

On bounded redundancy of universal codes*

Łukasz Dębowski

*Institute of Computer Science, Polish Academy of Sciences
ul. Jana Kazimierza 5, 01-248 Warszawa, Poland*

Abstract

Consider stationary ergodic measures for which the difference between the expected length of a uniquely decodable code and block entropy is asymptotically bounded by a constant. Using ergodic decomposition, it is shown that the number of such measures is less than the base of logarithm raised to the power of the constant. In consequence, an analogical statement is derived for excess lengths of universal codes. The latter was previously communicated without proof.

Keywords: uniquely decodable codes, entropy, ergodic decomposition

*The work was partly supported by the Polish Ministry of Scientific Research and Information Technology, grant no. 1/P03A/045/28.

1 Introduction

The aim of this note is to establish an impossibility result about uniquely decodable codes and stationary ergodic measures. Consider measures for which the difference between the expected length of a code and block entropy is asymptotically bounded by a constant. We will show that the number of such measures is less than the base of logarithm raised to the power of the constant. In other words, there cannot be too good codes. This simple but novel result is a showcase application of Shannon information measures for σ -algebras, a neat and powerful tool developed in [7, 2]. As a corollary, we will derive an analogical statement for excess lengths of universal codes, which was communicated in [2] in a weaker form without proof. We have been informed that these results may be also derived using channel capacity for universal codes [1, Section 13.1], however, we believe that sorting out formal details of the alternative proofs would not yield considerably shorter reasoning.

The preliminaries are as follows. Let $\mathbb{X} = \{0, 1, \dots, D_X - 1\}$ be a finite alphabet. For the measurable space $(\mathbb{X}^{\mathbb{Z}}, \mathfrak{X}^{\mathbb{Z}})$, consider the shift transformation $T : \mathbb{X}^{\mathbb{Z}} \ni (x_k)_{k \in \mathbb{Z}} \mapsto (x_{k+1})_{k \in \mathbb{Z}} \in \mathbb{X}^{\mathbb{Z}}$, where $x_k \in \mathbb{X}$. The shift-invariant σ -algebra is $\mathfrak{I} := \{A \in \mathfrak{X}^{\mathbb{Z}} : TA = A\}$. Let $(\mathbb{S}, \mathfrak{S})$ be the measurable space of stationary probability measures on $(\mathbb{X}^{\mathbb{Z}}, \mathfrak{X}^{\mathbb{Z}})$ (i.e., $\mu \circ T = \mu$ for $\mu \in \mathbb{S}$) and let $(\mathbb{E}, \mathfrak{E}) \subset (\mathbb{S}, \mathfrak{S})$ be the subspace of ergodic measures (i.e., $\mu(A) \in \{0, 1\}$ for $\mu \in \mathbb{E}$ and $A \in \mathfrak{I}$). Precisely, \mathfrak{S} and \mathfrak{E} are defined as the smallest σ -algebras containing all cylinder sets $\{\mu \in \mathbb{S} : \mu(A) \leq r\}$ and $\{\mu \in \mathbb{E} : \mu(A) \leq r\}$, $A \in \mathfrak{X}^{\mathbb{Z}}$, $r \in \mathbb{R}$, respectively. Since $\mathfrak{X}^{\mathbb{Z}}$ is countably generated, all respective singletons $\{\mu\}$ belong to \mathfrak{S} and \mathfrak{E} .

Define random variables $X_i((x_k)_{k \in \mathbb{Z}}) := x_i$ on $(\mathbb{X}^{\mathbb{Z}}, \mathfrak{X}^{\mathbb{Z}})$ and write the blocks as $X_{n:m} := (X_i)_{n \leq i \leq m}$. For a stationary measure $\mu \in \mathbb{S}$ consider block entropy

$$H_\mu(n) := H_\mu(X_{i+1:i+n}) := \mathbf{E}_\mu[-\log \mu(X_{i+1:i+n} = \cdot)],$$

where \mathbf{E}_μ is the expectation with respect to μ and \log is the natural logarithm. Moreover, we consider another finite alphabet $\mathbb{Y} = \{0, 1, \dots, D_Y - 1\}$ and a code $C : \mathbb{X}^+ \rightarrow \mathbb{Y}^+$. Denote the expected length of the code as

$$H_\mu^C(n) := \mathbf{E}_\mu |C(X_{1:n})| \log D_Y.$$

The code is called uniquely decodable if its extension $C^*(u_1, \dots, u_k) := C(u_1) \dots C(u_k)$, $u_i \in \mathbb{X}^n$, $k \in \mathbb{N}$, is an injection for any n . As shown, e.g., in [1], this property implies the source coding inequality

$$H_\mu^C(n) \geq H_\mu(n). \tag{1}$$

The main result of this paper is as follows.

Theorem 1 *Let C be a uniquely decodable code. Then*

$$\text{card} \left\{ \mu \in \mathbb{E} : \limsup_{n \rightarrow \infty} [H_\mu^C(n) - H_\mu(n)] \leq K \right\} \leq \exp(K). \tag{2}$$

This theorem will be proved in the following section. As we have said, the proposition states that there cannot be too good codes. Whereas there are uncountably many ergodic measures, the difference $H_\mu^C(n) - H_\mu(n)$ is bounded

only for countably many of them. Indeed, there exists a code satisfying the later condition for measures in an arbitrary countable subset $A \subset \mathbb{E}$. For instance, let C be the Shannon-Fano code for measure $P = \sum_{\mu \in A} c_\mu \mu$, where $\sum_{\mu \in A} c_\mu = 1$ and $c_\mu > 0$. Then $\limsup_n [H_\mu^C(n) - H_\mu(n)] \leq -\log c_\mu + \log D_Y < \infty$ for all $\mu \in A$. Here we show that this is not true if A is uncountable.

As another related result, Shields [8] demonstrated that for any $\beta \in [0, 1)$ and any universal code there exists such an ergodic measure that

$$\sup_{n \in \mathbb{N}} [H_\mu^C(n) - H_\mu(n)] / n^\beta = \infty. \quad (3)$$

The code is called here universal if it is uniquely decodable and

$$\lim_{n \rightarrow \infty} [H_\mu^C(n) - H_\mu(n)] / n = 0$$

holds for any stationary measure $\mu \in \mathbb{S}$, cf. [1]. Thus Theorem 1 strengthens Shields' result for $\beta = 0$.

Next, we discuss the result mentioned in [2]. Denote the mutual information between blocks of length n ,

$$\begin{aligned} E_\mu(n) &:= 2H_\mu(n) - H_\mu(2n) \\ &= I_\mu(X_{1:n}; X_{n+1:2n}) := H_\mu(X_{1:n}) + H_\mu(X_{n+1:2n}) - H_\mu(X_{1:2n}), \end{aligned}$$

and the expected excess length of the code C ,

$$\begin{aligned} E_\mu^C(n) &:= 2H_\mu^C(n) - H_\mu^C(2n) \\ &= \mathbf{E}_\mu(|C(X_{1:n})| + |C(X_{n+1:2n})| - |C(X_{1:2n})|) \log D_Y. \end{aligned}$$

There are a few universal codes for which excess lengths have a natural interpretation. Firstly, let $C(u)$ be the shortest program to generate string u for a prefix Turing machine. This code is universal and then $E_\mu^C(n)$ is the expectation of algorithmic information between blocks $X_{1:n}$ and $X_{n+1:2n}$, cf. [6]. While the shortest program to generate a string cannot be efficiently found, there exist also computable universal codes such as the Lempel-Ziv code [9] or grammar-based codes [5, 3]. In particular, excess length $E_\mu^C(n)$ of admissibly minimal grammar-based codes is bounded above by the number of distinct nonterminal symbols in the grammar used for compression [3].

We claim this proposition, announced without proof as [2, Theorem 6].

Theorem 2 *Let C be a universal code. We have*

$$\text{card} \left\{ \mu \in \mathbb{E} : \limsup_{n \rightarrow \infty} [E_\mu^C(n) - E_\mu(n)] \leq K \right\} \leq \exp(K).$$

The original statement in [2] was weaker, namely, the term $-E_\mu(n)$ was missing. ($E_\mu(n)$ is nonnegative.) Our former intention was to derive Theorem 2 first and then infer Theorem 1 as a corollary but we realized that the reversed order yields stronger bounds.

To prove the Theorem 2, we use a lemma which resembles [3, Lemma 1].

Lemma 1 (Excess-bounding lemma II) *Consider a function $G : \mathbb{N} \rightarrow \mathbb{R}$ such that $\lim_k G(k)/k = 0$ and $G(n) \geq 0$ for all n . For any $A \in \mathbb{N}$ and $\beta \in [0, 1)$ these statements are equivalent:*

(i) $G(n) \leq \gamma n^\beta$ holds for a $\gamma > 0$ and all but finitely many $n \geq N$.

(ii) $AG(n) - G(An) \leq \delta n^\beta$ holds for a $\delta > 0$ and all but finitely many $n \geq N$.

The exact relationship between constants γ and δ is given in the proof.

Proof of Lemma 1: If (i) holds then (ii) holds for $\delta = A\gamma$ because $G(n) \geq 0$. Conversely, if (ii) is true then for sufficiently large n , we obtain

$$G(n) = G(n) - n \lim_{k \rightarrow \infty} \frac{G(k)}{k} = \sum_{k=0}^{\infty} \frac{AG(A^k n) - G(A^{k+1} n)}{A^{k+1}} \leq \frac{\delta n^\beta}{A(1 - A^{\beta-1})}$$

so (i) holds with $\gamma = \delta/[A(1 - A^{\beta-1})]$. \square

Proof of Theorem 2: We apply Lemma 1 (ii) \implies (i) with $G(n) = H_\mu^C(n) - H_\mu(n)$ and $A = 2$. Consider an ergodic measure μ such that $\limsup_n [E_\mu^C(n) - E_\mu(n)] \leq K$. This implies $\limsup_n [H_\mu^C(n) - H_\mu(n)] \leq K + \epsilon$ for any $\epsilon > 0$. Thus the claim follows from Theorem 1. \square

Lemma 1 implies that redundancy $H_\mu^C(n) - H_\mu(n)$ of universal codes is always bounded in a similar fashion as excess redundancy $E_\mu^C(n) - E_\mu(n)$. In particular, the result (3) by Shields is equivalent to saying that for any $\beta \in [0, 1)$ and any universal code there exists an ergodic measure such that

$$\sup_{n \in \mathbb{N}} [E_\mu^C(n) - E_\mu(n)] / n^\beta = \infty.$$

This remark concludes the Introduction. In the remaining part of the paper we prove Theorem 1.

2 The proof of Theorem 1

Recall that \mathfrak{J} is the shift-invariant σ -algebra. Consider a stationary measure $P \in \mathbb{S}$. According to the ergodic decomposition theorem [4, Theorems 9.10-12], if \mathfrak{X} is countable then there exists a random ergodic measure $F : (\mathfrak{X}^{\mathbb{Z}}, \mathfrak{J}) \rightarrow (\mathbb{E}, \mathfrak{E})$ such that

$$F(A) = P(A|\mathfrak{J}) \tag{4}$$

P -almost surely for all $A \in \mathfrak{X}^{\mathbb{Z}}$. Because $\mathbf{E}_P P(A|\mathfrak{J}) = P(A)$, we have

$$H_P^C(n) = \mathbf{E}_P H_F^C(n). \tag{5}$$

Now we will discuss a similar decomposition for block entropy. By the generalized chain rule [2, Theorem 2(ii)], we have

$$H_P(n) = H_P(X_{1:n}) = I_P(X_{1:n}; \mathfrak{J}) + H_P(X_{1:n}|\mathfrak{J}),$$

where $I_\mu(\mathfrak{A}; \mathfrak{B})$ is the mutual information and $H_\mu(\mathfrak{A}|\mathfrak{B})$ is the conditional entropy as defined in [2]. (For the classical but less general treatment of information measures for σ -algebras, cf. [7].) Moreover, $H_P(X_{1:n}|\mathfrak{J}) = \mathbf{E}_P H_F(n)$ follows from (4). Resuming, it follows that

$$H_P(n) = I_P(X_{1:n}; \mathfrak{J}) + \mathbf{E}_P H_F(n). \tag{6}$$

The same formula was derived in [3, Theorem 7].

By the source coding inequality (1) for $\mu = P$, formulae (5) and (6) imply

$$\mathbf{E}_P [H_F^C(n) - H_F(n)] \geq I_P(X_{1:n}; \mathfrak{J}).$$

Because the P -completion of \mathfrak{J} is contained in the P -completion of the tail σ -algebra by [2, Lemma 3], we also have $\lim_n I_P(X_{1:n}; \mathfrak{J}) = I_P(\mathfrak{J}; \mathfrak{J})$ by [2, Theorems 1(iii)-(v) and 2(i)]. Hence

$$\limsup_{n \rightarrow \infty} \mathbf{E}_P [H_F^C(n) - H_F(n)] \geq I_P(\mathfrak{J}; \mathfrak{J}) = H_P(\mathfrak{J}), \quad (7)$$

where $H_\mu(\mathfrak{A}) := \sup \left[-\sum_{i=1}^I \mu(A_i) \log \mu(A_i) \right]$ is the entropy of a σ -algebra \mathfrak{A} with the supremum taken over all finite partitions $\{A_i\}_{i=1}^I$, $A_i \in \mathfrak{A}$. Inequality (7) is used in the further reasoning.

Write

$$N(K) := \text{card} \left\{ \mu \in \mathbb{E} : \limsup_{n \rightarrow \infty} [H_\mu^C(n) - H_\mu(n)] \leq K \right\}.$$

Observe that if $N(K) = 0$ then (2) holds trivially. Thus it suffices to prove (2) for $N(K) \geq 1$. Consider a natural number $M \geq 1$ such that $M \leq N(K)$. Let $A \subset \mathbb{E}$ be such a subset of M distinct ergodic measures μ that $\limsup_n [H_\mu^C(n) - H_\mu(n)] \leq K$. Put the measure $P = M^{-1} \sum_{\mu \in A} \mu$. By the uniqueness of its ergodic decomposition [4, Theorem 9.12], we have $P(F = \mu) = 1/M$ for $\mu \in A$ and $P(F = \mu) = 0$ else. Let \mathfrak{F} be the smallest σ -algebra against which F is measurable. This σ -algebra is generated by the finite partition $\{(F = \mu)\}_{\mu \in A}$ and $H_P(\mathfrak{F}) = \log M$. Because the P -completion of \mathfrak{F} equals the P -completion of \mathfrak{J} by [2, Lemma 3], we also have $H_P(\mathfrak{J}) = H_P(\mathfrak{F})$ by [2, Theorem 2(i)].

Take an $\epsilon > 0$. Random variables $K + \epsilon - [H_F^C(n) - H_F(n)]$ are nonnegative P -almost surely for sufficiently large n because F assumes only finitely many values. Thus, by the Fatou lemma, $K + \epsilon - \mathbf{E}_P \limsup_n [H_F^C(n) - H_F(n)] \leq K + \epsilon - \limsup_n \mathbf{E}_P [H_F^C(n) - H_F(n)]$. Hence from inequality (7) we obtain

$$\begin{aligned} \log M = H_P(\mathfrak{J}) &\leq \limsup_{n \rightarrow \infty} \mathbf{E}_P [H_F^C(n) - H_F(n)] \\ &\leq \mathbf{E}_P \limsup_{n \rightarrow \infty} [H_F^C(n) - H_F(n)] \leq K. \end{aligned}$$

Since this holds for any $M \leq N(K)$, inequality (2) follows.

Acknowledgment

I would like to thank Jan Mielniczuk for his remarks.

References

- [1] T. M. Cover and J. A. Thomas. *Elements of Information Theory, 2nd ed.* Wiley, 2006.

- [2] Ł. Dębowski. A general definition of conditional information and its application to ergodic decomposition. *Statist. Probab. Lett.*, 79:1260–1268, 2009.
- [3] Ł. Dębowski. On the vocabulary of grammar-based codes and the logical consistency of texts. *IEEE Trans. Inform. Theor.*, 57:4589–4599, 2011.
- [4] O. Kallenberg. *Foundations of Modern Probability*. Springer, 1997.
- [5] J. C. Kieffer and E. Yang. Grammar-based codes: A new class of universal lossless source codes. *IEEE Trans. Inform. Theor.*, 46:737–754, 2000.
- [6] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications, 3rd ed.* Springer, 2008.
- [7] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.
- [8] P. C. Shields. Universal redundancy rates don't exist. *IEEE Trans. Inform. Theor.*, IT-39:520–524, 1993.
- [9] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inform. Theor.*, 23:337–343, 1977.