A MONOTONE ISOMORPHISM THEOREM

TERRY SOO

Dedicated to Professor Andrés del Junco, September 21, 1948 – June 17, 2015

ABSTRACT. In the simple case of a Bernoulli shift on two symbols, zero and one, by permuting the symbols, it is obvious that any two equal entropy shifts are isomorphic. We show that the isomorphism can be realized by a factor that maps a binary sequence to another that is coordinatewise smaller than or equal to the original sequence.

1. INTRODUCTION

Let N be a positive integer, $[N] = \{0, 1, \ldots, N - 1\}$, and $\Omega = [N]^{\mathbb{Z}}$. Let $T : \Omega \to \Omega$ be the left-shift given by $(Tx)_i = x_{i+1}$ for all $i \in \mathbb{Z}$. Given a probability measure $\mathbf{p}$ on $[N]$, we call $B(\mathbf{p}) = (\Omega, \mathbf{p}^{\mathbb{Z}}, T)$ a **Bernoulli-shift** on N symbols. We say that a Bernoulli shift $B(\mathbf{q})$ is a **factor** of $B(\mathbf{p})$, if there exists a measurable map $\phi : \Omega \to \Omega$ such that the push-forward of $\mathbf{p}^{\mathbb{Z}}$ under ϕ is $\mathbf{q}^{\mathbb{Z}}$ and $\phi \circ T = T \circ \phi$ on a subset of Ω with $\mathbf{p}^{\mathbb{Z}}$-full measure; we also call the map ϕ a **factor** from $B(\mathbf{p})$ to $B(\mathbf{q})$. We say that the Bernoulli shifts $B(\mathbf{p})$ and $B(\mathbf{q})$ are **isomorphic** if there exists a factor map ϕ from $B(\mathbf{p})$ to $B(\mathbf{q})$ such that its inverse ϕ^{-1} serves as a factor map from $B(\mathbf{q})$ to $B(\mathbf{p})$; in this case, we call ϕ an **isomorphism** of $B(\mathbf{p})$ and $B(\mathbf{q})$. A factor map ϕ is **monotone** if for all $x \in \Omega$, we have $\phi(x)_i \leq x_i$ for all $i \in \mathbb{Z}$.

Theorem 1. *If $p \in (\frac{1}{2}, 1)$, then there exists a monotone isomorphism of $B(1 - p, p)$ and $B(p, 1 - p)$.*

Let us remark that the map defined by $\phi(x)_i = \mathbf{1}[x_i = 0]$ for all $i \in \mathbb{Z}$, which just swaps zeros and ones, is clearly an isomorphism of $B(1 - p, p)$ and $B(p, 1 - p)$. However, it is not monotone.

It is easy to determine when two Bernoulli shifts are isomorphic via an invariant introduced by Kolmogorov [9], which is non-increasing under factors and preserved under isomorphisms. The **entropy** of

a probability measure $\mathbf{p} = (p_0, \ldots, p_{N-1})$ on $[N]$ is given by $H(\mathbf{p}) := -\sum_{i=0}^{N-1} p_i \log p_i$. Sinai [21, 22] proved that if $H(\mathbf{p}) \geq H(\mathbf{q})$, then $B(\mathbf{q})$ is a factor of $B(\mathbf{p})$, and Ornstein [16, 17] proved that the entropies of two Bernoulli shifts are equal if and only if the two Bernoulli shifts are isomorphic.

Although it is easy to compute the entropy of a Bernoulli shift and to determine whether two Bernoulli shifts are isomorphic, the actual factor map which realizes the isomorphism is in general a much more complicated object. In some special cases, the factor map has a simple description [15, 2]. The first non-trivial example of an isomorphism is due to Melshalkin [15], which also gives a monotone isomorphism. I thank Zemer Kosloff for his help with the following example.

Example 1 (A classical example due to Melshalkin [15]). We will adjust the treatment given in [12] to ensure monotonicity. Let $\mathbf{p} = (\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2})$ and $\mathbf{q} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0)$, so that $N = 5$, and $\mathbf{p}$ and $\mathbf{q}$ are probability measures on $[N] = \{0, 1, 2, 3, 4\}$. Let $x \in \Omega = [5]^{\mathbb{Z}}$. We define a factor map $\phi : \Omega \to \Omega$ such that if $x_i = 4$, then $\phi(x)_i \in \{2, 3\}$, if $x_i \in \{2, 3\}$, then $\phi(x)_i = 1$, and if $x_i \in \{0, 1\}$, then $\phi(x)_i = 0$.

It remains to specify what happens when $x_i = 4$. Think of every $x_i = 4$ as a right parenthesis, and think of every $x_i \neq 4$ as a left parenthesis. Ergodicity implies that every parenthesis will be matched legally almost surely. If $x_i = 4$, then let j be the position of the corresponding left parenthesis. If x_j is odd, then we set $\phi(x)_i = 3$, if x_j is even, then we set $\phi(x)_i = 2$.

By definition, the map ϕ satisfies $\phi \circ T = T \circ \phi$ and is monotone. Melshalkin proved that ϕ is an isomorphism of $B(\mathbf{p})$ and $B(\mathbf{q})$.   ◇

It is easy to see that a necessary condition for the existence of a monotone factor from $B(\mathbf{p})$ to $B(\mathbf{q})$ is that there exists a ***monotone coupling*** of $\mathbf{p}$ and $\mathbf{q}$; that is, a probability measure ρ on $[N] \times [N]$ such that $\rho(\cdot, [N]) = \mathbf{p}$, $\rho([N], \cdot) = \mathbf{q}$, and $\rho\{(n, m) : n \geq m\} = 1$. By Strassen's theorem [23], the existence of a monotone coupling is equivalent to the condition that $\sum_{i=0}^{k} p_i \leq \sum_{i=0}^{k} q_i$ for all $0 \leq k < N$, in which case we say that $\mathbf{p}$ ***stochastically dominates*** $\mathbf{q}$.

Theorem 2 (Quas and Soo [18]). *Let $\mathbf{p}$ and $\mathbf{q}$ be probability measures on $[N]$. If $\mathbf{p}$ stochastically dominates $\mathbf{q}$ and $H(\mathbf{p})$ is strictly greater than $H(\mathbf{q})$, then there exists a monotone factor from $B(\mathbf{p})$ to $B(\mathbf{q})$.*

Karen Ball [1] proved Theorem 2 in the case that the measure $\mathbf{q}$ is supported on two symbols. In both of those papers, a strict entropy inequality is required. In this paper, we treat the case of equal entropy, in the special case where there are only two symbols, zero and one, in

each of the Bernoulli shifts. The methods used to prove Theorem 1 can also be adapted to produce monotone factors in other specific cases, but we do not know the answer to the following question.

Question 1. *Let* $\mathbf{p}$ *and* $\mathbf{q}$ *be probability measures on* $[N]$ *such that* $\mathbf{p}$ *stochastically dominates* $\mathbf{q}$ *and* $H(\mathbf{p}) = H(\mathbf{q})$. *Does there exists a monotone factor from* $B(\mathbf{p})$ *to* $B(\mathbf{q})$?

Russell Lyons [1] first posed the question of whether a monotone factor exists between two Bernoulli shifts. The requirement of monotonicity makes defining maps more difficult. In a related problem, Gurel-Gurevich and Peled [6, Theorem 1.3] proved that for $p \in (\frac{1}{2}, 1)$ there exists a monotone map $\phi : \{0,1\}^{\mathbb{Z}} \to \{0,1\}^{\mathbb{Z}}$ such that the product measure $(p, 1-p)^{\mathbb{Z}}$ is the push-forward of $(1-p, p)^{\mathbb{Z}}$ under ϕ; however, their map is *not* be equivariant; that is, it does not satisfy $\phi \circ T = T \circ \phi$.

See [24, 17, 9] for more information on entropy and the isomorphism problem in ergodic theory. See [18] and [13] for background on factors in probability theory.

The proof of Theorem 1 will involve some of the methods of [18], which in turn combines ideas from various treatments of the Ornstein and Sinai factor theorems given by Keane and Smorodinsky [10, 11], Burton and Rothstein [3, 4], del Junco [7, 8], and Ball [1]. We briefly summarize some of the main features and differences in their proofs. Keane and Smorodinsky, and Ball employed a marker-filler method and a version of Hall's marriage theorem (see Remark 9). Del Junco also employed a marker-filler method, but he replaced the marriage lemma with his star-coupling (see Section 4). These constructions are explicit and they exhibit factor maps that are finitary–an almost surely continuity property (see [20] for details). In a somewhat more abstract approach, Burton and Rothstein proved that in a suitably defined metric space, the set of all factors is a residual set, in the sense of the Baire category theorem. This was the approach taken in [18], and will also be the approach we take here.

DEDICATION

I never had the pleasure of meeting Professor del Junco, but I wrote to him in December 2013 about Theorem 2 with a preprint of [18]. He wrote back the same day saying he was glad that an old idea of his had found another application and that he always felt that the star-coupling was one of his best ideas.

His coupling was a key feature in our proof of Theorem 2, and will also be a star feature in the proof of Theorem 1.

2. Coupling and Stochastic domination

Strassen's theorem [23] holds in the much more general setting of a partially ordered Polish space. The proof, even in the case of a finite set is non-trivial, see for example [14, Theorem 10.4]. However, in the special case of real-valued random variables or random variables taking values on a finite totally ordered set, the proof is easily obtained using a simple coupling of random variables.

2.1. **Quantiles.** Let X be a real-valued random variable, with cumulative distribution function or law given by $F(z) = F_X(z) := \mathbb{P}(X \le z)$ for all $z \in \mathbb{R}$. Define the generalized inverse of F via $F^{-1}(y) := \sup \{x \in \mathbb{R} : F(x) < y\}$. Let U be uniformly distributed in $[0, 1]$, so that $F_U(z) = z$ for all $z \in [0, 1]$. We call $F_X^{-1}(U)$ the **quantile representation** of X. It is easy to see that the random variable $F_X^{-1}(U)$ has the same law as X. When we define random variables using the quantile representation sometimes we will refer to the random variable U as the **randomization**; often U will be chosen to be independent of any previously defined random variables.

If X and Y are two real-valued random variables, we say that X **stochastically dominates** Y if $\mathbb{P}(X \le z) \le \mathbb{P}(Y \le z)$ for all $z \in \mathbb{R}$. A **coupling** of X and Y is a pair of random variables (X', Y') defined on the same probability space such that X' has the same law as X and Y' has the same law as Y. Let U be uniformly distributed in $[0, 1]$. If we set $X' := F_X^{-1}(U)$ and $Y' := F_Y^{-1}(U)$, then the **quantile coupling** of X and Y is given by (X', Y'). We say that the coupling (X', Y') is **monotone** if $X' \ge Y'$. Strassen's theorem implies that X stochastically dominates Y if and only if there exists a monotone coupling of X and Y. Clearly, the existence of a monotone coupling implies stochastic domination; on the other hand, it is easy to see that the quantile coupling is monotone under the assumption of stochastic domination.

Let us remark that stochastic domination and the quantile coupling are also similarly defined in the case that the random variables take values in a finite totally ordered space.

Lemma 3 (Strassen's theorem via the quantile coupling). *Let X and Y be real-valued random variables or random variables taking values in a finite totally ordered space. If (X', Y') is a quantile coupling of X and Y, then X' is almost surely greater than or equal to Y' if and only if X stochastically dominates Y.*

In Section 4, we will discuss an ingenious variation of the quantile coupling due to del Junco [7, Section 4], which will be a key ingredient in our proof of Theorem 1.

2.2. **An simple application of Strassen's theorem.** Lemma 3 will be used to prove the following simple observation, which will serve as the starting point in our proof of Theorem 1. For two binary sequences x and y of the same length, we write $x \preceq y$ if and only if $x_i \leq y_i$ for all indices i. Thus the relation $\preceq$ defines a partial order on the set of binary sequences with the same length. We write $x = 1^n 0^\ell$ to mean a binary sequence of n ones followed by ℓ zeros.

Lemma 4. *Let $n \geq 1$ and $p \in (\frac{1}{2}, 1)$. Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ be an i.i.d. sequences of Bernoulli random variables with parameters p and $1-p$, respectively. Let B_n be the set of size $n+1$ of all binary sequences z of length n of the form $z = 1^{n-\ell} 0^\ell$ for some $\ell \in [0, n]$. Let X^* and Y^* be random variables that have laws X and Y conditioned to be in B_n, respectively. Then with respect to the order $\preceq$, defined on binary sequences, X^* stochastically dominates Y^*, and there is a monotone coupling of X^* and Y^*.*

Note that in Lemma 4, although the set of all binary sequences of a fixed length is only partially ordered by $\preceq$, the set B_n is totally ordered by $\preceq$. The set B_n can also be described as the set of binary sequences of length n that do not have a zero followed by a one. We will refer to B_n as a **filler** set.

Proof of Lemma 4. Lemma 4 is simple consequence of the duality between p and $1 - p$. For every integer $\ell \in [0, n]$, we have

$$\sum_{i=0}^{n-\ell} p^{n-\ell-i}(1-p)^i = \sum_{i=0}^{n-\ell}(1-p)^{n-\ell-i}p^i; \tag{1}$$

this implies, using $\ell = 0$, that $\mathbb{P}(X \in B_n) = \mathbb{P}(Y \in B_n)$. Thus in order to prove that X^* stochastically dominates Y^*, it suffices to show that for all $z \in B_n$, we have $\mathbb{P}(X \preceq z, X \in B_n) \leq \mathbb{P}(Y \preceq z, Y \in B_n)$. If $z = 1^{n-\ell} 0^\ell$, then since $p > 1 - p$, by equality (1) we have

$$
\begin{aligned}
\mathbb{P}(X \preceq z, X \in B_n) &= (1-p)^\ell \Big(\sum_{i=0}^{n-\ell} p^{n-\ell-i}(1-p)^i \Big) \\
&\leq p^\ell \Big(\sum_{i=0}^{n-\ell}(1-p)^{n-\ell-i}p^i \Big) \\
&= \mathbb{P}(Y \preceq z, Y \in B_n).
\end{aligned}
$$

The existence of a monotone coupling follows from Lemma 3.   □

3. Markers, fillers, and joinings

3.1. **Markers.** Let us fix $\Omega = \{0,1\}^{\mathbb{Z}}$. Let $x \in \Omega$. We call the interval $[i, i+1] \subset \mathbb{Z}$ a ***primary marker*** if $x_i = 0$ and $x_{i+1} = 1$. Later, we will define secondary and tertiary markers which will consist of consecutive primary markers. Note that two distinct primary markers have an empty intersection. We call an interval of $\mathbb{Z}$ a ***filler*** if it is nonempty and lies between two primary markers. Thus each $x \in \Omega$ partitions $\mathbb{Z}$ into intervals of primary markers and fillers.

Let $p \in (\frac{1}{2}, 1)$, and consider the product probability measures on $\Omega = \{0,1\}^{\mathbb{Z}}$ given by $\mu := (1-p, p)^{\mathbb{Z}}$ and $\nu := (p, 1-p)^{\mathbb{Z}}$. Thus the probability that the zeroth coordinate is a one under μ is p and is $1-p$ under ν. By conditioning, an instance of a random variable X with law μ can be given by first deciding on the locations of the primary markers, and then deciding on the content of the filler; the same observation holds for a random variable Y with law ν.

To be more precise, let $\mathsf{T} = \{\mathsf{M}, \mathsf{F}\}^{\mathbb{Z}}$, where M and F are two symbols that stand for 'marker' and 'filler.' For each $x \in \Omega$, define the ***hat map*** by setting

$$\hat{x}(i) = \begin{cases} \mathsf{M} \text{ if } i \in \mathbb{Z} \text{ is in a primary marker;} \\ \mathsf{F} \text{ otherwise.} \end{cases}$$

Let τ and τ' be push-forwards of the measures μ and ν via the hat map. Sometimes we will refer to τ as the ***marker measure***. We have the following disintegration. For τ-almost every $t \in \mathsf{T}$, there exists a probability measure, μ_t on Ω, such that

$$\int f(x)d\mu(x) = \int \left(\int f(x)d\mu_t(x) \right) d\tau(t)$$

for all measurable $f : \Omega \to [0, \infty)$.

Remark 5. Keane and Smorodinsky [10, Lemma 4] give a concrete description of μ_t. The measure μ_t assigns the sequence 01 to each primary marker interval of t, and is a product measure on the filler intervals, where on a filler interval of length n it is the law of n i.i.d. Bernoulli random variables with parameter p conditioned to be in the set B_n of sequences of consecutive ones followed by consecutive zeros (see Lemma 4). The analogous result holds of ν.   ◇

Remark 6. Notice that the probability that the origin is contained in a primary marker is same under μ and ν. Keane and Smorodinsky [10,

Lemma 3] proved that $\tau = \tau'$. Thus the marker measure τ is the same for μ and ν and depends only on the parameter p. This fact will also be important in our proof of Theorem 1. $\diamondsuit$

3.2. **Joinings.** A **coupling** of μ and ν is a probability measure ξ on $\Omega \times \Omega$ that has marginals μ and ν; a **joining** is a coupling that is invariant under the product shift $T \times T$, so that $\xi \circ (T \times T) = \xi$. A joining ξ is **ergodic** if all ξ-almost sure $(T \times T)$-invariant sets have measure zero or one. A coupling ξ is **monotone** if

$$\xi \{(x, y) \in \Omega \times \Omega : x_i \geq y_i \text{ for all } i \in \mathbb{Z}\} = 1.$$

A joining ξ is of **marker form** if for ξ-almost every $(x, y) \in \Omega \times \Omega$ the binary sequences x and y have the same primary markers. It follows from Remark 6 that there exists a joining of μ and ν in marker form. We will use a monotone version of this fact.

Proposition 7. *There exists a monotone joining of μ and ν of marker form.*

Proof. By Remark 6, we have $\tau = \tau'$. Hence we may assume that there exist random variables X and Y with laws μ and ν such that X and Y have the same primary markers and filler intervals. Consider a coupling of X and Y defined in the following way. By Remark 5, conditioned on the locations of the primary markers, for each filler interval I of X, we know that the law of restrictions of X to I is given by the law of a finite sequence of i.i.d. Bernoulli random variables with parameter p conditioned be in a filler set; furthermore, conditioned on the locations of primary markers, the restrictions of X to each filler interval give independent random variables. The analogous statement holds for Y. For each filler interval I, by Lemma 4, there exists a monotone coupling of the restriction of X to I and the restriction of Y to I. Hence by applying Lemma 4 to each of the filler intervals independently, and leaving the primary markers alone, we obtain a coupling (X', Y') of X and Y whose law is monotone and of marker form. $\square$

3.3. **The Baire category approach of Burton and Rothstein.** Let $p \in (\frac{1}{2}, 1)$ and $J = J(p)$ be the set of all monotone ergodic joinings of $\mu = (1-p, p)^{\mathbb{Z}}$ and $\nu = (p, 1-p)^{\mathbb{Z}}$ of marker form. Note that J is nonempty by Proposition 7. Following the approach of Burton and Rothstein [4], we will show the monotone isomorphisms are a residual set in J, when we endow J with a suitable topology. Following del Junco [8], we assign a complete metric to J as follows. For $i \geq 0$, let $\mathcal{C}_i$ be the set of measurable $C \subset \Omega \times \Omega$ that only depend on the

coordinates $j \in [-i, i]$; we will call such sets **cylinder sets**. For any two measures ζ and ξ on $\Omega \times \Omega$ (which may not be joinings), set

$$\mathrm{d}^*(\zeta, \xi) := \sum_{i=0}^{\infty} 2^{-(i+1)} \sup_{C \in \mathcal{C}_i} |\zeta(C) - \xi(C)|.$$

Thus d^* is the usual weak-star metric. For $\xi \in J$, let ξ_t be ξ conditioned to have the primary markers given by $t \in \mathsf{T}$. Let us remark that ξ_t is no longer a joining. Let τ be the common marker measure. For $\zeta, \xi \in J$, set

$$\mathrm{d}(\xi, \zeta) := \int \mathrm{d}^*(\xi_t, \zeta_t) d\tau(t). \tag{2}$$

Standard methods show that (J, d) is a Baire space (see for example [18, Lemma 17]). We will show that the set of monotone isomorphisms contains an intersection of open dense sets of J, and hence is nonempty by the Baire category theorem. To be more precise, let $\mathcal{F}$ denote the product sigma-algebra for Ω. Let $\mathcal{P} := \{P_0, P_1\}$ denote the partition of Ω according the zeroth coordinate so that $P_i := \{x \in \Omega : x_0 = i\}$. Let $\zeta \in J$ and let $\varepsilon > 0$. If there exists $\mathsf{T}' \subset \mathsf{T}$ with $\tau(\mathsf{T}') > 1 - \varepsilon$ such that for every $t \in \mathsf{T}'$, and each $P \in \sigma(\mathcal{P})$ there exists a $P' \in \mathcal{F}$ such that

$$\zeta_t\Big((P' \times \Omega) \triangle (\Omega \times P)\Big) < \varepsilon,$$

then we say that ζ is an ε-**almost factor** from $B(1-p, p)$ to $B(p, 1-p)$. For each $\varepsilon > 0$, let U_ε be the set of all ε-almost factors from $B(1-p, p)$ to $B(p, 1 - p)$. It is routine to verify that U_ε is an open set (see for example [5, page 123–24]) and that an element in the intersection of all the U_ε defines a monotone factor from $B(1 - p, p)$ to $B(p, 1 - p)$ (see for example [19, Theorem 2.8]). The real work lies in verifying that U_ε is dense; once this has been proved, the Baire category theorem gives that the set of monotone factors from $B(1-p, p)$ to $B(p, 1-p)$ contains an intersection of open dense sets, and hence is nonempty.

Theorem 1 asserts the existence of a monotone isomorphism which appears to be a much stronger statement the existence of a monotone factor. However, one of the advantages of the Baire category approach is that proving the existence of the isomorphism requires little additional work. We define an approximate factor from $B(p, 1 - p)$ to $B(1 - p, p)$ in the analogous way. Let $\zeta \in J$ and let $\varepsilon > 0$. If there exists $\mathsf{T}' \subset \mathsf{T}$ with $\tau(\mathsf{T}') > 1 - \varepsilon$ such that for every $t \in \mathsf{T}'$, and each $P \in \sigma(\mathcal{P})$ there exists a $P' \in \mathcal{F}$ such that $\zeta_t\big((P \times \Omega) \triangle (\Omega \times P')\big) < \varepsilon$, then we say that ζ is an ε-**almost factor** from $B(p, 1-p)$ to $B(1-p, p)$. For each $\varepsilon > 0$, let V_ε be the set of all ε-almost factors. Again, one can verify that V_ε is an open set, and that an element in the intersection

of all the V_ε defines a monotone factor from $B(p, 1 - p)$ to $B(1 - p, p)$. Moreover, any element in the grand intersection of all the U_ε and V_ε defines a monotone isomorphism. It will become apparent that the same proof that shows that U_ε is dense can be essentially copied to show that V_ε is dense. Thus the Baire category theorem shows that the grand intersection is nonempty.

It remains to verify that for each $\varepsilon > 0$, the set U_ε of ε-almost factors is dense. Given $\varepsilon > 0$ and $\xi \in J$, we need to find $\xi' \in U_\varepsilon$ with $d(\xi, \xi') < \varepsilon$. We will define ξ' as a certain perturbation of ξ which will be obtained using del Junco's star-coupling [7, Section 4 and Proposition 4.7].

4. THE STAR-COUPLING

Let X and Y be random variables taking values on finite sets A and B, respectively. In this section, we will discuss various couplings of X and Y; that is, random variables X' and Y' defined together on the same probability space with the same distribution as X and Y, respectively.

Let ρ be a joint probability mass function for X and Y. We say that an element $a \in A$ is **split** by ρ if there exist distinct $b, b' \in B$, such that $\rho(a, b) > 0$ and $\rho(a, b') > 0$. For the purposes of defining factors, we are interested in couplings that do not split many elements.

Remark 8. Let us remark that if we assign an arbitrary total ordering to A and B, then the law of a quantile coupling of X and Y will split at most $|B| - 1$ elements of A. $\diamondsuit$

Remark 9. Keane and Smorodinsky [10, Theorem 11] proved that there is a coupling of X and Y with law ρ' that will split at most $|B| - 1$ elements of A and in addition, ρ' is absolutely continuous with respect to ρ; that is, $\rho(a, b) = 0$ implies $\rho'(a, b) = 0$. A version of their theorem was used in the proof of Theorem 2, but we will not need to appeal to this result in our proof of Theorem 1. $\diamondsuit$

Let X and Y be jointly distributed random variables taking values on totally ordered finite sets $(A, \leq)$ and $(B, \leq)$, respectively. Let X' have the same law as X. One way to generate another random variable Y' so that (X', Y') has the same joint distribution as (X, Y) is to appeal to a quantile representation. Consider the set of conditional cumulative distribution functions given by $Q_a := \mathbb{P}(Y \leq b \mid X = a)$ for each $a \in A$. Let U be uniformly distributed in $[0, 1]$ and independent of X'. Set $Y' := Q_{X'}^{-1}(U)$. It is easy to verify that (X', Y') has the same

joint distribution as (X, Y); we call (X', Y') the **conditional quantile representation** of (X, Y).

The next coupling we discuss is due to del Junco [7, 8]. Let (X_1, Y_1) and (X_2, Y_2) be random variables taking values on the finite sets (A_1, B_1) and (A_2, B_2), respectively. Suppose that each of the sets A_1, A_2, B_1, and B_2 are totally ordered sets. We will define (X'_1, Y'_1) and (X'_2, Y'_2) such that (X'_i, Y'_i) has the same law as (X_i, Y_i) for $i = 1, 2$. Let U_1, U_2, and U be independent random variables uniformly distributed in the unit interval $[0, 1]$. Let X'_2 and Y'_1 be independent random variables that have the same laws as X_2 and Y_1, respectively; more specifically, we may assume that they are given by their respective quantile representations with sources of randomization given by U_2 and U_1. Next, using the *same* source of randomization U, let Y'_2 be such that (X'_2, Y'_2) is the conditional quantile representation of (X_2, Y_2), and let X'_1 be such that (Y'_1, X'_1) is the conditional quantile representation of (Y_1, X_1). We refer to $((X'_1, Y'_1), (X'_2, Y'_2))$ as the **star-coupling** of (X_1, Y_1) and (X_2, Y_2).

Remark 10. It is immediate from the definition the star-coupling that X'_2 is independent of (Y'_1, X'_1) and Y'_1 is independent of (X'_2, Y'_2). $\Diamond$

Remark 11. It follows from Remark 8, that the star-coupling of the random variables (X_1, Y_1) and (X_2, Y_2) taking values on (A_1, B_1) and (A_2, B_2), respectively, has the property that for a fixed $a_2 \in A_2$ and $b_1 \in B_1$, the number of $a_1 \in A_1$ such that there are distinct $b_2, b'_2 \in B_2$ with both (a_1, b_1, a_2, b_2) and (a_1, b_1, a_2, b'_2) receiving positive mass under the law of star-coupling (X'_1, Y'_1, X'_2, Y'_2) is at most $|B_2| - 1$. $\Diamond$

Remark 12. del Junco refers to his coupling as the $*$-joining [7, 8]. $\Diamond$

We may also iterate the star-coupling to more than two pairs of random variables. For example, if (X_i, Y_i) are finite-valued random variables taking values in totally ordered spaces (A_i, B_i) for $i = 1, 2, 3$, we define its **iterated star-coupling** in the following way. Take the star-coupling of (X_1, Y_1) and (X_2, Y_2) to be given by $((X'_1, Y'_1), (X'_2, Y'_2))$. Assign a lexicographic ordering to the set $A_1 \times A_2$, and take the star-coupling of $((X'_1, X'_2), (Y'_1, Y'_2))$ and (X_3, Y_3), to obtain random variables $((X''_1, X''_2, X''_3), (Y''_1, Y''_2, Y''_3))$. Notice that by definition of the star-coupling, (X''_i, Y''_i) has the same law as (X_i, Y_i) for $i = 1, 2, 3$.

We will make use of the following variation of the iterated star-coupling. Let ρ be a probability measure on the finite set $A \times B$ which has projections α and β on the sets A and B, respectively. For every $k \in \mathbb{Z}^+$, let α^k and β^k denote the k-fold product measures on A^k and B^k, respectively. Let n_{initial} and $k_{\text{group}} \geq 2$ be integers. Let $Z_i = (X_i, Y_i)$ be random variables with the following **grouping property** with law

ρ and constants n_{initial} and k_{group}. For each $i \geq 1$, the random variable Z_i takes values on $(A \times B)^{k_{\text{group}}} \equiv A^{k_{\text{group}}} \times B^{k_{\text{group}}}$ and has a law that has projections $\alpha^{k_{\text{group}}}$ and $\beta^{k_{\text{group}}}$ on $A^{k_{\text{group}}}$ and $B^{k_{\text{group}}}$, respectively, and a projection ρ on each copy of $A \times B$. Similarly, for $i = 0$, the random variable Z_0 takes values on $(A \times B)^{n_{\text{initial}}}$ and has a law that has projections $\alpha^{n_{\text{initial}}}$ and $\beta^{n_{\text{initial}}}$ on $A^{n_{\text{initial}}}$ and $B^{n_{\text{initial}}}$, respectively, and a projection ρ on each copy of $A \times B$.

For $i \geq 1$, write $X_i = (X_i^1, \ldots, X_i^{k_{\text{group}}})$, $Y_i = (Y_i^1, \ldots, Y_i^{k_{\text{group}}})$, $Z_i^j = (X_i^j, Y_i^j)$, $Y_i^{\text{miss}} = (Y_i^1, \ldots, Y_i^{k_{\text{group}}-1})$, and $Z_i^{\text{miss}} = (X_i, Y_i^{\text{miss}})$.

First, consider the following coupling of X_0, X_1, Y_0, and Y_1. The resulting coupling will *not* be a coupling of $Z_0 = (X_0, Y_0)$ and $Z_1 = (X_1, Y_1)$, but the resulting coupling as a measure on $(A \times B)^{n_{\text{initial}}+k_{\text{group}}}$ will have ρ as a projection on each copy of $A \times B$. Let $W = (E, F)$ have law given by the product measure $\rho^{k_{\text{group}}}$. Let $(\underline{Z_0}, \underline{Z_1}^{\text{miss}})$ be a star-coupling of $Z_0 = (X_0, Y_0)$ and $W^{\text{miss}} = (E, F^{\text{miss}})$. Using independent randomization, let Y_1^{rep} be such that the pair $(\underline{X_1}^{k_{\text{group}}}, Y_1^{\text{rep}})$ is the conditional quantile representation of $W^{k_{\text{group}}} = (E^{k_{\text{group}}}, F^{k_{\text{group}}})$. It is easy to verify from the properties of the star-coupling and the independence of $(W^1, \ldots, W^{k_{\text{group}}})$ that Y_1^{rep} is independent of $\underline{Y_1}^{\text{miss}}$; moreover, $\left(\underline{Z_0}, (\underline{X_1}, (\underline{Y_1}^{\text{miss}}, Y_1^{\text{rep}}))\right)$ is a coupling of Z_0 and W such that $(\underline{X_0}, \underline{X_1})$ has law $\alpha^{n_{\text{initial}}+k_{\text{group}}}$ and $(\underline{Y_0}, \underline{Y_1}^{\text{miss}}, Y_1^{\text{rep}})$ has law $\beta^{n_{\text{initial}}+k_{\text{group}}}$. We will refer to this coupling as the **star-coupling with replacement** of Z_0 and Z_1.

Here, two 'replacements' take place, Z_1 was replaced by $W = (E, F)$ which has the product measure $\rho^{k_{\text{group}}}$ as its law, and we only applied the star-coupling to Z_0 and W^{miss}, where in the final construction, the 'missing' value is replaced with a conditional quantile representation.

We iterate this construction as follows. First, let (Z_0', Z_1') be the star-coupling with replacement of Z_0 and Z_1. Next, we take the star-coupling with replacement of $\left((X_0', X_1'), (Y_0', Y_1')\right)$ and Z_2; to obtain random variables $\left((\underline{X_0'}, \underline{X_1'}, \underline{X_2}), (\underline{Y_0'}, \underline{Y_1'}, \underline{Y_2}^{\text{miss}}, Y_2^{\text{rep}})\right)$ taking values on $A^{n_{\text{initial}}+2k_{\text{group}}} \times B^{n_{\text{initial}}+2k_{\text{group}}}$ with a law that has projections $\alpha^{n_{\text{initial}}+2k_{\text{group}}}$ and $\beta^{n_{\text{initial}}+2k_{\text{group}}}$, respectively. Finally, it is clear that this construction can be extended an arbitrary number of times in the obvious way. We call this construction the **iterated-star coupling with replacement** of $Z_0, Z_1, \ldots Z_n$.

The importance of the star-coupling can be summarized in Proposition 13, below; it is a version of del Junco's [8, Proposition 4.7].

Proposition 13 (del Junco). *Let ρ be a probability measure on the finite set $A \times B$ and have marginals α and β, on A and B, respectively.*

Assume that $H(\alpha) = H(\beta)$. Let $k_{\mathrm{group}} \geq 2$. For $\eta > 0$, there exists $n_{\mathrm{initial}} = n_{\mathrm{initial}}(\eta, k_{\mathrm{group}}) \in \mathbb{Z}^+$ such that the following holds.

Let $n \in \mathbb{Z}^+$. Let $Z_i = (X_i, Y_i)$, for $i = 0, 1, \ldots, n$, have the grouping property with the law ρ and constants n_{initial} and k_{group}. Define the following product spaces

$$\mathbf{I}_j := A^{n_{\mathrm{initial}}} \times A^{k_{\mathrm{group}}j} \equiv A^{n_{\mathrm{initial}}+k_{\mathrm{group}}j},$$

$$\mathbf{J}_j := B^{n_{\mathrm{initial}}} \times B^{k_{\mathrm{group}}j} \equiv B^{n_{\mathrm{initial}}+k_{\mathrm{group}}j},$$

and

$$\bar{\mathbf{J}}_j := B^{(k_{\mathrm{group}}-1)j}.$$

For $\mathbf{y} = (y_0, (y_1, y_1'), \ldots, (y_j, y_j')) \in \mathbf{J}_j = B^{n_{\mathrm{initial}}} \times B^{k_{\mathrm{group}}j} = B^{n_{\mathrm{initial}}} \times (B^{k_{\mathrm{group}}-1} \times B) \times \cdots \times (B^{k_{\mathrm{group}}-1} \times B)$, let

$$\bar{\mathbf{y}} = (y_1, \ldots, y_j) \in \bar{\mathbf{J}}_j.$$

Let $\mathbf{W}_n = (\mathbf{X}_n, \mathbf{Y}_n)$ be a random variable given by the iterative star-coupling with replacement of $Z_0, Z_1, \ldots, Z_n$. There exists a deterministic function $\Psi : \mathbf{I}_n \to \bar{\mathbf{J}}_n$ such that $\mathbb{P}\big(\bar{\mathbf{Y}}_n = \Psi(\mathbf{X}_n)\big) > 1 - \eta$.

The proof of Proposition 13 uses the Shannon-McMillan-Breiman theorem and Remark 11. A version of Proposition 13 is also used the proof of Theorem 2 of Quas and Soo, see [18, Proposition 14].

5. The Proof of Proposition 13

Proof of Proposition 13. We will place conditions on n_{initial} later. Let $h := H(\alpha) = H(\beta)$. Let $\varepsilon > 0$ such that

$$h - 2\varepsilon > \big(1 - \tfrac{1}{k_{\mathrm{group}}}\big)(h + \varepsilon). \tag{3}$$

Set

$$\mathbf{L}_j := n_{\mathrm{initial}} + k_{\mathrm{group}}j, \text{ for } 0 \leq j \leq n.$$

Let $\mathbf{x} \in \mathbf{I}_j$ be given by $\mathbf{x} = (x_0, \ldots, x_j)$. We say that $\mathbf{x}$ is α-***good*** if

$$\alpha^{\mathbf{L}_j}(\mathbf{x}) < e^{-(h-\varepsilon)\mathbf{L_j}}, \tag{4}$$

and is α-***completely good*** if for all $0 \leq i \leq j$, we have $(x_0, \ldots, x_i) \in \mathbf{I}_i$ is good.

The corresponding definition for β is more complicated. We remark that in the presence of a strict entropy gap, $H(\alpha) > H(\beta)$, the definition could be more simple and symmetric (see for example, [10, page 366] or [18, Proof of Proposition 14]. We declare that every $\mathbf{y} \in \mathbf{J}_0$ is β-***good***. Set

$$\bar{\mathbf{L}}_j := (k_{\mathrm{group}} - 1)j \text{ for } 0 \leq j \leq n,$$

so that $\mathbf{L}_j = \bar{\mathbf{L}}_j + n_{\text{initial}} + j$. We say that $\mathbf{y} = (y_0, (y_1, y_1'), \ldots, (y_j, y_j')) \in \mathbf{J}_j$ is β-**good** if

$$\beta^{\bar{\mathbf{L}}_j}(\bar{\mathbf{y}}) > e^{-(h-2\varepsilon)\mathbf{L_j}}. \tag{5}$$

Note that being β-good does not depend on the behavior of $(y_0, y_1', \ldots, y_j')$ and $\mathbf{L}_j$ appears in the exponent rather than $\bar{\mathbf{L}}_j$ on the right hand side of (5). We say that $\mathbf{y}$ is β-**completely good** if for all $0 \leq i \leq j$, we have $\mathbf{y}_j = (y_0, (y_1, y_1'), \ldots, (y_i, y_i')) \in \mathbf{J}_i$ is good.

Note that if $\mathbf{y} \in \mathbf{J}_n$ is not completely good, then for some $j \geq 1$, we have $\beta^{\bar{\mathbf{L}}_j}(\bar{\mathbf{y}}_j) < e^{-(h-2\varepsilon)\mathbf{L}_j}$, and by (3),

$$\beta^{\bar{\mathbf{L}}_j}(\bar{\mathbf{y}}_j) < e^{-(h-2\varepsilon)\mathbf{L}_j} \leq e^{-(1-1/k_{\text{group}})(h+\varepsilon)\mathbf{L}_j} \leq e^{-(h+\varepsilon)\bar{\mathbf{L}}_j}. \tag{6}$$

For two elements $\mathbf{y}, \mathbf{z} \in \mathbf{J}_j$, we say that they are **equivalent** if $\bar{\mathbf{y}} = \bar{\mathbf{z}}$. We let $[\mathbf{y}] \subset \mathbf{J}_i$ be the equivalence class of $\mathbf{y}$. Given a measure on $\mathbf{I}_j \times \mathbf{J}_j$ we say it **finely splits** an element $\mathbf{x} \in \mathbf{I}_j$ if there exists $\mathbf{y}, \mathbf{z} \in \mathbf{J}_i$ such that $[\mathbf{y}] \neq [\mathbf{z}]$ and for which the measure assigns positive mass to both $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{x}, \mathbf{z})$.

For $j \geq 0$, let $\mathbf{W}_j = (\mathbf{X}_j, \mathbf{Y}_j)$ be a random variable given by the iterative star-coupling with replacement of $Z_0, Z_1, \ldots, Z_j$, where we set $\mathbf{W}_0 := Z_0$; thus $(\mathbf{X}_j, \mathbf{Y}_j)$ takes values in $\mathbf{I}_j \times \mathbf{J}_j$. We say that $\mathbf{x} \in \mathbf{I}_j$ is **desirable** if the following properties are satisfied.

(a) The element $\mathbf{x}$ is α-completely good.
(b) The element $\mathbf{x}$ is not finely split by (the law of) $\mathbf{W}_j = (\mathbf{X}_j, \mathbf{Y}_j)$.
(c) Furthermore, up to equivalence, there is a unique β-completely good $\mathbf{y} \in \mathbf{J}_j$ for which $(\mathbf{x}, \mathbf{y})$ receives positive mass under (the law of) $\mathbf{W}_j$.

For desirable $\mathbf{x} \in \mathbf{I}_j$, set $\Psi_j(\mathbf{x}) = \bar{\mathbf{y}}$, where $\mathbf{y}$ is determined by condition (c); otherwise if $\mathbf{x}$ is not desirable simply set $\Psi_j(\mathbf{x}) = \mathbf{y}'$ for some predetermined fixed $\mathbf{y}' \in \mathbf{J}_j$. Note that

$$\mathbb{P}\big(\bar{\mathbf{Y}}_j = \Psi_j(\mathbf{X}_j)\big) \geq \mathbb{P}(\mathbf{X}_j \text{ is desirable}).$$

Remark 11 and del Junco's inductive argument [8, Lemma 4.6] will be used to show that for all $j \geq 0$,

$$\mathbb{P}(\mathbf{X}_j \text{ is not desirable}) \quad \leq \quad \mathbb{P}(\mathbf{X}_j \text{ is not c.g.}) + \mathbb{P}(\mathbf{Y}_j \text{ is not c.g.}) +$$

$$|B|^{k_{\text{group}}} \sum_{i=0}^{j-1} e^{-\varepsilon\mathbf{L}_i}, \tag{7}$$

where "c.g." is short for completely good.

The case $j = 0$ is vacuous, since being good implies being completely good, and under Z_0 no elements are finely split.

Assume (7) for the case $j-1 \geq 0$. We show that (7) holds for the case j. Let E be the event that $\mathbf{X}_{j-1}$ is desirable, but $\mathbf{X}_j$ is not desirable. Clearly,

$$\mathbb{P}(\mathbf{X}_j \text{ is not desirable}) \leq \mathbb{P}(\mathbf{X}_{j-1} \text{ is not desirable}) + \mathbb{P}(E). \qquad (8)$$

Note that on the event E, the random variables $\mathbf{X}_{j-1}$ and $\mathbf{Y}_{j-1}$ are completely good. Observe that the event E is contained in the following three events

(I) $E_1 :=$ The random variable $\mathbf{X}_j$ is not good, but $\mathbf{X}_{j-1}$ is completely good.

(II) $E_2 :=$ The random variable $\mathbf{X_j}$ is completely good, but is finely split under the iterative star-coupling $\mathbf{W}_j$, even though $\mathbf{X}_{j-1}$ is desirable.

(III) $E_3 :=$ The random variable $\mathbf{Y_j}$ is not good, but $\mathbf{Y}_{j-1}$ is completely good.

Clearly,

$$\mathbb{P}(E_1) + \mathbb{P}(\mathbf{X}_{j-1} \text{ is not c.g.}) = \mathbb{P}(\mathbf{X}_j \text{ is not c.g.}). \qquad (9)$$

Similarly,

$$\mathbb{P}(E_3) + \mathbb{P}(\mathbf{Y}_{j-1} \text{ is not c.g.}) = \mathbb{P}(\mathbf{Y}_j \text{ is not c.g.}). \qquad (10)$$

Let us focus on the event E_2. Let $\mathbf{X}_j = (\mathbf{X}_{j-1}, X)$, so that X takes values in $A^{k_{\text{group}}}$. We show that for any $x \in A^{k_{\text{group}}}$ and any completely good $\mathbf{y} \in \mathbf{J}_{j-1}$ that

$$\mathbb{P}(E_2 \mid X = x, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}) \leq |B|^{k_{\text{group}}} e^{-\varepsilon \mathbf{L}_{j-1}}, \qquad (11)$$

so that $\mathbb{P}(E_2) \leq |B|^{k_{\text{group}}} e^{-\varepsilon \mathbf{L}_{j-1}}$ and it follows that (7) holds by (8), (9), (10), and the inductive hypothesis.

Note that if $\mathbf{x}$ and $\mathbf{y}$ are good, then

$$\begin{aligned}
\mathbb{P}(\mathbf{X}_{j-1} = \mathbf{x} \mid X = x, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}) &= \frac{\mathbb{P}(\mathbf{X}_{j-1} = \mathbf{x}, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}, X = x)}{\mathbb{P}(\bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}, X = x)} \\
&= \frac{\mathbb{P}(\mathbf{X}_{j-1} = \mathbf{x}, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}})}{\mathbb{P}(\mathbf{Y}_{j-1} = \bar{\mathbf{y}})} \qquad (12) \\
&\leq \frac{\mathbb{P}(\mathbf{X}_{j-1} = \mathbf{x})}{\mathbb{P}(\bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}})} \\
&\leq e^{-\varepsilon \mathbf{L}_{j-1}}, \qquad (13)
\end{aligned}$$

where (12) follows from the independence properties of the star-coupling (with replacement) and (13) follows from (4) and (5). Also note that if $\mathbf{x}$ is desirable, then if $(\mathbf{x}, x)$ is finely split under $\mathbf{W}_j$, then for the unique, up to equivalence, $\mathbf{y}$ for which $(\mathbf{x}, \mathbf{y})$ receives positive mass

under $\mathbf{W}_{j-1}$ there exist $(y, u), (y', u') \in B^{k_{\text{group}}} = B^{k_{\text{group}}-1} \times B$ such that $y \neq y'$ and both $((\mathbf{x}, x), (\mathbf{y}, y, u))$ and $((\mathbf{x}, x), (\mathbf{y}, y', u'))$ receive positive mass under $\mathbf{W}_j$. By Remark 11 and the definition of the star-coupling with replacement, for a fixed $x \in A^{k_{\text{group}}}$ and $\mathbf{y} \in \mathbf{J}_{j-1}$ the set of all $\mathbf{x}$ such that there exists distinct $y, y' \in B^{k_{\text{group}}-1}$ for which there are $u, u' \in B$ such that both $((\mathbf{x}, x), (\mathbf{y}, y, u))$ and $((\mathbf{x}, x), (\mathbf{y}, y', u'))$ receive positive mass under $\mathbf{W}_j$ has at most $|B|^{k_{\text{group}}-1} - 1$ elements; thus summing over all such $\mathbf{x}$ yields (11).

The Shannon–McMillan–Breiman theorem implies that n_{initial} can be chosen so that all three terms in (7) can be made smaller than $\eta/3$. This is done in the following way. Set

$$S_A(k, K) := \left\{ \mathbf{a} \in A^K : \alpha^\ell(\mathbf{a}) < e^{-(h-\varepsilon)\ell} \text{ for all } k \leq \ell \leq K \right\}$$

and

$$S_B(k, K) := \left\{ \mathbf{b} \in B^K : \beta^\ell(\mathbf{b}) > e^{-(h+\varepsilon)\ell} \text{ for all } k \leq \ell \leq K \right\},$$

where we have the slight abuse of notation that if $\mathbf{a} = (a_1, \ldots, a_K)$, then $\alpha^\ell(\mathbf{a}) = \alpha^\ell(a_1, \ldots, a_\ell)$.

First, by the Shannon–McMillan–Breiman theorem choose κ so that for all $K \geq \kappa$, we have

$$\beta^K(S_B(\kappa, K)) > 1 - \eta/3. \tag{14}$$

Next, using the Shannon–McMillian–Brieman theorem again, choose n_{initial} sufficiently large so that the following three inequalities are satisfied:

$$\alpha^K(S_A(n_{\text{initial}}, K)) > 1 - \eta/3 \text{ for all } K \geq n_{\text{initial}}, \tag{15}$$

$$\min \left\{ \beta^\ell(\mathbf{y}) > 0 : \mathbf{y} \in B^\ell, 0 \leq \ell \leq \kappa \right\} > e^{-(h-2\varepsilon)n_{\text{initial}}}, \tag{16}$$

and

$$|B|^{k_{\text{group}}} \sum_{i=n_{\text{initial}}}^{\infty} e^{-\varepsilon i} < \eta/3. \tag{17}$$

Finally, we will verify that this choice of n_{initial} is sufficient. Condition (15) gives that $\mathbb{P}(\mathbf{X}_j \text{ is not c.g.}) < \eta/3$. Recall that by definition, $\mathbf{L}_j \geq n_{\text{initial}}$ for all $j \geq 0$, so that (17) ensures that $|B|^{k_{\text{group}}} \sum_{i=0}^{j-1} e^{-\varepsilon \mathbf{L}_j} < \eta/3$. It remains to verify that $\mathbb{P}(\mathbf{Y}_j \text{ is not c.g.}) < \eta/3$.

The definition of completely good, (5), gives that if

$$\mathbf{y} = (y_0, (y_1, y_1'), \ldots, (y_j, y_j')) \in \mathbf{J}_n$$

is not completely good, then for some $i > 0$, we have

$$\beta^{\bar{\mathbf{L}}_i}(\bar{\mathbf{y}}_i) < e^{-(h-2\varepsilon)\mathbf{L}_i} < e^{-(h-2\varepsilon)n_{\text{initial}}}, \tag{18}$$

where $\mathbf{y}_i = (y_0, (y_1, y_1'), \ldots, (y_i, y_i'))$; inequalities (18) and (16) imply that

$$\bar{\mathbf{L}}_i > \kappa; \tag{19}$$

moreover, (6) gives that

$$\beta^{\bar{\mathbf{L}}_i}(\bar{\mathbf{y}}_i) \le e^{-(h+\varepsilon)\bar{\mathbf{L}}_i}. \tag{20}$$

Hence if $\mathbf{y}$ is not completely good, then by (19) and (20) it belongs to the complement of $S_B(\kappa, K)$ for all $K \ge \kappa$. Thus (14) gives that $\mathbb{P}(\mathbf{Y}_j$ is not c.g.$) < \eta/3$.   □

In Proposition 13, we have that given $\mathbf{x} \in \mathbf{I}_n$, with high probability, up to equivalence, it determines a corresponding

$$\mathbf{y} = (y_0, (y_1, y_1'), \ldots, (y_n, y_j')) \in \mathbf{J}_n.$$

It will be useful to refer to $(y_0, y_1', \ldots, y_n')$ as the **undetermined coordinates**, and $(y_1, \ldots, y_n)$ as the **destined coordinates**. We say that there are $n_{\text{initial}} + n$ undetermined coordinates, since $y_0 \in B^{n_{\text{initial}}}$, and $y_i \in B$ for $1 \le i \le n$, and there are $(k_{\text{group}} - 1)n$ destined coordinates.

6. Perturbing the joining

Let $\xi \in J$ be a monotone joining of marker form. We will define a perturbation ξ' of ξ using the iterated star-coupling with replacement. The perturbation will depend on a few parameters. With the help of Proposition 13, we will be able to make a choice of these parameters so that ξ' will be an almost factor and close to ξ in the metric defined in (2).

6.1. **Defining the perturbation.** Let $k_{\text{mark}} < r_{\text{mark}}$ be large integers to be chosen later. A **secondary marker** is the maximal union of at least k_{mark} consecutive primary markers, so that secondary markers have no filler between them and if the interval $[i, j]$ is a secondary marker for $x \in \Omega$, then x restricted to $[i, j]$ has the form $0101 \cdots 0101$.

Similarly, a **tertiary marker** is the maximal union of at least r_{mark} consecutive primary markers. We call the set of integers between but not including two secondary markers a **block**, and the set of integers between but not including two tertiary markers a **city**. Thus within a city there are blocks, which we consider ordered from left to right. Note that a block may contain primary markers.

Let $p \in (\frac{1}{2}, 1)$. Let $\xi \in J(p)$. Let $Z = (X, Y)$ have law ξ. Suppose that we are given that Z has primary markers given by $t \in \mathsf{T}$. Let $I \subset \mathbb{Z}$ be a block of length n. We are interested in the distribution of the random variable taking values in $\{0, 1\}^n \times \{0, 1\}^n$ given by the distribution of Z, conditioned on t, restricted to I. The **type** of the

block I is defined to be the vector containing an alternating sequence of integers that are the lengths of the filler and marker intervals in I and the **length** of the type is simply the sum of the integers in the type which give the length of the block. The distribution of this random variable is determined by the parameter p and the type of I. There are a countable number of types. Fix an enumeration $(\text{type}_i)_{i \in \mathbb{N}}$ of the types and let ρ_i be the corresponding law. Associate to each type-i block a large integer n^i_{initial} which will be chosen later; here i is an index that is not an exponent. A **census** of a city is the sequence of nonnegative integers c_i, where each c_i is the number of type-i blocks in the city.

Remark 14. Note that for every $i \in \mathbb{N}$, if the length of the type-i is n, then ρ_i is a probability measure on $\{0,1\}^n \times \{0,1\}^n$ with projections α_i and β_i that have equal entropy. The equal entropy assertion also follows from the duality between p and $1-p$. ◊

A **modification** of $Z = (X, Y)$ on a subset of $\mathbb{Z}$ is a coupling of X and Y given by $Z' = (X', Y')$ such that Z' is equal to Z off the subset and has the same primary markers as Z. We will define a modification Z' of Z so that the law of Z' will be a member of J. The modifications will be made independently on each city, so that we need only define what changes occur on a city. On each city the modifications will be made independently on each set of types, so that we need only define what changes occur on each set of types.

Suppose that the primary markers of Z are given by $t \in \mathsf{T}$. Fix $i \in \mathbb{N}$. Let us focus on the type-i blocks in a single fixed city. We will refer to this modification as the **star-modification of type-i** on a city. Suppose that the census c is such that we may write

$$c_i = n^i_{\text{initial}} + q_i k_{\text{group}} + r_i, \tag{21}$$

where $0 \leq r_i < k_{\text{group}}$ and q_i is an nonnegative integer. We will not make modifications on the last r_i blocks. Suppose that length of the type-i block is n. It may be helpful to think of two different copies of $\{0,1\}^n$ by setting $A = B = \{0,1\}^n$. Let $W = (W_j)_{j=1}^{c_i}$ be the set of random variables taking values in $A \times B$ obtained by taking the restriction of Z, conditioned to have primary markers given by t, to each block of type-i in the city. Although W gives an identical sequence, where each W_j has law ρ_j, it may not be independent. However, the projections on A and B are independent; if we write $W_j = (X_j, Y_j)$, then by Remark 5, we have that $X = (X_j)$ and $Y = (Y_j)$ are i.i.d. sequences. Consider the first n^i_{initial} random variables together as a single random variable

taking values in $(A \times B)^{n^i_{\text{initial}}}$, and each subsequent k_{group} random variables together as random variables taking values in $(A \times B)^{k_{\text{group}}}$. We obtain a sequence of random variables $M = (M_0, M_1, \ldots, M_{q_i})$. Thus M takes values on

$$(A \times B)^{n^i_{\text{initial}} + q_i k_{\text{group}}} \equiv A^{n^i_{\text{initial}} + q_i k_{\text{group}}} \times B^{n^i_{\text{initial}} + q_i k_{\text{group}}}.$$

Take the iterative star-coupling with replacement of these random variables to obtain new random variables $M' = (M'_0, \ldots, M'_{q_i})$; furthermore, using independent randomization, we may stipulate that these random variables are independent of Z. We define a modification Z' of Z by replacing the values of M with those of M', so that $Z = Z'$ off the type-i blocks in the city, and if Z restricted to the type-i blocks, then it is given by M, then Z' restricted to the type-i blocks is given by M'. The iterated star-coupling with replacement gives that the law of each M'_j projected onto each of the $n_{\text{initial}} + q_i k_{\text{group}}$ copies of $A \times B$ is ρ_i, so that monotonicity is preserved and the primary markers remain unchanged. Also, the projections of M and M' on $A^{c_i - r_i}$ have the same law. Similarly, the projections of M and M' on $B^{c_i - r_i}$ have the same law, so that by Remark 5, the random variable Z' gives the required coupling.

Note we have only defined the star modification of type-i when $c_i \geq n^i_{\text{initial}} + k_{\text{group}}$. In the case that c_i is not sufficiently large, we simply do nothing, that is, we stipulate that the star modification of type-i leaves everything unchanged.

For a single type-i, if we apply the star modification of the type on each city, independently, then we obtain a modification of Z that has law that belonging to J. We call this the **star-modification of type-i** of Z. We summarize our construction in the following proposition.

Proposition 15. *Let $p \in (\frac{1}{2}, 1)$ and $\xi \in J(p)$. If Z has law ξ and if Z' is a star modification of Z of a particular type, then the law of Z' is also a member of $J(p)$.*

Given a finite set of types, the **star-modification** of Z on the set of types is obtained by applying the star-modification in succession, starting with the smallest type.

Remark 16. In our construction of the star modification of type-i on a city, we relied on the fact that the law of M'_j still has a projection of ρ_i on each copy of $A \times B$ to ensure that primary markers and monotonicity are preserved. This fact will also be important for us later in proving that the parameters of the star-modification Z' of Z can be chosen so that it is a small perturbation in the d-metric, since on the event that the origin is contained in a block, and the coordinates of a cylinder set

C lie in that block, we have that the probabilities of C under Z and Z' are not only close, they are equal! This is another one of nice features of del Junco's star coupling.   ◊

6.2. **Choosing the parameters.** From the discussion in Section 3.3, it remains to show that given a joining $\xi \in J(p)$, we can choose parameters so that the star-modification of a random variable with law Z on a finite set of type results in a random variable with law ξ' that is close to ξ in the metric defined by (2) and is also an almost factor.

Proof of Theorem 1. Let $p \in (\frac{1}{2}, 1)$. Let $\varepsilon > 0$. As discussed in Section 3.3, it suffices to show that U_ε, the set of ε-almost factors from $B(1 - p, p)$ to $B(p, 1 - p)$ is dense. The proof that V_ε, the set of ε-almost factors from $B(p, 1 - p)$ to $B(1 - p, p)$ is similar with the roles of p and $1 - p$ reversed.

Let $\xi \in J(p)$ and Z have law ξ. Let $\varepsilon > 0$. We will choose the parameters for the star-modification Z' of Z as follows. The modification will occur on a finite set of types $\mathcal{T}$, which will be specified later. Recall that in the star-modification, some blocks are left **unchanged**, so that Z equals Z' on those blocks, and whereas some blocks are **modified** via the iterated star-coupling with replacement, so that Z may not equal Z' on those blocks. Note that Z' and Z always share the same primary markers, and although markers may lie in the modified coordinates they are always preserved. If ξ' is the law of Z', then these parameters will be chosen so that $d(\xi, \xi') < \varepsilon$ and $\xi' \in U_\varepsilon$. We choose the parameters as follows.

(i) Set $\varepsilon' := \varepsilon/100$.
(ii) Let $\delta > 0$ be small enough and ℓ^* be large enough so that two measures ζ and ζ' on $\{0,1\}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}$ are ε' close in the metric d^*, if for all cylinder sets $C \in \mathcal{C}_{\ell^*}$, we have $|\zeta(C) - \zeta'(C)| < \delta$.
(iii) Choose k_{mark} sufficiently large so that with probability at least $1 - \varepsilon'$ the origin is in a block and the interval $[-2\ell^*, 2\ell^*]$ is in the block.
(iv) With this choice of k_{mark}, there exists $L > 0$ such that with probability at least $1 - 2\varepsilon'$ the origin will in a block and the length of the block will be between ℓ^* and L.
(v) In particular, there exists a finite set of types $\mathcal{T}$, those with lengths between ℓ^* and L, such that with probability at least $1 - 2\varepsilon'$, each block will be of type $\mathcal{T}$. Since we have a fixed enumeration of the types, we will view $\mathcal{T}$ as a subset of $\mathbb{N}$.
(vi) Set $k_{\mathrm{group}} = \lceil 1/\varepsilon' \rceil + 1$.

(vii) For each $i \in \mathbb{N}$, choose n^i_{initial} via Proposition 13, by substituting $\rho = \rho_i$, $n_{\text{initial}} = n^i_{\text{initial}}$, and $\eta = \varepsilon'$.

(viii) Let c be the census of the city containing the origin. If c_i is sufficiently large, define q_i as in (21). Choose r_{mark} sufficiently large so that with probability at least $1 - \varepsilon'$, the origin is in a city, and for all $i \in \mathcal{T}$ the census will satisfy $n^i_{\text{initial}}/q_i < \varepsilon'$.

Applying Proposition 15 a finite number of times gives that $\zeta' \in J$. By Remark 16, conditions (i), (ii), and (iii), imply that $\mathrm{d}(\zeta', \zeta) < \varepsilon$.

It remains to verify that $\zeta' \in U_\varepsilon$. Call $t \in \mathsf{T}$ a **model marker** if the block containing the origin is a modified block and the origin lies in a destined coordinate. Property (vii) and Proposition 13 imply that for all model markers $t \in \mathsf{T}$ there exists a deterministic measurable $\psi : \Omega \to \{0, 1\}$ such that

$$\zeta'_t \left\{ (x, y) : (x, y_0) = (x, \psi(x)) \right\} > 1 - \varepsilon'. \tag{22}$$

For a particular type-i, with $c_i = n^i_{\text{initial}} + q_i k_{\text{group}} + r_i$ as in (21) the ratio of undetermined coordinates plus those that are unchanged to destined coordinates is $(n^i_{\text{initial}} + q_i + r_i)/q_i(k_{\text{group}} - 1)$. Recall that $r_i < k_{\text{group}}$. Conditions (iv), (v), (vi), and (viii), ensure us that the set of model markers has probability at least $1 - 7\varepsilon'$; this fact together with (22) and (i) imply that $\zeta' \in J$. □

7. SOME OTHER EXAMPLES

One of the key observations of Keane and Smorodinsky [11, Lemmas 2 and 3] that allowed the definition of markers in their proof of that two Bernoulli shifts $B(\mathbf{p})$ and $B(\mathbf{q})$ of equal entropy are isomorphic was that one could assume without loss of generality that $p_0 = q_0$ in the case where $\mathbf{p}$ and $\mathbf{q}$ give non-zero mass to three or more symbols, and in the case where $\mathbf{p}$ gives non-zero mass to only two symbols, then one can assume that $p_0^k p_1 = q_0^k q_1$ from some k. In general, in the construction of monotone factors, we may not make this reduction since monotonicity may not preserved. However by a straightforward adaptation of the proof of Theorem 1, the following monotone versions of the Keane and Smorodinsky reductions are enough to prove the existence of a monotone isomorphism.

Theorem 17. *Let $N \geq 2$. Let $\mathbf{p}$ and $\mathbf{q}$ be probability measures on $[N]$ of equal entropy. Suppose $\mathbf{p}$ stochastically dominates $\mathbf{q}$, and furthermore there exists $i \geq j$ such that $p_i = q_j$ and $\mathbf{p}^*$ stochastically dominates $\mathbf{q}^*$, where $\mathbf{p}^*$ is the law of a random variable with law $\mathbf{p}$*

conditioned not to take the value i, and $\mathbf{q}^*$ is the law of a random variable with law $\mathbf{q}$ conditioned not to take the value j. Then there exists a monotone isomorphism of $B(\mathbf{p})$ and $B(\mathbf{q})$.

Theorem 18. *Let $N \geq 2$. Let $\mathbf{p}$ and $\mathbf{q}$ be probability measures on $[N]$ of equal entropy. Suppose $\mathbf{p}$ stochastically dominates $\mathbf{q}$, and furthermore there exists $i \geq j$ and $k \geq \ell$ such that $p_i p_k = q_j q_\ell$ and for all $n \geq 1$, we have that $\mathbf{p}^{n*}$ stochastically dominates $\mathbf{q}^{n*}$, where $\mathbf{p}^{n*}$ is the law of a random vector with law $\mathbf{p}^n$ conditioned so that an occurrence of an i is never immediately followed by an occurrence of a k, and $\mathbf{q}^{n*}$ is the law of a random vector with law $\mathbf{q^n}$ conditioned so that an occurrence of a j is never followed by an occurrence of an ℓ. Then there exists a monotone isomorphism of $B(\mathbf{p})$ and $B(\mathbf{q})$.*

Acknowledgement

References

[1] K. Ball. Monotone factors of i.i.d. processes. *Israel J. Math.*, 150:205–227, 2005.

[2] J. R. Blum and D. L. Hanson. On the isomorphism problem for Bernoulli schemes. *Bull. Amer. Math. Soc.*, 69:221–223, 1963.

[3] R. Burton, M. Keane, and J. Serafin. Residuality of dynamical morphisms. *Colloq. Math.*, 85:307–317, 2000.

[4] R. Burton and A. Rothstein. Isomorphism theorems in ergodic theory. Technical report, Oregon State University, 1977.

[5] T. Downarowicz. *Entropy in dynamical systems*, volume 18 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2011.

[6] O. Gurel-Gurevich and R. Peled. Poisson thickening. *Israel J. Math.*, 196(1):215–234, 2013.

[7] A. del Junco. Finitary codes between one-sided Bernoulli shifts. *Ergodic Theory Dynamical Systems*, 1(3):285–301 (1982), 1981.

[8] A. del Junco. Bernoulli shifts of the same entropy are finitarily and unilaterally isomorphic. *Ergodic Theory Dynam. Systems*, 10(4):687–715, 1990.

[9] A. Katok. Fifty years of entropy in dynamics: 1958–2007. *J. Mod. Dyn.*, 1(4):545–596, 2007.

[10] M. Keane and M. Smorodinsky. A class of finitary codes. *Israel J. Math.*, 26:352–371, 1977.

[11] M. Keane and M. Smorodinsky. Bernoulli schemes of the same entropy are finitarily isomorphic. *Ann. of Math. (2)*, 109:397–406, 1979.

[12] J. L. F. King. Entropy in ergodic theory. In *Mathematics of complexity and dynamical systems. Vols. 1–3*, pages 205–224. Springer, New York, 2012.

[13] R. Lyons. Factors of iid on trees. *Combin. Probab. Comput., to appear.* http://arxiv.org/abs/1401.4197.

[14] R. Lyons *with* Y. Peres. *Probability on Trees and Networks*. Cambridge University Press, 2014. In preparation. Current version available at `http://mypage.iu.edu/~rdlyons/`.

[15] L. D. Mešalkin. A case of isomorphism of Bernoulli schemes. *Dokl. Akad. Nauk SSSR*, 128:41–44, 1959.

[16] D. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Math.*, 4:337–352, 1970.

[17] D. Ornstein. Newton's laws and coin tossing. *Notices Amer. Math. Soc.*, 60(4):450–459, 2013.

[18] A. Quas and T. Soo. A monotone Sinai theorem. *Ann. Probab.*, 44(1):107–130, 2016.

[19] T. de la Rue. An introduction to joinings in ergodic theory. *Discrete Contin. Dyn. Syst.*, 15:121–142, 2006.

[20] J. Serafin. Finitary codes, a short survey. In *Dynamics & stochastics*, volume 48 of *IMS Lecture Notes Monogr. Ser.*, pages 262–273. Inst. Math. Statist., Beachwood, OH, 2006.

[21] J. G. Sinaĭ. On a weak isomorphism of transformations with invariant measure. *Mat. Sb. (N.S.)*, 63 (105):23–42, 1964.

[22] Y. G. Sinai. *Selecta. Volume I. Ergodic theory and dynamical systems*. Springer, New York, 2010.

[23] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

[24] B. Weiss. The isomorphism problem in ergodic theory. *Bull. Amer. Math. Soc.*, 78:668–684, 1972.

Department of Mathematics, University of Kansas, 405 Snow Hall, 1460 Jayhawk Blvd, Lawrence, Kansas 66045-7594
Email address: `math@terrysoo.com`
URL: `http://www.terrysoo.com`