and at this level some properties of various B-processes may vary. We can distinguish here several classes with particularly "nice" behavior. The most restrictive is, of course, the class of *Bernoulli processes*, *i.e.,* finite state i.i.d. processes. A wider (and the most extensively investigated) such class is *mixing Markov processes* proved to be dynamically Bernoulli by Friedman and Ornstein [F-O]. More general classes are: *weak Bernoulli processes* [F-O] and *block independent processes* [Sh1]. We refer the reader to [Sh2] for a fairly complete exposition on B-processes.

The work presented in this note has been inspired by the following problem:

(∗) Characterize the shift invariant measures which are output measures in the sense that they are images of λ_d via **topological** factor maps (*i.e.*, via finite block codes).

In a recent survey paper [F-J], Freiling and Jackson isolate another class of dynamically Bernoulli processes which they call "generalized Bernoulli". This class is interesting to us because, as mentioned in [F-J] (and shown in Section 4 of this note), all output measures via finite block codes belong to this class. The original proof of [F-J] that "generalized Bernoulli" processes are dynamically Bernoulli uses prefix codes and synchronizing words and leads to a non-finitary factor map. Unfortunately, as their paper is an abstract of a conference talk, it only sketches the proofs and does not refer to a more complete version. In Section 3 below we hope to provide a more detailed interpretation of certain notions introduced in [F-J], proofs of some facts presented there, and we answer some of the questions posed in that article. In particular, we provide a method for entropy calculation and we establish a criterion for such a process to be an output measure via a finitary coding.

For finite code output measures, there exist in the literature criteria for a more general class: finite block code images of Markov measures over a finite state space (with a transition probability mechanism), see Heller [H]. These criteria are given in a difficult language of algebraic properties of the map induced by the measure on the free algebra generated by the symbols of the subshift. It is desirable to find a simpler, combinatorial condition, one that distinguishes particularly the images of the uniform Bernoulli measure providing an answer to question $(*)$. In this vein, our work aims toward answering the following question formulated by Daniel Goldstein in February 2001 (we quote it with slight adaptation of the notation):

Question. *Let G be a directed graph whose adjacency matrix M satisfies the following:*

(a) M has row sums equal to an integer d;
(b) $\mathrm{tr} M > 0$;
(c) for some k all entries of $M + M^2 + \cdots + M^k$ are positive;
(d) M has a unique (complex) nonzero eigenvalue.

Suppose further that G is equipped with a labeling π with finitely many symbols. Define a function μ on words B over these symbols by the formula:

$$\mu(B) = d^{-|B|} \sum_{\tau:\pi(\tau)=B} a_{i(\tau)},$$

where τ ranges over all paths labeled as B, $i(\tau)$ denotes the initial vertex of τ, and a is the unique left probability eigenvector of M. This function extends to a shift invariant probability measure on the shift space. Is every such μ an output measure via a finite block code?

Notice that (a) implies that the unique right eigenvector of M is uniform. In section 5 we will answer this question positively even dropping the condition (a) – this will impose another term in the formula for μ; this term equals 1 if (a) is fulfilled. Our result does not pretend to be a satisfactory solution of question $(*)$ for one major reason: the language of our condition is not so different from that of the definition of a finite code output measure. Namely, it is obvious that a walk on a labeled full graph with d vertices (whose adjacency matrix is $\mathbf{1}_{d\times d}$ with all entries 1), is a code image of the full shift on d symbols and hence the image of the maximal entropy measure of such a walk is an output measure. Note that the matrix $\mathbf{1}_{d\times d}$ is positive and has rank one. Let us note that in answering Goldstein's question, we extend the class of allowable matrices to roots of positive matrices of rank one (we call them *root matrices*, see Lemma 6). Our proof follows almost directly from a special case of a deep and nontrivial result of Ashley that every such walk is in fact a code image of a full shift. We have decided to present our argument simply because the question has been around for quite a while and assembling the ingredients together has escaped the attention of specialists.

Let us also mention that one can attempt to relax the conditions on the matrix M as follows: it remains primitive, but it is allowed to have (in addition to a unique nonzero simple real eigenvalue) some complex eigenvalues. We do not know the answer to the analogous question involving such matrices.

The authors would like to thank Mike Boyle, Mike Keane and Karl Petersen for sharing their ideas and acquaintance with the literature of the matter. We also thank the referee for valuable remarks.

2. Notation

Let $\mathcal{A}$ be a finite or countable set. We endow the space $X := \mathcal{A}^{\mathbb{Z}}$ of doubly infinite sequences $x = (\dots x(-1)\underline{x(0)}x(1)\dots)$ $(x(i) \in \mathcal{A})$ with the product sigma field Σ and the product topology, where $\mathcal{A}$ is considered discrete. We underline the zero coordinate entry of a two-sided sequence. The shift map S given by $(Sx)(i) = x(i+1)$ is a homeomorphism. By a *block* of length l over $\mathcal{A}$ we shall mean any finite string $B = b(0)b(1)\dots b(l-1) \in \mathcal{A}^l$. By a *cylinder* $[B]$ *over a block* B in $\mathcal{B}^{\mathbb{Z}}$ we mean the set of all sequences reading as B at coordinates 0 through $l-1$:

$$[B] = \{x \in \mathcal{A}^z : x[0, l-1] = B\}.$$

Because the cylinders generate (via the shift map) the Borel sigma field Σ, every shift-invariant measure μ on X is determined by its values $\mu([B])$ on cylinders. By a *process* we will mean the system (X, Σ, μ, S) with its measurable and topological structures. For example, a *Bernoulli process* is obtained as the shift space on an alphabet $\mathcal{A}$, with the product (Bernoulli) measure $\mu = p^{\times\mathbb{Z}}$, where p is a probability distribution (vector) on $\mathcal{A}$, *i.e.*, $\mu([B]) = p(b(0))p(b(1))\dots p(b(l-1))$, where $B = b(0)b(1)\dots b(l-1)$.

If P is a not shift-invariant measure on Σ, we can still define an invariant measure $\mu = A(P)$ by averaging as follows:

$$\mu([B]) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P(S^{-i}([B])),$$

provided the limit exists for every cylinder $[B]$. In fact, it suffices to require this convergence for a smaller family of blocks as long as the corresponding cylinders generate.

If $\mathcal{D}$ and $\mathcal{A}$ are finite sets then by a *finite code* we will understand any map $\pi : \mathcal{D}^{2r+1} \to \mathcal{A}$. The parameter r is often called the *radius* of the code. This map extends to a map (denoted by the same letter) $\pi : \mathcal{D}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ as follows:

$(\pi x)(i) = \pi(x[i-r, i+r])$. The map π is shift equivariant: $\pi \circ S = S \circ \pi$. Similarly, we extend the original finite code π to a map sending blocks of length $2r + l$ over $\mathcal{D}$ to blocks of length l over $\mathcal{A}$.

3. Generalized Bernoulli measures

The notion of a generalized Bernoulli measure has been introduced by C. Freiling and S. Jackson in [F-J]. They proved that any process on two symbols endowed with such measure is an image of the Bernoulli process $\{\frac{1}{2}, \frac{1}{2}\}^{\times\mathbb{Z}}$ via what they call an "infinite block map". But in fact any Borel measurable equivariant map between subshifts is an "infinite block map" (see Lemma 1 below), therefore their statement is equivalent to saying that every generalized Bernoulli measure is a measure-theoretic factor of a Bernoulli measure, and it is known that such factors are themselves dynamically Bernoulli, *i.e.*, isomorphic to Bernoulli measures.

Lemma 1. *Every Borel measurable shift-equivariant map $\phi : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}'^{\mathbb{Z}}$ ($\mathcal{A}$ and $\mathcal{A}'$ are finite alphabets) is an "infinite block map" in the sense of [F-J], i.e., there is a Borel measurable finite partition $\alpha = \{U_a : a \in \mathcal{A}'\}$ of $\mathcal{A}^{\mathbb{Z}}$ such that $\phi(x)(n) = a \iff S^n(x) \in U_a$.*

Proof. Assign $U_a = \phi^{-1}([a])$. The rest is straightforward. □

As already mentioned, the paper of Freiling and Jackson presents only sketches of the proofs. In view of this, in this section we generalize their definition and provide a complete proof of shift invariance, Within this context, we are then able to derive the main result of [F-J] from the general theory. We hope this approach will cast more light on the main problem. Also, using Rudolph's result (generalizing the Keane-Smorodinsky technique of finitary coding), we answer a question posed in [J-F] about the existence of a finitary code.

Let $\mathcal{B} = \{B_1, B_2, B_3, \dots\}$ be a finite or countable collection of blocks of various lengths $l_i = |B_i| \in \mathbb{N}$, *i.e.*, $B_i = b_i(0)b_i(1)\dots b_i(l_i - 1)$ ($b_i(j) \in \mathcal{A}$) over a finite set of symbols $\mathcal{A}$. Let us extend the alphabet with some left and right markers. Set $\overline{\mathcal{A}} = \{a, {}^{\ulcorner}a, a^{\urcorner} : a \in \mathcal{A}\}$. We define the *marked unpacking map* ϕ^* from $\mathcal{B}^{\mathbb{Z}}$ into $X^* = \overline{\mathcal{A}}^{\mathbb{Z}}$ as follows:

$$\phi^*(\dots \underline{B_{i_0}} B_{i_1} \dots) = \dots \underline{{}^{\ulcorner}b_{i_0}(1)} b_{i_0}(2) \dots b_{i_0}(l_{i_0})^{\urcorner}\, {}^{\ulcorner}b_{i_1}(1) b_{i_1}(2) \dots b_{i_1}(l_{i_1})^{\urcorner} \dots$$

The above map consists in reading the "$\mathcal{A}$ contents" of the sequence of blocks and marking the beginning (with the "left end marker" $^{\ulcorner}$) and end (with the "right end marker" $^{\urcorner}$) of each source block. Notice that the markers appear in the image sequence in pairs allowing us to locate the breaking points between the original (source) blocks B_{i_j}. The map ϕ^* is hence invertible on its image $X_0^* = \phi^*(\mathcal{B}^{\mathbb{Z}})$.

Now suppose P is a shift-invariant measure on $\mathcal{B}^{\mathbb{Z}}$. The measure P is sent by ϕ^* to a measure P^* on X^*. Note that this new measure is not shift-invariant, because it is supported by the non-invariant set X_0^* of sequences in X^* having a symbol with the left end marker at coordinate zero. Nonetheless, if the needed limits exist, we can generate an invariant measure $\mu^* = A(P^*)$ by averaging. In order for this to succeed we need to assume that the P-expected length of the zero-coordinate block in $\mathcal{B}^{\mathbb{Z}}$, $\mathrm{E} = \mathrm{E}_P(l_{i_0})$, is finite. We will soon provide an appropriate lemma. We will also prove that the measure μ^* determines P^* as the conditional measure on X_0^*.

Once this is done, we can remove the markers and obtain a shift-invariant measure μ on $X = \mathcal{A}^{\mathbb{Z}}$. Formally, we define a one-block code (code of radius zero) by $\psi(\sigma) = a$, if $\sigma \in \{a, {}^{\lceil}a, a^{\rceil}\}$. Again, ψ will denote the generated map from X^* to X. We let $\mu = \psi(\mu^*)$. Note that the map ψ is no longer invertible.

Definition 1. The measures μ^* on X^* and μ on X will be called the *marked unpacked measure* and the *unpacked measure* generated by P.

The "generalized Bernoulli" measures of [F-J] are exactly the unpacked measures with $P = p^{\times\mathbb{Z}}$ (a Bernoulli measure on $\mathcal{B}^{\mathbb{Z}}$) with the additional property that the largest common divisor $\gcd(l_i)$ of the lengths l_i (such that $p_i := p(B_i) > 0$) is 1. This condition guarantees that the derived unpacked measure is mixing. Notice that in the generalized Bernoulli case we have $\mathrm{E} = \mathrm{E}_p(l_i) = \sum_i p_i l_i$ (the p-expected length of a block in $\mathcal{B}$).

We now proceed with the promised two lemmas concerning the general case.

Lemma 2. (comp. Theorem 12 in [F-J]) *If P is ergodic and $\mathrm{E} < \infty$, then μ^* is a well defined probability measure and it is ergodic.*

Proof. Let C^* be a finite block over $\overline{\mathcal{A}}$ which begins with a symbol with the left end marker and ends with a symbol with the right end marker, and which appears in an element of X^*. Note that every block appearing in the support of P^* extends (in both directions if necessary) to a block C^* of this form. In other words, cylinders over such blocks C^* generate the sigma algebra in X^*. Thus it suffices to prove the following: every block C^* of the above form appears in P^*-almost every sequence $x^* \in X^*$ with a density independent from the choice of x^* (this will automatically imply ergodicity of μ^*). But each such block C^* is determined in a 1-1 way (via the unpacking map ϕ^*) by a finite block D over $\mathcal{B}$. By ergodicity of P, D appears in P-almost every sequence y in $\mathcal{B}^{\mathbb{Z}}$ with density $P([D])$. The finite expected length statement implies, that for large n and sequences y from a set of nearly full measure P, the initial block $y[0, n-1]$ of y unpacks to a block of length approximately $n\mathrm{E}$. It is now immediately seen that C^* appears in P^*-almost every x^* with the density $\frac{P([D])}{\mathrm{E}}$ (which does not depend on x^*). $\square$

Lemma 3. *With assumptions of Lemma 2, P^* coincides with the conditional measure $\mu^*|X_0^*$.*

Proof. First notice that $\mu^*(X_0^*) = \frac{1}{\mathrm{E}} > 0$. Again, it suffices to examine blocks C^* starting and ending with the appropriate markers, and since every corresponding cylinder is contained in X_0^* we need to show that $\mu^*([C^*])/\mu(X_0^*) = P^*([C^*])$. As shown in the preceding proof, the first number is $\frac{P([D])}{\mathrm{E}}/\frac{1}{\mathrm{E}} = P([D])$. The second number is also $P([D])$, which follows directly because $P^* = \phi^*(P)$, $[C^*] = \phi^*([D])$, and ϕ^* is invertible. $\square$

We now pass to the case of a generalized Bernoulli measure. The main result of [F-J] asserts that every generalized Bernoulli measure is dynamically Bernoulli. Below we provide a short proof based on the general theory. Recall that an equivariant map between two processes is called *finitary* if it is continuous after discarding (from both spaces) sets of measure zero. Equivalently, for almost every x in the larger process the zero coordinate of its image can be determined by viewing a certain finite block in x. The key observation is that generalized Bernoulli measures are directly related to countable state Markov processes, a fairly well studied class. A good exposition on this subject can be found in Kitchens [K].

Lemma 4. *If $P = p^{\times\mathbb{Z}}$ is a Bernoulli measure on $\mathcal{B}^{\mathbb{Z}}$ with finite expected length $\mathrm{E}_p(l_i)$ then $(X^*, \Sigma^*, \mu^*, S)$ is finitarily isomorphic to a finite entropy countable state Markov process.*

Proof. Finite entropy is obvious, because μ^* is represented over a finite alphabet $\overline{\mathcal{A}}$. The process defined by μ^* is isomorphic to the following countable state Markov process: The state space is $\{(i,j) : i \in \mathbb{N},\ 1 \le j \le l_i\}$. The transition probabilities are: if $j < l_i$ then (i,j) is followed by $(i, j+1)$ with probability 1, each (i, l_i) is followed by $(i', 1)$ $(i' \in \mathbb{N})$ with probabilities $p_{i'}$. The isomorphism is by the one-block code replacing each symbol of the form $(i,1)$ by $\lceil b_i(1)$, each symbol (i, l_i) by $b_i(l_i)\rceil$, and other symbols (i,j) by $b_i(j)$. The inverse map is finitary; the block $\lceil b_i(1) b_i(2) \dots b_i(l_i)\rceil$ is replaced by $(i,1)(i,2)\dots(i,l_i)$. The verification that the image of the described Markov measure is indeed μ^* can be done as follows: the set X_0^* corresponds in the Markov process representation to the set $V = [(i,1)]$ of all sequences starting with $(i,1)$ for some i. The induced process on X_0^* is isomorphic to the Bernoulli process $(\mathcal{B}^{\mathbb{Z}}, P)$ and so is the corresponding induced process on V. $\square$

The result of [F-J] now follows directly from the fact that every finite entropy countable state mixing Markov process is weakly Bernoulli, hence dynamically Bernoulli (see [K]; Dan Rudolph has pointed out to us that even a more general class of mixing *renewal processes* is Bernoulli).

Freiling and Jackson ask about a quick way of calculating the entropy of a generalized Bernoulli process ([F-J], Question 4). In this direction we now note that there are some easily computable lower and upper bounds. Namely, the process induced on X_0^* is Bernoulli and has entropy $h(p) := \sum_i p_i \log p_i$, so the entropy of μ^* is $h_{\mu^*}(S) = h(p)\mu(X_0^*) = \frac{h(p)}{\mathrm{E}}$. Thus $h_\mu(S) \le \frac{h(p)}{\mathrm{E}}$ because μ is a factor of μ^*. On the other hand, the marked unpacked process is a subsystem of the direct product of the unpacked process with the process of markers. So $h_{\mu^*}(S) \le h_\mu(S) + h_\nu(S)$, where the latter is the entropy of the marker process. The marker process has two symbols $\{\text{marker}, \text{no marker}\}$ (we now interpret that our marker looks like $\rceil\lceil$), and the probability of seeing a marker is $\frac{1}{\mathrm{E}}$. Thus $h_\nu(S) \le h(\frac{1}{\mathrm{E}}) := \frac{\log \mathrm{E}}{\mathrm{E}} - (1 - \frac{1}{\mathrm{E}})\log(1 - \frac{1}{\mathrm{E}})$. As a result, we have proved

Theorem 1.

$$\frac{h(p)}{\mathrm{E}} - h(\tfrac{1}{\mathrm{E}}) \le h_\mu(S) \le \frac{h(p)}{\mathrm{E}}.$$

□

Notice that replacing the family of blocks $\mathcal{B}$ by their finite concatenations (with appropriately calculated probabilities) we can represent the same measure μ so that E is arbitrarily large. Then the estimates of Theorem 1 become very tight (which also implies that the value of $\frac{h(p)}{\mathrm{E}}$ (the entropy of μ^*) will not be affected too much by such change, if E is already large).

We will now answer Question 5 of [F-J] about the existence of a finitary map from a Bernoulli process to a generalized Bernoulli process. In the case of Markov processes, the existence of such map is equivalent to the following property defined by M. Smorodinsky:

Definition 2. A process has *exponentially decaying return times* if for every open set U eventually $\mu(U_n) < c^n$, for some $c < 1$ (depending on U), where

$$U_n = U \setminus \bigcup_{i=1}^{n} T^{-i}(U).$$

Since finitary maps preserve (by preimage) open sets up to measure zero, it is clear that the above property passes to finitary factors. A countable state mixing Markov process is finitarily isomorphic to a Bernoulli process if and only if it has exponentially decaying return times (see [R]).

Lemma 5. *Let p be a distribution on a family $\mathcal{B}$ of blocks over a finite alphabet $\mathcal{A}$. The following are equivalent:*

(a) the marked unpacked process $(X^*, \Sigma^*, \mu^*, S)$ has exponentially decaying return times;
(b) $(X^*, \Sigma^*, \mu^*, S)$ is finitarily isomorphic to a Bernoulli shift over a finite alphabet;
(c) $(X^*, \Sigma^*, \mu^*, S)$ is a finitary factor of the Bernoulli process on n symbols with the measure $\{\frac{1}{n}, \cdots, \frac{1}{n}\}^{\times\mathbb{Z}}$ for every $n > e^{h_{\mu^*}(S)}$;
(d) the distribution of lengths l_i is exponential, i.e.,

$$\sum_{\{i: l_i > n\}} p_i \le c^n$$

eventually, for some $c < 1$.

Proof. Equivalences (a) $\iff$ (b) $\iff$ (c) follow from Lemma 4, the remark following Definition 2, and the quoted result from [R]. We now prove (d) $\implies$ (a). Because every open set U contains a (perhaps shifted) cylinder of the form $[C^*]$ (see the proof of Lemma 2), it suffices to estimate the measure of the set $V_n = X \setminus \bigcup_{i=1}^{n} T^{-i}([C^*])$. The block C^* is a concatenation of some m blocks from $\mathcal{B}$ (with markers). The marked unpacked process defined by the family $\mathcal{B}^m$ of all m-concatenations over $\mathcal{B}$ (maintaining the markers and with appropriate product probabilities) is isomorphic to μ^* and also has exponential distribution of lengths. Thus, without loss of generality, we can assume that $C^* = \lceil B \rceil$ for some $B \in \mathcal{B}$. Let $l = |B|$ and let $q = p(B)$. The set V consists of sequences x for which this block $\lceil B \rceil$ does not occur in $x[1, n-l]$. Let k denote the number of unpacked marked blocks occurring entirely in $x[1, n-l]$. Then

$$\mu(V_n) \leq \sum_{k=1}^{n-l} (1-q)^k \cdot P'\{L_k \geq n-l\},$$

where L_k is the random variable denoting the joint length of the first k (complete) blocks in x, and P' is the conditional process deprived of the block B. Obviously, this conditional process also has exponential distribution of lengths with some (new) constant $c < 1$. Each L_k is a sum of independent copies of the single length variable, so it is exponential with the constant $c^{\frac{1}{k}}$. Hence we have the estimate

$$\mu(V_n) \leq \cdot \sum_{k=1}^{n-l} (1-q)^k \cdot c^{\frac{n-l}{k}} \leq (n-l)(c')^{n-l} \leq (c'')^n,$$

eventually, with $1 > c'' > c' = \max\{c, 1-q\}$. Finally we prove that (a) $\implies$ (d). The exponentially decaying return times condition of a fixed cylinder $[\lceil B \rceil]$ ($B \in \mathcal{B}$) implies that the probability that the block $\lceil B \rceil$ is followed by a block $\lceil B_i \rceil$ longer than n is, for large n, smaller than c^n. This directly implies (d). ☐

Theorem 2. (see [F-J], Question 5) *A generalized Bernoulli measure μ obtained from a distribution p with exponential distribution of block lengths is an image by a finitary map of a Bernoulli $\{\frac{1}{n}, \dots, \frac{1}{n}\}^{\times \mathbb{Z}}$ measure with $n > e^{h_\mu(S)}$. The exponential distribution assumption is essential.*

Proof. The positive statement follows directly from Lemma 5, the remarks following Theorem 1, and the fact that the generalized Bernoulli process is a one-block code image of μ^*. For a counterexample take any polynomial distribution of lengths on $\mathcal{B}$ containing blocks which already include the end markers and apply Lemma 5. ☐

Obviously, for the unmarked process there can't exist a length-related necessary criterion, because, by passing to concatenations, we can create a representation of the same process with a slowly decaying distribution of lengths.

We conclude this section with a short discussion referring to other questions formulated at the end of [F-J]. Question 1 is whether all "output measures" via "infinite block codes" are generalized Bernoulli. We can answer this question negatively using general facts from ergodic theory: As noted before, an "output measure" in the above meaning is synonymous to dynamically Bernoulli measure. Now applying the Jewett-Krieger theorem we obtain a strictly ergodic (minimal and uniquely ergodic) subshift on two symbols whose unique invariant measure μ is dynamically Bernoulli (and has entropy strictly between 0 and $\log 2$). We claim that μ does not provide a generalized Bernoulli process. This follows immediately from the simple observation that the topological support of any generalized Bernoulli measure is not minimal, because it contains an abundance of periodic sequences (periodic repetitions of the blocks from the family $\mathcal{B}$).

Question 2 of [F-J] – are two generalized Bernoulli measures of the same entropy necessarily isomorphic – is a misunderstanding. The positive answer follows directly from Ornstein's theory.

We would like to strengthen their last question: Is every dynamically Bernoulli generalized Bernoulli measure an "output measure" via a finitary map? We have answered this question in Theorem 2, but one could also ask the following

Question. With exponential distribution of lengths, can the expected coding length obtained be finite?

We remark that the expected coding length has been considered by Keane-Smorodinsky [K-S] and later by Serafin [Se] and Iwanik-Serafin [I-S] for codes between finite state Markov processes, but, to our knowledge, it has not been studied for countable state processes. Nonetheless, in our case the countable state process appears only as an auxiliary tool, while the question still concerns finite state processes.

We would also like to pose an inverted problem about strengthening the statement of Theorem 3 of the next section:

Question. Is every output measure via a finitary map a generalized Bernoulli measure? Does the same hold at least for codes with finite expected coding length?

4. Output measures are generalized Bernoulli

Finite block codes are of special interest because they represent topological factor maps between subshifts. Throughout this and the following section, by an *output measure* we will understand a shift-invariant measure $\mu = \pi(\lambda_d)$ on $X = \mathcal{A}^{\mathbb{Z}}$ obtained as the image of the standard Bernoulli measure $\lambda_d = \{\frac{1}{d}, \dots, \frac{1}{d}\}^{\times\mathbb{Z}}$ defined on $\mathcal{D}^{\mathbb{Z}}$, (where $\#\mathcal{D} = d$) via a finite block code $\pi : \mathcal{D}^{2r+1} \to \mathcal{A}$.

In this section we verify the relation between the output measures and generalized Bernoulli measures. The fact stated below appears in a similar formulation in [F-J], without a proof:

Theorem 3. *Every output measure $\nu = \pi(\lambda_d)$ is a generalized Bernoulli measure, generated by a sequence of blocks $\mathcal{B}$ with lengths l_i assuming all values larger than or equal to $2r$ with positive probabilities.*

Proof. Denote by Θ the block $0^r1^r = 00\dots011\dots1$ of r zeros followed by r ones. Define a new code $\pi^* : \mathcal{D}^{2r+1} \to \overline{\mathcal{A}}$ (and a corresponding map $\pi^* : X \to X^*$) by:

$$\pi^*(B) = \begin{cases} {}^{\ulcorner}\pi(B) & \text{if } B = \Theta\sigma \ (\sigma \in \mathcal{D}); \\ \pi(B)^{\urcorner} & \text{if } B = \sigma\Theta; \\ \pi(B) & \text{if } B \in \mathcal{D}^{2r+1}, \ B \neq \Theta\sigma \ \text{ and } \ B \neq \sigma\Theta. \end{cases}$$

Define $\mathcal{B}^*$ as the family of all blocks appearing in $\pi^*(X)$, beginning with a symbol with the left end marker, ending with a symbol with the right end marker and having no other markers inside. Then, let $\mathcal{B} = \psi(\mathcal{B}^*)$ (remove the markers). Or, in other words, $\mathcal{B}$ contains all images via the code π (extended to blocks) of blocks of lengths at least $4r$ beginning and ending with Θ and where Θ does not occur otherwise. It is important to notice that since the structure of Θ prevents it from having overlapping occurrences in a sequence in X, the shortest length of a block in B is $2r$ and there is exactly one block of this length; it is the π-image of $\Theta\Theta$). All larger lengths are represented in $\mathcal{B}$. Recall that the unpacking map ϕ^* from $\mathcal{B}^{\mathbb{Z}}$ to X^* is invertible on its image X_0^*. Notice that X_0^* consists of sequences having a symbol with the left end marker at coordinate 0, and it admits all possible infinite concatenations of the blocks from $\mathcal{B}^*$. Denote $\nu^* = \pi^*(\lambda_d)$ (a measure on X^*). Let P^* be the conditional measure $\nu^*|X_0^*$, and let $P = (\phi^*)^{-1}(P^*)$ (using invertibility of ϕ^*). The measure P is defined on $\mathcal{B}^{\mathbb{Z}}$. We need to show three facts:

1) P is a shift-invariant product measure (in particular, P is ergodic),
2) the expected length in $\mathcal{B}$ is finite,
3) ν^* coincides with the measure μ^* obtained from P^* via the limit (*), and
4) for each $l \geq 0$, P assigns positive values to some block of length $2r + l$.

This will end the proof, because then $\nu = \pi(\lambda_d) = \psi\pi^*(\lambda_d) = \psi(\nu^*) = \psi(\mu^*)$, which is generalized Bernoulli by definition.

For 1), we first calculate the probability $P\{z(0) = B_i\}$ ($B_i \in \mathcal{B}, \ z \in \mathcal{B}^Z$). By definition, it equals

$$P^*\{y^*[0, l_i - 1] = B_i^*\} = \nu^*\{y^*[0, l_i - 1] = B_i^* \mid y^*(0) = {}^{\ulcorner}a, \ (a \in \mathcal{A})\} = \\ \lambda_d\{x : \pi^*(x[-r, l_i + r - 1]) = B_i^* \mid x[-r, r-1] = \Theta\}.$$

Next we calculate the conditional probability $P\{z(k+1) = B_i \mid z(-\infty, k]\}$ for some $k \geq 0$. The condition is equivalent to fixing $y^*(-\infty, n-1]$ for some $n > 0$. In addition we know that $y^*(0) = {}^{\ulcorner}a$ and that $y^*(n-1) = a^{\urcorner}$. So, similarly as before, the requested probability equals

$$P^*\{y^*[n,n+l_i-1]=B_i^* \mid y^*(-\infty,n-1]\} =$$
$$\nu^*\{y^*[n,n+l_i-1]=B_i^* \mid y^*(-\infty,n-1]\}$$
$$\text{(the condition on } y^*(0) \text{ is already included)} \ =$$
$$\lambda_d\{x:\pi^*(x[n{-}r,n{+}l_i{+}r{-}1])=B_i^* \mid \pi^*(x(-\infty,n{+}r{-}1])\,\&\,x[n{-}r,n{+}r{-}1]=\Theta\}.$$

Because λ_d is Bernoulli, $x(-\infty,n+r-1]$ and $x[n-r,n+l_i+r-1]$ behave independently given $x[n-r,n+r-1]$, so the first part of the condition can be dropped. Then, by shift invariance of λ_d, we can also replace n by 0. Finally, we obtain

$$\lambda_d\{x:\pi^*(x[-r,l_i+r-1])=B_i^* \mid x[-r,r-1]=\Theta\},$$

which coincides with the formerly evaluated unconditional probability. A similar calculation yields the same result for $P\{z(k-1)=B_i \mid z[k,\infty)\}$ for $k\le 0$. This proves that P is a shift invariant product measure.

For 2) just note that the P-expected length E of B_i coincides with $2r$ plus the λ_d-expected length of the gap between two consecutive blocks Θ in X, which is obviously finite.

The condition 3) now follows automatically, because on one hand, by definition, $P^*=\nu^*|X_0^*$, on the other, by ergodicity of P and Lemma 3, $P^*=\mu^*|X_0^*$, both μ^* and ν^* are shift-invariant and ergodic (ν^* as a factor of the ergodic measure λ_d; μ^* by Lemma 2), and both these measures assign positive value $\frac{1}{\mathrm{E}}$ to X_0^*. Because an ergodic measure is determined by its restriction to a positive set, $\nu^*=\mu^*$.

To see that 4) holds observe that all lengths of gaps between consecutive occurrences of Θ appear with positive probability λ_d. $\square$

5. Subshifts of finite type associated to root matrices

In this section we provide a characterization of output measures as images via one-block maps (codes of radius zero) of maximal entropy measures supported by subshift of finite type whose adjacency matrices are root matrices, *i.e.*, have a rank one strictly positive power. Clearly, replacing the full shift by the corresponding subshift of finite type over $(2r+1)$-blocks, the output measures are immediately seen to be of the above described form. Only the converse observation is nontrivial. Let us start with the relevant definitions:

Definition 3. Let M be a non-negative integer-valued square $d\times d$ matrix. Let $G=G_M$ be the directed graph with the set of vertices $V=V_M=\{1,2,\dots,d\}$, and with the set of edges $E=E_M$ such that $M_{i,j}$ equals the number of edges in E with starting vertex i and ending vertex j. We call G *the graph of* M and we call M *the adjacency matrix of* G. The graph G determines M up to permutation of vertices.

Given a finite directed graph G, the subshift of finite type defined as the shift-invariant subset $X = X_G$ of $E^{\mathbb{Z}}$ by the rule

$$x = (x(n))_{n\in\mathbb{Z}} \in X \iff (\forall_{n\in\mathbb{Z}})\,\text{ending vertex of } x(n) = \text{ starting vertex of } x(n+1),$$

will be called *the edge walk on* G. Note that the above condition describes X by restricting only blocks of length 2.

Conversely, every SFT X, represented so that the forbidden blocks have length 2 (which is always possible up to topological conjugacy), is a walk on some graph (and hence is associated to some matrix), however, these objects are not uniquely determined.

Given a non-negative integer-valued square matrix M, we consider a *labeled directed graph* (G_M, π), where π is any map from E_G into a finite set of labels $\mathcal{A}$. The walk X_π on such labeled graph (obtained by only tracing the labels while walking on the graph) corresponds to a subshift obtained as the image of X_G by the one-block code π. Such walk needn't be a SFT, it belongs to the family of *sofic systems*, the topological factors of SFT's. (In fact every sofic system can be represented in such a way.)

Definition 4. A non-negative square $n \times n$ matrix M is called a *root matrix* if, for some integer r, M^r is strictly positive and has rank one.

Clearly, a power of a root matrix is again a root matrix. It is elementary (nonetheless we sketch the proof below) that root matrices are exactly the ones invoked in Goldstein's question:

Lemma 5. *Root matrices are characterized by the conditions (b), (c) and (d).*

Proof. Root matrices satisfy (c) by definition. Consider M as a linear map on $\mathbb{C}^n$. The rank one power M^r has a one-dimensional (over $\mathbb{C}$) range, which is the span of all eigenvectors. Since every eigenvector of M is an eigenvector of M^r, the span of all eigenvectors of M is also one-dimensional, so M satisfies (d). In other words, the characteristic polynomial of M is $t^{n-1}(d-t)$ (where d is the eigenvalue and n is the size of M). The trace of a matrix always equals the coefficient at t^{n-1}, in this case it equals d. So the trace is nonzero, and since the matrix has nonnegative integer entries, (b) holds. Moreover, the unique nonzero eigenvalue is a positive integer.

For the opposite implication note that (b) means that there is a looped vertex in the graph. By (c), from and to this looped vertex there lead paths connecting with all other vertices. This easily implies that a sufficiently high power M^s has all entries positive (*i.e.,* M is primitive). Every primitive matrix has a simple eigenvalue. By (d), it must be also unique, which imposes the above written form of the characteristic polynomial, and thus a one-dimensional span of all eigenvectors. Consider the action of M on any invariant subspace V of $\mathbb{C}^n$. This action also has a one-dimensional span of eigenvectors, which means a unique and simple eigenvalue d. Thus, unless the subspace is one-dimensional, zero is a solution of the characteristic equation of the matrix representing the action on V, *i.e.,* such action is singular. In particular, it reduces the dimension. This proves that the range of M^r is one-dimensional for some $r < n$. If M^r had a zero entry, it would have either a zero column or a zero row, which would then persist in higher powers s, contradicting primitivity. So M^r is strictly positive. $\square$

Many root matrices arise from conjugate representations of the full shift. So called *succession matrices*, the adjacency matrices in the higher-block representation of the full shift) are root matrices. Root matrices are also related to full shifts via a special case of a famous *Williams' conjecture* [W] – a still open problem whether a walk on a graph with a root adjacency matrix is topologically conjugate to a full shift (the original conjecture concerned *primitive matrices*, *i.e.*, roots of strictly positive matrices, but in this generality it has been proved false). A partial answer to this is contained in the following special case of a deep result of J. Ashley [A] (it seems this particular case existed before Ashley, but we failed to locate an earlier reference or identify the first author):

Theorem 4. *The edge walk X_G on a graph G associated to a root matrix M with eigenvalue d is the image of the full shift on d symbols via a bounded-to-one finite block code ϕ.* $\square$

Because bounded-to-one maps do not decrease entropy, the unique measure μ of maximal entropy of the above edge walk is the image via this code of the Bernoulli measure λ_d, hence is an output measure. The measure μ is identified as follows: let $a = (a_i)$ and $b = (b_i)$ denote the unique (up to rescaling) left and right eigenvectors of M. Choose a to be a probability vector and b to have the inner product with a equal to 1. By strict positivity of M^r, a and b have no zero entries. Now, if τ is a path of length k in G (a block appearing in the generated SFT) starting at a vertex $i(\tau)$ and ending a vertex $j(\tau)$ then

$$\mu([\tau]) = d^{-|\tau|} a_{i(\tau)} b_{j(\tau)}$$

(see e.g. [K]). In other words, μ defines a Markov process with the following transition mechanism: at a vertex i we choose an edge heading toward j with probability $(db_i)^{-1} b_j$ (it is easy to verify that the sum over the vertices j available from i of such probabilities is 1). In this interpretation a is the unique stationary initial distribution on the vertices.

Now consider a labeling π of G with labels in a set $\mathcal{A}$. The image μ_π of μ via π is obviously again an output measure. It can be identified by the values on cylinders $[B]$ over the alphabet $\mathcal{A}$:

$$(*) \qquad \mu_\pi([B]) = d^{-|B|} \sum_{\tau : \pi(\tau) = B} a_{i(\tau)} b_{j(\tau)}.$$

The theorem provided below is now an immediate consequence of the observations described above, yet the corresponding question was around without answer for quite a while, which, in our opinion, justifies publication.

Theorem 5. *A shift-invariant measure ν supported by $\mathcal{A}^{\mathbb{Z}}$ (where $\mathcal{A}$ is a finite set) is an output measure if and only if it has the form μ_π given in (*) for some labeling π of a directed graph G associated to a root matrix M. The non-zero eigenvalue d of M specifies the size of the alphabet $\mathcal{D}$ in the full shift whose standard Bernoulli measure λ_d codes down to μ.*

Proof. Sufficiency has been derived above. Necessity does not require any deep theorems (and has always been obvious): If ν is an output measure, *i.e.*, an image of the standard Bernoulli measure λ_d on d symbols via a finite code π of radius r, then π is a labeling of the graph associated with the $(d, 2r)$-succession matrix (the adjacency graph in the $2r$-block representation of the full shift), which is a root matrix, has eigenvalue d, and $\nu = \mu_\pi$. $\square$

References

[A] J. Ashley, *Resolving factor maps for shifts of finite type with equal entropy*, Ergodic Theory Dynam. Systems **11** (1991), 219–240.

[F] U.R. Fiebig, *A return time invariant for finitary isomorphisms*, Ergodic Theory Dynam. Sys. **4** (1984), 225–231.

[F-J] C. Freiling and S. Jackson, *Infinite block maps and generalized Bernoulli measures*, Real Analysis Exchange, Summer Symposium 2002, Lexington, 11–34.

[F-O] N.A. Friedman and D.S. Ornstein, *On isomorphism of weak Bernoulli transformations*, Adv. in Math. **5,** (1970), 365–394.

[H] A. Heller, *On stochastic processes derived from Markov chains*, Ann. Math. Stat. **36** (1965), 1286–1291.

[I-S] A. Iwanik and J. Serafin, *Code length between Markov processes*, Israel J. Math. **111** (1999), 29–51.

[K-S] M. Keane and M. Smorodinsky, *A class of finitary codes*, Israel J. Math. **26** (1977), 352–371.

[K] B.P. Kitchens, *Symbolic dynamics. One-sided, two-sided and countable state Markov shifts*, Springer, Universitext. Berlin, 1998.

[O1] D.S. Ornstein, *Bernoulli shifts with the same entropy are isomorphic*, Adv. in Math. **4,** (1970), 337–352.

[O2] D.S. Ornstein, *Imbedding Bernoulli shifts in flows*, Contributions to Ergodic Theory and Probability, Lecture Notes in Mathematics **160** (1970), Springer-Verlag, New York, 178–218.

[O3] D.S. Ornstein, *Ergodic theory, randomness, and dynamical systems*, Yale Mathematics Monographs 5, Yale Univ. Press, New Haven, CT, 1974.

[O-W] D.S. Ornstein and B. Weiss, *Finitely determined implies very weak Bernoulli*, Israel J. Math. **17** (1974), 94-104.

[R] D.J. Rudolph, *A mixing Markov chain with exponentially decaying return times is finitarily Bernoulli*, Ergodic Theory Dynam. Sys. **2** (1982), 85–97.

[Se] J. Serafin, *The finitary coding of two Bernoulli schemes with unequal entropies has finite expectation*, Indag. Math. vol 7 (1996), 503–519.

[Sh1] P.C. Shields, *Almost block independence*, Z. fur Wahr. **49** (1979), 119-123.

[Sh2] P.C. Shields, *The Ergodic Theory of Discrete Sample Paths,* Graduate Studies in Mathematics, vol. 13, AMS, 1996.
[W] R.F. Williams, *Classification of subshifts of finite type*, Ann. of Math. **98** (1973), 120–153.

Institute of Mathematics, Technical University, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland.

Department of Mathematics, Box 311430, University of North Texas, Denton, TX 76203 USA
E-mail address: downar@im.pwr.wroc.pl, mauldin@unt.edu