An invariant of finitary codes with finite expected square root coding length

Nate Harvey* Yuval Peres†

July 7, 2003

Abstract

Let p and q be probability vectors with the same entropy h. Denote by $B(p)$ the Bernoulli shift indexed by $\mathbb{Z}$ with marginal distribution p. Suppose that φ is a measure preserving homomorphism from $B(p)$ to $B(q)$. We prove that if the coding length of φ has a finite 1/2 moment, then $\sigma_p^2 = \sigma_q^2$, where $\sigma_p^2 = \sum_i p_i(-\log p_i - h)^2$ is the *informational variance* of p. In this result, the 1/2 moment cannot be replaced by a lower moment. On the other hand, for any $\theta < 1$, we exhibit probability vectors p and q that are not permutations of each other, such that there exists a finitary isomorphism Φ from $B(p)$ to $B(q)$ where the coding lengths of Φ and of its inverse have a finite θ moment. We also present an extension to ergodic Markov chains.

1 Introduction

Let $\mathbf{A} = \{\alpha_0, \ldots, \alpha_{a-1}\}$ be a finite alphabet and $p = (p_0, \ldots, p_{a-1})$ a probability vector with entropy $h(p) = \sum_{i=0}^{a-1} -p_i \log(p_i)$. Consider the Bernoulli shift $B(p) = (X, \mathcal{A}, \mathbf{P}, T)$, where $X = \mathbf{A}^{\mathbb{Z}}$ is equipped with the product σ-algebra $\mathcal{A}$, the product measure $\mathbf{P} = p^{\mathbb{Z}}$ and the left shift T. Let $\mathbf{B} = \{\beta_0, \ldots, \beta_{b-1}\}$ be another finite alphabet, and $q = (q_0, \ldots, q_{b-1})$ a probability vector; denote by $B(q) = (Y, \mathcal{B}, \mathbf{Q}, T)$ the corresponding Bernoulli

*Department of Mathematics, UC Berkeley, CA 94720, USA.

†Departments of Statistics and Mathematics, UC Berkeley, CA 94720, USA. peres@stat.berkeley.edu. Research supported by NSF Grants DMS-0104073, CCR-0121555 and a Miller Professorship at UC Berkeley.

shift. A **homomorphism** φ from $B(p)$ to $B(q)$ is a measurable map from X to Y, defined $\mathbf{P}$-a.e., such that $\mathbf{P}\varphi^{-1} = \mathbf{Q}$ and $\varphi T = T\varphi$ $\mathbf{P}$-a.e.. An **isomorphism** is an invertible homomorphism. A homomorphism φ from $B(p)$ to $B(q)$ is **finitary** if there exists a set $W \subseteq X$ with $\mathbf{P}(W) = 1$, that has the following property: for all $x \in W$ there exists $n = n(x)$ such that if $\widetilde{x} \in W$ and $\widetilde{x}_i = x_i$ for all $-n \leq i \leq n$, then $(\varphi(x))_0 = (\varphi(\widetilde{x}))_0$. We write $N_\varphi(x)$ for the minimal such n, and call $N_\varphi(x)$ the **coding length** of φ. A **finitary isomorphism** is an invertible finitary homomorphism whose inverse is also finitary.

By the Kolmogorov-Sinai Theorem (see, e.g., [10]), if $B(p)$ and $B(q)$ are isomorphic, then $h(p) = h(q)$. The converse was established by Ornstein [6]. Keane and Smorodinsky [3] proved that if $h(p) = h(q)$, then there exists a finitary isomorphism from $B(p)$ to $B(q)$. Parry [8] and Schmidt [11] showed that if a finitary isomorphism from $B(p)$ to $B(q)$ has finite expected coding length in both directions, then p and q must be permutations of each other.

In this paper, we prove that the **informational variance** of p,

$$\sigma_p^2 = \sum_{i=0}^{a-1} p_i \Big(-\log(p_i) - h(p) \Big)^2$$

is an invariant of isomorphisms φ that satisfy $\mathbf{E}\Big(N_\varphi^{1/2}\Big) < \infty$. More precisely:

Theorem 1 *Let p and q be probability vectors that satisfy $h(p) = h(q)$ and $\sigma_p^2 \neq \sigma_q^2$. Then there exists a constant $c_{p,q} > 0$ such that for any finitary homomorphism φ from $B(p)$ to $B(q)$, we have*

$$\liminf_{n\to\infty} \frac{\mathbf{E}(N_\varphi \wedge n)}{\sqrt{n}} \geq c_{p,q}$$

and consequently, $\mathbf{E}\Big(N_\varphi^{1/2}\Big) = \infty$.

(Here and throughout, $\mathbf{E}$ denotes expectation with respect to $\mathbf{P} = p^{\mathbb{Z}}$.)

The exponent $1/2$ in the theorem is sharp, since Meshalkin [5] (see §3) constructed a finitary isomorphism φ from $B(p)$ for $p = \Big(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\Big)$ to $B(q)$ for $q = \Big(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\Big)$, where $\mathbf{P}[N_\varphi > k]$ equals the probability that a simple random walk remains positive for k steps. Thus for Meshalkin's code,

$0 < \lim_k \mathbf{P}[N_\varphi > k]\sqrt{k} < \infty$, whence $\mathbf{E}\Big(N_\varphi^\theta\Big) < \infty$ for all $\theta < 1/2$. Clearly $\sigma_q < \sigma_p$ in this case, so Meshalkin's code is essentially optimal.

The assumption that $\sigma_p^2 \neq \sigma_q^2$ in Theorem 1 cannot be dropped, as shown by our next result.

Theorem 2 *For any $0 < \theta < 1$, there are probability vectors p and q where p is not a permutation of q, such that there exists a finitary isomorphism Φ from $B(p)$ to $B(q)$ that satisfies $\mathbf{E}(N_\Phi^\theta) < \infty$ and $\mathbf{E}_{\mathbf{Q}}(N_{\Phi^{-1}}^\theta) < \infty$.*

Theorem 1 is proved in the next section. In §3 we recall Meshalkin's isomorphism, and describe an adaptation of Meshalkin's code which motivates Theorem 2. In §4 we define a class of matchings useful for the proof of Theorem 2, and in §5 we prove the theorem. In §6 we define informational variance for ergodic Markov chains, and present an extension of Theorem 1 to this setting.

2 Proof of Theorem 1

With the notation of the introduction in force, we may assume that the probability vectors p and q satisfy $p_i > 0$ for all $0 \leq i < a$ and $q_j > 0$ for all $0 \leq j < b$. Let φ be a finitary homomorphism from $B(p)$ to $B(q)$. For $x = (x_k)_{k\in\mathbb{Z}} \in X$, write $X_i(x) = -\log(p(x_i)) - h(p)$, where $p(\alpha_j) = p_j$ for any j. Similarly, if $\varphi(x) = y = (y_k)_{k\in\mathbb{Z}} \in Y$, let $Y_i(x) = -\log(q(y_i)) - h(q)$. Since $\varphi\mathbf{P}^{-1} = \mathbf{Q}$, it follows that $\mathbf{E}(X_i) = \mathbf{E}(Y_i) = 0$. Let $S_n = \sum_{i=1}^n X_i$ and $R_n = \sum_{i=1}^n Y_i$. Write $t^+ = \max\{t, 0\}$.

Lemma 3 *If $\sigma_p^2 \neq \sigma_q^2$, then*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}}\mathbf{E}(R_n - S_n)^+ \geq \frac{|\sigma_q - \sigma_p|}{\sqrt{2\pi}}\,.$$

Proof. By a version of the central limit theorem (see [12], Cor. 2.1.9),

$$\lim_{n\to\infty} \mathbf{E}\Big(\frac{R_n^+}{\sqrt{n}}\Big) = \frac{\sigma_q}{\sqrt{2\pi}}\int_0^\infty te^{-\frac{t^2}{2}}\,dt = \frac{\sigma_q}{\sqrt{2\pi}}\,,$$

and similarly

$$\lim_{n\to\infty} \mathbf{E}\Big(\frac{S_n^+}{\sqrt{n}}\Big) = \frac{\sigma_p}{\sqrt{2\pi}}\,.$$

Since $(R_n - S_n)^+ \geq R_n^+ - S_n^+$, we infer that

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \mathbf{E}(R_n - S_n)^+ \geq \frac{\sigma_q - \sigma_p}{\sqrt{2\pi}} \tag{1}$$

and similarly

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \mathbf{E}(S_n - R_n)^+ \geq \frac{\sigma_p - \sigma_q}{\sqrt{2\pi}}\,. \tag{2}$$

If $\sigma_q > \sigma_p$, then (1) proves the lemma. In the remaining case, $\sigma_p > \sigma_q$, the assertion of the lemma follows from (2) by taking expectations in the identity

$$(R_n - S_n)^+ = (R_n - S_n) + (S_n - R_n)^+\,.$$

□

Lemma 4 *Let φ be a finitary homomorphism from $B(p)$ to $B(q)$. Denote $\lambda_q = \max\{-\log(q_j) \,:\, 0 \leq j \leq b-1\}$. Then for all n,*

$$\mathbf{E}(R_n - S_n)^+ \leq 2\lambda_q \mathbf{E}(N_\varphi \wedge n)\,.$$

Proof. Let

$$I_n = I_n(x) = \Big\{ i \in \{1, \ldots, n\} : N_\varphi(T^i x) > \min\{i, n+1-i\} \Big\}$$

and denote $J_n = \{1, \ldots, n\} \setminus I_n$. Observe that

$$\mathbf{E}|I_n| = \sum_{i=1}^n \mathbf{P}(i \in I_n) \leq 2 \sum_{i=1}^n \mathbf{P}(N_\varphi \geq i) \leq 2\mathbf{E}(N_\varphi \wedge n)\,. \tag{3}$$

Fix $x \in X$ and let $y = \varphi(x)$. Since

$$\Big\{ \widetilde{x} \in X : (\widetilde{x}_1, \ldots, \widetilde{x}_n) = (x_1, \ldots, x_n) \Big\} \subset \varphi^{-1} \Big\{ \widetilde{y} \in Y : \widetilde{y}_j = y_j \ \forall j \in J_n \Big\}\,,$$

it follows that

$$\mathbf{P}\Big\{ \widetilde{x} \in X : (\widetilde{x}_1, \ldots, \widetilde{x}_n) = (x_1, \ldots, x_n) \Big\} \leq \mathbf{Q}\Big\{ \widetilde{y} \in Y : \widetilde{y}_j = y_j \ \forall j \in J_n \Big\}\,.$$

Taking logarithms, this implies that

$$\sum_{k=1}^n \log p(x_k) \leq \sum_{k=1}^n \log q(y_k) - \sum_{i \in I_n} \log q(y_i) \leq \sum_{k=1}^n \log q(y_k) + \lambda_q |I_n|\,.$$

Since $h(p) = h(q)$, we deduce from the last equation and the definitions of R_n and S_n that $R_n - S_n \le \lambda_q |I_n|$, whence by (3),

$$\mathbf{E}(R_n - S_n)^+ \le \lambda_q \mathbf{E}|I_n| \le 2\lambda_q \mathbf{E}(N_\varphi \wedge n)\,.$$

□

Proof of Theorem 1. Lemmas 3 and 4 imply that

$$\liminf_{n\to\infty} \frac{\mathbf{E}(N_\varphi \wedge n)}{\sqrt{n}} \ge \frac{|\sigma_q - \sigma_p|}{2\lambda_q\sqrt{2\pi}} > 0\,, \tag{4}$$

so it only remains to verify the final assertion of the theorem.

Observe that $N_\varphi \wedge n \le \sqrt{N_\varphi n}$ and $(N_\varphi \wedge n)/\sqrt{n} \to 0$ $\mathbf{P}$-a.e. If we had $\mathbf{E}\Big(\sqrt{N_\varphi}\Big) < \infty$, then we could deduce by dominated convergence that $\mathbf{E}(N_\varphi \wedge n)/\sqrt{n} \to 0$, which contradicts (4). Thus $\mathbf{E}\Big(\sqrt{N_\varphi}\Big) = \infty$. □

A similar idea was used in a different context by Liggett [4].

3 Motivating examples and heuristics

Meshalkin's coding

First, we briefly recall the Meshalkin isomorphism [5]. Let $B(r)$ be the Bernoulli shift on the alphabet $\mathbf{A}_1 = \{\alpha_1, \dots, \alpha_5\}$ for $r = \left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$ and let $B(s)$ be the Bernoulli shift on the alphabet $\mathbf{B}_1 = \{\beta_1, \dots, \beta_4\}$ for $s = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. We represent the symbols of $\mathbf{A}_1$ as

$$\alpha_1 = 0\ ,\ \ \alpha_2 = \begin{matrix}1\\0\\0\end{matrix}\ ,\ \ \alpha_3 = \begin{matrix}1\\0\\1\end{matrix}\ ,\ \ \alpha_4 = \begin{matrix}1\\1\\0\end{matrix}\ ,\ \ \alpha_5 = \begin{matrix}1\\1\\1\end{matrix}\ ,$$

The symbols of $\mathbf{B}_1$ are represented as:

$$\beta_1 = \begin{matrix}0\\0\end{matrix}\ ,\ \ \beta_2 = \begin{matrix}0\\1\end{matrix}\ ,\ \ \beta_3 = \begin{matrix}1\\0\end{matrix}\ ,\ \ \beta_4 = \begin{matrix}1\\1\end{matrix}\ ,$$

The Meshalkin finitary isomorphism φ from $B(r)$ to $B(s)$ can be described in two equivalent ways. Given a sequence $x = (x_j)_{j\in\mathbb{Z}} \in \mathbf{A}_1^{\mathbb{Z}}$, denote by ℓ_i the length of the binary representation of $x_i \in \mathbf{A}_1$. The **random walk**

description of φ is obtained by defining, for each i with $\ell_i = 1$,

$$m(i) = \min\Big\{m \geq i : \sum_{j=i}^{m}(\ell_i - 2) = 0\Big\}. \tag{5}$$

Observe that $m(\cdot)$ is an injective map from $\{i \in \mathbb{Z} : \ell_i = 1\}$ onto $\{j \in \mathbb{Z} : \ell_j = 3\}$. For each $i \in \mathbb{Z}$ with $\ell_i = 1$, remove the bottom bit from $x_{m(i)}$ and append it at the bottom of x_i. This produces two symbols from $\mathbf{B}_1$ that are denoted $y_{m(i)}$ and y_i, respectively. Set $\varphi(x) = y = (y_j)_{j\in\mathbb{Z}}$.

Alternatively, we have an equivalent **inductive construction** of φ:
Step 1: For each $i \in \mathbb{Z}$ such that $\ell_i = 1$ and $\ell_{i+1} = 3$, send the bottom bit of x_{i+1} below x_i, output the resulting $\mathbf{B}_1$ symbols and **remove from consideration** both i and $i + 1$.

For each $n \geq 2$, perform:
Step n: For all $i \in \mathbb{Z}$ such that $\ell_i = 1, \ell_{i+n} = 3$ and $i, i + n$ have not been removed from consideration, send the bottom bit of x_{i+n} below x_i, output the corresponding $\mathbf{B}_1$ symbols and **remove from consideration** both i and $i + n$.

An adaptation of Meshalkin's coding

Next we describe informally a variant of the coding above, which we will generalize in §5 to prove Theorem 2. Consider the random walk where each increment X_i has $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = 3) = \frac{1}{2}$. The moment generating function is

$$\Gamma(z) = \mathbf{E}(z^{X_i}) = \frac{z + z^3}{2}.$$

Consider also the walk where each increment Y_i equals 2 with probability 1. This has moment generating function

$$\Delta(z) = \mathbf{E}(z^{Y_i}) = z^2.$$

These walks count the accumulated information for the Bernoulli shifts $B(r)$ and $B(s)$, where $r = \left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$ and $s = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ as in Meshalkin's coding. The entropy equality $h(r) = h(s)$ corresponds to the identity $\Gamma'(1) = \Delta'(1)$ while the inequality of informational variance corresponds to the inequality $\Gamma''(1) \neq \Delta''(1)$. The identity

$$\Gamma^2(z) - \Delta^2(z) = \frac{1}{2}\Big(\Gamma\Big(z^2\Big) - \Delta\Big(z^2\Big)\Big)$$

underlies the construction below. We add markers α_0 and β_0, respectively, to the alphabets $\mathbf{A}_1$ and $\mathbf{B}_1$ described above. Let $B(p)$ be the Bernoulli shift on the alphabet $\mathbf{A} = \{\alpha_0, \ldots, \alpha_5\} = \{\alpha_0\} \cup \mathbf{A}_1$, with associated probability vector $p = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}\right)$. Let $B(q)$ be the Bernoulli shift on the alphabet $\mathbf{B} = \{\beta_0, \ldots, \beta_4\}$ with associated probability vector $q = \left(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$.

Next we construct Φ, a finitary isomorphism from $B(p)$ to $B(q)$:

Step 0: If $x_i = \alpha_0$, let $(\Phi(x))_i = \beta_0$; that is, send markers to markers.

Step 1: Match the non-marker locations in pairs. Suppose that i is paired with j. If $\ell_i \neq \ell_j$, we can assume that $\ell_i = 1$ and $\ell_j = 3$ (otherwise reverse the roles). Remove the bottom bit of x_j and append it below x_i, output the resulting $\mathbf{B}$ symbols, and **remove from consideration** both i and j. If $\ell_i = \ell_j$, then do not remove i and j from consideration.

For each $n \geq 2$, perform:

Step n: The locations which we have not removed from consideration are grouped in 2^{n-1}- tuples. Each such 2^{n-1}-tuple is either of type 3 (which we define to mean that for every location i within the tuple $\ell_i = 3$), or of type 1. Using the markers, match the 2^{n-1}-tuples which have not been removed from consideration in pairs to form 2^n-tuples. If a 2^{n-1}-tuple ξ_3 of type 3 is matched with a 2^{n-1}-tuple ξ_1 of type 1, remove the bottom bit from each x_i in ξ_3, and append it to the corresponding symbol in ξ_1. Finally, output the symbols of $\mathbf{B}$ thus generated, and remove these locations from consideration.

The coding length for the isomorphism described above has essentially the same tails as Meshalkin's. To explain this, observe that the probability F_k that a symbol at the origin is not coded during the first k pairing stages is approximately 2^{-k} (the approximation is due to parity problems caused by markers.) After the k^{th} pairing stage, only about $1/2^k$ of the symbols remain uncoded, and these symbols are grouped into 2^k-tuples. Thus heuristically, the event F_k corresponds to an expected coding distance of order 4^k. This suggests that $\mathbf{P}(N_\Phi > t) \approx t^{-\frac{1}{2}}$. Indeed, for this example, Theorem 1 implies that $\mathbf{E}(N_\Phi^{1/2}) = \infty$ and the proof of Theorem 2 will show that $\mathbf{E}(N_\Phi^\theta) < \infty$ for all $\theta < 1/2$.

An example with $3/4 - \epsilon$ moments: heuristics. Consider different probability vectors p and q, chosen so that the random walks counting the accumulated information of non-marker symbols have moment generating functions

$$\Gamma(z) = \left(\left(\frac{1+z}{2} \right)^4 + \left(\frac{1-z}{2} \right)^4 \right) z^3 \tag{6}$$

and

$$\Delta(z) = \left(\left(\frac{1+z}{2} \right)^4 - \left(\frac{1-z}{2} \right)^4 \right) z^3 , \tag{7}$$

respectively. Then

$$\Gamma^2(z) - \Delta^2(z) = \frac{1}{8}\Big(\Gamma\Big(z^2\Big) - \Delta\Big(z^2\Big)\Big).$$

This example is the case $n = 2$ of the sequence of examples analyzed in §5; see (13) and (14).

Define a finitary coding Φ from $B(p)$ to $B(q)$ by adapting the recipe above (see §4 and §5 for details). To estimate the tails of N_Φ, start by observing that the probability that a symbol is not coded during the first k pairing stages is about 8^{-k}. At that stage, symbols are grouped into 2^k-tuples, and only $1/8^k$ of them remain uncoded, so heuristically, this event corresponds to an expected coding distance of order 16^k. This suggests that

$$\mathbf{P}(N_\Phi > t) \approx t^{-\frac{3}{4}}.$$

Indeed, for this example we will show in §5 that $\mathbf{E}(N_\Phi^\theta) < \infty$ for all $\theta < 3/4$. This is consistent with Theorem 1, since the identities $\Gamma'(1) = \Delta'(1)$ and $\Gamma''(1) = \Delta''(1)$ indicate that p and q have the same entropy and the same informational variance.

4 Ordered measure preserving matchings

In this section, we define a type of matching which we will employ in our constructions in §5, and derive some useful properties of these matchings. Let $\mathbf{C} = \{\gamma_1, \ldots, \gamma_c\}$ and $\mathbf{D} = \{\delta_1, \ldots, \delta_d\}$ be finite alphabets, and let $r = (r(\gamma_1), \ldots, r(\gamma_c))$ and $s = (s(\delta_1), \ldots, s(\delta_d))$ be probability vectors. Let

$\Gamma_k^* = \Gamma(k, \mathbf{C}, r) = \sum_{\mathbf{C}}\{r(\gamma_i) : r(\gamma_i) = 2^{-k}\}$

and

$\Delta_k^* = \Delta(k, \mathbf{D}, s) = \sum_{\mathbf{D}}\{s(\delta_j) : s(\delta_j) = 2^{-k}\}.$

Define an order relation $\prec$ on $\mathbf{C}$ such that $\gamma_1 \prec \cdots \prec \gamma_c$ and an order relation $\prec$ on $\mathbf{D}$ such that $\delta_1 \prec \cdots \prec \delta_d$. Endow $\mathbf{C}\times\mathbf{C}$ with the lexicographic ordering, i.e., define $\gamma_i\gamma_j \prec \gamma_m\gamma_n$ if $\gamma_i \prec \gamma_m$ or if $\gamma_i = \gamma_m$ and $\gamma_j \prec \gamma_n$. Similarly, endow $\mathbf{D} \times \mathbf{D}$ with the lexicographic ordering $\prec$.

Let $r(\gamma_i\gamma_j) = r(\gamma_i)r(\gamma_j)$ and $s(\delta_i\delta_j) = s(\delta_i)s(\delta_j)$. We define the **maximal ordered measure preserving matching (mompm)** $\psi = \psi_{(\mathbf{C},\mathbf{D},r,s)}$ from $\mathbf{C} \times \mathbf{C}$ to $\mathbf{D} \times \mathbf{D}$ given (r, s) as follows:

For all $t \in \mathbb{R}$, write the ordered set $\{x \in \mathbf{C} \times \mathbf{C} : r(x) = t\}$ in increasing order as $\{x_t(i) : 1 \le i \le \ell_t\}$, and similarly, write the ordered set $\{y \in \mathbf{D}\times\mathbf{D} : s(y) = t\}$ in increasing order as $\{y_t(i) : 1 \le i \le m_t\}$, assuming these sets are non-empty. Define $\psi(x_t(i)) = y_t(i)$ for $1 \le i \le \min\{\ell_t, m_t\}$.

Let $E = E(\mathbf{C}, \mathbf{D}, r, s)$ be the set in $\mathbf{C} \times \mathbf{C}$ where ψ is defined. Let $F = F(\mathbf{C}, \mathbf{D}, r, s) = \psi(E)$. Let $G = G(\mathbf{C}, \mathbf{D}, r, s) = \mathbf{C} \times \mathbf{C} - E$. Let $H = H(\mathbf{C}, \mathbf{D}, r, s) = \mathbf{D} \times \mathbf{D} - F$.

Let $\widetilde{r} = \widetilde{r}_{(\mathbf{C},\mathbf{D},r,s)} = (\widetilde{r}(x) : x \in G)$, where $\widetilde{r}(x) = \frac{r(x)}{\sum_{\widetilde{x}\in G} r(\widetilde{x})}$ and $\widetilde{s} = \widetilde{s}_{(\mathbf{C},\mathbf{D},r,s)} = (\widetilde{s}(y) : y \in H)$, where $\widetilde{s}(y) = \frac{s(y)}{\sum_{\widetilde{y}\in H} s(\widetilde{y})}$ be the **probability vectors induced** by (r, s) on G and H.

Let

$\Upsilon^*_k = \Upsilon(k, \mathbf{C}, r) = \sum_{x\in\mathbf{C}\times\mathbf{C}}\{r(x) : r(x) = 2^{-k}\}$,

$\Omega^*_k = \Omega(k, \mathbf{D}, s) = \sum_{y\in\mathbf{D}\times\mathbf{D}}\{s(y) : s(y) = 2^{-k}\}$,

$\Lambda^*_k = \Lambda(k, \mathbf{C}, \mathbf{D}, r, s) = \sum_{x\in G}\{r(x) : r(x) = 2^{-k}\}$,

and let

$\Xi^*_k = \Xi(k, \mathbf{C}, \mathbf{D}, r, s) = \sum_{y\in H}\{s(y) : s(y) = 2^{-k}\}$.

We say that ψ **reduces mass** by a factor of t if $\sum_{k=0}^{\infty} \Lambda^*_k = t$.

Let

$$\Gamma(z) = \Gamma(\mathbf{C}, \mathbf{D}, r, s, z) = \sum_{k=0}^{\infty} \Gamma^*_k z^k. \tag{8}$$

Define $\Delta(z)$, $\Upsilon(z)$, $\Omega(z)$, $\Lambda(z)$, and $\Xi(z)$ analogously. Then $\Upsilon(z) = \Gamma^2(z)$ and $\Omega(z) = \Delta^2(z)$. Also, $\Lambda(z) - \Xi(z) = \Upsilon(z) - \Omega(z)$.

Lemma 5 *Suppose* $\Gamma^2(z) - \Delta^2(z) = t\Gamma(z^2) - t\Delta(z^2)$. *Then:*

(i) $\Lambda(z) = t\Gamma(z^2)$ *and* $\Xi(z) = t\Delta(z^2)$

(ii) ψ reduces mass by a factor of t.

PROOF.

(i) $\Lambda(z) - \Xi(z) = \Upsilon(z) - \Omega(z) = \Gamma^2(z) - \Delta^2(z) = t\Gamma(z^2) - t\Delta(z^2)$, hence

$$\Lambda(z) = t\Gamma(z^2) \tag{9}$$

and

$$\Xi(z) = t\Delta(z^2). \tag{10}$$

(ii) By (9),

$$\sum_{k=0}^{\infty} \Lambda_k^* = t \sum_{k=0}^{\infty} \Gamma_k^* = t.$$

□

Let $C_1 = \mathbf{C}$, let $D_1 = \mathbf{D}$, let $r_1 = r$, and let $s_1 = s$. Inductively, let $C_{i+1} = G(C_i, D_i, r_i, s_i)$, let $D_{i+1} = H(C_i, D_i, r_i, s_i)$, let $r_{i+1} = \widetilde{r}_{(C_i,D_i,r_i,s_i)}$, and let $s_{i+1} = \widetilde{s}_{(C_i,D_i,r_i,s_i)}$. Let $\psi_i = \psi(C_i, D_i, r_i, s_i)$. Note that ψ_i matches 2^i-tuples to 2^i-tuples. We call $\{\psi_i\}_{i\geq 1}$ the **sequence of mompm's associated to** $(\mathbf{C}, \mathbf{D}, r, s)$. Let $\Gamma_i(z) = \Gamma(C_i, D_i, r_i, s_i, z)$. In particular, $\Gamma_1(z) = \Gamma(z)$ as defined in equation (8). Define $\Delta_i(z)$, $\Upsilon_i(z)$, $\Omega_i(z)$, $\Lambda_i(z)$, and $\Xi_i(z)$ analogously.

Inductive application of Lemma 5 gives:

Corollary 6 *Suppose* $\Gamma^2(z) - \Delta^2(z) = t\Gamma_1(z^2) - t\Delta_1(z^2)$.

(i) If $i \in \mathbb{Z}_+$, then $\Gamma_i^2(z) - \Delta_i^2(z) = t\Gamma_i(z^2) - t\Delta_i(z^2)$

(ii) If $i \in \mathbb{Z}_+$, then ψ_i reduces mass by a factor of t.

5 A class of codes with finite moments

Finally, we construct a class of examples to prove Theorem 2.

Fix $n \in \mathbb{Z}_+$.

Let $p_0 = q_0 = \frac{1}{2}$. Construct $p = (\frac{1}{2}, p_1, \ldots, p_{a-1})$ such that for each integer $m \in [0, n]$, exactly $2^{2m}\binom{2n}{2m}$ of the p_i take the value 2^{-2m-2n}. Thus if for $i \geq 1$, we denote $r(\alpha_i) = 2p_i$, then

$$\Gamma^*_{2m+2n-1} = \sum_i \{r(\alpha_i) : r(\alpha_i) = 2^{-(2m+2n-1)}\} = \frac{1}{2^{2n-1}}\binom{2n}{2m} \tag{11}$$

for all $m \in \mathbb{Z}$ such that $0 \leq m \leq n$. Define $\Gamma^*_k = 0$ for all other k.

Similarly, for $j \geq 1$, denote $s(\beta_j) = 2q_j$ and construct $q = (\frac{1}{2}, q_1, \ldots, q_{b-1})$ such that

$$\Delta^*_{2m+2n} = \sum_j \{s(\beta_j) : s(\beta_j) = 2^{-(2m+2n)}\} = \frac{1}{2^{2n-1}}\binom{2n}{2m+1} \tag{12}$$

for all $m \in \mathbb{Z}$ such that $0 \leq m \leq n-1$, and define $\Delta^*_k = 0$ for other k.

Let $B(p)$ be the Bernoulli shift with probability vector p on the alphabet $\mathbf{A} = \{\alpha_0, \ldots, \alpha_{a-1}\}$. Let $B(q)$ be the Bernoulli shift with probability vector q on the alphabet $\mathbf{B} = \{\beta_0, \ldots, \beta_{b-1}\}$.

Let $\mathbf{C} = \{\alpha_1, \ldots, \alpha_{a-1}\}$ and let $\mathbf{D} = \{\beta_1, \ldots, \beta_{b-1}\}$. Consider the probability vectors $r = (r(\alpha_i) : 1 \leq i \leq a-1)$ and let $s = (s(\beta_j) : 1 \leq j \leq b-1)$. Relative to these, define all other terms as in §4.

Lemma 7 *If $i \in \mathbb{Z}_+$, then ψ_i reduces mass by a factor of $\frac{1}{2^{2n-1}}$.*

Proof. Recall that $\Gamma(z) = \sum_{k=0}^{\infty} \Gamma^*_k z^k$ and $\Delta(z) = \sum_{k=0}^{\infty} \Delta^*_k z^k$. By the binomial theorem and equations (11) and (12), we find that

$$\begin{aligned}
\Gamma^2(z) - \Delta^2(z) &= (\Gamma(z) - \Delta(z))(\Gamma(z) + \Delta(z)) \\
&= 2\Big(\frac{1-z}{2}\Big)^{2n} z^{2n-1} 2\Big(\frac{1+z}{2}\Big)^{2n} z^{2n-1} \\
&= \frac{1}{2^{2n-1}} 2\Big(\frac{1-z^2}{2}\Big)^{2n} z^{4n-2} \\
&= \frac{1}{2^{2n-1}}\Big(\Gamma\Big(z^2\Big) - \Delta\Big(z^2\Big)\Big),
\end{aligned}$$

so the desired result holds by Corollary 6. □

Example

When $n = 2$, we may let $\mathbf{A} = \{\alpha_0, \dots, \alpha_{41}\}$; $\mathbf{B} = \{\beta_0, \dots, \beta_{40}\}$; $p = (p_0, \dots, p_{41})$ such that $p_0 = 2^{-1}$, $p_1 = 2^{-4}$, $p_2 = \cdots = p_{25} = 2^{-6}$, and $p_{26} = \cdots = p_{41} = 2^{-8}$; and $q = (q_0, \dots, q_{40})$ such that $q_0 = 2^{-1}$, $q_1 = \cdots = q_8 = 2^{-5}$, and $q_9 = \cdots = q_{40} = 2^{-7}$.

Taking logarithms to base 2, $h(p) = h(q) = \frac{7}{2}$ and $\sigma_p^2 = \sigma_q^2 = \frac{27}{4}$, hence Theorem 1 does not apply. These vectors correspond to the generating functions in (6) and (7). We find that

$$\left(\Gamma_3^*, \Gamma_5^*, \Gamma_7^*\right) = \left(\frac{1}{8}, \frac{3}{4}, \frac{1}{8}\right) \tag{13}$$

$$\left(\Delta_4^*, \Delta_6^*\right) = \left(\frac{1}{2}, \frac{1}{2}\right) \tag{14}$$

$$\left(\Upsilon_6^*, \Upsilon_8^*, \Upsilon_{10}^*, \Upsilon_{12}^*, \Upsilon_{14}^*\right) = \left(\frac{1}{64}, \frac{3}{16}, \frac{19}{32}, \frac{3}{16}, \frac{1}{64}\right) \tag{15}$$

$$\left(\Omega_8^*, \Omega_{10}^*, \Omega_{12}^*\right) = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) \tag{16}$$

$$\left(\Lambda_6^*, \Lambda_{10}^*, \Lambda_{12}^*\right) = \left(\frac{1}{64}, \frac{3}{32}, \frac{1}{64}\right) \tag{17}$$

$$\left(\Xi_8^*, \Xi_{12}^*\right) = \left(\frac{1}{16}, \frac{1}{16}\right). \tag{18}$$

Definition of Φ

For $x = (x_k)_{k\in\mathbb{Z}} \in X = \mathbf{A}^{\mathbb{Z}}$, define a ***j*-marker** as a run of at least $2nj$ consecutive α_0 symbols. Define a ***j*-gap** as the location of the non-α_0 symbols between neighboring j-markers.

Let $G_{j,0} = \{g(j,0,1), \dots, g_(j,0,\ell_{j,0})\}$ be the ordered elements (from left to right) of the j-gap containing $\min\{i \geq 0 : x_i \neq \alpha_0\}$. More generally, let $G_{j,i} = \{g(j,i,1), \dots, g(j,i,\ell_{j,i})\}$ be the ordered elements of the i^{th} j-gap to the right of $G_{j,0}$ (to the left if $i < 0$).

Step 0: If $x_i = \alpha_0$, let $(\Phi(x))_i = \beta_0$.

Step 1: Within each 1-gap, match the elements in pairs, starting from the left ($g(1,i,1)$ with $g(1,i,2)$, $g(1,i,3)$ with $g(1,i,4)$, etc.). All the elements will be paired except possibly $g(1,i,\ell_{1,i})$.

If $\psi_1(x_{g(1,i,2k+1)}x_{g(1,i,2k+2)})$ is defined, then let

$$(\Phi(x))_{g(1,i,2k+1)}(\Phi(x))_{g(1,i,2k+2)} = \psi_1(x_{g(1,i,2k+1)}x_{g(1,i,2k+2)}),$$

and **remove from consideration** $g(1, i, 2k + 1)$ and $g(1, i, 2k + 2)$.

Starting from the left, match the pairs which have not been removed from consideration into quartets. If ψ_2 of the symbols at the position of a quartet is defined, output the result in the position of the quartet and remove the elements of the quartet from consideration.

Iterate, matching 2^{k-1}-tuples which have not been removed from consideration into 2^k-tuples and applying ψ_k, until $2^k > \ell_{1,i}$.

For each $j \geq 2$, do the following:

Step j**:** Within each j-gap, starting from the left, match into pairs any elements in $G_{j,i}$ which were not paired in any of the previous steps, and apply ψ_1 as in Step 1.

Match into quartets any previously unmatched pairs (including the pairs just created) which have not been removed from consideration, and apply ψ_2, etc., iterating until $2^k > \ell_{j,i}$.

When $n = 1$, this is the code described in §3. The next two lemmas are needed as preparation for bounding the tails of N_Φ.

Lemma 8 *If* $f(x) = \ell_{j,0}^{-1}\mathbf{1}_{[x_0 \neq \alpha_0]}$*, then* $\mathbf{E}(f) = 2^{-2nj-1}$.

Proof. The sum $\sum_{m=1}^{M} f(T^m x)$ differs from the number of j-gaps in $[1, M]$ by at most 2. Counting j-gaps in $[1, M]$ is equivalent to counting runs of $2nj$ marker symbols followed by a non-marker symbol; such strings have asymptotic frequency 2^{-2nj-1}. Taking the limit of $\frac{1}{M}\sum_{m=1}^{M} f(T^m x)$ as $M \to \infty$, the ergodic theorem yields the assertion. □

Lemma 9 *For* $j \geq 1$ *Let* $L_{j,i} = 2nj + g(j, i, \ell_{j,i}) - g(j, i, 1)$ *denote the "span" of the* i^{th} j*-gap. If* $\theta < 1$*, then*

$$\mathbf{E}\Big((L_{j,0} - L_{j-1,0})^\theta \mid x_0 \neq \alpha_0\Big) \leq 2^{(2+2nj)\theta}\,.$$

Proof. The expected distance between the beginnings of successive j-gaps is 2^{1+2nj} by Kac's Theorem (see [10], p. 46), whence

$$\mathbf{E}\Big(L_{j,0} - L_{j-1,0} \mid x_0 \neq \alpha_0\Big) \leq 2^{2+2nj}\,.$$

The assertion of the lemma follows by Jensen's inequality. □

Lemma 10 *If* $\theta < 1 - \frac{1}{2n}$, *then* $\mathbf{E}(N_\Phi(x))^\theta < \infty$.

PROOF. Recall $L_{j,i}$ from the previous lemma and define $L_{0,i} = 0$. If Step j determines $(\Phi(x))_0$, then $N_\Phi(x) \leq L_{j,0}$. Let A_j be the event that Steps 1 to j do not determine $(\Phi(x))_0$.

Let B_j be the event that the 0^{th} coordinate is matched at least j times by the end of Step j, but $(\Phi(x))_0$ has not yet been determined. Let C_j be the event that at the end of Step j, the 0^{th} coordinate has been matched at most $j-1$ times (so it is not part of a 2^j-tuple). Clearly, for each $j \geq 1$,

$$\mathbf{P}(A_j) \leq \mathbf{P}(B_j) + \mathbf{P}(C_j)\,. \tag{19}$$

Every time an undetermined coordinate is matched, the probability that it remains undetermined is $\frac{1}{2^{2n-1}}$, whence

$$\mathbf{P}(B_j) \leq \Big(\frac{1}{2^{2n-1}}\Big)^j\,. \tag{20}$$

Since, for all k and j, at most one 2^k-tuple in $G_{j,0}$ is unmatched at the end of Step j, it follows that

$$\mathbf{P}(C_j) \leq \mathbf{E}\Big(\frac{\sum_{k=0}^{j-1} 2^k}{\ell_{j,0}} \mid x_0 \neq \alpha_0\Big) \leq 2^j \Big(\frac{1}{2^{2n}}\Big)^j = \Big(\frac{1}{2^{2n-1}}\Big)^j$$

by Lemma 8. Thus

$$\mathbf{P}(A_j) \quad \leq \quad 2\Big(\frac{1}{2^{2n-1}}\Big)^j.$$

Therefore

$$\begin{aligned}
\mathbf{E}(N_\Phi(x))^\theta \quad &\leq \quad \sum_{j=1}^{\infty} \mathbf{P}(A_{j-1}) \mathbf{E}\left(L_{j,0}^\theta - L_{j-1,0}^\theta \mid A_{j-1}\right) \\
&\leq \quad \sum_{j=1}^{\infty} 2\left(\frac{1}{2^{2n-1}}\right)^j \mathbf{E}\left((L_{j,0} - L_{j-1,0})^\theta \mid A_{j-1}\right)
\end{aligned}$$

Conditional on the event that $x_0 \neq \alpha_0$, the random variable $(L_{j,0} - L_{j-1,0})$ is independent of the event A_{j-1}, hence by Lemma 9,

$$\mathbf{E}(N_\Phi(x))^\theta \quad \leq \quad \sum_{j=1}^{\infty} 2\Big(\frac{1}{2^{2n-1}}\Big)^j \mathbf{E}\Big((L_{j,0} - L_{j-1,0})^\theta \mid x_0 \neq \alpha_0\Big)$$

$$\leq \sum_{j=1}^{\infty} 2\Big(\frac{1}{2^{2n-1}}\Big)^j 4(2^{2n})^{j\theta} = 8\sum_{j=1}^{\infty}\Big(\frac{1}{2^{2n-1}}\Big)^{j\left(1-\frac{2n\theta}{2n-1}\right)} < \infty.$$

□

A similar argument gives:

Lemma 11 *If $\theta < 1 - \frac{1}{2n}$, then $\mathbf{E}_Q(N_{\Phi^{-1}}(x))^\theta < \infty$.*

Proof of Theorem 2. By Lemmas 10 and 11, it only remains to verify that Φ is an isomorphism. Since Φ is finitary, it gives an a.e. defined map from $B(p)$ to $B(q)$. As our definition of $(\Phi(x))_i$ depends only on the position of i within its j-blocks, Φ is translation invariant. Since each ψ_k is a one-to-one measure preserving matching from previously uncoded sequences to previously uncoded sequences, it follows that Φ is measure preserving and invertible. More precisely, for $\mathbf{P}$-a.e. $x \in X$ and any $n \geq 1$, all the symbols in the string $(x_{-n}, \ldots, x_n)$ get coded within a finite distance. This means that the cylinder set $\Big\{\widetilde{x} \in X : (\widetilde{x}_{-n}, \ldots, \widetilde{x}_n) = (x_{-n}, \ldots, x_n)\Big\}$ is partitioned into countably many cylinder sets $\mathcal{C}_j$ (and a set of measure zero); each $\mathcal{C}_j$ is mapped, using one of our matchings $\psi_{k(j)}$, to a cylinder set $\Phi(\mathcal{C}_j)$ in Y with $\mathbf{Q}(\Phi(\mathcal{C}_j)) = \mathbf{P}(\mathcal{C}_j)$. This completes the proof.

□

6 Extension to ergodic Markov chains

Let $\mathbf{A} = \{\alpha_0, \ldots, \alpha_{a-1}\}$ be a finite alphabet and let $p = (p(\alpha_i, \alpha_j))_{0\leq i,j\leq a-1}$ be an irreducible stochastic matrix. The associated Markov chain $M(p)$ is ergodic and has a (strictly positive) unique stationary distribution $\widetilde{p} = (\widetilde{p}(\alpha_0), \ldots, \widetilde{p}(\alpha_{a-1}))$. Similarly, let $\mathbf{B} = \{\beta_0, \ldots, \beta_{b-1}\}$ be a finite alphabet and let $q = (q(\beta_i, \beta_j))_{0\leq i,j\leq b-1}$ be a stochastic matrix such that $M(q)$ is ergodic with unique stationary distribution $\widetilde{q} = (\widetilde{q}(\beta_0), \ldots, \widetilde{q}(\beta_{b-1}))$. The Markov chain $M(p)$ has entropy

$$h(p) = \sum_{0\leq i,j\leq a-1} -\widetilde{p}(\alpha_i)p(\alpha_i, \alpha_j)\log p(\alpha_i, \alpha_j)\,.$$

We will assume that $h(p) = h(q)$. Let φ be a finitary homomorphism from $M(p)$ to $M(q)$. For $x = (x_k)_{k\in\mathbb{Z}} \in \mathbf{A}^{\mathbb{Z}}$, let $X_i(x) = -\log(p(x_{i-1}, x_i)) - h(p)$. Similarly, if $\varphi(x) = y = (y_k)_{k\in\mathbb{Z}} \in \mathbf{B}^{\mathbb{Z}}$, let $Y_i(x) = -\log(q(y_{i-1}, y_i)) - h(q)$. Let $S_{m,n} = \sum_{i=m+1}^{n} X_i$ and let $R_{m,n} = \sum_{i=m+1}^{n} Y_i$. Since φ is measure preserving, it follows that $\mathbf{E}(X_i) = \mathbf{E}(Y_i) = 0$. Let

$$\lambda_p = \max_{0\le i,j\le a-1} \{-\log(p(\alpha_i, \alpha_j)) : p(\alpha_i, \alpha_j) \neq 0\},$$

and let $\gamma_p = \max_{0\le i\le a-1}\{-\log(\widetilde{p}(\alpha_i))\}$.

The following central limit theorem can be found, e.g., in [1], p. 422 under an additional aperiodicity assumption, and in [2] in much greater generality. For the reader's convenience, we include a brief proof.

Lemma 12 *If $M(p)$ is an ergodic Markov chain on a finite alphabet, then there exists a constant $\sigma_p \ge 0$ depending only on p such that $\frac{S_{0,n}}{\sqrt{n}} \Rightarrow \chi\sigma_p$ in law, where χ denotes a standard normal variable.*

We define σ_p^2 to be the **asymptotic informational variance** of p.

Proof. For any $x \in \mathbf{A}^{\mathbb{Z}}$, let $T_0 = \min\{t > 0 : x_t = \alpha_0\}$. Inductively, for $i \ge 0$, let $T_{i+1} = \min\{t > T_i : x_t = \alpha_0\}$. The increments $T_i - T_{i-1}$ are i.i.d. and have exponential tails. The partial sums $\{S_{T_{i-1},T_i}\}_{i\ge1}$ are also i.i.d. Let $d_p = \mathbf{E}(T_1 - T_0) > 0$. By an application of the ergodic theorem and the law of large numbers, $\mathbf{E}(S_{T_0,T_1}) = 0$. Since $|X_i| \le \lambda_p$, it follows that $S_{T_0,T_1}^2 \le (T_1 - T_0)^2\lambda_p^2$, whence $\mathbf{E}(S_{T_0,T_1}^2) = c_p^2 < \infty$. Let $N_n = \min\{m > 0 : T_m \ge n\}$. Since $\frac{N_n}{n/d_p} \to 1$ in probability, the random index central limit theorem (see [1], p. 116) states that

$$\frac{S_{T_0,T_{N_n}}}{\sqrt{n}} \Rightarrow \frac{c_p\chi}{\sqrt{d_p}}. \tag{21}$$

Define $\sigma_p^2 = \frac{c_p^2}{d_p}$. Since $\mathbf{E}(T_{N_n} - n) \le \max_{0\le i\le a-1} \mathbf{E}(T_0 \mid x_0 = \alpha_i)$ for all $n \in \mathbb{Z}_+$, it follows that

$$\mathbf{E}(|S_{0,n} - S_{T_0,T_{N_n}}|) \le \lambda_p\mathbf{E}(T_0) + \lambda_p \max_{0\le i\le a-1} \mathbf{E}(T_0 \mid x_0 = \alpha_i).$$

In conjunction with (21), this gives $\frac{S_{0,n}}{\sqrt{n}} \Rightarrow \chi\sigma_p$. □

Let $J_n = \{i \in \{1,\ldots,n\} : n_\varphi(T^i x) > \min\{i, n+1-i\}$ or $n_\varphi(T^{i-1}x) > \min\{i-1, n+2-i\}\}$. Let $I_n = \{1,\ldots,n\} - J_n$.

As in §2, we deduce from the CLT and uniform integrability:

Lemma 13 *If* $\sigma_p^2 \neq \sigma_q^2$*, then* $\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \mathbf{E}(R_{0,n} - S_{0,n})^+ \geq \frac{|\sigma_q - \sigma_p|}{\sqrt{2\pi}}$.

The proofs of Lemma 4 and Theorem 1 adapt to prove the following.

Lemma 14 *Suppose* $M(p)$ *and* $M(q)$ *are ergodic Markov chains and* φ *is a finitary homomorphism from* $M(p)$ *to* $M(q)$*, Then for all* n*,*

$$\mathbf{E}(R_{0,n} - S_{0,n})^+ \leq \gamma_p + 4\lambda_q(\mathbf{E}(N_\varphi \wedge n) + 1)$$

Theorem 15 *Let* $M(p)$ *and* $M(q)$ *be ergodic Markov chains such that* $h(p) = h(q)$ *and* $\sigma_p^2 \neq \sigma_q^2$*. If* φ *is a finitary homomorphism from* $M(p)$ *to* $M(q)$*, then* $\mathbf{E}\left(\sqrt{N_\varphi(x)}\right) = \infty$*. More precisely,* $\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \mathbf{E}(N_\varphi(x) \wedge n) \geq c_{p,q} > 0$*.*

7 Higher moments: a problem

Theorem 1 and our constructions in §5 suggest the following:
Question. Let p and q be probability vectors with $h(p) = h(q)$. Fix an integer $k > 2$. Suppose that φ is a finitary homomorphism from $B(p)$ to $B(q)$, that satisfies $\mathbf{E}\left(N_\varphi^{1-1/k}\right) < \infty$. Does it follow that

$$\sum_i p_i (\log p_i)^k = \sum_j q_j (\log q_j)^k \ \ ?$$

Acknowledgment

We thank Alexander Holroyd for helpful discussions, Serban Nacu for comments and Ben-Zion Rubshtein and Jeff Steif for references.

References

[1] R. Durrett (1996), *Probability: Theory and Examples*, Second Edition. Duxbury Press, New York.

[2] M. I. Gordin and B. A. Lifšic (1978), The central limit theorem for stationary Markov processes. *Soviet Math. Dokl.* **19**, 392–394.

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

[4] T. Liggett (2002), Tagged particle distributions or how to choose a head at random. *In and out of equilibrium* (Mambucaba, 2000), Progr. Probab. **51**, Birkhäuser Boston, 133–162.

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

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

[7] W. Parry (1979), Finitary isomorphisms with finite expected code lengths. *Bull. London Math. Soc.* **11**, 170–176.

[8] W. Parry (1979), An information obstruction to finite expected coding length. *Ergodic theory* (Proc. Conf., Math. Forschungsinst., Oberwolfach, 1978), pp. 163–168, Lecture Notes in Math. **729**, Springer, Berlin.

[9] W. Parry and K. Schmidt (1984), Invariants of finitary isomorphisms with finite expected code-lengths. *Conference in modern analysis and probability* (New Haven, Conn., 1982), 301–307, Contemp. Math. **26**, Amer. Math. Soc., Providence, RI.

[10] K. Petersen (1983), *Ergodic Theory.* Cambridge University Press.

[11] K. Schmidt (1984), Invariants for finitary isomorphisms with finite expected code lengths. *Invent. Math.* **76**, 33–40.

[12] D. W. Stroock (1993), *Probability theory, an analytic view.* Cambridge University Press.