Ergod. Th. & Dynam. Sys. (1981), **1**, 285–301
Printed in Great Britain

Finitary codes between one-sided Bernoulli shifts

ANDRÉS DEL JUNCO
Department of Mathematics, Ohio State University, Columbus, Ohio 43210, *USA*

(*Received* 16 *February* 1981 *and revised* 7 *July* 1981)

Abstract. If p and q are probability vectors of equal entropy each having at least three non-zero components then there exists a finitary homomorphism between the corresponding one-sided Bernoulli shifts.

1. *Introduction*

Let $C=\{1,\ldots,n\}$ and $D=\{1,\ldots,m\}$ be finite sets. $X=C^{\mathbb{N}}$ (similarly $Y=D^{\mathbb{N}}$) denotes the space of one-sided sequences indexed by $\mathbb{N}=\{1,2,\ldots\}$ with values in C. Suppose that p and q are probability measures on C and D respectively and denote by μ (respectively ν) the infinite product measure $p^{\mathbb{N}}$ (respectively $q^{\mathbb{N}}$). The shift transformation σ acting on (X,μ) is called a one-sided Bernoulli shift. We also denote by τ the shift on (Y,ν). A homomorphism (or code) from (X,μ) to (Y,ν) is a measure-preserving map ϕ which commutes with the shifts, that is $\phi\circ\sigma=\tau\circ\phi$. ϕ is said to be finitary if, for a.a. $x\in X$, $\phi(x)(1)$ depends only on finitely many coordinates in x. More precisely, if $Q_i=\{y\in Y: y(1)=i\}$, then for each $i\in D$, $\phi^{-1}(Q_i)$ agrees a.e. with a countable union of cylinder sets in X (and by shift invariance the same is true of the inverse image of any finite cylinder in Y). The entropy $h(p)$ of p is defined by

$$h(p)=-\sum_{i\in C}p(i)\log p(i).$$

(We use $\log_2$ throughout.) The purpose of this paper is to prove the following result.

THEOREM 1. *If the probability vectors p and q each have at least three non-zero components and $h(p)=h(q)$, then there is a finitary homomorphism $\phi:(X,\mu)\to(Y,\nu)$.*

Theorem 1 is equivalent to the assertion that there exists a finitary and non-anticipating (one-sided) code Ψ between the two-sided Bernoulli shifts $(C^{\mathbf{Z}},p^{\mathbf{Z}})$ and $(D^{\mathbf{Z}},q^{\mathbf{Z}})$, that is for $p^{\mathbf{Z}}$-a.a. $x\in C^{\mathbf{Z}}$, $\Psi(x)(0)$ depends only on finitely many of the coordinates $x(0), x(1),\ldots$. If we drop the finitariness requirement, this result was proved by Sinai [**7**] without any restriction on p and q and for a general ergodic shift-invariant measure in place of $p^{\mathbf{Z}}$. Ornstein & Weiss [**6**] have given another proof of Sinai's theorem. Although the new feature here is the finitariness, the reader will notice that our argument gives the easiest and most elementary proof of Sinai's theorem (albeit in a less general setting). Meshalkin [5] constructed an isomorphism between the two-sided Bernoulli shifts based on $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ which is both finitary and non-anticipating, with an inverse which is finitary and forgetful (one-sided in the opposite direction). It is well known [8] that an isomorphism and its inverse cannot both be non-anticipating unless $p = q$.

Keane & Smorodinsky [4] have established the existence of a finitary isomorphism between any two Bernoulli shifts of equal entropy (also without restrictions on p and q). Our approach uses the Keane–Smorodinsky marker and skeleton technique, but with some essential differences which seem to be forced by the one-sidedness. In particular, it seems that one cannot use the marriage lemma of [4] in this context. This is a virtue as well as a necessity, for it makes the present approach almost completely elementary.

In § 2 we describe our choice of markers (the same as in [4]). § 3 defines skeleta in a one-sided way and defines their filler sets. § 4 describes our method of constructing joint measures, which we call the $*$-joining. Instead of the marriage lemma we use the elementary fact that one can always join two partitions P and Q so that fewer than $\#Q$ atoms of P are split. In § 5 we construct the superpositions (following the terminology of [2]) corresponding to each skeleton. Unlike previous work, these superpositions actually form a consistent system of measures. We conclude the proof by using these superpositions to construct directly a joint measure on $X \times Y$ which corresponds to the graph of a homomorphism.

We have not as yet been able to obtain theorem 1 in the case when p or q has only two components. In fact, it is not hard to see that it is impossible to construct the desired code by exactly matching markers, as in this paper, if the distribution of fillers after a marker is affected by that marker occurrence. (This is the reason for the three-state assumption – it allows us to use a word of length one as a marker, so that this does not happen.) However, we do have the following result.

THEOREM 2. *If* $(C^{\mathbb{N}}, \mu)$ *is a (one-sided) stationary ergodic Markov process and* $(D^{\mathbb{N}}, \nu)$ *is a Bernoulli shift with strictly lower entropy then there is a finitary homomorphism* $\phi : C^{\mathbb{N}} \to D^{\mathbb{N}}$.

Of course theorem 2 includes the case of general one-sided Bernoulli shifts with unequal entropies. Theorem 2 is proved by combining the methods of the present paper with those of [3]. For this purpose, the Rohlin-type marker of [3] should be chosen to have bounded return time, which is possible if one allows it to be finitary, instead of actually finite, as in [3]. We shall not give any details here.

As an application of theorem 2 we mention the following result which is a consequence of theorem 2 and the construction of Markov partitions for toral automorphisms [1].

THEOREM 3. *If* Φ *is an ergodic automorphism of the* 2*-torus and* $h(p) < h(\Phi)$, *then there is an independent partition for* Φ *with distribution p, each atom of which consists of a set of fibres in the expanding direction, that set being identified* (a.e.) *by a countable union of intervals in the contracting direction.*

https://doi.org/10.1017/S0143385700001267 Published online by Cambridge University Press

2. *Markers*

[**4**, lemma 2] enables us to assume that $p(1) = q(1)$. The symbol 1 will be used as a marker in both X and Y just as in [**4**], so we briefly review some facts from [**4**]. X and Y are fibred by the positions of marker occurrences as follows. For $x \in X$, $\hat{x} \in \hat{X} = \{0, 1\}^{\mathbb{N}}$ is defined by

$$\hat{x}(i) = \begin{cases} 1 & \text{if } x(i) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

For $\hat{x} \in \hat{X}$, $X(\hat{x})$ denotes the fibre over $\hat{x}$, that is $\{\xi \in X : \hat{\xi} = \hat{x}\}$. The projection of μ on $\hat{X}$, denoted by $\hat{\mu}$, is $\hat{p}^{\mathbb{N}}$, where $\hat{p}$ denotes the projection of p on $\{0, 1\}$ ($\hat{p}(1) = p(1)$, $\hat{p}(0) = 1 - p(1)$). Similar definitions apply to Y and evidently $\hat{\mu} = \hat{\nu}$. $\mu_{\hat{x}}$ (respectively $\nu_{\hat{x}}$) denotes the conditional measure on $X(\hat{x})$ (respectively $Y(\hat{x})$), so that

$$\mu = \int \mu_{\hat{x}}\, d\hat{\mu}(\hat{x}).$$

Setting $A = \{2, \ldots, n\}$, $B = \{2, \ldots, m\}$ and $I(\hat{x}) = \{i : \hat{x}(i) = 0\}$, $X(\hat{x})$ is naturally identified with $A^{I(\hat{x})}$ and with this identification $\mu_{\hat{x}}$ is $p_0^{I(\hat{x})}$, where p_0 denotes p conditioned on A. Similarly, $\nu_{\hat{x}}$ is $q_0^{I(\hat{x})}$, where q_0 is q conditioned on B.

3. *Skeleta*

Let $N_0 < N_1 < N_2 < \cdots$ be a sequence of positive integers to be specified later. For $r \geq 0$ a skeleton $\mathcal{S}$ of rank r consists of the integer r together with a sequence of 0s (blanks) and 1s indexed by some finite interval I in $\mathbb{N}$ (so that $\mathcal{S} \in \{0, 1\}^I$) which has the form

$$0^{m_1}1^{n_1}0^{m_2}1^{n_2} \cdots 0^{m_k}1^{n_k}, \tag{3.1}$$

where $m_i > 0$, $n_i > 0$ and

$$\max_{1 \leq i < k} n_i < N_r \leq n_k.$$

Note that the configuration (3.1) may have several possible ranks. We distinguish between subskeleta of different rank with the same configuration. $\{i \in I : \mathcal{S}(i) = 0\}$, the set of blank indices of $\mathcal{S}$, is denoted by $|\mathcal{S}|$.

A subskeleton $\bar{\mathcal{S}}$ of $\mathcal{S}$ is the restriction of $\mathcal{S}$ to some subinterval J of I such that $\bar{\mathcal{S}}$ ends with a full block of markers from $\mathcal{S}$ and $\bar{\mathcal{S}}$ is itself a skeleton with a rank not greater than that of $\mathcal{S}$. If $\bar{\mathcal{S}}$ is a subskeleton of $\mathcal{S}$ we write $\bar{\mathcal{S}} < \mathcal{S}$. Note that, if $j \in |\mathcal{S}|$, the restriction of $\mathcal{S}$ to $I \cap [j, \infty)$ is always a subskeleton of $\mathcal{S}$ which also has rank r and every subskeleton of full rank is of this form. This is the main difference between one- and two-sided skeleta, which gives rise to the new ideas needed here. Note also that subskeleta $\bar{\mathcal{S}}_1$ and $\bar{\mathcal{S}}_2$ of $\mathcal{S}$ may overlap without one containing the other. However, if we define a subskeleton of rank s to be rank-maximal if it is maximal among the subskeleta of $\mathcal{S}$ which have rank s (that is, it cannot be extended to the left in $\mathcal{S}$), we have the following lemma.

LEMMA 3.1. *If $\bar{\mathcal{S}}_1$ and $\bar{\mathcal{S}}_2$ are rank-maximal subskeleta of $\mathcal{S}$ then either one is a subskeleton of the other or $|\bar{\mathcal{S}}_1| \cap |\bar{\mathcal{S}}_2| = \emptyset$.*

Proof. If $|\bar{\mathcal{S}}_1| \cap |\bar{\mathcal{S}}_2|$ is not empty, then (interchanging $\bar{\mathcal{S}}_1$ and $\bar{\mathcal{S}}_2$ if necessary) the final marker block in $\bar{\mathcal{S}}_1$ is a marker block of $\bar{\mathcal{S}}_2$ so that $\bar{\mathcal{S}}_2$ may be extended to the left to cover $\bar{\mathcal{S}}_1$. Since $\bar{\mathcal{S}}_2$ is rank-maximal, it follows that $\bar{\mathcal{S}}_1 < \bar{\mathcal{S}}_2$. □

If $\mathcal{S}_t, \ldots, \mathcal{S}_0$ are the rank-maximal subskeleta of rank $r-1$ of a skeleton of rank r, listed in order of appearance from left to right, then the $\mathcal{S}_i$ are pairwise disjoint and $\bigcup_{i=0}^{t} |\mathcal{S}_i| = |\mathcal{S}|$. We write $\mathcal{S} = \mathcal{S}_t \times \cdots \times \mathcal{S}_0$ and refer to this as the rank-decomposition of $\mathcal{S}$. Note that the rank decomposition may consist of $\mathcal{S}$ alone. We will make frequent use of induction on rank $\mathcal{S}$ and the fact that, if $\bar{\mathcal{S}} < \mathcal{S}$ and rank $\bar{\mathcal{S}} <$ rank $\mathcal{S}$, then $\bar{\mathcal{S}} < \mathcal{S}_i$ for some i.

We write $l(\mathcal{S})$ for $\#|\mathcal{S}|$. For $i \in |\mathcal{S}|$ and $r \le \text{rank}\, \mathcal{S}$, $i \in |\bar{\mathcal{S}}|$ for a unique rank-maximal subskeleton $\bar{\mathcal{S}}$ of rank r which we denote by $\mathcal{S}(i, r)$. If the domain of definition of $\mathcal{S}(i, r)$ is I, say, we set

$$l(i, r, \mathcal{S}) = l(\mathcal{S}'),$$

where $\mathcal{S}'$ is the restriction of $\mathcal{S}(i, r)$ to $I \cap [i, \infty)$. Informally, $l(i, r, \mathcal{S})$ is the distance, measured in $|\mathcal{S}(i, r)|$, from i to the beginning of the rightmost marker run in $\mathcal{S}(i, r)$. We write $L(\mathcal{S}) = \min |\mathcal{S}|$ and $R(\mathcal{S}) = \max |\mathcal{S}|$.

Our next task in this section is to define suitable subsets $I(\mathcal{S})$ and $J(\mathcal{S})$ so that $A^{I(\mathcal{S})}$ and $B^{J(\mathcal{S})}$ can play the role of filler sets. Because of the method we will use to define superpositions on $A^{I(\mathcal{S})} \times B^{J(\mathcal{S})}$, it is crucial that $\#J(\mathcal{S})$ be bounded for skeletons of a given rank (see lemma 4.5). This accounts to a large extent for the complications in the definition of $I(\mathcal{S})$ and $J(\mathcal{S})$.

First we define a method of truncating skeleta. Let

$$0 < C_0 < C_1 < C_2 < \cdots$$

be a sequence of positive integers to be specified later. We define a set of 'stopping times' $\Gamma(\mathcal{S})$ depending on the skeleton $\mathcal{S}$ (but not on the filler, as in [**4**]) by

$$\Gamma(\mathcal{S}) = \{i \in |\mathcal{S}| : \text{for } 0 \le r \le \text{rank}\, \mathcal{S}, \quad l(i, r, \mathcal{S}) \ge C_r \quad \text{or} \quad i = L(\mathcal{S}(i, r))\}.$$

The above definition is obviously equivalent to the following inductive one, which is more useful in practice: for rank $\mathcal{S} = 0$,

$$\Gamma(\mathcal{S}) = \{i \in |\mathcal{S}| : l(i, 0, \mathcal{S}) \ge C_0 \quad \text{or} \quad i = L(\mathcal{S})\}$$

and for rank $\mathcal{S} = r$, $\mathcal{S} = \mathcal{S}_t \times \cdots \times \mathcal{S}_0$, $\Gamma(\mathcal{S})$ consists of those $i \in \bigcup_{j=1}^{t} \Gamma(\mathcal{S}_j)$ such that $l(i, r, \mathcal{S}) \ge C_r$, together with $L(\mathcal{S})$. We set

$$C(\mathcal{S}) = [\max \Gamma(\mathcal{S}), \infty) \cap |\mathcal{S}|.$$

It is immediate by induction that

$$\# C(\mathcal{S}) \le C_0 + C_1 + \cdots + C_r \quad \text{if rank } (\mathcal{S}) = r.$$

To illustrate this definition and the ones to follow choose $N_0 = 1$, $N_1 = 2$, $N_2 = 3$, $C_0 = 2$, $C_1 = 3$, $C_2 = 17$ and consider the following skeleton $\mathcal{S}$ of rank two and its $\Gamma(\mathcal{S})$ indicated by dots below the appropriate zeros. ($I(\mathcal{S})$ and $J(\mathcal{S})$ will be defined

https://doi.org/10.1017/S0143385700001267 Published online by Cambridge University Press

later.) The row below $\Gamma(\mathscr{S})$ gives the Γs for the 1-skeleta in the rank decomposition of $\mathscr{S}$.

```
   S  = 00100100001100100100001100101011001000100110111   (3.2)
Γ(S)    .  .  ..    .  .
        .  .  ..    .  .  ..    .       .  ..      .
I(S)                       ...  .. . .          .
J(S)                       ..   .. .
```

Note that

$$\mathscr{S} = \mathscr{S}_4 \times \mathscr{S}_3 \times \mathscr{S}_2 \times \mathscr{S}_1 \times \mathscr{S}_0,$$

where the $\mathscr{S}_i$ lie between runs of two or more 1s and

$$\mathscr{S}_i = \mathscr{S}_{i,2} \times \mathscr{S}_{i,1} \times \mathscr{S}_{i,0} \quad \text{for } i > 0,$$

while $\mathscr{S}_0$ has a single 0-skeleton $\mathscr{S}_{00}$ in its rank decomposition. (We cannot write $\mathscr{S}_0 = \mathscr{S}_{00}$ because they have different ranks.)

It will be crucial for us that the definition of $C(\mathscr{S})$ is non-anticipating in the sense that if $\bar{\mathscr{S}}$ is a full-rank subskeleton of $\mathscr{S}$ then

$$C(\bar{\mathscr{S}}) = C(\mathscr{S}) \cap |\bar{\mathscr{S}}|,$$

that is, one only has to look to the right of $i \in \mathbb{N}$ in order to know what $C(\mathscr{S})$ looks like to the right of i. This follows from a corresponding non-anticipating property of Γ, namely

$$\Gamma(\bar{\mathscr{S}}) = (\Gamma(\mathscr{S}) \cap |\bar{\mathscr{S}}|) \cup \{L(\mathscr{S})\},$$

which is immediate by induction on rank $\mathscr{S}$, using the inductive definition of $\Gamma(\mathscr{S})$. In the following lemma and throughout this paper $I < J$, where I and J are subsets of $\mathbb{N}$, means $i < j$ for $i \in I$, $j \in J$.

LEMMA 3.2. *If $\bar{\mathscr{S}} < \mathscr{S}$ then $C(\bar{\mathscr{S}})$ intersects $\Gamma(\mathscr{S})$ in at most* $\min C(\bar{\mathscr{S}})$. *In addition, $C(\bar{\mathscr{S}}) < C(\mathscr{S})$ or $C(\bar{\mathscr{S}}) \subset C(\mathscr{S})$.*

Proof. The first assertion says that $C(\bar{\mathscr{S}}) - \min C(\bar{\mathscr{S}})$ is caught between successive intervals of $\Gamma(\mathscr{S})$ (by which we mean to include the possibility that $C(\bar{\mathscr{S}}) \geq \max \Gamma(\mathscr{S})$). This is clearly true when rank $\mathscr{S} = 0$. To prove it in general we may assume that rank $\bar{\mathscr{S}} <$ rank $\mathscr{S}$ (otherwise $C(\bar{\mathscr{S}}) \geq \max \Gamma(\mathscr{S})$), so that $\bar{\mathscr{S}} < \mathscr{S}_i$ for some $\mathscr{S}_i$ in the rank decomposition of $\mathscr{S}$. By induction, $C(\bar{\mathscr{S}}) - \min C(\bar{\mathscr{S}})$ is caught between successive intervals of $\Gamma(\mathscr{S}_i)$, which implies the same for $\Gamma(\mathscr{S})$. This establishes the first assertion and the second is an immediate consequence. □

Let $m_1 < m_2 < \ldots$ be another sequence of positive integers to be specified later. If $\mathscr{S} = \mathscr{S}_t \times \cdots \times \mathscr{S}_0$ has rank r, say that $\mathscr{S}_i$ is principal if $i > m_r$ and $C(\mathscr{S}_i) \subset C(\mathscr{S})$. (There may be no principal $\mathscr{S}_i$.) Say that $\mathscr{S}_i$ is initial if $i \leq m_r$. Set

$$\dot{C}(\mathscr{S}) = C(\mathscr{S}) - R(\mathscr{S}).$$

We are now ready to define $I(\mathscr{S})$ and $J(\mathscr{S})$. If rank $\mathscr{S} = 0$ we set

$$I(\mathscr{S}) = J(\mathscr{S}) = \dot{C}(\mathscr{S}).$$

https://doi.org/10.1017/S0143385700001267 Published online by Cambridge University Press

If $\mathcal{S} = \mathcal{S}_t \times \cdots \times \mathcal{S}_0$ has rank r,

$$\bar{m} = \max\{i: C(\mathcal{S}_i) \subset C(\mathcal{S})\}$$

$$I(\mathcal{S}) = \bigcup\{C(\mathcal{S}_i): \mathcal{S}_i \text{ principal}\} \cup \{R(\mathcal{S}_i): 1 \le i \le \bar{m}\},$$

$$J(\mathcal{S}) = \bigcup\{\dot{C}(\mathcal{S}_i): \mathcal{S}_i \text{ principal}\}.$$

Note that these definitions are non-anticipating. Observe also that $I(\mathcal{S})$ and $J(\mathcal{S})$ are both contained in $\dot{C}(\mathcal{S})$. See (3.2) for an example of $I(\mathcal{S})$ and $J(\mathcal{S})$ with $m_1 = 0$, $m_2 = 1$. Note that in (3.2) $\mathcal{S}_3$ and $\mathcal{S}_2$ are principal and $\mathcal{S}_1$ and $\mathcal{S}_0$ are initial.

LEMMA 3.3. *If $\bar{\mathcal{S}} < \mathcal{S}$ then either $I(\bar{\mathcal{S}}) \subset I(\mathcal{S})$ or $I(\bar{\mathcal{S}}) \cap I(\mathcal{S}) = \varnothing$. In the first case $J(\bar{\mathcal{S}}) \subset J(\mathcal{S})$ and in the second case $J(\bar{\mathcal{S}}) \cap J(\mathcal{S}) = \varnothing$.*

Proof. We can assume that $I(\bar{\mathcal{S}}) \neq \varnothing$ (otherwise the lemma is vacuously true). We may also assume that rank $\bar{\mathcal{S}} <$ rank $\mathcal{S}$, otherwise $I(\bar{\mathcal{S}}) \subset I(\mathcal{S})$ and $J(\bar{\mathcal{S}}) \subset J(\mathcal{S})$ so we have finished. Thus $\bar{\mathcal{S}} < \mathcal{S}_i$ for some $\mathcal{S}_i$ in the rank decomposition of $\mathcal{S}$. We look at three cases:

(i) $C(\bar{\mathcal{S}}) \subset C(\mathcal{S}_i)$, $\mathcal{S}_i$ principal;
(ii) $C(\bar{\mathcal{S}}) \subset C(\mathcal{S}_i)$, $\mathcal{S}_i$ not principal;
(iii) $C(\bar{\mathcal{S}}) < C(\mathcal{S}_i)$.

In the first case

$$J(\bar{\mathcal{S}}) \subset I(\bar{\mathcal{S}}) \subset \dot{C}(\bar{\mathcal{S}}) \subset \dot{C}(\mathcal{S}_i)$$

so

$$I(\bar{\mathcal{S}}) \subset I(\mathcal{S}) \quad \text{and} \quad J(\bar{\mathcal{S}}) \subset J(\mathcal{S}).$$

In the second case

$$J(\bar{\mathcal{S}}) \subset I(\bar{\mathcal{S}}) \subset \dot{C}(\mathcal{S}_i)$$

so

$$I(\bar{\mathcal{S}}) \cap I(\mathcal{S}) = \varnothing \quad \text{and} \quad J(\bar{\mathcal{S}}) \cap J(\mathcal{S}) = \varnothing.$$

In the third case

$$I(\bar{\mathcal{S}}) \subset C(\bar{\mathcal{S}}) < C(\mathcal{S}_i)$$

so

$$I(\bar{\mathcal{S}}) \cap I(\mathcal{S}) = \varnothing \quad \text{and} \quad J(\bar{\mathcal{S}}) \cap J(\mathcal{S}) = \varnothing. \qquad \square$$

We conclude this section by defining two families of subskeleta of $\mathcal{S}$. The first of these, denoted by $\Delta(\mathcal{S})$, might be called the C-decomposition of $\mathcal{S}$ and has the property that the sets $C(\bar{\mathcal{S}})$, $\bar{\mathcal{S}} \in \Delta(\mathcal{S})$ are precisely the gaps in $\Gamma(\mathcal{S})$ (ignoring marker indices). It is defined inductively as follows. If rank $\mathcal{S} = 0$, then $\Delta(\mathcal{S}) = \mathcal{S}$; if $\mathcal{S} = \mathcal{S}_t \times \cdots \times \mathcal{S}_0$ has rank r, then

$$\Delta(\mathcal{S}) = \{\mathcal{S}\} \cup \left\{\bar{\mathcal{S}} \in \bigcup_{i=0}^{t} \Delta(\mathcal{S}_i): C(\bar{\mathcal{S}}) < C(\mathcal{S})\right\}.$$

$\Delta(\mathcal{S})$ is non-anticipating in the sense that, if $\mathcal{S}^*$ denotes the restriction of $\mathcal{S}$ to $[j, \infty) \cap |\mathcal{S}|$ for any fixed $j \in \mathbb{N}$, then

$$\Delta(\mathcal{S}^*) = \{\bar{\mathcal{S}}^*: \bar{\mathcal{S}} \in \Delta(\mathcal{S})\}.$$

Evidently, the sets $C(\bar{\mathcal{S}})$, $\bar{\mathcal{S}}\in\Delta(\mathcal{S})$, are just the sets

$$[\max I, \min I')\cap|\mathcal{S}|,$$

where I is an interval in $\Gamma(\mathcal{S})$ and I' is the next interval to the right. It follows from this and lemma 3.2 that, if $\bar{\mathcal{S}}<\mathcal{S}$, then $C(\bar{\mathcal{S}})\subset C(\hat{\mathcal{S}})$ for some $\hat{\mathcal{S}}\in\Delta(\mathcal{S})$.

Next we define $D(\mathcal{S})$, a family of subskeleta of $\mathcal{S}$ which could be called the I-decomposition of $C(\mathcal{S})$. For rank $\mathcal{S}=0$ set $D(\mathcal{S})=\{\mathcal{S}\}$. Now suppose $D(\mathcal{S})$ has been defined for rank $\mathcal{S}<r$ and that $\mathcal{S}=\mathcal{S}_t\times\cdots\times\mathcal{S}_0$ has rank r. For principal $\mathcal{S}_i$ let $\Lambda(\mathcal{S}_i)$ be the collection of skeleta $\bar{\mathcal{S}}$ in $\Delta(\mathcal{S}_i)$ such that $C(\bar{\mathcal{S}})<C(\mathcal{S}_i)$ and $C(\bar{\mathcal{S}})\subset C(\mathcal{S})$; for initial $\mathcal{S}_i$ let $\Lambda(\mathcal{S}_i)$ be $\Delta(\mathcal{S}_i)$ and set

$$D(\mathcal{S})=\{\mathcal{S}\}\cup\bigcup\{D(\bar{\mathcal{S}}): \bar{\mathcal{S}}\in\Lambda(\mathcal{S}_i),\ \mathcal{S}_i \text{ initial or principal}\}.$$

Notice that the definition of $D(\mathcal{S})$ is non-anticipating in the same sense as $\Delta(\mathcal{S})$.

To check his understanding of $\Delta(\mathcal{S})$ and $D(\mathcal{S})$ the reader may verify that in (3.2)

$$\Delta(\mathcal{S})=\{\mathcal{S}, \mathcal{S}_4, \mathcal{S}_{4,2}, \mathcal{S}_{4,1}, \mathcal{S}_{3,2}\},$$
$$D(\mathcal{S})=\{\mathcal{S}, \mathcal{S}_{3,1}, \mathcal{S}_1, \mathcal{S}_{1,2}, \mathcal{S}_{1,0}, \mathcal{S}_{0,0}\}.$$

In connection with the following lemma note that in (3.2) $I(\mathcal{S}_1)$ consists of the middle zero in $\mathcal{S}_1$, while I of any 0-skeleton is just the second zero from the right in that 0-skeleton.

LEMMA 3.4 *The sets* $I(\bar{\mathcal{S}})$, $\bar{\mathcal{S}}\in D(\mathcal{S})$ *are disjoint and contained in* $\dot{C}(\mathcal{S})$. *The* $J(\bar{\mathcal{S}})$, $\bar{\mathcal{S}}\in D(\mathcal{S})$, *are disjoint and contained in* $\dot{C}(\mathcal{S})$. *If* $\bar{\mathcal{S}}<\mathcal{S}$ *and* $I(\bar{\mathcal{S}})\subset C(\mathcal{S})$, *then* $I(\bar{\mathcal{S}})\subset I(\mathcal{S}')$ *for some* $\mathcal{S}'\in D(\mathcal{S})$.

Proof. The first assertion is easily proved by induction on rank $\mathcal{S}$ and the second assertion follows from the first. For the third assertion, if rank $\bar{\mathcal{S}}=$ rank $\mathcal{S}$, then $I(\bar{\mathcal{S}})\subset I(\mathcal{S})$ and $\mathcal{S}\in D(\mathcal{S})$, so we may assume that rank $\bar{\mathcal{S}}<$ rank $\mathcal{S}$. Then $\bar{\mathcal{S}}<\mathcal{S}_i$ for some initial or principal $\mathcal{S}_i$ in the rank decomposition of $\mathcal{S}$, and we consider three cases:

(i) $C(\bar{\mathcal{S}})\subset C(\mathcal{S}_i)$, $\mathcal{S}_i$ principal;

(ii) $C(\bar{\mathcal{S}})<C(\mathcal{S}_i)$, $\mathcal{S}_i$ principal;

(iii) $\mathcal{S}_i$ initial.

In case (i) $I(\bar{\mathcal{S}})\subset I(\mathcal{S})$ and $\mathcal{S}\in D(\mathcal{S})$. In both cases (ii) and (iii)

$$C(\bar{\mathcal{S}})\subset C(\hat{\mathcal{S}})\quad\text{for some } \hat{\mathcal{S}}\in\Delta(\mathcal{S}_i)$$

and then, by induction,

$$I(\bar{\mathcal{S}})\subset I(\mathcal{S}')\quad\text{for some } \mathcal{S}'\in D(\hat{\mathcal{S}}).$$

In case (ii) because $C(\bar{\mathcal{S}})<C(\mathcal{S}_i)$ we see that also

$$C(\hat{\mathcal{S}})<C(\mathcal{S}_i)\quad\text{so}\quad \hat{\mathcal{S}}\in\Lambda(\mathcal{S}_i).$$

In case (iii) $\hat{\mathcal{S}}\in\Lambda(\mathcal{S}_i)$ is immediate. Since $\mathcal{S}'\in D(\hat{\mathcal{S}})$ in both cases we have $\mathcal{S}'\in D(\mathcal{S})$ as desired. □

4. **-joinings and superpositions*

Suppose ρ and σ are probability measures on finite sets E and F each of which is totally ordered so that

$$E=\{e_1<e_2<\cdots<e_r\},\quad F=\{f_1<f_2<\cdots<f_s\}.$$

We define a joining $\rho \cdot \sigma$ of ρ and σ (that is, a measure on $E \times F$ with marginals ρ and σ) as follows. Let

$$0 = x_0 < x_1 < \cdots < x_r = 1$$

be points in [0, 1] such that

$$\lambda(x_{i-1}, x_i) = \rho(e_i) \quad \text{for } 1 \leq i \leq r,$$

where λ denotes Lebesgue measure. Similarly,

$$0 = y_0 < y_1 < \cdots < y_s = 1$$

are such that

$$\lambda(y_{j-1}, y_j) = \sigma(f_j) \quad \text{for } 1 \leq j \leq s.$$

Set

$$\rho \cdot \sigma(e_i, f_j) = \lambda((x_{i-1}, x_i) \cap (y_{j-1}, y_j)).$$

The proof of the following lemma is immediate from the definition.

LEMMA 4.1. *Suppose E, F and G are finite totally ordered sets and $E \times F$ is given the reverse lexicographic ordering:*

$$(e_1, f_1) < (e_2, f_2) \Leftrightarrow f_1 < f_2 \text{ or } f_1 = f_2 \text{ and } e_1 < e_2.$$

If ρ and σ are probability measures on $E \times F$ and G respectively and $\bar{\rho}$ is the marginal of ρ on F, then the marginal of $\rho \cdot \sigma$ on $F \times G$ is $\bar{\rho} \cdot \sigma$.

Next suppose π_2 and π_1 are probability measures on $E_2 \times F_2$ and $E_1 \times F_1$ respectively, where E_1 and F_2 are totally ordered. For $i = 2, 1$ we denote by P_i the partition of $E_i \times F_i$ according to the E_i coordinate and by Q_i the partition of $E_i \times F_i$ according to the F_i coordinate. By a slight abuse of notation P_i (respectively Q_i) will also be used for the partition of $(E_2 \times E_1) \times (F_2 \times F_1)$ according to the E_i (respectively F_i) coordinate. For $q_1 \in Q_1$, $d_{\pi_1}(P_1|q_1)$ denotes the conditional distribution of P_1 on q_1, with respect to π_1. $d_{\pi_1}(P_1|q_1)$ is a measure on the totally ordered set E_1 and similarly for $p_2 \in P_2$, $d_{\pi_2}(Q_2|p_2)$ is a measure on F_2, so that

$$d_{\pi_2}(Q_2|p_2) \cdot d_{\pi_1}(P_1|q_1)$$

is meaningful. Setting $E = E_2 \times E_1$ and $F = F_2 \times F_1$, we now define a joining $\pi = \pi_2 * \pi_1$ on $E \times F$ by decreeing that

$$d_\pi(P_2) = d_{\pi_2}(P_2), \quad d_\pi(Q_1) = d_{\pi_1}(Q_1),$$

P_2 and Q_1 are independent on $E \times F$ and, for each atom $p_2 q_1$ of $P_2 \vee Q_1$ on $E \times F$,

$$d_\pi(Q_2 P_1 | p_2 q_1) = d_{\pi_2}(Q_2|p_2) \cdot d_{\pi_1}(P_1|q_1).$$

Note that $\pi_2 * \pi_1 \neq \pi_1 * \pi_2$ (in fact $\pi_1 * \pi_2$ is not defined in general).

LEMMA 4.2. *π is a joining of π_2 and π_1. Moreover, with the above notation P_2 is independent of $P_1 \vee Q_1$ with respect to π and Q_1 is independent of $P_2 \vee Q_2$ with respect to π.*

Proof. Since

$$d_\pi(Q_2 P_1 | p_2 q_1) = d_{\pi_2}(Q_2|p_2) \cdot d_{\pi_1}(P_1|q_1)$$

we have

$$d_\pi(Q_2|p_2q_1)=d_{\pi_2}(Q_2|p_2)$$

and hence

$$d_\pi(Q_2|p_2)=d_{\pi_2}(Q_2|p_2).$$

Since $d_\pi(P_2)=d_{\pi_2}(P_2)$ it follows that

$$d_\pi(Q_2\vee P_2)=d_{\pi_2}(Q_2\vee P_2).$$

Similarly, π has marginal π_1. Moreover, we have just observed that (with respect to π) Q_2 and Q_1 are independent given P_2. Since Q_1 and P_2 are independent by construction, $Q_2\vee P_2$ is independent of Q_1. Similarly, P_2 is independent of $P_1\vee Q_1$. □

If I and J are finite subsets of $\mathbb{N}$ we denote by μ_I and ν_J the measures p_0^I and q_0^J on A^I and B^J respectively. For a finite set $I\subset I'\subset\mathbb{N}$ we write P^I for the partition of $A^{I'}$ according to coordinates in I. Q^I has the analogous meaning for $B^{I'}$. A measure π on $A^I\times B^J$ with marginals μ_I and ν_J will be called a superposition. π will be called non-anticipating if, for each $t\in\mathbb{N}$,

$$P^{I\cap[1,t)}\perp P^{I\cap[t,\infty)}\vee Q^{J\cap[t,\infty)}\quad\text{(with respect to }\pi).$$

We now fix once and for all total orderings of A and B. For finite $I\subset\mathbb{N}$, A^I (similarly B^I) is totally ordered by the reverse lexicographic ordering:

for $x,\bar{x}\in A^I$, $x<\bar{x}\Leftrightarrow$for some $i_0\in I$, $x(i_0)<\bar{x}(i_0)$ and $x(i)=\bar{x}(i)$ for all $i\in I$, $i>i_0$.

Thus, if π_2 and π_1 are superpositions, $\pi_2*\pi_1$ is meaningful. In the following lemma, and throughout the rest of the paper, $\pi_2\times\pi_1$ denotes product measure.

LEMMA 4.3. *Suppose I_2, I_1, J_2 and J_1 are finite subsets of $\mathbb{N}$ such that $I_2\cap I_1=\varnothing$, $J_2\subset I_2$ and $J_1\subset I_1$, and suppose that π_i is a superposition on $A^{I_i}\times B^{J_i}$ $(i=2,1)$. Then we have*

(*a*) *$\pi_2\times\pi_1$ and $\pi_2*\pi_1$ are superpositions on $A^{I_2\cup I_1}\times B^{J_2\cup J_1}$.*
(*b*) *If π_2 and π_1 are non-anticipating then so is $\pi_2\times\pi_1$.*
(*c*) *If π_2 and π_1 are non-anticipating and $I_2<I_1$ then $\pi_2*\pi_1$ is non-anticipating.*

Proof. $\pi_2\times\pi_1$ is obviously a superposition and $\pi_2*\pi_1$ is a superposition by lemma 4.2. It is easy to see that $\pi_2\times\pi_1$ is non-anticipating if π_2 and π_1 are. Setting $\pi=\pi_2*\pi_1$, to prove (c) we must show that (with respect to π)

$$P^{(I_2\cup I_1)\cap[1,t)}\perp P^{(I_2\cup I_1)\cap[t,\infty)}\vee Q^{(J_2\cup J_1)\cap[t,\infty)},\tag{1}$$

and we may as well assume that $t\in I_2$ or $t\in I_1$. If $t\in I_1$, (1) reduces to

$$P^{I_2\cup(I_1\cap[1,t))}\perp P^{I_1\cap[t,\infty)}\vee Q^{J_1\cap[t,\infty)},$$

which is true since π_1 is non-anticipating and $P^{I_2}\perp P^{I_1}\vee Q^{J_1}$, by lemma 4.2. If $t\in I_2$, (1) becomes

$$P^{I_2\cap[1,t)}\perp P^{(I_2\cap[t,\infty))\cup I_1}\vee Q^{(J_2\cap[t,\infty))\cup J_1}.\tag{2}$$

Let $p_2'\in P^{I_2\cap[1,t)}$, $p_2^*\in P^{I_2\cap[t,\infty)}$ and $q_1\in Q^{J_1}$. Because

$$d_\pi(Q^{J_2}P^{I_1}|p_2'p_2^*q_1)=d_{\pi_2}(Q^{J_2}|p_2'p_2^*)\cdot d_{\pi_1}(P^{I_1}|q_1),$$

https://doi.org/10.1017/S0143385700001267 Published online by Cambridge University Press

because of lemma 4.1 and because of the way B^{J_2} is ordered we see that

$$d_\pi(Q^{J_2\cap[t,\infty)}P^{I_1}|p_2'p_2^*q_1)=d_{\pi_2}(Q^{J_2\cap[t,\infty)}|p_2'p_2^*)\cdot d_{\pi_1}(P^{I_1}|q_1). \qquad (3)$$

Since π_2 is non-anticipating we see that the distribution on the right of (3) does not depend on p_2', so the same is true of the distribution on the left. On the other hand, p_2' and $p_2^*q_1$ are independent (since $p_2'p_2^* \perp q_1$ and $p_2' \perp p_2^*$) so (2) follows. □

LEMMA 4.4. *Suppose $I_2<I_1$, $J_2\subset I_2$ and $J_1\subset I_1$ are all finite subsets of* ℕ *and π_2 and π_1 are superpositions on $A^{I_2}\times B^{J_2}$ and $A^{I_1}\times B^{J_1}$ respectively. For $t\in I_2$ set*

$$I_2^*=I_2\cap[t,\infty), \quad J_2^*=J_2\cap[t,\infty)$$

and let π_2^* denote the marginal of π_2 on $A^{I_2^*}\times B^{J_2^*}$. If π_2 is non-anticipating then the marginal of $\pi_2*\pi_1$ on $A^{I_2^*\cup I_1}\times B^{J_2^*\cup J_1}$ is $\pi_2^**\pi_1$.

Proof. Adopting the notation in the proof of lemma 4.3(*c*), it follows from (3) in that proof that

$$d_\pi(Q^{J_2^*}P^{I_1}|p^*q_1)=d_{\pi_2}(Q^{J_2^*}|p_2^*)\cdot d_{\pi_1}(P^{I_1}|q_1).$$

Moreover, p_2^* and q_1 are independent with respect to π (since $P^{I_2}\perp Q^{J_1}$), which establishes the lemma. □

We say a superposition π on $A^I\times B^J$ splits $p\in P^I$ if $\pi(p,q)\neq 0$ and $\pi(p,q')\neq 0$ for distinct $q, q'\in Q^J$. Say p is contained (with respect to π) in $q\in Q^J$ if $\pi(pq)=\mu_I(p)$.

LEMMA 4.5. *Suppose $I_n<I_{n-1}<\cdots<I_0$ are finite subsets of* ℕ, *$J_i\subset I_i$ for $i=n,\ldots,1$, π_i is a superposition on $A^{I_i}\times B^{J_i}$ for $i=n,\ldots,1$ and set $\pi_0=\mu_{I_0}$. Write*

$$l_j=\sum_{i=0}^{j}\#I_i,$$

$$P^j=P^{I_j\cup\cdots\cup I_0}, \quad \mu_j=\mu_{I_j\cup\cdots\cup I_0},$$

for $j=n,\ldots,0$ and

$$Q^j=Q^{J_j\cup\cdots\cup J_1}, \quad \nu_j=\nu_{J_j\cup\cdots\cup J_1}$$

for $j=n,\ldots,1$. Set $h=h(p_0)=h(q_0)$ and fix $\varepsilon>0$. If $p^i\in P^{I_i}$ for $j\geq i\geq 0$, $p=p^j\cdots p^0\in P^j$ is called good if

$$\mu_j(p)<2^{-(h-\varepsilon)l_j}.$$

p is called completely good (c.g.) *if $p^i\ldots p^0$ is good for all $j\geq i\geq 0$. For $j\geq 1$ call $q=q^j\cdots q^1\in Q^j$ good if*

$$\nu_j(q)>2^{-(h-2\varepsilon)l_j}$$

and completely good if $q^i\ldots q^1$ is good for all $j\geq i\geq 1$. Finally, say that $p\in P^j$ is desirable if p is not split by

$$\Pi_j=\pi_j*(\pi_{j-1}*(\cdots*(\pi_1*\pi_0)\cdots)),$$

p is contained (with respect to Π_j) in a completely good $q\in Q^j$ and p is completely good. Then, setting

$$\rho_j=\mu_j\{p\in P^j: p \text{ is not desirable}\},$$

https://doi.org/10.1017/S0143385700001267 Published online by Cambridge University Press

we have

$$\rho_i \leq \mu_j\{p \in P^j : p \text{ is not c.g.}\} + \nu_j\{q \in Q^j : q \text{ is not c.g.}\}$$
$$+ M \sum_{i=0}^{j-1} 2^{-\varepsilon(i+m)},$$

where $M = \max_{1 \leq i \leq n} \# B^{J_i}$, $m = \# I_0$.

Proof. Notice that the definition of desirable is meaningful for $p \in P^0$ and that we trivially have the estimate

$$\rho_0 < \mu_0\{p \in P^0 : p \text{ is not c.g.}\}.$$

To prove the lemma by induction it suffices to assume that the estimate on ρ_j holds for $j = n-1$ and prove it holds for $j = n$. For $p \in P^n$ write

$$p = p^n p^*, \quad p^n \in P^{I_n}, \quad p^* \in P^{n-1}.$$

We obviously have

$$\rho_n \leq \rho_{n-1} + \mu_n(\Gamma), \tag{1}$$

where

$$\Gamma = \{p^n p^* \in P^n : p^* \text{ is desirable but } p^n p^* \text{ is not}\}.$$

We claim that

$$\mu_n(\Gamma) < M 2^{-\varepsilon(m+n-1)} + \mu_n\{\text{bad } p^n p^* \in P^n : p^* \text{ is c.g.}\}$$
$$+ \nu_n\{\text{bad } q^n q^* \in Q^n : q^* \text{ is c.g.}\}. \tag{2}$$

To see this first observe that if $p \in \Gamma$ then p belongs to one of the following sets:

$$E_1 = \Gamma \cap \{p \in P^n : p \text{ is split}\},$$
$$E_2 = \Gamma \cap \{p \in P^n : p \text{ is contained in a good } q \in Q^n\},$$
$$E_3 = \Gamma \cap \{p \in P^n : p \text{ is contained in a bad } q \in Q^n\}.$$

We estimate the measure of E_1 by regarding it as a subset of $P^n \times Q^n$ and conditioning it on sets of the form $p_1^n q^*$, where $p_1^n \in P^{I_n}$, $q^* \in Q^{n-1}$. Since E_1 is contained in the union of completely good Q^{n-1} atoms, we may assume that q^* is completely good. Fixing $p_1^n q^*$, if $p^n p^* \in E_1$, then $p^n p^* \cap p_1^n q^*$ is either empty or equal to $p^n p^*$ (since p^* is not split). In the second case

$$p^* \cap p_1^n p^* = p^n p^* = p_1^n p^*$$

is split by $Q^{J_n} \cap p_1^n q^*$ (otherwise $p^n p^*$ would not be split by Q^n). Thus to estimate $\Pi_n(E_1 | p_1^n q^*)$ it suffices to estimate the $p_1^n q^*$-conditional measure of desirable $p^* \in P^{n-1}$ such that $p^* \subset q^*$ and $p^* \cap p_1^n q^*$ is split by $Q^{J_n} \cap p_1^n q^*$. By the definition of $\Pi_n = \pi_n * \Pi_{n-1}$, there are fewer than $\# B^{J_n}$ such p^*, and the conditional measure $\Pi_n(p^* | p_1^n q^*)$ of such a p^* is $\Pi_n(p^* | q^*)$ (lemma 4.2) which is estimated by

$$\Pi_n(p^* | q^*) \leq 2^{-\varepsilon l_{n-1}},$$

since $p^* \subset q^*$ and both are completely good. Thus

$$\Pi_n(E_1 | p_1^n q^*) \leq \# B^{J_n} 2^{-\varepsilon l_{n-1}} \leq M 2^{-\varepsilon(n-1+m)},$$

whence also

$$\mu_n(E_1) = \Pi_n(E_1) \leq M 2^{-\varepsilon(n-1+m)}.$$

Now, if $p \in E_2$, p is bad (otherwise p would be desirable) so $\pi(E_2)$ is less than the second term on the right of (2), while $\pi(E_3)$ is clearly less than the third term. This establishes (2).

Now by (1), (2) and our induction hypothesis we have

$$\begin{aligned}\rho_n \le &\mu_{n-1}\{p \in P^{n-1}: p \text{ is not c.g.}\} + \mu_n \{\text{bad } p^n p^* \in P^n : p^* \text{ is c.g.}\}\\ &+ \nu_{n-1}\{q \in Q^{n-1}: q \text{ is not c.g.}\} + \nu_n \{\text{bad } q^n q^* \in Q^n : q^* \text{ is c.g.}\}\\ &+ M \sum_{i=0}^{n-2} 2^{-\varepsilon(i+m)} + M 2^{-\varepsilon(n-1+m)}\\ = &\mu_n\{p \in P^n : p \text{ is not c.g.}\} + \nu_n\{q \in Q^n : q \text{ is not c.g.}\} + M \sum_{i=0}^{n-1} 2^{-\varepsilon(i+m)}. \quad \square\end{aligned}$$

5. *Construction of superpositions and proof of theorem 1*

We now define for each skeleton $\mathcal{S}$ a superposition $\pi_{\mathcal{S}}$ on $A^{I(\mathcal{S})} \times B^{J(\mathcal{S})}$. For 0-skeleta $\mathcal{S}$,

$$\pi_{\mathcal{S}} = \mu_{I(\mathcal{S})} \times \nu_{J(\mathcal{S})}.$$

Now suppose $\pi_{\mathcal{S}}$ has been defined for rank $\mathcal{S} < r$ and $\mathcal{S} = \mathcal{S}_t \times \cdots \times \mathcal{S}_0$ has rank r. For each $i \ge 1$ such that $\mathcal{S}_{i+m_r}$ is principal set

$$I_i = C(\mathcal{S}_{i+m_r}), \quad \bar{I}_i = I_i - \bigcup\{I(\bar{\mathcal{S}}): \bar{\mathcal{S}} \in D(\mathcal{S}_{i+m_r})\};$$
$$J_i = \dot{C}(\mathcal{S}_{i+m_r}), \quad \bar{J}_i = J_i - \bigcup\{J(\bar{\mathcal{S}}): \bar{\mathcal{S}} \in D(\mathcal{S}_{i+m_r})\}.$$

Define a superposition π_i on $A^{I_i} \times B^{J_i}$ by

$$\pi_i = \Pi\{\pi_{\bar{\mathcal{S}}}: \bar{\mathcal{S}} \in D(\mathcal{S}_{i+m_r})\} \times \mu_{\bar{I}_i} \times \nu_{\bar{J}_i}.$$

(This makes sense in view of lemma 3.4.) Also set

$$I_0 = \bigcup\{R(\mathcal{S}_i): 1 \le i \le m_r\}, \quad \pi_0 = \mu_{I_0},$$

and define

$$\pi_{\mathcal{S}} = \pi_{\bar{t}} * (\pi_{\bar{t}-1} * (\cdots * (\pi_1 * \pi_0) \cdots)),$$

where $\bar{t}$ is the largest i such that $\mathcal{S}_{i+m_r}$ is principal.

LEMMA 5.1. *The superpositions $\pi_{\mathcal{S}}$ are consistent in the sense that if $\bar{\mathcal{S}} \prec \mathcal{S}$ then either $I(\bar{\mathcal{S}}) \cap I(\mathcal{S}) = \varnothing$ and $J(\bar{\mathcal{S}}) \cap J(\mathcal{S}) = \varnothing$ or $I(\bar{\mathcal{S}}) \subset I(\mathcal{S})$, $J(\bar{\mathcal{S}}) \subset J(\mathcal{S})$ and the marginal of $\pi_{\mathcal{S}}$ on $A^{I(\bar{\mathcal{S}})} \times B^{J(\bar{\mathcal{S}})}$ is $\pi_{\bar{\mathcal{S}}}$. Moreover, if $\mathcal{S}'$ is the translate of $\mathcal{S}$ by $t \in \mathbb{N}$ (formally $\mathcal{S}'(i+t) = \mathcal{S}(i)$ and* rank $\mathcal{S}'$ = rank $\mathcal{S}$*), then $\pi_{\mathcal{S}'}$ is the translate of $\pi_{\mathcal{S}}$ by t.*

Proof. First observe that $\pi_{\mathcal{S}}$ is non-anticipating for all $\mathcal{S}$, as can be seen by induction on rank $\mathcal{S}$ using lemma 4.3. Now, by lemma 3.3, $I(\bar{\mathcal{S}})$ and $J(\bar{\mathcal{S}})$ are both contained in, or both disjoint from, $I(\mathcal{S})$ and $J(\mathcal{S})$ respectively. If $I(\bar{\mathcal{S}}) \subset I(\mathcal{S})$ and rank $\bar{\mathcal{S}} <$ rank $\mathcal{S}$, then $\bar{\mathcal{S}} \prec \mathcal{S}_i$ for some principal $\mathcal{S}_i$ in the rank decomposition of $\mathcal{S}$ and $I(\bar{\mathcal{S}}) \subset C(\mathcal{S}_i)$, so by lemma 3.4, $I(\bar{\mathcal{S}}) \subset I(\mathcal{S}')$ for some $\mathcal{S}' \in D(\mathcal{S}_i)$. By induction we then have that $\pi_{\mathcal{S}'}$ has marginal $\pi_{\bar{\mathcal{S}}}$. But since $\pi_{\mathcal{S}}$ has marginal $\pi_{\mathcal{S}'}$ we see that $\pi_{\mathcal{S}}$ has marginal $\pi_{\bar{\mathcal{S}}}$.

Now suppose that rank $\bar{\mathcal{S}}$ = rank $\mathcal{S}$, so that $\bar{\mathcal{S}}$ is the restriction of $\mathcal{S}$ to $[j, \infty)$ for some $j \in \mathbb{N}$. For $I \subset \mathbb{N}$ write

$$I^* = I \cap [j, \infty),$$

for a skeleton $\tilde{\mathscr{S}}$ write $\tilde{\mathscr{S}}^*$ for the restriction of $\tilde{\mathscr{S}}$ to $[j, \infty)$ and for a measure π on $A^I \times B^J$, write π^* for the marginal of π on $A^{I^*} \times B^{J^*}$. What we want to show is that $\pi_{\mathscr{S}^*} = (\pi_{\mathscr{S}})^*$. (Note $\pi_{\mathscr{S}^*}$ is a measure on $A^{I(\mathscr{S})^*} \times B^{J(\mathscr{S})^*}$ since $I(\mathscr{S}^*) = I(\mathscr{S})^*$ and $J(\mathscr{S}^*) = J(\mathscr{S})^*$.) Observe that $\pi_{\mathscr{S}^*} = (\pi_{\mathscr{S}})^*$ if $\mathscr{S}$ is a 0-skeleton and assume that this is also so for rank $\mathscr{S} < r$. Suppose $\mathscr{S} = \mathscr{S}_t \times \cdots \times \mathscr{S}_0$ has rank r and $j \in |\mathscr{S}_{i+m_r}|$, $i \geq 1$. (If $j \in |\mathscr{S}_k|$ for an initial $\mathscr{S}_k$ what we are trying to show is obvious.) Recall that

$$\pi_{\mathscr{S}} = \pi_{\bar{t}} * (\pi_{\bar{t}-1} * (\cdots * (\pi_1 * \pi_0) \cdots)),$$

where we adopt all the notation introduced in the definition of $\pi_{\mathscr{S}}$. Thus

$$\mathscr{S}^* = \mathscr{S}^*_{i+m_r} \times \mathscr{S}_{i-1+m_r} \times \cdots \times \mathscr{S}_0.$$

Since

$D(\mathscr{S}^*_{i+m_r}) = \{\bar{\mathscr{S}}^* : \bar{\mathscr{S}} \in D(\mathscr{S}_{i+m_r})\}$, $C(\mathscr{S}^*_{i+m_r}) = C(\mathscr{S}_{i+m_r})^*$ and $\dot{C}(\mathscr{S}^*_{i+m_r}) = \dot{C}(\mathscr{S}_{i+m_r})^*$,

we have

$$\pi_{\mathscr{S}^*} = \pi'_i * (\pi_{i-1} * (\cdots * (\pi_1 * \pi_0) \cdots)),$$

where

$$\pi'_i = \Pi\{\pi_{\bar{\mathscr{S}}^*} : \bar{\mathscr{S}} \in D(\mathscr{S}_{i+m_r})\} \times \mu_{I^*_i} \times \nu_{J^*_i}.$$

Now by our induction hypothesis

$$\pi_{\bar{\mathscr{S}}^*} = (\pi_{\bar{\mathscr{S}}})^* \quad \text{for } \bar{\mathscr{S}} \in D(\mathscr{S}_{i+m_r})$$

so

$$\Pi\{\pi_{\bar{\mathscr{S}}^*} : \bar{\mathscr{S}} \in D(\mathscr{S}_{i+m_r})\} = \Pi\{(\pi_{\bar{\mathscr{S}}})^* : \bar{\mathscr{S}} \in D(\mathscr{S}_{i+m_r})\}$$
$$= (\Pi\{\pi_{\bar{\mathscr{S}}} : \bar{\mathscr{S}} \in D(\mathscr{S}_{i+m_r})\})^*,$$

and evidently

$$\mu_{\bar{I}^*_i} = (\mu_{\bar{I}_i})^* \quad \text{and} \quad \nu_{\bar{J}^*_i} = (\nu_{\bar{J}_i})^*.$$

Thus $\pi'_i = \pi^*_i$. Now $\pi_{\mathscr{S}}$ has marginal

$$\pi_i * (\pi_{i-1} * (\cdots * (\pi_1 * \pi_0) \cdots))$$

and by lemma 4.4, since π_i is non-anticipating, this measure in turn has marginal

$$\pi^*_i * (\pi_{i-1} * (\cdots * (\pi_1 * \pi_0) \cdots)) = (\pi_{\mathscr{S}})^*.$$

This completes the inductive argument.

The assertion about translation invariance is clear. □

LEMMA 5.2. *Suppose N_i, C_i and m_i have been chosen for $i < r$. If m_r is chosen sufficiently large then for all skeleta $\mathscr{S}$ of* rank r

$$\mu_{I(\mathscr{S})}\{x \in A^{I(\mathscr{S})} : \pi_{\mathscr{S}} \text{ splits } x\} < \eta.$$

Proof. Again adopting the notation in the definition of $\pi_{\mathscr{S}}$,

$$\pi_{\mathscr{S}} = \pi_{\bar{t}} * (\pi_{\bar{t}-1} * (\cdots * (\pi_1 * \pi_0) \cdots)),$$

so the hypotheses of lemma 4.5 are satisfied with

$$I_i = C(\mathscr{S}_{i+m_r}), \quad J_i = \dot{C}(\mathscr{S}_{i+m_r}) \quad \text{for } i \geq 1$$

and

$$I_0 = \bigcup\{R(\mathscr{S}_i) : 1 \leq i \leq m_r\}.$$

https://doi.org/10.1017/S0143385700001267 Published online by Cambridge University Press

We adopt all the notation and terminology of lemma 4.4 and, in addition, we will write

$$\bar{l}_j = \sum_{i=1}^{j} \#J_i.$$

Recall that

$$\#J_i = \#C(\mathscr{S}_{i+m_r}) - 1 \le C_0 + \cdots + C_{r-1} - 1 = C - 1,$$

so

$$\#J_i \le \left(1 - \frac{1}{C}\right)(\#J_i + 1) \le \left(1 - \frac{1}{C}\right) \#I_i.$$

It follows that

$$\bar{l}_j < \left(1 - \frac{1}{C}\right) l_j \quad \text{for all } j,$$

and

$$\#B^{J_i} \le \#B^{C-1} = M.$$

It follows that

$$\bar{l}_j < \left(1 - \frac{1}{M}\right) l_j \quad \text{for all } j.$$

Fix $\varepsilon > 0$ such that

$$h - 2\varepsilon > \left(1 - \frac{1}{C}\right)(h + \varepsilon). \tag{1}$$

By the strong law of large numbers, given $\eta > 0$ we may choose k such that, for all $K > k$,

$$\nu_{[1,K]}\{y \in B^{[1,K]}: \nu_{[1,\bar{l}]}(y[1, \bar{l}]) > 2^{-(h+\varepsilon)\bar{l}} \quad \text{for } k \le \bar{l} \le K\} > 1 - \eta. \tag{2}$$

(Here $y[1, \bar{l}]$ denotes the restriction of y to $[1, \bar{l}]$.) Choose m_r so that

$$\nu_{[1,\bar{l}]}(y) > 2^{-(h-2\varepsilon)m_r} \quad \text{for all } y \in B^{[1,\bar{l}]},\ 0 \le \bar{l} \le k, \tag{3}$$

$$\mu_{[1,K]}\{x \in A^{[1,K]}: \mu_{[1,l]}(x[1, l]) < 2^{-(h-\varepsilon)l} \quad \text{for } m_r \le l \le K\} > 1 - \eta \quad \text{for } K \ge m_r, \tag{4}$$

$$M \sum_{i=m_r}^{\infty} 2^{-\varepsilon i} < \eta. \tag{5}$$

Now if $q = q^n \ldots q^1 \in Q^n$ is not completely good then for some $j \ge 1$

$$\nu_j(q^j \cdots q^1) < 2^{-(h-2\varepsilon)l_j} < 2^{-(h-2\varepsilon)m_r},$$

so that, by (3), $\bar{l}_j > k$. Moreover,

$$\nu_j(q^j \cdots q^1) < 2^{-(h-2\varepsilon)l_j} < 2^{-(1-1/C)(h+\varepsilon)l_j} < 2^{-(h+\varepsilon)\bar{l}_j}. \tag{6}$$

By (2), the ν_n-measure of q's in Q^n such that (6) occurs for some $\bar{l}_j > k$ is less than η, so we have

$$\nu_n\{q \in Q^n: q \text{ is not c.g.}\} < \eta. \tag{7}$$

(4) implies that

$$\mu_n\{p \in P^n : p \text{ is not c.g.}\} < \eta. \tag{8}$$

The result now follows from (5), (7), (8) and lemma 4.5. □

If $I = [i, j]$ is an interval in $\mathbb{N}$, $\hat{x} \in \hat{X}$ and $\mathcal{S}$ is an r-skeleton, we say that $\mathcal{S}$ occurs in $\hat{x}$ on I if $\hat{x}(j+1) = 0$ and the restriction of $\hat{x}$ to I is $\mathcal{S}$. We also say that $\mathcal{S}$ is the (unique) r-skeleton beginning at i in $\hat{x}$. It is clear that, for $\hat{\mu}$-a.a. $\hat{x} \in \hat{X}$ such that $\hat{x}(i) = 0$, there is a unique r-skeleton beginning at time i in $\hat{x}$. For $\hat{x} \in \hat{X}$ we denote by $\mathcal{S}_r(\hat{x})$ the unique r-skeleton beginning at time i_0 in $\hat{x}$, where i_0 is the least i such that $\hat{x}(i) = 0$. We also write

$$I_r(\hat{x}) = I(\mathcal{S}_r(\hat{x})) \quad \text{and} \quad J_r(\hat{x}) = J(\mathcal{S}_r(\hat{x})).$$

We say an r-skeleton $\mathcal{S}$ is caught if $C_r > \#|\mathcal{S}|$, which implies that $C(\mathcal{S}) = |\mathcal{S}|$. We also say that $\mathcal{S}$ is substantial if it contains at least one principal $r-1$ skeleton.

Now fix a sequence $\{\eta_r\}$ decreasing to 0 such that $\sum \eta_r < \infty$. Suppose that N_i, C_i and m_i have been chosen for $i < r$. By lemma 5.2 we may choose m_r so that for all skeleta $\mathcal{S}$ of rank r

$$\mu_{I(\mathcal{S})}\{x \in A^{I(\mathcal{S})} : \pi_{\mathcal{S}} \text{ splits } x\} < \eta_r. \tag{5.1}$$

Next we can choose N_r and C_r so that

$$\hat{\mu}\{\hat{x} \in \hat{X} : \mathcal{S}_r(\hat{x}) \text{ is substantial}\} > 1 - \eta_r, \tag{5.2}$$

$$\hat{\mu}\{\hat{x} \in \hat{X} : \mathcal{S}_r(\hat{x}) \text{ is caught}\} > 1 - \eta_r. \tag{5.3}$$

To see that this is possible set

$$G = \{\hat{x} \in \hat{X} : \hat{x}[1, N_{r-1}+2] = 01^{N_{r-1}}0\},$$

and observe that K can be chosen so large that the $\hat{\mu}$-measure of

$$H_1 = \left\{\hat{x} \in \hat{X} : \sum_{i=0}^{K} 1_G \hat{\sigma}^i(\hat{x}) \geq m_r + 2\right\}$$

is greater than $1 - \frac{1}{2}\eta_r$ ($\hat{\sigma}$ denotes the shift on $\hat{X}$). Then, setting

$$F = \{\hat{x} \in \hat{X} : \hat{x}[1, N_r] = 1^{N_r}\},$$

N_r can be chosen so large that the $\hat{\mu}$-measure of

$$H_2 = \{\hat{x} \in \hat{X} : \hat{\sigma}^i(\hat{x}) \notin F \quad \text{for } 0 \leq i \leq K\}$$

is greater than $1 - \frac{1}{2}\eta_r$. If $\hat{x} \in H_1 \cap H_2$, then $\mathcal{S}_r(\hat{x})$ is substantial, which takes care of (5.2). Then C_r can clearly be chosen so that (5.3) holds. We now assume that the N_r, C_r and m_r have all been chosen so that (5.1), (5.2), and (5.3) hold for all r.

By (5.2), (5.3) and the Borel–Cantelli lemma, there is a set $\hat{X}^* \subset \hat{X}$ such that $\hat{\mu}(\hat{X}^*) = 1$ and, for each $\hat{x} \in \hat{X}^*$, there is an integer $r_0(\hat{x})$ such that, for $r \geq r_0(\hat{x})$, $\mathcal{S}_r(\hat{x})$ is substantial and caught. If $\mathcal{S}_r(\hat{x})$ is substantial and caught and $\mathcal{S}_{r-1}(\hat{x})$ is caught, then

$$J_r(\hat{x}) \supset \dot{C}(\mathcal{S}_{r-1}(\hat{x})) = |\mathcal{S}_{r-1}(\hat{x})| - \{R(\mathcal{S}_{r-1}(\hat{x}))\}.$$

Also, for $r' < r$,

$$I_{r'}(\hat{x}) \subset I_r(\hat{x}) \quad \text{and} \quad J_{r'}(\hat{x}) \subset J_r(\hat{x}).$$

Thus we see that for $\hat{x} \in \hat{X}^*$

$$\bigcup_{r \geq 0} J_r(\hat{x}) = \bigcup_{r \geq 0} I_r(\hat{x}) = \{i: \hat{x}(i) = 0\} = I(\hat{x}),$$

and the sequences $I_r(\hat{x})$ and $J_r(\hat{x})$ are eventually increasing. In view of these remarks and lemma 5.1, we can define $\pi_{\hat{x}}$ to be the measure on $A^{I(\hat{x})} \times B^{I(\hat{x})}$ whose projection on each $A^{I_r(\hat{x})} \times B^{J_r(\hat{x})}$ is $\pi_{\mathscr{S}_r(\hat{x})}$. Using the natural identification of $A^{I(\hat{x})} \times B^{I(\hat{x})}$ with

$$X(\hat{x}) \times Y(\hat{x}) \subset X \times Y$$

we define

$$\pi = \int \pi_{\hat{x}} \, d\hat{\mu}(\hat{x}),$$

a measure on $X \times Y$ supported on $\{(x, y): \hat{x} = \hat{y}\}$. The proof of theorem 1 is now concluded by the following lemma.

LEMMA 5.3. *There exists a finitary homomorphism* ϕ: $X \to Y$ *such that* $\pi(B) = \mu\{x \in X: (x, \phi(x)) \in B\}$.

Proof. For $\hat{x} \in \hat{X}^*$ define $\phi_{\hat{x}}$: $X(\hat{x}) \to Y(\hat{x})$ by requiring that $\phi_{\hat{x}}(x)(J_r(\hat{x}))$ (the restriction to $J_r(\hat{x})$) be y_r, whenever $x(I_r(\hat{x})) \subset y_r$ with respect to $\pi_{\mathscr{S}_r(\hat{x})}$. Define $\phi(x) = \phi_{\hat{x}}(x)$. To see that this definition is unambiguous suppose that

$$r' > r, \quad I_{r'}(\hat{x}) \supset I_r(\hat{x}) \quad \text{and} \quad J_{r'}(\hat{x}) \supset J_r(\hat{x})$$

(if this is not the case $I_{r'}(\hat{x}) \cap I_r(\hat{x}) = \varnothing$ and $J_{r'}(\hat{x}) \cap J_r(\hat{x}) = \varnothing$). By lemma 5.1, $\pi_{\mathscr{S}_{r'}(\hat{x})}$ has marginal $\pi_{\mathscr{S}_r(\hat{x})}$ so we see that, if $x(I_{r'}(\hat{x})) \subset y_{r'}$ with respect to $\pi_{\mathscr{S}_{r'}(\hat{x})}$, and $x(I_r(\hat{x})) \subset y_r$ with respect to $\pi_{\mathscr{S}_r(\hat{x})}$, then we must have

$$y_{r'}(J_r(\hat{x})) = y_r.$$

To see that $\phi_{\hat{x}}(x)$ is defined on all of $I(\hat{x})$ observe that by inequality (5.1), for $\mu_{\hat{x}}$-a.a. $x \in X(\hat{x})$, $x(I_r(\hat{x}))$ is split by $\pi_{\mathscr{S}_r(\hat{x})}$ only for finitely many r. Thus, for large r, $\phi_{\hat{x}}(x)$ is defined on $J_r(\hat{x})$, hence it is defined on $I(\hat{x})$.

ϕ is finitary because, by the remarks preceding this lemma, for sufficiently large r, $\phi(x)$ is determined on all of

$$|\mathscr{S}_{r-1}(\hat{x})| - \{R(\mathscr{S}_{r-1}(\hat{x}))\}$$

by $x(I_r(\hat{x}))$ and $\mathscr{S}_r(\hat{x})$, which are both determined by a finite segment of x. To see that ϕ is shift invariant first note that, for $\hat{\mu}$-a.a. $\hat{x}$,

$$\mathscr{S}_r(\hat{\sigma}(\hat{x})) = S(\mathscr{S}_r(\hat{x})),$$

where $S(\mathscr{S})$ denotes $\mathscr{S}$ shifted one unit to the left (if $\mathscr{S}$ is indexed by $[1, n]$ then one must also delete the initial zero and if, in addition, $\mathscr{S}$ has a single initial zero one must delete the following run of 1's as well). By lemma 5.1

$$\pi_{\mathscr{S}_r(\hat{\sigma}(\hat{x}))} = \pi_{S(\mathscr{S}_r(\hat{x}))} = S\pi_{\mathscr{S}_r(\hat{x})},$$

where S again denotes the left shift, now acting on measures. It follows that, if $x(\mathscr{S}_r(\hat{x})) \subset y$ with respect to $\pi_{\mathscr{S}_r(\hat{x})}$, and also

$$(\sigma x)(\mathscr{S}_r(\sigma(\hat{x}))) = \sigma x(S(\mathscr{S}_r(\hat{x}))) \subset y'$$

then $y' = \tau y$. Thus $\phi\sigma = \tau\phi$.

To show that

$$\mu\{x: (x, \phi(x)) \in B\} = \pi(B)$$

it suffices to show that, for $\hat{x} \in \hat{X}^*$,

$$\mu_{\hat{x}}\{x \in X(\hat{x}): (x, \phi_{\hat{x}}(x) \in B\} = \pi_{\hat{x}}(B).$$

Suppose that $p_0 \in P^{I_{r_0}(\hat{x})}$ and $q_0 \in Q^{I_{r_0}(\hat{x})}$. By the definition of $\phi_{\hat{x}}$ one sees that $\mu_{\hat{x}}\{x \in X: (x, \phi_{\hat{x}}(x)) \in p_0 q_0\}$

$$= \lim_{r\to\infty} \mu_{\hat{x}} \bigcup \{p \in P^{I_r(\hat{x})}: p \subset p_0 \quad \text{and} \quad \exists q \in Q^{J_r(\hat{x})} \text{ s.t. } p \subset q \subset q_0 \text{ w.r.t. } \pi_{\mathcal{S}_r(\hat{x})}\}$$

$$\leq \lim_{r\to\infty} \mu_{\hat{x}} \bigcup \{p \in P^{I_r(\hat{x})}: p \subset p_0 \quad \text{and } p \subset q_0 \text{ w.r.t. } \pi_{\mathcal{S}_r(\hat{x})}\}$$

$$\leq \lim_{r\to\infty} \pi_{\mathcal{S}_r(\hat{x})}(p_0 q_0) = \pi_{\mathcal{S}_{r_0}(\hat{x})}(p_0 q_0) = \pi_{\hat{x}}(p_0 q_0).$$

Thus, whenever B is a cylinder in $X(\hat{x}) \times Y(\hat{x})$, and hence for all B, we have

$$\mu_{\hat{x}}\{x \in X(\hat{x}): (x, \phi(x)) \in B\} \leq \pi_{\hat{x}}(B).$$

Since both sides of the above inequality are probability measures we can replace the inequality by equality. Thus we have

$$\mu\{x \in X: (x, \phi(x)) \in B\} = \pi(B),$$

and, in particular, ϕ is measure-preserving. □

REFERENCES

[1] R. Adler & B. Weiss. Similarity of automorphisms of the torus. *A.M.S. Memoirs* **98** (1970).

[2] M. A. Akcoglu, A. del Junco & M. Rahe. Finitary codes between Markov processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **47** (1979), 305–314.

[3] A. del Junco. Finitary coding of Markov random fields. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **52** (1980), 193–202.

[4] M. Keane & M. Smorodinsky. Bernoulli shifts of the same entropy are finitarily isomorphic. *Annals of Math.* **109** (1979), 397–406.

[5] L. Meshalkin. A case of isomorphy of Bernoulli schemes. *Dokl. Akad. Nauk SSSR* **128** (1959), 41–44.

[6] D. Ornstein & B. Weiss. Unilateral codings of Bernoulli systems. *Israel J. Math.* **21** (1975), 159–166.

[7] Ja. G. Sinai. A weak isomorphism of transformations with an invariant measure. *Dokl. Akad. Nauk SSSR* **147** (1962), 797–800.

[8] P. Walters. *Some Results on the Classification of Non-invertible Measure-preserving Transformations.* Lecture Notes in Math. Vol. 318, pp. 266–276. Springer-Verlag: Berlin, 1973.