RECONSTRUCTION SEQUENCES AND EQUIPARTITION MEASURES: AN EXAMINATION OF THE ASYMPTOTIC EQUIPARTITION PROPERTY

J.T. Lewis[a], C.-E. Pfister[b], R. Russell[a], W.G. Sullivan[ac]

[a]Dublin Institute for Advanced Studies
10 Burlington Road
Dublin 4, Ireland

[b]Ecole Polytechnique Fédérale
Département de Mathématiques
CH-1015 Lausanne, Switzerland

[c]University College
Department of Mathematics
Belfield, Dublin 4, Ireland

Abstract: We consider a stationary source emitting letters from a finite alphabet $\mathtt{A}$. The source is described by a stationary probability measure α on the space $\Omega := \mathtt{A}^{\mathbb{N}}$ of sequences of letters. Denote by Ω_n the set of words of length n and by α_n the probability measure induced on Ω_n by α. We consider sequences $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ having special properties. Call $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ a **supporting sequence** for α if $\lim_n \alpha_n[\Gamma_n] = 1$. It is well-known that the exponential growth-rate of a supporting sequence is bounded below by $h_{Sh}(\alpha)$, the Shannon entropy of the source α. For efficient simulation, we require Γ_n to be as large as possible, subject to the condition that the measure α_n is approximated by the **equipartition measure** $\beta_n^{\Gamma_n}$, the probability measure on Ω_n which gives equal weight to the words in Γ_n and zero weight to words outside it. We say that a sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is a **reconstruction sequence** for α if each Γ_n is invariant under cyclic permutations and $\lim_n \beta_m^{\Gamma_n} = \alpha_m$ for each $m \in \mathbb{N}$. We prove that the exponential growth-rate of a reconstruction sequence is bounded above by $h_{Sh}(\alpha)$. We use a large–deviation property of the **cyclic empirical measure** to give a constructive proof of an existence theorem: if α is a stationary source, then there exists a reconstruction sequence for α having maximal exponential growth-rate; if α is ergodic, then the reconstruction sequence may be chosen so as to be supporting for α. We prove also a characterization of ergodic measures which appears to be new.

Key Words: asymptotic equipartition property, empirical measure, stationary, ergodic, Kolmogorov, reconstruction, large deviations

1 Introduction

Let Ω_n denote the set of words of length n formed using letters taken from a finite alphabet $\mathtt{A}$ of size r; if Γ_n is a subset of Ω_n, then we denote the number of elements in Γ_n by $\#\Gamma_n$. Let $\Omega := \mathtt{A}^{\mathbb{N}}$ denote the set of all sequences of letters from $\mathtt{A}$ – the set of words of infinite length. Suppose the source emitting the letters which form the words is described by a stationary probability measure α on the the space Ω; can we find a sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ of sets of words of increasing length from which we can reconstruct the measure α ?

Take $\mathtt{A} = \{1, 0\}$ with α the Bernoulli $(\frac{1}{3}, \frac{2}{3})$ measure on $\Omega = \mathtt{A}^{\mathbb{N}}$. Define, for n such that $\frac{n}{3}$ is an integer,

$$\Gamma_n := \{\mathbf{a} \in \Omega_n : \frac{1}{n}\sum_{j=1}^{n} \mathbf{a}_j = \frac{1}{3}\} ; \tag{1.1}$$

these are the words of length n in which the relative frequencies of ones and zeroes are $\frac{1}{3}$ and $\frac{2}{3}$. We claim that the measure α is determined completely by the sequence $\{\Gamma_n \subset \Omega_n : \frac{n}{3} \in \mathbb{N}\}$.

The first step is to construct a sequence of **equipartition measures**. Define $\beta_n^{\Gamma_n}$ to be the probability measure on Ω_n which gives equal weight to the words in Γ_n and zero weight to words outside it: for each subset Δ_n of Ω_n, put

$$\beta_n^{\Gamma_n}[\Delta_n] = \frac{\#(\Delta_n \cap \Gamma_n)}{\#\Gamma_n} . \tag{1.2}$$

For $m < n$, every measure λ_n on Ω_n induces a measure λ_m on Ω_m via the projection $X_m^n : \Omega_n \to \Omega_m$ which selects the first m letters from a word of length n. We claim that

$$\lim_{\substack{n\to\infty \\ \frac{n}{3}\in\mathbb{N}}} \beta_m^{\Gamma_n}[a] = \alpha_m[a] \tag{1.3}$$

for every $a \in \Omega_m$ and every $m \in \mathbb{N}$; here α_m is the measure on Ω_m induced by the measure α on Ω via the projection $X_m : \Omega \to \Omega_m$ which selects the first m letters in an infinite sequence. But the set $\{\alpha_m[a] : a \in \Omega_m, m \in \mathbb{N}\}$ is precisely the data required, according to Kolmogorov's Reconstruction Theorem [K], to determine the measure α completely. Our claim (1.3) can be proved using a conditional limit theorem of van Campenhout and Cover [CC]; see (6.13) of Section 6.

We say that a sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is a **reconstruction sequence** for α if each Γ_n is invariant under cyclic permutations and

$$\lim_n \beta_m^{\Gamma_n} = \alpha_m \tag{1.4}$$

for each $m \in \mathbb{N}$; an alternative definition of the concept is discussed in Section 6. The concept of a **reconstruction sequence** for α is illustrated by the example of the sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ defined by (1.1).

For efficient simulation, we would like the sequence to grow as fast as possible so that we have large samples of words of reasonable length. Consider the sequence constructed using a thickened shell:

$$\Gamma_n^\delta := \{\mathbf{a} \in \Omega_n : |\frac{1}{n}\sum_{j=1}^{n} \mathbf{a}_j - \frac{1}{3}| \le \delta\} ; \tag{1.5}$$

this sequence has a faster growth-rate than the sequence defined by (1.1); it is a reconstruction sequence, but not for α: for $0 < \delta < \frac{1}{6}$, the sequence $\{\beta^{\Gamma_n^\delta}\}$ converges to α^δ, the Bernoulli $(\frac{1}{3} + \delta,\ \frac{2}{3} - \delta)$ measure on Ω. This can be deduced from a conditional limit theorem proved in [LPS1] (see also [LPS2]). (For $\delta \geq \frac{1}{6}$, we recover the Bernoulli $(\frac{1}{2}, \frac{1}{2})$ measure.)

These examples illustrate a property of reconstruction sequences: they cannot grow too quickly. In fact, we have the following upper bound on the growth-rate:

- If $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is a reconstruction sequence for α, then

$$\limsup_n \frac{1}{n} \log \#\Gamma_n \leq h_{Sh}(\alpha)\,, \tag{1.6}$$

  where $h_{Sh}(\alpha)$ is the **Shannon entropy** of α.

There are reconstruction sequences which grow very slowly; in Section 6, we give a proof of the following result:

> Let α be a stationary source; then there exists a reconstruction sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ for α which has zero growth-rate:
>
> $$\limsup_n \frac{1}{n} \log \#\Gamma_n \;=\; 0\,. \tag{1.7}$$

We have the following existence theorem:

- Let α be a stationary source; then there exists a reconstruction sequence for α having maximal growth-rate.

We turn our attention to another property which a sequence of sets of words may have: we call a sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ a **supporting sequence** for α if

$$\lim_n \alpha_n[\Gamma_n] \;=\; 1\,, \tag{1.8}$$

where α_n is the probability measure induced on Ω_n.

The sequence defined by (1.5) is, for all values of $\delta > 0$, an example of a supporting sequence for the Bernoulli $(\frac{1}{3}, \frac{2}{3})$ measure while that defined by (1.1) fails to be.

A supporting sequence cannot grow too slowly. We have the following lower bound to the growth-rate:

- If $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is a supporting sequence for α, then

$$\liminf_n \frac{1}{n} \log \#\Gamma_n \;\geq\; h_{Sh}(\alpha)\,. \tag{1.9}$$

For economical coding, it is important to have a supporting sequence which grows as slowly as possible. We have the following existence theorem:

- Let α be an ergodic source; then there exists a supporting sequence for α having minimal growth-rate.

Since the Shannon entropy is a lower bound on the growth-rate of a supporting sequence and an upper bound on the growth-rate of a reconstruction sequence, a sequence which has both properties has a growth-rate equal to the Shannon entropy.

Let us examine how we can modify the construction (1.5) in order to get a sequence which is both a reconstruction sequence for α and a supporting sequence for α. Define

$$\Gamma'_n := \{\mathrm{a} \in \Omega_n : |\frac{1}{n}\sum_{j=1}^{n} \mathrm{a}_j - \frac{1}{3}| \leq \log n/\sqrt{n}\}. \tag{1.10}$$

We can use the conditional limit theorem in [LPS1] to prove that the sequence $\{\Gamma'_n\}$ has the reconstruction property and the Central Limit Theorem to prove that it has the supporting property for the Bernoulli $(\frac{1}{3}, \frac{2}{3})$ measure.

Let us examine this construction more closely. It selects those words of length n for which the relative frequency of ones lies in a closed neighbourhood of $\frac{1}{3}$ (and hence the relative frequency of zeroes lies in a closed neighbourhood of $\frac{2}{3}$). We can think of the measure α as being described by a vector $(\frac{1}{3}, \frac{2}{3})$. Introducing a relative frequency vector

$$R_n(\mathrm{a}) := (\frac{1}{n}\sum_{j=1}^{n} \mathrm{a}_j,\ 1 - \frac{1}{n}\sum_{j=1}^{n} \mathrm{a}_j)\ , \tag{1.11}$$

we can re-write (1.10) as

$$\Gamma'_n := R_n^{-1} F_n \tag{1.12}$$

where F_n is the closed ball of radius $\log n/\sqrt{n}$ centred on the point $(\frac{1}{3}, \frac{2}{3})$. In other words, what we have done is to define a mapping R_n from Ω_n to the space of Bernoulli measures and a decreasing sequence $\{F_n\}$ of closed neighbourhoods of α in the space of Bernoulli measures whose intersection is α, and taken $\Gamma_n{}'$ to be those words $\mathrm{a} \in \Omega_n$ for which $R_n(\mathrm{a})$ lies in F_n. This choice has some nice properties:

1. the set $\Gamma_n{}'$ is invariant under cyclic permutations of the letters in the words — this is important because the measure α which we are attempting to approximate is stationary;

2. the set $\Gamma_n{}'$ is nonempty for all n sufficiently large — this is important because we want to condition on it.

In order to prove our existence theorems, we need to generalize the construction which produced the sequence $\{\Gamma_n{}'\}$. We introduce a class of sequences called **canonical sequences**; to define these, we make use of the **cyclic empirical measure**, a mapping T_n from Ω to $\mathcal{M}_1^+(\Omega)$, the space of probability measures on Ω.

The cyclic empirical measure is a generalization of the relative frequency vector which will do what we want in the general case — its precise definition will be given later. For the present, we will describe it in terms of its marginals. For each $\omega \in \Omega$, we have a measure $T_n(\omega)[\cdot]$ defined on subsets of Ω; the projection $X_m : \Omega \to \Omega_m$ induces a measure $T_{n,m}(\omega)[\cdot]$ on the subsets of Ω_m:

$$T_{n,m}(\omega)[\Delta_m] \;:=\; T_n(\omega)[X_m^{-1}\Delta_m] \,. \tag{1.13}$$

For $m \leq n$, and $\mathsf{a} = (\mathsf{a}_1, \cdots, \mathsf{a}_m) \in \Omega_m$, we can describe $T_{n,m}(\omega)[\mathsf{a}]$ directly. Consider the n cyclic permutations of the word $X_n\omega = (\omega_1, \cdots, \omega_n)$:

$$(\omega_1, \cdots, \omega_n),\ (\omega_2, \cdots, \omega_n, \omega_1), \cdots, (\omega_n, \omega_1, \cdots, \omega_{n-1}) \;; \tag{1.14}$$

then $T_{n,m}(\omega)[\mathsf{a}]$ is the fraction of these in which the first m entries coincide with $\mathsf{a} = (\mathsf{a}_1, ..., \mathsf{a}_m)$. Thus $T_{n,1}(\omega)[\mathsf{a}_1]$ is just the relative frequency of the letter a_1 in the word $X_n\omega$, $T_{n,2}(\omega)[(\mathsf{a}_1, \mathsf{a}_2)]$ is the relative frequency of the adjacent pair $(\mathsf{a}_1, \mathsf{a}_2)$ in the (cyclic) word $X_n\omega$, and so on. We take Γ_n to be the set of words $X_n\omega = (\omega_1, \cdots, \omega_n)$ in Ω_n for which $T_{n,m}(\omega)[\mathsf{a}]$ is close to $\alpha_m[\mathsf{a}]$ for all $m \leq n$ and all $\mathsf{a} \in \Omega_m$. The sequence $\{\Gamma_n\}$ is a canonical sequence.

Of course, it is necessary to say what we mean by 'close to'; that is what is accomplished by the formal definition: let $\{F_n\}$ be a decreasing sequence of closed neighbourhoods of α in the space of measures whose intersection is α; for each n, the measure $T_n(\omega)$ depends only on the first n coordinates of ω and so $T_n^{-1}F_n$ determines a subset Γ_n of Ω_n; a sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ constructed in this way with Γ_n nonempty for all n sufficiently large, is called the **canonical sequence** based on $\{F_n\}$. The definition of T_n ensures that the set Γ_n is cyclically invariant.

Our reason for introducing the concept of a canonical sequence is the following result which holds for an arbitrary stationary source α:

- Every canonical sequence for α is a reconstruction sequence for α.

All we have done so far is to push the problem of existence one stage back: does there exist a canonical sequence for an arbitrary stationary source α ? There is no difficulty in finding a sequence of neighbourhoods which contract to α; the problem is to prove that the subsets Γ_n which they determine are non-empty — at least for all n sufficiently large. One way of doing this is to show that the growth-rate of $\{\Gamma_n\}$ is strictly positive; this will be the case if the sequence of neighbourhoods of α contracts sufficiently slowly. Our strategy is to start with an arbitrary sequence of closed neighbourhoods contracting to α and slow its rate of contraction until we are sure that the corresponding subsets Γ_n are growing fast enough; to check on this, we use large–deviation theory. (In fact, we use only the most basic result of the theory: the large–deviation lower bound, a direct consequence of the existence of the rate–function; see [LP], for example. A derivation of the large–deviation properties of the cyclical empirical measure which we require can be found in [LPS].) We prove the following result:

- Let α be a stationary source; then there exists a canonical sequence for α having maximal growth-rate.

A canonical sequence is not necessarily supporting. In the case of a Bernoulli measure, we were able to use the Central Limit Theorem to find a rate which makes the sequence $\{R_n^{-1}F_n\}$ supporting; in the general case, we do not have such a precise estimate available. Nevertheless, when the measure is ergodic we are able to use the Ergodic Theorem to prove the existence of a canonical sequence which is supporting. The converse also holds so that we have the following characterization of ergodic measures:

- Let α be a stationary source; then α is ergodic if and only if there exists a canonical sequence which is supporting for α.

We have seen that canonical sequences of subsets are useful and arise naturally in the reconstruction problem for stationary sources. It is instructive to compare them with the set of 'typical' sequences of letters associated with an ergodic source. Let α be an ergodic source; there exists a set $\Delta(\alpha) \subset \Omega$ with $\alpha[\Delta(\alpha)] = 1$ such that each sequence ω in $\Delta(\alpha)$ determines α uniquely (see Section 6) . For a stationary source α, a canonical sequence plays an analogous rôle: let $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ be a canonical sequence for α; any sequence $\{\mathbf{a}_n \in \Gamma_n : n \in \mathbb{N}\}$ of words determines α uniquely (this is proved in Section 6). A canonical sequence has some advantages over the typical set: one is that every stationary source has a canonical sequence — it is not necessary that the source be ergodic; another is that a canonical sequence is associated with an increasing sequence $\{\mathcal{F}_n : n \in \mathbb{N}\}$ of σ-algebras, where $\mathcal{F}_n$ is generated by the first n coordinate functions, while the typical set $\Delta(\alpha)$ is in the tail σ-algebra so that the first n coordinates of an element of $\Delta(\alpha)$ are irrelevant.

To put our results in context, it may be useful to recall the Asymptotic Equipartition Property: in terms of the concepts used here, the conclusion of the theorem of Shannon–McMillan–Breiman ([S], [M], [B]), may be stated:

> Let α be an ergodic source; then for each $\delta > 0$ there exists a sequence $\{\Gamma_n^\delta\}$ which is supporting for α and whose growth–rate satisfies

$$h_{Sh}(\alpha) \leq \liminf_n \frac{1}{n} \log \#\Gamma_n^\delta \leq \limsup_n \frac{1}{n} \log \#\Gamma_n^\delta \leq h_{Sh}(\alpha) + \delta \,. \tag{1.15}$$

It follows from the construction used in the proof that each word $\mathbf{a} \in \Gamma_n$ satisfies

$$b^{-n(h_{Sh}(\alpha)+\delta)} \leq \alpha_n[\mathbf{a}] \leq b^{-n(h_{Sh}(\alpha)-\delta)}\,, \tag{1.16}$$

where b is the base of logarithms used in the definition of the Shannon entropy (see (2.15) below); this is the origin of the name **asymptotic equipartition property**. The sequence $\{\Gamma_n^\delta\}$ is a not a reconstruction sequence for α. In Section 6, we discuss how this construction may be refined to yield a sequence which has both properties.

2 Statement of Results

In Section 2, we make precise the concepts introduced informally in Section 1 and sketch proofs of our main theorems. The main result of this paper is an existence theorem:

Theorem 2.1 *Let α be a stationary source; then there exists a reconstruction sequence for α having maximal growth-rate. If, in addition, α is ergodic, then the reconstruction sequence may be chosen so as to be a supporting sequence for α.*

We will give a constructive proof of this theorem. A by-product of this investigation is a characterization of ergodic measures:

Theorem 2.2 *Let α be a stationary source; then α is ergodic if and only if there exists a canonical sequence for α which is supporting for α.*

This section is, to some extent, self-contained: we recall the definitions and results required to understand the concepts defined here. We state six lemmas, indicating roughly on what their proofs depend; we prove our existence theorem using the first five — the sixth is used to complete the proof of the characterization of ergodic measures. The reader who is prepared to accept the lemmas need read no further.

The first two lemmas are proved in Section 3 using properties of the **specific information gain** defined there. The third lemma is crucial: it states that every canonical sequence is a reconstruction sequence; it is proved in Section 4.

To construct sequences with the required properties, we make use of the **cyclic empirical measure** to define **canonical sequences** of subsets. The sequence of probability distributions of the cyclic empirical measure with respect to the uniform product measure on Ω satisfies a large deviation principle with the specific information gain as rate-function. This is exploited in Section 5 where the fourth, fifth and sixth lemmas are proved.

Some of the ideas have their origin in statistical mechanics; some readers will find reference to this confusing while others will find it enlightening. Having in mind those in the first category, we make no reference to statistical mechanics in the body of the paper; for the others, we provide in the final section, Section 6, a commentary on the concepts and results.

We now make precise the structures we are considering. The space $\Omega = \mathtt{A}^{\mathbb{N}}$ is the space of infinite sequences with entries taken from a finite alphabet $\mathtt{A} = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(r)}\}$ having $r > 1$ letters; the map $x_j : \Omega \to \mathtt{A}$ is the coordinate projection onto the jth factor in the product. Let $\mathcal{F}_n = \sigma(x_1, \ldots, x_n)$ be the σ-algebra generated by the first n coordinate functions and let $\mathcal{F} = \sigma(x_n : n \in \mathbb{N})$ be the σ-algebra generated by all coordinate functions. Since $\mathtt{A}$ contains r elements, the σ-algebra $\mathcal{F}_n$ is generated by the r^n atoms $\{A_{\mathbf{a}} = X_n^{-1}\mathbf{a} : \mathbf{a} \in \mathtt{A}^n\}$, where $X_n : \Omega \to \Omega_n := \mathtt{A}^n$ is the projection onto the first n coordinates. Sometimes the discussion can be clarified by working on Ω_n rather than Ω. Let α be a probability measure defined on $\mathcal{F}$. On Ω_n

we have α_n, the image measure defined on the subsets of Ω_n by $\alpha_n[B] = \alpha[X_n^{-1}B]$. Equivalently, one could consider α on Ω restricted to the σ–algebra $\mathcal{F}_n$. The two viewpoints are complementary.

Recall that, for $m < n$, every measure λ_n on Ω_n induces a measure λ_m on Ω_m via the projection $X_m^n : \Omega_n \to \Omega_m$ which selects the first m letters from a word of length n; since $X_m = X_m^n \circ X_n$, it follows that if the measures $\{\lambda_n : n \in \mathbb{N}\}$ are induced from a probability measure λ on $\mathcal{F}$ so that for all $n \in \mathbb{N}$ we have

$$\lambda_n = \lambda \circ X_n^{-1}, \tag{2.1}$$

then they satisfy the compatibility conditions

$$\lambda_m = \lambda_n \circ (X_m^n)^{-1} \tag{2.2}$$

for all $m \in \mathbb{N}$ and all $n > m$. Conversely, Kolmogorov's Reconstruction Theorem [K] implies that given a sequence $\{\lambda_n : n \in \mathbb{N}\}$ of probability measures satisfying the compatibility conditions (2.2), there exists a unique probability measure λ on $\mathcal{F}$ such that for all $n \in \mathbb{N}$ the probability measures λ_n are given by (2.1).

For a function $f : \Omega \to \mathbb{R}$, we write $f \in \mathcal{F}_n$ to mean that f is $\mathcal{F}_n$–measurable and bounded; we write $f \in \mathcal{F}_{\text{loc}}$ to mean that there exists a finite n with $f \in \mathcal{F}_n$. We use the notation $\mathcal{M}_1^+$ to denote the space of probability measures on $(\Omega, \mathcal{F})$ with the coarsest topology for which each mapping

$$\mathcal{M}_1^+ \ni \lambda \mapsto \int f \, d\lambda \in \mathbb{R} \tag{2.3}$$

is continuous whenever $f \in \mathcal{F}_{\text{loc}}$: this is called the **bounded local topology**. In Section 1, we encountered the following notion of convergence: a sequence $\{\lambda^{(n)} : n \in \mathbb{N}\}$ of probability measures on $\mathcal{F}$ converges to the probability measure ν in the sense of **convergence of finite–dimensional marginals** if, for every $m \in \mathbb{N}$ and every $a \in \Omega_m$,

$$\lim_n \lambda_m^{(n)}[a] = \nu_m[a]. \tag{2.4}$$

In the present set–up, convergence of finite–dimensional marginals is equivalent to convergence in the bounded local topology; this can be seen from the following considerations: $\lambda_m^{(n)}[\mathbf{a}]$ is the integral of the indicator function

$$1_{\mathbf{a}}(\omega) = \begin{cases} 1 & \text{if } \omega_1 = \mathbf{a}_1, \ldots, \omega_n = \mathbf{a}_n \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

of the atom $X_m^{-1}\mathbf{a}$ of $\mathcal{F}_m$ and the set $\{1_{\mathbf{a}} \in \mathcal{F}_m : \mathbf{a} \in \Omega_m, m \in \mathbb{N}\}$ spans $\mathcal{F}_{\text{loc}}$.

We use a product probability measure β on $(\Omega, \mathcal{F})$ as a reference measure; we take β to be the measure on $\mathcal{F}$ which, for all $n \in \mathbb{N}$, assigns equal probability to each of the r^n atoms of $\mathcal{F}_n$, so that $\beta[A_{\mathbf{a}}] = r^{-n}$ for each $\mathbf{a} \in \Omega_n$. Notice that for $\Gamma_n \subset \Omega_n$, we have

$$\beta_n[\Gamma_n] = \frac{\#\Gamma_n}{\#\Omega_n}. \tag{2.6}$$

Let α be a probability measure on $(\Omega, \mathcal{F})$; we may think of α as characterizing the statistical properties of the source of the words, and we shall refer to α itself as the **source**.

We recall the definitions of **stationary measure** and **ergodic measure**. We define the **shift operator** S on Ω by

$$(S\omega)_k := \omega_{k+1} \qquad k \in \mathbb{N}. \tag{2.7}$$

The shift S acts on functions $f : \Omega \to \mathbb{R}$ by composition:

$$Sf := f \circ S. \tag{2.8}$$

We define the action of S on a measure α by

$$\int f \, d(S\alpha) := \int (Sf) \, d\alpha. \tag{2.9}$$

From the shift operator, we construct the **averaging operator**:

$$\mathcal{A}_k := \frac{1}{k} \sum_{j=0}^{k-1} S^j. \tag{2.10}$$

A source α is **stationary** if it is invariant under the shift: for all $B \in \mathcal{F}$,

$$\alpha[B] \;=\; \alpha[S^{-1}B]\,. \tag{2.11}$$

A stationary source α satisfies the Ergodic Theorem:

Let α be a stationary probability measure; for $f \in L^1(\alpha)$, the limit

$$\overline{f}(\omega) \;:=\; \lim_n (\mathcal{A}_n f)(\omega) \tag{2.12}$$

exists α–almost surely and the function $\omega \mapsto \overline{f}(\omega)$ is shift-invariant and satisfies

$$\int_\Omega \overline{f}\, d\alpha \;=\; \int_\Omega f\, d\alpha\,. \tag{2.13}$$

A source α is **ergodic** if it is stationary and it assigns probability zero or one to each invariant subset: for all $B \in \mathcal{F}$ such that $S^{-1}B = B$, either $\alpha[B] = 0$ or $\alpha[B] = 1$. We have the following Corollary to the Ergodic Theorem:

If α is ergodic, then the limit (2.12) is constant for α–every ω and hence

$$\overline{f}(\omega) \;=\; \int_\Omega f\, d\alpha\,, \qquad \alpha\text{–a.e.}\,. \tag{2.14}$$

Recall that $h_{Sh}(\alpha)$, the **Shannon entropy** of a stationary source α, is non-negative and given by

$$h_{Sh}(\alpha) \;=\; -\lim_n \frac{1}{n} \sum_{\mathsf{a} \in \Omega_n} \alpha_n[\mathsf{a}] \log \alpha_n[\mathsf{a}]\,, \tag{2.15}$$

where the logarithm is taken in some fixed base $b > 1$.

Definition 2.1 *Let α be a stationary source. A sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is said to be a* **supporting sequence** *for α if and only if the condition*

$$\lim_{n\to\infty} \alpha_n[\Gamma_n] = 1 \tag{2.16}$$

holds.

Example A simple example of a supporting sequence is $\{\Gamma_n = \Omega_n : n \in \mathbb{N}\}$. The elements of this supporting sequence are too big: in the context of data compression, the goal is to choose the sets Γ_n to be as small as possible, consistent with condition (2.16) holding.

A lower bound on the exponential growth-rate of a supporting sequence is provided by the following result, a consequence of elementary properties of the specific information gain; it will be proved in Section 3, Proposition 3.1.

Lemma 2.1 *Let α be a stationary source. If $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is a supporting sequence for α, then*

$$\liminf_n \frac{1}{n} \log \#\Gamma_n \geq h_{Sh}(\alpha). \tag{2.17}$$

This result motivates the following definition.

Definition 2.2 *Let α be a stationary source. A sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is said to have* **entropic growth-rate** *for α if and only if*

$$\lim_n \frac{1}{n} \log \#\Gamma_n = h_{Sh}(\alpha). \tag{2.18}$$

As a first step in the definition of a reconstruction sequence, we define a class of probability measures, the equipartion measures. In Section 1, we defined them 'downstairs' on Ω_n: the equipartition measure $\beta_n^{\Gamma_n}$ determined by the subset Γ_n of Ω_n is the probability measure on Ω_n which gives equal weight to the words in Γ_n and zero weight to words outside it. Here we define them 'upstairs' on Ω with the aid of the reference measure β.

Definition 2.3 *Let Γ_n be a subset of Ω_n with $\beta_n[\Gamma_n] > 0$; we call the probability measure given on $\mathcal{F}$ by*

$$\beta^{\Gamma_n}[\,\cdot\,] := \beta[\,\cdot \mid X_n^{-1}\Gamma_n] \tag{2.19}$$

the **equipartition measure** *determined by Γ_n.*

Notice that, for each subset Δ_n of Ω_n, we have

$$\beta_n^{\Gamma_n}[\Delta_n] = \frac{\#(\Delta_n \cap \Gamma_n)}{\#\Gamma_n}. \tag{2.20}$$

Although the original measure β is stationary, the equipartition measure β^{Γ_n} is *not* stationary unless $\Gamma_n = \Omega_n$. Since we wish to use equipartition measures to approximate a stationary measure α, we have to do something about this. The most elegant solution is to define a reconstruction sequence with the aid of the averaging operator (2.10): a sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is a reconstruction sequence for α if $\lim_n \mathcal{A}_n \beta^{\Gamma_n} = \alpha$. While readers familiar with ergodic theory may find this definition natural, others may find it puzzling. For this reason, we prefer to adopt a definition in which the averaging is performed 'downstairs' on Ω_n rather than 'upstairs' on Ω; the connection between the two definitions is discussed in Section 6. We use σ_n, the **cyclic permutation operator** acting on Ω_n :

$$\mathsf{a} = (\mathsf{a}_1, \mathsf{a}_2, \ldots, \mathsf{a}_n) \mapsto \sigma_n \mathsf{a} := (\mathsf{a}_2, \ldots, \mathsf{a}_n, \mathsf{a}_1). \tag{2.21}$$

Definition 2.4 *Let α be a stationary source. A sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is said to be a* **reconstruction sequence** *for α if and only if*

1. *for all n sufficiently large, $\beta_n[\Gamma_n] > 0$;*
2. *each Γ_n is invariant under the cyclic permutation σ_n;*
3. *the corresponding sequence $\{\beta^{\Gamma_n}\}$ of equipartition measures converges to α:*

$$\lim_n \beta_m^{\Gamma_n} = \alpha_m \tag{2.22}$$

for each $m \in \mathbb{N}$.

For reconstruction sequences, we have the following upper bound on the exponential growth-rate; it will be proved in Section 3, Proposition 3.2, using the lower semicontinuity of the specific information gain.

Lemma 2.2 *Let α be a stationary source. If $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is a reconstruction sequence for α, then*

$$\limsup_n \frac{1}{n} \log \#\Gamma_n \leq h_{Sh}(\alpha)\,. \tag{2.23}$$

We have the following obvious corollary:

Corollary 2.1 *Let α be a stationary source. If a supporting sequence for α is also a reconstruction sequence for α, then it has entropic growth-rate.*

Examples:

Take $\Omega = \{1,0\}^{\mathbb{N}}$ with β the Bernoulli $(\frac{1}{2},\frac{1}{2})$ probability measure. Let α be the Bernoulli $(\frac{1}{3},\frac{2}{3})$ measure. Define

$$\Gamma_n := \{\mathbf{a} \in \Omega_n : |\frac{1}{n}\sum_{j=1}^{n} \mathbf{a}_j - \frac{1}{3}| \leq \delta_n\}. \tag{2.24}$$

- If $\delta_n < \dfrac{1}{3n}$, then $\Gamma_n = \emptyset$ unless n is divisible by 3; $\lim_n \frac{1}{n} \log \#\Gamma_n$ does not exist.
- If $\delta_n = \dfrac{1}{3n}$, then for each $n \in \mathbb{N}$ there is exactly one k so that $\mathbf{a} \in \Gamma_n$ implies $\sum_1^n \mathbf{a}_j = k$, and $\#\Gamma_n = \binom{n}{k}$. A direct calculation shows that $\{\Gamma_n\}$ is a reconstruction sequence for α. A simple computation using Stirling's formula shows that $\{\Gamma_n\}$ has entropic growth-rate. It is *not* supporting.
- If $\delta_n = \log n/\sqrt{n}$, then the Central Limit Theorem shows that $\{\Gamma_n\}$ is supporting for α. Direct arguments show that it is also a reconstruction sequence, hence has entropic growth-rate.
- If $\delta_n = \delta$, where $0 < \delta \leq \frac{1}{6}$ is a constant, then the sequence is supporting, but not a reconstruction sequence, for α. It is a reconstruction sequence with entropic growth-rate for the Bernoulli p–measure with $p = \frac{1}{3} + \delta$. However, it is *not* supporting for this p–measure.

One may also ask if a sequence can be both supporting and have entropic growth-rate for two distinct measures. To see that this is the case, let $\overline{\alpha}$ be the Bernoulli $(\frac{2}{3},\frac{1}{3})$ measure and let $\alpha^* := \frac{1}{2}\alpha + \frac{1}{2}\overline{\alpha}$; define

$$\overline{\Gamma}_n := \{\mathsf{a} \in \Omega_n : |\frac{1}{n}\sum_{j=1}^{n} \mathsf{a}_j - \frac{2}{3}| \leq \delta_n\} \tag{2.25}$$

and let $\Gamma_n^* := \Gamma_n \cup \overline{\Gamma}_n$. With $\delta_n = \log n/\sqrt{n}$, the sequence $\{\Gamma_n^*\}$ is supporting and of entropic growth-rate for α, $\overline{\alpha}$ and α^*. It is a reconstruction sequence for the *non–ergodic* α^*.

Sets forming a reconstruction sequence may grow very slowly; for examples with zero exponential growth-rate, see Section 6.

Our existence proofs make use of a construction which generalises that used for a Bernoulli measure in the above examples. Define the **blocking operator** $\mathcal{P}_n$:

$$\mathcal{P}_n(\omega) = (\omega_1,\ldots,\omega_n,\omega_1,\ldots,\omega_n,\ldots). \tag{2.26}$$

which is $\mathcal{F}_n$–measurable since $\mathcal{P}_n(\omega)$ depends only on $\omega_1,\ldots,\omega_n$. Next we define the **cyclic empirical measure**

$$T_n(\omega) := \mathcal{A}_n\delta_{\mathcal{P}_n(\omega)} \tag{2.27}$$

in the space $\mathcal{M}_1^+(\Omega)$ of probability measures on $(\Omega,\mathcal{F})$. Since the map $T_n : \Omega \to \mathcal{M}_1^+$ is $\mathcal{F}_n$-measurable, the inverse image $T_n^{-1}A$ of a subset A of $\mathcal{M}_1^+$ is determined completly by the first n coordinates.

Definition 2.5 *Let $T_n : \Omega \to \mathcal{M}_1^+(\Omega)$ be the cyclic empirical measure. A sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is said to be a* **canonical sequence** *for α if and only if*

1. *there exists a decreasing sequence $\{F_n\}$ of closed neighbourhoods of α whose intersection is $\{\alpha\}$;*
2. *each set Γ_n is given by*
$$\Gamma_n = X_nT_n^{-1}F_n\,; \tag{2.28}$$
3. *for all n sufficiently large, $\beta_n[\Gamma_n] > 0$.*

In this case, we shall say that the canonical sequence $\{\Gamma_n\}$ is **based on** *the sequence $\{F_n\}$.*

The key to the proof of our main theorem is a conditional limit theorem; this is the subject of Section 4. It says that if $\{\Gamma_n\}$ is a canonical sequence for the stationary measure α, then the sequence $\{\beta^{\Gamma_n}\}$ of conditioned measures converges to α:

Lemma 2.3 *Let α be a stationary source; every canonical sequence for α is a reconstruction sequence for α.*

This result is an easy consequence of the cyclical invariance of the sets Γ_n and the compactness of the space $\mathcal{M}_1^+(\Omega)$.

A great advantage which comes from working 'upstairs' on Ω is that we have available results on the large deviation properties of the cyclic empirical measure. The results we need are summarized in Section 5; proofs can be found in [LPS]. We use the large-deviation lower bound to prove the existence of a canonical sequence for a stationary measure.

Lemma 2.4 *Let α be a stationary source; then there exists a canonical sequence for α having entropic growth–rate.*

Since the alphabet is assumed to be finite, the existence of a decreasing sequence of closed neighbourhoods contracting to α is easily established; the large-deviation lower bound is used to control the rate at which the sequence contracts to α so as to ensure that, at least for all n sufficiently large, the sets $\Gamma_n = X_n T_n^{-1} F_n$ satisfy $\beta_n[\Gamma_n] > 0$. We do so by exhibiting a sequence $\{\Gamma_n\}$ whose growth–rate is bounded *below* by the Shannon entropy of α; it then follows from Lemma2.3 and Lemma 2.2 that the growth–rate is entropic.

A canonical sequence for α is not necessarily supporting; however, when the source is ergodic, we can use the Ergodic Theorem in place of the large-deviation lower bound to control the sequence of contracting neighbourhoods. In this way, we can ensure that the sequence we construct is supporting and, by Lemma 2.1, canonical. This is done in Section 5, Proposition 5.2, establishing the following result:

Lemma 2.5 *Let α be an ergodic source; then there exists a canonical sequence for α which is a supporting sequence for α.*

In Section 5, we use the compactness of $\mathcal{M}_1^+$ to prove the converse of Lemma 2.5:

Lemma 2.6 *If there exists a sequence which is both canonical and supporting for a stationary source α, then α must be ergodic.*

We are ready to prove Theorem 2.1:

Lemmas 2.4 and 2.3 together prove that if α is a stationary source, then there exists a reconstruction sequence for α having entropic growth–rate; Lemma 2.5 proves that if the source α is ergodic, then the reconstruction sequence may be chosen so as to be supporting for α.

Lemma 2.6 is the converse of Lemma 2.5; together they prove Theorem 2.2.

3 Information Gain

The principal tool in the proofs of Lemma 2.1, the lower bound on the growth-rate of a supporting sequence, and Lemma 2.2, the upper bound on the growth-rate of a reconstruction sequence, is the specific information gain.

Definition 3.1 *The* **information gain** *of the probability measure λ with respect to the probability measure ρ is given by*

$$D(\lambda \,||\, \rho) := \int d\lambda \log \frac{d\lambda}{d\rho} \tag{3.1}$$

when λ is absolutely continuous with respect to ρ; otherwise $D(\lambda \,||\, \rho) := +\infty$.

Definition 3.2 *The* **specific information gain** *of the probability measure λ with respect to ρ is given by*

$$h(\lambda \,|\, \rho) := \limsup_n \frac{1}{n} D(\lambda_{|\mathcal{F}_n} \,||\, \rho_{|\mathcal{F}_n}), \tag{3.2}$$

where $\lambda_{|\mathcal{F}_n}$ is the restriction of λ to $\mathcal{F}_n$. (Note: $D(\lambda_{|\mathcal{F}_n} \,||\, \rho_{|\mathcal{F}_n}) = D(\lambda_n \,||\, \rho_n)$).

We always have $D(\lambda \,||\, \rho) \geq 0$, $h(\lambda \,|\, \beta) \geq 0$; if $D(\lambda \,||\, \rho) = 0$, then $\lambda = \rho$, but the corresponding result for $h(\lambda \,|\, \rho)$ does not always hold. However, if α is stationary and ρ is a stationary product measure, then

$$h(\alpha \,|\, \rho) = \lim_n \frac{1}{n} D(\alpha|_{\mathcal{F}_n} \,||\, \rho|_{\mathcal{F}_n}). \tag{3.3}$$

When $\mathtt{A}$ is a finite alphabet with r letters, and β is the uniform product measure, this yields

$$h(\alpha \,|\, \beta) = \log r - h_{Sh}(\alpha). \tag{3.4}$$

Likewise, when $\mathtt{A}$ is a finite alphabet and $\Gamma_n \subset \Omega_n$, we have

$$\log \beta_n[\Gamma_n] = \log \#\Gamma_n - n \log r. \tag{3.5}$$

Henceforth we will replace $\#\Gamma_n$ by $\beta_n[\Gamma_n]$ and $h_{Sh}(\alpha)$ by $-h(\alpha \mid \beta)$ in the statements of the propositions. Modified in this way they hold in greater generality; this is discussed in Section 6.

Proposition 3.1 *Let α be a stationary source. If $\{\Gamma_n\}$ is a supporting sequence for α, then*

$$\liminf_n \frac{1}{n} \log \beta_n[\Gamma_n] \geq -h(\alpha|\beta). \tag{3.6}$$

Proof: Since $\tilde{\Gamma}_n := X_n^{-1}(\Gamma_n)$ is in $\mathcal{F}_n$, we have

$$D(\alpha_{|\mathcal{F}_n} \,||\, \beta_{|\mathcal{F}_n}) \geq \alpha[\tilde{\Gamma}_n] \log \frac{\alpha[\tilde{\Gamma}_n]}{\beta[\tilde{\Gamma}_n]} + \alpha[\Omega \setminus \tilde{\Gamma}_n]) \log \frac{\alpha[\Omega \setminus \tilde{\Gamma}_n]}{\beta[\Omega \setminus \tilde{\Gamma}_n]} \tag{3.7}$$

$$\geq \alpha[\tilde{\Gamma}_n] \log \alpha[\tilde{\Gamma}_n] + \alpha[\Omega \setminus \tilde{\Gamma}_n] \log \alpha[\Omega \setminus \tilde{\Gamma}_n] - \alpha[\tilde{\Gamma}_n] \log \beta[\tilde{\Gamma}_n]. \qquad (3.8)$$

The inequality follows by dividing by n and taking $\limsup_n$, using $\alpha[\tilde{\Gamma}_n] \to 1$. □

Lemma 2.1 follows using (3.4), (3.5) and (3.6).

There are some results of a more technical character which we require concerning cyclic symmetrization and the specific information gain. We collect them in a lemma; they are proved in Section 8 of [LPS]. The space Ω has a natural decomposition into a product space

$$\Omega = \Omega_n \times \Omega_n^c, \qquad (3.9)$$

and the measure β^{Γ_n} is the product measure of $\beta_n^{\Gamma_n}$ on Ω_n and β on Ω_n^c. Notice that, for each subset Δ_n of Ω_n, we have

$$\beta_n^{\Gamma_n}[\Delta_n] = \frac{\#(\Delta_n \cap \Gamma_n)}{\#\Gamma_n}. \qquad (3.10)$$

Lemma 3.1 *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. Let $\Omega = \Omega_1 \times \Omega_2$ with $\mathcal{F}$ the corresponding product σ-algebra. Let λ and β be probability measures on $(\Omega, \mathcal{F})$ with λ_1, λ_2 and β_1, β_2 denoting the restrictions to $\mathcal{F}_1, \mathcal{F}_2$ considered as sub-σ-algebras of $\mathcal{F}$. Assume $\beta = \beta_1 \otimes \beta_2$. Then we have*

$$D(\lambda||\beta) = D(\lambda||\lambda_1 \otimes \lambda_2) + D(\lambda_1||\beta_1) + D(\lambda_2||\beta_2). \qquad (3.11)$$

We are now in a position to prove the upper bound on the growth-rate of a reconstruction sequence.

Proposition 3.2 *Let α be a stationary source. If $\{\Gamma_n\}$ is a reconstruction sequence for α, then*

$$\limsup_n \frac{1}{n} \log \beta_n[\Gamma_n] \leq -h(\alpha|\beta). \qquad (3.12)$$

Proof: By direct calculation, we have

$$\begin{aligned} D(\beta^{\Gamma_n}{}_{|\mathcal{F}_n} \,||\, \beta_{|\mathcal{F}_n}) &= \int_{\Gamma_n} \log \frac{d}{d\beta}\beta^{\Gamma_n} \, d\beta^{\Gamma_n} \qquad (3.13) \\ &= -\log \beta[\Gamma_n], \end{aligned}$$

so that

$$\liminf_n \frac{1}{n} D(\beta^{\Gamma_n}{}_{|\mathcal{F}_n} \,||\, \beta_{|\mathcal{F}_n}) = -\limsup_n \frac{1}{n} \log \beta[\Gamma_n]. \qquad (3.14)$$

Next we make use of the cyclical invariance of Γ_n: for any integer k such that $k+m \leq n$, the projections of β^{Γ_n} on the σ-algebras $\sigma(x_1, \ldots, x_m)$ and $\sigma(x_{k+1}, \ldots, x_{k+m})$ are the same. Let $m < n$ and $q(n|m)$ be the largest integer smaller than n/m. From Lemma 3.1 we have

$$\frac{1}{n} D(\beta^{\Gamma_n}{}_{|\mathcal{F}_n} \,||\, \beta_{|\mathcal{F}_n}) \geq \frac{q(n|m)}{n} D(\beta^{\Gamma_n}{}_{|\mathcal{F}_m} \,||\, \beta_{|\mathcal{F}_m}). \qquad (3.15)$$

Since $\lim_n \beta^{\Gamma_n} = \alpha$, it follows from the lower semicontinuity of $D(\,\cdot\,||\,\beta_{|\mathcal{F}_m}\,)$ on the space of measures on Ω_m that

$$-\limsup_n \frac{1}{n}\log\beta_n[\Gamma_n] \;\geq\; \liminf_n \frac{1}{n} D(\,\beta^{\Gamma_n}{}_{|\mathcal{F}_n}\,||\,\beta_{|\mathcal{F}_n}\,) \qquad (3.16)$$

$$\geq\; \frac{1}{m} D(\,\alpha_{|\mathcal{F}_m}\,||\,\beta_{|\mathcal{F}_m}\,)\,.$$

Hence

$$\limsup_n \frac{1}{n}\log\beta_n[\Gamma_n] \leq -h(\alpha|\beta)\,. \qquad (3.17)$$

□

Lemma 2.2 follows using (3.4), (3.5) and (3.12).

4 A Conditional Limit Theorem

We state and prove a conditional limit theorem. We shall need the following lemma which exploits the invariance of the reference measure β under the the **cyclic shift operator** $\overset{\circ}{S}_n$, defined by:

$$(\overset{\circ}{S}_n\,\omega)_k := \begin{cases} \omega_{k+1} & \text{if } k \bmod n \neq 0; \\ \omega_{k-n+1} & \text{if } k \bmod n = 0. \end{cases} \tag{4.1}$$

We sometimes find the following notation useful: let $\lambda \in \mathcal{M}_1^+$ and $f \in \mathcal{F}_{\text{loc}}$; we set

$$\langle f, \lambda \rangle := \int f\, d\lambda. \tag{4.2}$$

Lemma 4.1 *Let $\widetilde{\Gamma}_n$ be $T_n^{-1}\mathcal{B}$-measurable with $\beta_n[\Gamma_n] > 0$ and let $f \in \mathcal{F}_k$ with $k \leq n$; then*

$$\int_\Omega f(\omega)\,\beta^{\Gamma_n}[d\omega] \;=\; \int_\Omega \langle f, T_n(\omega)\rangle\,\beta^{\Gamma_n}[d\omega]\,. \tag{4.3}$$

Proof: Since $\widetilde{\Gamma}_n$ is $T_n^{-1}\mathcal{B}$-measurable, there exists $C \in \mathcal{B}$ such that $\widetilde{\Gamma}_n = \{\omega : T_n(\omega) \in C\}$. Note that $\overset{\circ}{S}_n$ is bijective, $S\mathcal{P}_n = \overset{\circ}{S}_n \mathcal{P}_n = \mathcal{P}_n \overset{\circ}{S}_n$, $\overset{\circ}{S}{}_n^{n+j} = \overset{\circ}{S}{}_n^{j}$, and

$$T_n(\omega) = \overset{\circ}{\mathcal{A}}_n\,\delta_{\mathcal{P}_n(\omega)} \quad \text{with} \quad \overset{\circ}{\mathcal{A}}_n := \frac{1}{n}\sum_{j=0}^{n-1} \overset{\circ}{S}{}_n^{j}. \tag{4.4}$$

Also $\overset{\circ}{S}_n\, T_n(\omega) = T_n(\omega)$, $\overset{\circ}{S}_n\,\beta = \beta$, which imply

$$\overset{\circ}{S}{}_n^{-1}\{T_n \in C\} = \{T_n \in C\} \quad \text{and} \quad \overset{\circ}{S}_n\,\beta[\,\cdot\,|\{T_n \in C\}] = \beta[\,\cdot\,|\{T_n \in C\}]. \tag{4.5}$$

Since $f \in \mathcal{F}_k$ with $k \leq n$, we have $f = f \circ \mathcal{P}_n$ so

$$\begin{aligned} \int_\Omega f(\omega)\,\beta[d\omega|\{T_n \in C\}] &= \int_\Omega f \circ \mathcal{P}_n(\omega)\;\overset{\circ}{\mathcal{A}}_n\,\beta[d\omega|\{T_n \in C\}] \\ &= \int_\Omega \langle f, T_n(\omega)\rangle\,\beta[d\omega|\{T_n \in C\}]. \end{aligned} \tag{4.6}$$

□

A second ingredient in the proof of our conditional limit theorem is a lemma which states a simple consequence of the compactness of $\mathcal{M}_1^+$; we shall need it again in Section 5.

Lemma 4.2 *Let α be a stationary source and let $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ be a canonical sequence for α based on the sequence $\{F_n\}$. Then for each $f \in \mathcal{F}_{\text{loc}}$ and each $\varepsilon > 0$ there exists $N(f,\varepsilon)$ such that whenever $n \geq N(f,\varepsilon)$*

$$F_n \subset \{\lambda \in \mathcal{M}_1^+ : |\langle f,\lambda\rangle - \langle f,\alpha\rangle| < \varepsilon\}. \tag{4.7}$$

Proof: $\{F_n\}$ is a decreasing sequence of closed neighbourhoods of α whose intersection is $\{\alpha\}$. We then have

$$\bigcap_n (F_n \setminus \{\lambda \in \mathcal{M}_1^+ : |\langle f,\lambda\rangle - \langle f,\alpha\rangle| < \varepsilon\}) = \emptyset. \tag{4.8}$$

We deduce there exists N so that $F_N \setminus \{\lambda \in \mathcal{M}_1^+ : |\langle f,\lambda\rangle - \langle f,\alpha\rangle| < \varepsilon\}) = \emptyset$ because $\mathcal{M}_1^+$ is compact, so we have (4.7). □

Theorem 4.1 *Let α be a stationary source; every canonical sequence for α is a reconstruction sequence for α.*

Proof: Let $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ be a canonical sequence for α based on the sequence $\{F_n\}$. Then, by Lemma 4.2, for each $f \in \mathcal{F}_{\text{loc}}$ and each $\varepsilon > 0$ there exists $N(f,\varepsilon)$ such that $f \in \mathcal{F}_{N(f,\varepsilon)}$ and whenever $n \geq N(f,\varepsilon)$ we have

$$F_n \subset \{\lambda \in \mathcal{M}_1^+ : |\langle f,\lambda\rangle - \langle f,\alpha\rangle| < \varepsilon\}. \tag{4.9}$$

It follows that $n \geq N(f,\varepsilon)$ and $\omega \in \tilde{\Gamma}_n$ imply

$$|\langle f, T_n(\omega)\rangle - \langle f,\alpha\rangle| < \varepsilon. \tag{4.10}$$

Since β^{Γ_n} is supported by $\tilde{\Gamma}_n$, we have

$$\int_\Omega f(\omega)\,\beta^{\Gamma_n}[d\omega] = \int_{\tilde{\Gamma}_n} \langle f, T_n(\omega)\rangle\,\beta^{\Gamma_n}[d\omega]\,. \tag{4.11}$$

It follows from the above and (4.3) that

$$|\int_\Omega f(\omega)\,\beta^{\Gamma_n}[d\omega] - \langle f,\alpha\rangle| < \varepsilon \tag{4.12}$$

whenever $n \geq N(f,\varepsilon)$. This proves that the sequence $\{\beta^{\Gamma_n}\}$ converges to α. □

5 Canonical Sequences

We begin by summarizing the ideas of large deviation theory and the single result we shall require; proofs can be found in [LPS].

Denote by $\mathbb{M}_n$ the distribution of the cyclic empirical measure $T_n : \Omega \to \mathcal{M}_1^+(\Omega)$ defined on the probability space $(\Omega, \mathcal{F}, \beta)$, where β is our reference measure — the uniform product measure:

$$\mathbb{M}_n := \beta \circ T_n^{-1} . \tag{5.1}$$

For each open set G, define

$$\overline{m}[G] := \limsup_n \frac{1}{n} \log \mathbb{M}_n[G] , \tag{5.2}$$

$$\underline{m}[G] := \liminf_n \frac{1}{n} \log \mathbb{M}_n[G] ; \tag{5.3}$$

the following result is Lemma 8.3 of [LPS]: for each $\lambda \in \mathcal{M}_1^+$, we have

$$\inf\{\underline{m}[G] : G \ni \lambda\} \;=\; \inf\{\overline{m}[G] : G \ni \lambda\} . \tag{5.4}$$

Definition 5.1 ([LP]) *The* **Ruelle–Lanford function** μ *is defined on* $\mathcal{M}_1^+$ *by*

$$\begin{aligned} \mu(\lambda) &:= \inf\{\overline{m}[G] : G \ni \lambda\} && (5.5)\\ &= \inf\{\underline{m}[G] : G \ni \lambda\} . && (5.6) \end{aligned}$$

It is a basic result in Large–Deviation Theory (see [LP]) that the existence of the Ruelle–Lanford function implies that the large–deviation lower bound holds for open subsets:

- for each open set G, we have

$$\sup_{\lambda \in G} \mu(\lambda) \;\leq\; \liminf_n \frac{1}{n} \log \mathbb{M}_n[G]; \tag{5.7}$$

In the present case, we have:

- the Ruelle–Lanford function (RL–function) is given explicitly by

$$\mu(\lambda) = \begin{cases} -h(\lambda \mid \beta), & \text{if } \lambda \text{ is stationary,} \\ -\infty, & \text{otherwise.} \end{cases} \tag{5.8}$$

The following result is fundamental: it establishes the existence of the sequences of closed neighbourhoods of α on which our construction of canonical sequences is based.

Lemma 5.1 *Let α be a stationary source; then there exists a decreasing sequence $\{F_n\}$ of closed neighbourhoods of α in $\mathcal{M}_1^+$ such that*

$$\bigcap_n F_n = \{\alpha\} . \tag{5.9}$$

Proof: The statement is a consequence of our hypothesis that the factor spaces of Ω are finite sets, copies of a finite alphabet; it holds whenever the factor spaces are standard Borel spaces, for then there exists a sequence $\{g_m\}$ in $\mathcal{F}_{\text{loc}}$ which separates $\mathcal{M}_1^+$: if α, λ are any two probability measures which satisfy

$$\int g_m \, d\alpha = \int g_m \, d\lambda, \tag{5.10}$$

for all m, then $\lambda = \alpha$. In the finite alphabet case, we can take for $\{g_m\}$ the set $\{1_{\mathsf{a}} \in \mathcal{F}_n : \mathsf{a} \in \Omega_n, n \in \mathbb{N}\}$ of all indicator functions of atoms determined by finite words:

$$1_{\mathsf{a}}(\omega) = \begin{cases} 1 & \text{if } \omega_1 = \mathsf{a}_1, \ldots, \omega_n = \mathsf{a}_n \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

We can choose

$$F_n := \{\lambda \in \mathcal{M}_1^+ : |\int g_k \, d\lambda - \int g_k \, d\alpha| \leq \frac{1}{n}, k = 1, \ldots, n\}. \tag{5.12}$$

□

We now use the large deviation lower bound (5.7) to prove the existence of a canonical sequence for which a *lower* bound on the exponential growth-rate holds. The proof employs a construction which we call stretching: Let $\{F_k\}$ be a decreasing sequence in $\mathcal{M}_1^+$ of closed neighbourhoods of α whose intersection is $\{\alpha\}$ and let $\{N_m\}$ be a strictly increasing sequence of positive integers; the decreasing sequence $\{F'_n\}$ defined, for each $n \in \mathbb{N}$, by

$$F'_n := \begin{cases} \mathcal{M}_1^+ & \text{if } n < N_1 \,; \\ F_m & \text{if } N_m \leq n < N_{m+1}, \end{cases} \tag{5.13}$$

is called the **stretching** of $\{F_n\}$ by $\{N_m\}$; note that the intersection of the stretched sequence is again $\{\alpha\}$.

Proposition 5.1 *Let α be a stationary source; then there exists a canonical sequence $\{\Gamma_n\}$ for which*

$$\lim_n \frac{1}{n} \log \beta_n[\Gamma_n] = -h(\alpha \mid \beta)\,. \tag{5.14}$$

Proof: Let $\{F_k\}$ be a decreasing sequence in $\mathcal{M}_1^+$ of closed neighbourhoods of α whose intersection is $\{\alpha\}$; consider F_m for fixed m. Since F_m is a neighbourhood of α, there exists an open set G such that $\alpha \in G \subset F_m$. It follows from the large deviation lower bound (5.7) that

$$\mu(\alpha) \leq \sup_{\lambda \in G} \mu(\lambda) \leq \liminf_n \frac{1}{n} \log \mathbb{M}_n[G] \leq \liminf_n \frac{1}{n} \log \mathbb{M}_n[F_m]\,; \tag{5.15}$$

hence, for each $m \in \mathbb{N}$, there is an integer N_m such that

$$\mu(\alpha) - \frac{1}{m} \leq \frac{1}{p} \log \mathbb{M}_p[F_m] \tag{5.16}$$

for all $p \geq N_m$. We may choose the sequence $\{N_m\}$ to be strictly increasing. Let $\{F'_n\}$ be the stretching of $\{F_n\}$ by $\{N_m\}$; define Γ_n by

$$\Gamma_n = X_n T_n^{-1} F'_n\,. \tag{5.17}$$

By construction, for all n such that $N_m \leq n < N_{m+1}$, we have

$$\frac{1}{n}\log \beta_n[\Gamma_n] = \frac{1}{n}\log \mathbb{M}_n[F'_n] \geq \mu(\alpha) - \frac{1}{m} \; ; \tag{5.18}$$

it follows that

$$\liminf_n \frac{1}{n}\log \beta_n[\Gamma_n] \geq \mu(\alpha) \, . \tag{5.19}$$

But α is stationary, so that

$$\mu(\alpha) \;=\; -h(\alpha \mid \beta) \; ; \tag{5.20}$$

hence we have

$$\liminf_n \frac{1}{n}\log \beta_n[\Gamma_n] \;\geq\; -h(\alpha \mid \beta) \, . \tag{5.21}$$

In particular, $\beta_n[\Gamma_n] > 0$ for all n sufficiently large; it follows that $\{\Gamma_n\}$ is a canonical sequence and that the lower bound holds. The equality (5.14) for the growth–rate now follows from Theorem 4.1 and Proposition 3.2. □

When the source is ergodic, we use the stretching construction together with the Ergodic Theorem to prove the existence of a canonical sequence $\{\Gamma_n\}$ which is supporting. First we need two lemmas:

Lemma 5.2 *Let α be an ergodic source and let $f \in \mathcal{F}_{\text{loc}}$; then*

$$\lim_{n\to\infty} \langle f, T_n(\omega)\rangle \;=\; \langle f, \alpha\rangle \; , \qquad \alpha\text{–a.e.} \, . \tag{5.22}$$

Proof: For $f \in \mathcal{F}_m$, an elementary calculation shows that

$$\sup_{\omega\in\Omega} |\langle f, T_n(\omega)\rangle \;-\; (\mathcal{A}_n f)(\omega)| \leq \frac{2(m-1)}{n} \sup_{\omega\in\Omega} |f(\omega)| \, , \tag{5.23}$$

so that the sequence $\{\langle f, T_n(\omega)\rangle\}$ converges whenever the sequence $\{(\mathcal{A}_n f)(\omega)\}$ converges and they have the same limit; since α is ergodic, it follows from the Ergodic Theorem that their common limit is $\langle f, \alpha\rangle$. □

Lemma 5.3 *Let α be an ergodic source and let G be an open subset of $\mathcal{M}_1^+$ containing α; then*

$$\lim_{n\to\infty} \alpha[T_n^{-1}G] = 1. \tag{5.24}$$

Proof: Since G is an open set containing α, there exist $f_1, \ldots, f_m$ and positive numbers $\varepsilon_1, \ldots, \varepsilon_m$ such that

$$\{\lambda \in \mathcal{M}_1^+ : |\langle f_k, \lambda\rangle \;-\; \langle f_k, \alpha\rangle| \;<\; \varepsilon_k \, , \, k = 1, \ldots, m\} \subset G. \tag{5.25}$$

Since α is ergodic and $f_k \in \mathcal{F}_{\text{loc}}$, it follows from Lemma 5.2 that

$$\lim_{n\to\infty} \langle f_k, T_n(\omega)\rangle \;=\; \langle f_k, \alpha\rangle \;\; \alpha\text{–a.e.} \, , \tag{5.26}$$

so

$$\lim_{n\to\infty} \alpha[\{\omega \in \Omega : |\langle f_k, T_n(\omega)\rangle \;-\; \langle f_k, \alpha\rangle| \;<\; \varepsilon_k \, , \, k = 1, \ldots, m\}] \;=\; 1 \, . \tag{5.27}$$

It now follows from (5.25) that

$$\lim_{n\to\infty} \alpha[T_n^{-1}G] = 1. \tag{5.28}$$

□

Proposition 5.2 *Let α be an ergodic source; then there exists a canonical sequence for α which is supporting for α.*

Proof: Let $\{F_k\}$ be a decreasing sequence in $\mathcal{M}_1^+$ of closed neighbourhoods of α whose intersection is α; consider F_m for fixed m. Since F_m is a neighbourhood of α, there exists an open set G such that $\alpha \in G \subset F_m$. Suppose that the source α is ergodic; then, by Lemma 5.3, we have

$$\lim_n \alpha[T_n^{-1}G] = 1 \tag{5.29}$$

so there exists N_m such that, for all $n \geq N_m$,

$$\alpha[T_n^{-1}F_m] \;\geq\; \alpha[T_n^{-1}G] \;\geq\; 1 - 1/m\,. \tag{5.30}$$

The sequence $\{N_m\}$ may be chosen to be strictly increasing. Let $\{F_n'\}$ be the stretching of $\{\mathcal{F}_n\}$ by $\{N_m\}$; put

$$\Gamma_n \;=\; X_n T_n^{-1} F_n'\,. \tag{5.31}$$

Then

$$\alpha_n[\Gamma_n] \;=\; \alpha[T_n^{-1}F_n']\,, \tag{5.32}$$

so that

$$\lim_n \alpha_n[\Gamma_n] \;=\; \lim_n \alpha[T_n^{-1}F_n'] \;=\; 1\,. \tag{5.33}$$

Hence $\{\Gamma_n\}$ is supporting for α; in particular, by Proposition 3.1, $\beta_n[\Gamma_n] > 0$ for all n sufficiently large so that $\{\Gamma_n\}$ is a canonical sequence. □

We conclude this section by proving a theorem of relevance to the coding problem:

Theorem 5.1 *Let α be a stationary source. If there exists a canonical sequence for α which is supporting for α, then the source α is ergodic.*

We make use of two lemmas.

Lemma 5.4 *Let α be a stationary source. Let $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ be a canonical sequence for α and suppose there exists a sequence $\{f_k \in \mathcal{F}_{\rm loc} : k \in \mathbb{N}\}$ which separates $\mathcal{M}_1^+$ and*

$$\lim_n \sup_{\omega \in \tilde{\Gamma}_n} |\langle T_n(\omega), f_k\rangle - \langle \alpha, f_k\rangle| \;=\; 0 \tag{5.34}$$

for each $k \in \mathbb{N}$. If $\{\Gamma_n\}$ is supporting for α, then the source α is ergodic.

Proof: For simplicity, we replace each f_k by $f_k - \langle \alpha, f_k\rangle$ so that

$$\langle \lambda, f_k\rangle \;=\; 0 \quad \text{for all } k \in \mathbb{N} \iff \lambda \;=\; \alpha\,. \tag{5.35}$$

Since $f_k \in \mathcal{F}_{\rm loc}$, we deduce from (5.34) that

$$\lim_n \sup_{\omega \in \tilde{\Gamma}_n} |A_n f_k(\omega)| \;=\; 0\,. \tag{5.36}$$

Since f_k is bounded and $\lim_n \alpha[\tilde{\Gamma}_n] \;=\; 1$ by hypothesis, it follows that

$$\lim_n \int_\Omega |A_n f_k(\omega)|\, d\alpha \;=\; 0\,. \tag{5.37}$$

Now let A be any shift-invariant set with $\alpha[A] > 0$. We have

$$|\int_A f_k \, d\alpha| = |\int_A S f_k \, d\alpha| = |\int_A \mathcal{A}_n f_k \, d\alpha| \le \int_\Omega |\mathcal{A}_n f_k| \, d\alpha \qquad (5.38)$$

because both α and A are shift-invariant. Then (5.37) and (5.38) imply that

$$\int_\Omega f_k \, \alpha[\, d\omega \,|A] = 0 \qquad (5.39)$$

for all $k \in \mathbb{N}$. From (5.35), we conclude that

$$\alpha[\,\cdot\,|A] = \alpha[\,\cdot\,] \qquad (5.40)$$

which implies that $\alpha[A] = 1$. Since each shift-invariant set A has either $\alpha[A] = 0$ or $\alpha[A] = 1$, we deduce that α is ergodic. □

We use compactness to prove the existence of a separating sequence having property (5.34).

Lemma 5.5 *Let α be a stationary source and let $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ be a canonical sequence for α based on the sequence $\{F_n\}$; then condition (5.34) holds for each f_k in $\mathcal{F}_{\text{loc}}$.*

Proof: $\{F_n\}$ is a decreasing sequence of closed neighbourhoods of α whose intersection is $\{\alpha\}$. It follows from Lemma 4.2 that for each $f \in \mathcal{F}_{\text{loc}}$ and each $\varepsilon > 0$ there exists $N(f, \varepsilon)$ such that whenever $n \ge N(f, \varepsilon)$

$$F_n \subset \{\lambda \in \mathcal{M}_1^+ : |\langle f, \lambda \rangle - \langle f, \alpha \rangle| < \varepsilon\}. \qquad (5.41)$$

It follows that $n \ge N(f, \varepsilon)$ and $\omega \in \tilde{\Gamma}_n$ imply

$$|\langle f, T_n(\omega) \rangle - \langle f, \alpha \rangle| < \varepsilon. \qquad (5.42)$$

This implies (5.34). □

Since $\mathtt{A}$ is finite, the collection of indicator functions of atoms from Ω_n for all n is a countable separating set. Taken together, Lemma 5.4 and Lemma 5.5 prove Theorem 5.1.

6 Commentary

1. To simplify the exposition, we have assumed A to be a finite set with β_1 the equiprobable distribution on $\Omega_1 = \mathtt{A}$. Our results extend with some modification to the case in which A is a compact metric space and β_1 a probability measure on A with β_n on Ω_n and β on Ω being the product of copies of β_1. Here are the modifications which must be made:

   - $\#\Gamma_n$ must be replaced by $\beta_n[\Gamma_n]$ and $h_{Sh}(\alpha)$ by $-h(\alpha \mid \beta)$ in the statements of the propositions; this is done when we come to prove them in Sections 3, 4 and 5;
   - the hypothesis 'Let α be a stationary source' must be amplified to read 'Let α be a stationary source with $h(\alpha \mid \beta)$ finite'.

   The results may be further extended to the case in which A is a standard Borel space and β_1 a probability measure on A, with β_n and β the corresponding product probability measures. In this case we need to modify the definition of "canonical sequence" so that with $\widetilde{\Gamma}_n = T_n^{-1}F_n$ we require not only that $\{F_n\}$ be a decreasing sequence of closed neighbourhoods of α with $\cap F_n = \{\alpha\}$, but also that there exists a sequence $\{f_k\}$ with each $f_k \in \mathcal{F}_{\mathrm{loc}}$, such that the topology on $\mathcal{M}_1^+$ determined by $\{f_k\}$ separates the points of $\mathcal{M}_1^+$, and that $\{F_n\}$ is a neighbourhood basis for α in this topology. In the non-compact case, in general, Lemma 4.2 is no longer valid. However, the conclusions of Lemma 4.2 hold when $f = f_j \in \{f_k\}$, the separating sequence. One proves Theorem 4.1 by noting that the level sets of h are compact so that the sequence $\{\beta^{\Gamma_n}\}$ has limit points in $\mathcal{M}_1^+$. The modified Lemma 4.2 shows uniqueness of the limit point of $\{\beta^{\Gamma_n}\}$, which implies convergence.

2. The concept of an equipartition measure is inspired by that of a microcanonical measure in statistical mechanics. For Gibbs, the microcanonical measure was fundamental; the canonical measure, an approximation to the microcanonical, was useful by virtue of being more tractable analytically. The idea of bounded local convergence is fore-shadowed in the statement of his 'general theorem':

   > If a system of a great number of degrees of freedom is microcanonically distributed in phase, any very small part of it may be regarded as canonically distributed. ([G],p.183)

   In the concept of a reconstruction sequence, we turn Gibbs' idea on its head: the stationary source corresponds to his canonical measure; our equipartition measure corresponds to his microcanonical measure. For us, the stationary source is fundamental; it can be approximated by an equipartition measure.

3. From the point of view of digital computation, a reconstruction sequence is more tractable than a stationary measure. Reconstruction sequences may prove useful in providing efficient ways of simulating stationary measures. We will pursue these ideas elsewhere.

4. The distinction between average and cyclic average becomes negligible as $n \to \infty$ because the limit employs $\mathcal{F}_{loc}$: for $f \in \mathcal{F}_m$ and $n > m$, an elementary computation shows that

$$\sup_{\omega\in\Omega} |((\mathcal{A}_n f) \circ \mathcal{P}_n)(\omega) - (\mathcal{A}_n f)(\omega)| \leq \frac{2(m-1)}{n} \sup_{\omega\in\Omega} |f(\omega)|, \tag{6.1}$$

so the results of this paper hold with the following alternative definition of reconstruction sequence:

Definition 6.1 *Let α be a stationary source. A sequence $\{\Gamma_n \subset \Omega_n : n \in \mathbb{N}\}$ is said to be a* **reconstruction sequence** *for α if and only if*

(a) for all n sufficiently large, $\beta[\Gamma_n] > 0$;

(b) the corresponding sequence $\{\mathcal{A}_n\beta^{\Gamma_n}\}$ of averaged equipartition measures converges to α:

$$\lim_n \mathcal{A}_n\beta^{\Gamma_n} = \alpha. \tag{6.2}$$

If we use this definition, then canonical sequences have the following important property:

Corollary 6.1 (to Theorem 4.1) *Let $\{\Gamma_n\}$ be a canonical sequence for the stationary source α. Let $\{\Gamma'_n\}$ satisfy $\Gamma'_n \subset \Gamma_n$ and $\beta[\Gamma'_n] > 0$ for all n sufficiently large. Then $\{\Gamma'_n\}$ is a reconstruction sequence in the sense of Definition 6.1:*

$$\lim_n \mathcal{A}_n\beta_n^{\Gamma'_n} = \alpha. \tag{6.3}$$

Proof: Let $f \in \mathcal{F}_m$; from the proof of Theorem 4.1, for each $\varepsilon > 0$, there exists $N(f,\varepsilon)$ so that $n \geq N(f,\varepsilon)$ implies

$$\sup_{\omega\in\tilde{\Gamma}_n} |\langle f, T_n(\omega)\rangle - \langle f,\alpha\rangle| \leq \varepsilon. \tag{6.4}$$

By hypothesis we have for all large n

$$\beta^{\Gamma'_n}[\tilde{\Gamma}_n] = 1. \tag{6.5}$$

It follows from (6.1) that

$$\begin{aligned} |\langle f, \mathcal{A}_n\beta^{\Gamma'_n}\rangle - \langle f,\alpha\rangle| &= |\mathcal{A}_n f, \beta^{\Gamma'_n}\rangle - \langle f,\alpha\rangle| \\ &\leq \varepsilon + \frac{2(m-1)}{n} \sup_{\omega\in\Omega} |f(\omega)|. \end{aligned} \tag{6.6}$$

Hence

$$\limsup_n |\langle f, \mathcal{A}_n\beta^{\Gamma'_n}\rangle - \langle f,\alpha\rangle| \leq \varepsilon. \tag{6.7}$$

Since $f \in \mathcal{F}_{\rm loc}$ and $\varepsilon > 0$ are arbitrary, it follows that $\{\mathcal{A}_n\beta^{\Gamma'_n}\}$ converges to α. □

Remarks:

(a) If each Γ'_n has cyclic symmetry, then the conclusion of the corollary holds with the original definition of reconstruction sequence, Definition 2.4.

(b) The set Γ'_n can be a singleton or, in the case of cyclic symmetry, contain at most n elements. For such $\{\Gamma'_n\}$, we have

$$\lim \frac{1}{n} \log \#\Gamma'_n = 0 . \tag{6.8}$$

5. Examples of 'small' reconstruction sequences are provided also by the Ergodic Theorem. Let α be an ergodic measure on $(\Omega, \mathcal{F})$; we give an example of a reconstruction sequence $\{\Gamma_n\}$ for α which grows very slowly:

$$\lim_n \frac{1}{n} \log \#\Gamma_n = 0 . \tag{6.9}$$

Let $1_{\mathsf{a}} \in \mathcal{F}_n$ denote the indicator function of the atom $X_n^{-1}\mathsf{a}$ of $\mathcal{F}_n$, where a is a word in Ω_n;

$$1_{\mathsf{a}}(\omega) = \begin{cases} 1 & \text{if } \omega_1 = \mathsf{a}_1, \ldots, \omega_n = \mathsf{a}_n \\ 0 & \text{otherwise.} \end{cases} \tag{6.10}$$

The Ergodic Theorem implies that

$$\lim_{n\to\infty} \mathcal{A}_n 1_{\mathsf{a}}(\omega) = \alpha[X_n^{-1}\mathsf{a}] = \alpha_n[\mathsf{a}] \quad \alpha\text{-}a.e. \tag{6.11}$$

Let $\Delta(\alpha; \mathsf{a})$ be the set on which the above limit holds; let $\Delta(\alpha)$ denote the intersection of $\Delta(\alpha, \mathsf{a})$ over all words a in Ω_n and all $n = 1, 2, \ldots$. We have $\alpha[\Delta(\alpha)] = 1$; hence $\Delta(\alpha)$ is non-empty. Choose a sequence $\omega^* \in \Delta(\alpha)$; for each $n \in \mathbb{N}$, define Γ_n to be the set formed by the distinct cyclic permutations of the word $X_n\omega^*$; then

$$\lim_n \beta^{\Gamma_n} = \alpha , \tag{6.12}$$

so that $\{\Gamma_n\}$ is a reconstruction sequence for α.

6. The approach to large deviation theory sketched in Section 1 is described fully in [LP]; it has its origins in Ruelle's treatment [R] of thermodynamic entropy and Lanford's proof [L] of Cramér's Theorem.

7. Our conditional limit theorem has antecedents; the earliest we are aware of is due to van Campenhout and Cover [CC]:

   Let $Y_1, Y_2, \ldots$ be i.i.d. random variables having uniform probability mass on the range $\{1, 2, \ldots, m\}$. Then, for $1 \le \alpha \le m$ and for all $x \in \{1, 2, \ldots m\}$, we have

$$\lim_{\substack{n\to\infty \\ n\alpha \text{ integer}}} \text{Prob}\{Y_1 = x \mid \frac{1}{n}\sum_{i=1}^{n} Y_i = \alpha\} = \beta^\lambda(x), \tag{6.13}$$

   where

$$\beta^\lambda(x) = \frac{e^{\lambda x}}{\sum_{k=1}^{m} e^{\lambda k}} \tag{6.14}$$

   and the constant λ is chosen to satisfy the constraint $\sum_k k\beta^\lambda(k) = \alpha$.

A landmark in the development of such theorems is the paper by Csiszár [C], in which several important concepts are introduced.

8. We have shown that canonical sequences have the reconstruction property. In the case where α is a product measure, we have given other examples of reconstruction sequences. In the literature, sequences of the form

$$\Gamma_n^m := \{\mathsf{a} \in \Omega_n : \left| \frac{1}{n} \log \frac{\alpha_n[\mathsf{a}]}{\beta_n[\mathsf{a}]} - h(\alpha|\beta) \right| \leq \frac{1}{m}\}, \tag{6.15}$$

are used frequently. When α is ergodic, the Shannon–McMillan–Breiman Theorem implies that $\{\Gamma_n^m\}$ is a supporting sequence:

$$\lim_{n\to\infty} \alpha_n[\Gamma_n^m] = 1. \tag{6.16}$$

We can choose the strictly increasing sequence N_m so that $n \geq N_m$ implies

$$\alpha_n[\Gamma_n^m] \geq 1 - \frac{1}{m}. \tag{6.17}$$

Define

$$\Gamma_n' := \begin{cases} \Omega_n & \text{if } n < N_1, \\ \Gamma_n^m & \text{if } N_m \leq n < N_{m+1}; \end{cases} \tag{6.18}$$

in the case of ergodic α, we can use sets of the form (6.15) to generate a supporting sequence for α. Straightforward estimates show $\{\Gamma_n'\}$ to have entropic growth–rate. This suggests asking whether a supporting sequence with entropic growth–rate is a reconstruction sequence. The example near (2.25) shows that, in general, a supporting sequence for α with entropic growth–rate need not be a reconstruction sequence for α. We need additional hypothesis so that Γ_n does not include points of low α–probability. Here is a result of this kind:

Proposition 6.1 *Let α be a stationary product measure and let $\{\Gamma_n\}$ be a supporting sequence for α. For each $\varepsilon > 0$, define*

$$\Gamma_n^\varepsilon := \{\mathsf{a} \in \Gamma_n : \frac{1}{n} \log \frac{\alpha_n[\mathsf{a}]}{\beta_n[\mathsf{a}]} < h(\alpha|\beta) - \varepsilon\}. \tag{6.19}$$

If for each $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{\beta_n[\Gamma_n^\varepsilon]}{\beta_n[\Gamma_n]} = 0, \tag{6.20}$$

then $\{\Gamma_n\}$ is a reconstruction sequence for α.

Proof: We have

$$\beta^{\Gamma_n} = \frac{\beta_n[\Gamma_n \setminus \Gamma_n^\varepsilon]}{\beta_n[\Gamma_n]} \beta^{\Gamma_n \setminus \Gamma_n^\varepsilon} + \frac{\beta_n[\Gamma_n^\varepsilon]}{\beta_n[\Gamma_n]} \beta^{\Gamma_n^\varepsilon}, \tag{6.21}$$

so (6.20) implies that the sequence $\{\mathcal{A}_n \beta^{\Gamma_n}\}$ converges to α if, and only if, the sequence $\{\mathcal{A}_n \beta^{\Gamma_n \setminus \Gamma_n^\varepsilon}\}$ converges to α. Thus it suffices to consider the case in which $\Gamma_n^\varepsilon = \emptyset$ for all $\varepsilon > 0$ and all $n \in \mathbb{N}$. For $\mathsf{a} \in \Omega_n$, we have

$$f_n(\mathsf{a})\, \beta_n[\Gamma_n]\, \beta_n^{\Gamma_n}[\mathsf{a}] = \alpha_n[\mathsf{a}], \tag{6.22}$$

where f_n is given by

$$f_n(\mathrm{a}) := \alpha_n[\mathrm{a}]/\beta_n[\mathrm{a}]. \tag{6.23}$$

Assuming $\Gamma_n^\varepsilon = \emptyset$, we have

$$\log f_n \geq n\,(h(\alpha|\beta) - \varepsilon)\,. \tag{6.24}$$

Then

$$D(\beta^{\Gamma_n}\|\alpha_n) = \frac{1}{\beta_n[\Gamma_n]}\int_{\Gamma_n} -\log(f_n\beta_n[\Gamma_n])\,d\beta_n, \tag{6.25}$$

so

$$\limsup_{n\to\infty}\frac{1}{n}D(\beta^{\Gamma_n}\|\alpha_n) \leq -\liminf_{n\to\infty}\frac{1}{n}\log\beta_n[\Gamma_n] - h(\alpha|\beta) + \varepsilon \leq \varepsilon \tag{6.26}$$

for every $\varepsilon > 0$, because $\{\Gamma_n\}$ is a supporting sequence. Any limit point λ of the sequence $\{\mathcal{A}_n\beta^{\Gamma_n}\}$ is stationary. Lemma 8.1 of [LPS] and the lower semi-continuity of the specific information gain imply

$$h(\lambda|\alpha) = 0. \tag{6.27}$$

Since α is assumed to be a product measure, this implies $\lambda = \alpha$. But the level sets of h are compact, so the sequence $\{\mathcal{A}_n\beta^{\Gamma_n}\}$ converges to α. □

Remarks:

(a) If $\{\Gamma_n^*\}$ is a supporting sequence for α, then the sequence $\{\Gamma_n\}$ given by

$$\Gamma_n := \Gamma_n^* \bigcap \Gamma_n', \tag{6.28}$$

as given in (6.18), is a supporting sequence for α, which satisfies (6.20).

(b) Under the condition that α is *weakly dependent* (see [LPS]), one may also deduce (6.27). One can then conclude that λ is a *Gibbs state* for the interaction associated with α; this does *not*, in general, imply that $\lambda = \alpha$.

Acknowledgements:
We thank Frank den Hollander for a careful reading of an earlier draft of this paper. This work was partially supported by the European Commission under the Human Capital and Mobility Scheme (EU contract CHRX-CT93-0411).

References

[B] L.Breiman, The Individual Ergodic Theorem of Information Theory, *Ann.Math.Stat.* **28** 809-811 (1957)

[C] I.Csiszár, Sanov property, generalized I-projection and a conditional limit theorem, *Ann.Prob.* **12**, 768-793 (1984)

[CC] J.M.van Campenhout and T.M.Cover, Maximum Entropy and Conditional Probability, *IEEE Trans. Inform. Theory* **4**, 483-489 (1981)

[G] J.W.Gibbs, *Elementary Principles of Statistical Mechanics* New Haven, Connecticut: Yale University Press (1902)

[K] A.N.Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung (Ergebnisse der Math.)*, Berlin: Springer (1933) Transl. as *Foundations of Probability* New York: Chelsea (1956)

[L] O.E.Lanford, Entropy and equilibrium states in classical statistical mechanics, in Lecture Notes in Physics **20**, 1-113 Berlin: Springer (1973)

[LP] J.T.Lewis, C.-E.Pfister, Thermodynamic Probability Theory: *Russian Math. Surveys* **50**:2, 279-317 (1995)

[LPS1] J.T.Lewis, C.-E.Pfister and W.G.Sullivan, Large Deviations and the Thermodynamic Formalism: a new proof of the equivalence of ensembles, in On Three Levels, M. Fannes, C. Maes, A. Verbeure eds., Plenum Press 183-193, (1994)

[LPS2] J.T.Lewis, C.-E.Pfister and W.G.Sullivan, The Equivalence of Ensembles for Lattice Systems: Some examples and a Counterexample

[LPS] J.T.Lewis, C.-E.Pfister and W.G.Sullivan, Entropy, Concentration of Probability and Conditional Limit Theorems, *Markov Processes and Related Fields*, **1** 319–386 (1995)

[M] B.McMillan, The Basic Theorems of Information Theory, *Ann.Math.Stat.* **24** 196–219 (1953)

[R] D.Ruelle, Correlation functionals, *J. Math. Phys.* **6**, 201-220 (1965)

[S] C.E.Shannon, A mathematical theory of communications, *Bell System Technical Journal* **27** 379-423, 623-656 (1948)