

Generating Random Bits from an Arbitrary Source: Fundamental Limits

Sridhar Vembu, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

Abstract—Suppose we are given a random source and want to use it as a random number generator; at what rate can we generate fair bits from it? We address this question in an information-theoretic setting by allowing for some arbitrarily small but nonzero deviation from “ideal” random bits. We prove our results with three different measures of approximation between the ideal and the obtained probability distributions: the variational distance, the d -bar distance, and the normalized divergence. Two different contexts are studied: fixed-length and variable-length random number generation. The fixed-length results of this paper provide an operational characterization of the *inf-entropy rate* of a source, defined in Han and Verdú [1] and the variable-length results characterize the *liminf* of the entropy rate, thereby establishing a pleasing duality with the fundamental limits of source coding. A feature of our results is that we do not restrict ourselves to ergodic or to stationary sources.

Index Terms—Shannon theory, random number generation, entropy, fixed-length source coding, variable-length source coding, approximation theory.

1. INTRODUCTION

CANONICAL random number generators produce independent unbiased bits. They operate by transforming deterministically a given random source. In this paper, we investigate the maximum rate at which random bits from an arbitrary random source can be extracted.

In the special case where the random source is a sequence of independent flips of a biased coin, von Neumann [4] suggested a very simple method for this problem which is as follows: Flip the biased coin repeatedly and split the resulting sequence into pairs of consecutive coin flips. Output 1 when HT occurs, output 0 when TH occurs, and output nothing otherwise. This scheme generates random bits at the rate of $p(1-p)$ bits per coin flip, where p is the coin bias, a rate which is suboptimal as shown by Elias [5]. He shows that for a general stationary source, the entropy rate is an upper bound on the rate at which we can generate independent unbiased bits from the source. Elias also shows how to achieve this for the special class of stationary sources consisting of finite-state sources. In particular, this solves the problem for the case of independent and identically distributed (i.i.d.) sources. These are basically the most general results known so far.

Manuscript received April 4, 1994; revised March 1, 1995. This work was supported in part by the National Science Foundation under PYI Grant ECSE-8857689. The work of S. Vembu was also supported by an IBM Graduate Fellowship.

S. Vembu is with Qualcomm Inc., San Diego, CA 92121.

S. Verdú is with the Department of Electrical Engineering Princeton University, Princeton NJ 08544 USA.

IEEE Log Number 9413636.

There are some further results on the algorithmic aspects of random number generation, where computational efficiency, rather than the extraction of the maximum number of bits, is the main goal. Iterating the von Neumann procedure, Peres [6] gave a computationally efficient method for the i.i.d. and finite-state case, that also achieves the optimal rate. Many other existing methods do not achieve the optimal rate and they will not concern us here.

The fundamental approach in generating fair bits in the schemes of [4]–[6] is to exploit the symmetries in the source and consider equally-likely events. For example, if we take n independent flips of a biased coin, all sequences that have exactly k heads are equiprobable and there are

$$\binom{n}{k}$$

such sequences. So we can generate an equiprobable bit string of length

$$\lfloor \log \binom{n}{k} \rfloor$$

whenever the coin flips produce one of these k -head sequences. Such schemes are called variable-length schemes, a term we will elaborate on in the succeeding paragraphs. We can exploit similar symmetries for finite-order sources and finite-state Markov sources. If we require exactly equiprobable bits, the Elias upper bound may not be achievable for infinite-memory sources (even if they are stationary and ergodic) because such symmetries need not exist. In general, we can only hope for almost equally-likely events.

In many situations requiring pure coin flips (randomized algorithms, for example) such approximate random bits have been shown to be sufficient. An extensive literature, much of it fairly recent, exists in computer science which deals with “derandomization”—substituting almost pure randomness instead of pure randomness, see [16] and references therein.

One feature of the von Neumann procedure and much of the work in the computer science literature (see [18]) is that they do not require the exact statistics of the source to be known—they only need structural information about the source; for example, whether it is i.i.d. or if it is Markov, the order of the Markov process. In this sense these algorithms are universal. In contrast we are interested in characterizing the maximum number of random bits we can generate from a source whose statistics are known. However, the statistics can be arbitrary—no stationarity or ergodicity assumptions are imposed.

To reiterate, the basic approach taken in this paper is to relax the requirement on exactly equiprobable bits, by asking for an arbitrarily accurate approximation. This is very much in tune with the classical information theoretic approach [7], where a nonzero (but arbitrarily small) error is allowed, for example, in transmission of information, fixed-length source coding, etc. Once we do that, the Elias upper bound is achievable and we are able to show both achievability and converse results that are applicable in general. In particular, the converse result is a stronger version of the Elias bound, because we prove that even if we are content with asymptotically pure bits, we cannot achieve more than the entropy rate.

To motivate the solution found in this paper, it is useful to recall Sinai's ergodic theoretic result [9]. Consider a stationary ergodic process Z and an i.i.d. process B such that the entropy rate of Z is greater than or equal to the entropy rate of B . Then there exists a sliding block encoding (time-invariant deterministic transformation that looks at the infinite past and the infinite future and outputs a single symbol and then shifts the input by one position to the left and repeats the above operation) which takes in as input the process Z and outputs a process statistically identical to B .

Our setting is very different from the ergodic theory approach in that we consider a "one-shot" finite-dimensional coding problem: we take an n -symbol string of the input process and output a binary sequence of length r where r may or may not depend on the particular n -symbol realization and prove results about the asymptotic behavior of the rate $\frac{r}{n}$ when the distribution of the output sequence is required to become asymptotically equiprobable. In contrast to the ergodic theory setting, we do not consider mappings that take infinite strings to infinite strings. Nevertheless, our results are consistent with Sinai's result, when specialized to the stationary ergodic case and properly interpreted.

The largest asymptotic rate at which almost independent equiprobable bits can be generated by a deterministic transformation of the source will be referred to as the Intrinsic Randomness (IR) rate of the source. We deal with this problem in two settings: a) fixed-length random number generation, where every source n -symbol realization is deterministically transformed into an r -bit sequence where r depends only on n and not on the particular realization of the source sequence and b) variable-length random number generation, where the length of the output string (which could be zero) depends on the particular realization of the source sequence. The natural performance measure in the variable-length case is the asymptotic *average* rate of the bit sequence generated. This situation of fixed length versus variable length is reminiscent of (almost noiseless) source coding—an analogy that will become more apparent when we describe the formulas for random number generation. However, the proofs of our direct and converse results are quite different from those of source coding.

What do we mean by approximately equiprobable bits? We prove our results using three measures of approximation of probability distributions—variational distance, the d -bar distance, and normalized divergence. Of these measures the d -bar distance is known to be weaker than the variational

	Source	Channel
Sup	$\bar{H}(Y)$ Minimum Source Coding Rate	$\sup_x \bar{I}(X; Y)$ Resolvability
Inf	$\underline{H}(Y)$ Intrinsic Randomness	$\sup_x \underline{I}(X; Y)$ Channel Capacity

Fig. 1. Operational characterizations.

distance, while the normalized divergence is neither stronger nor weaker than the others. Nevertheless, in all three cases the Intrinsic Randomness turns out to be the same. We show that for fixed-length random number generation the maximal achievable rate is the *inf-entropy rate*, defined in [1], and for the variable-length case it is the *liminf* of the per-symbol entropy. The results of this paper are a pleasing counterpart to the general source coding results of [1] where the minimum achievable fixed-length encoding rate is shown to be the *sup-entropy rate* and the minimum achievable variable-length source coding rate is shown to be the *limsup* of the per-symbol entropy for an arbitrary finite alphabet source. The inf- and sup-entropy rates are defined in the next section.

For the fixed-length problem, the analogy is complete in the sense of Fig. 1: intrinsic randomness plays the counterpart to the source-coding rate, analogously to channel resolvability and channel capacity [2]. The problem of resolvability is introduced in [1] and it is the dual of the channel capacity problem.

This paper is organized as follows. In Section II we provide the definitions and illustrative examples. Section III gives some examples that illustrate the computation of the Intrinsic Randomness rate in the fixed-length and variable-length modes. Section IV is devoted to two lemmas that form the core of the proofs in the remaining two sections. In Sections V and VI we prove the results on fixed-length and variable-length random number generation, respectively.

II. DEFINITIONS AND STATEMENT OF RESULTS

In this paper we deal with general discrete random sources, characterized by their sequence of finite-dimensional distributions

$$Z = \{P_{Z^n}\}_{n=1}^{\infty}$$

where Z^n takes values on A^n , and A is a finite set.¹ Note that the sources we allow include but are not restricted to random processes, because the finite-dimensional distributions are not required to be consistent.

To help illustrate the distinction between fixed- and variable-length random number generation we provide the following example. Consider the probability distribution on $\{0, 1\}^n$ such that the all-zero string has probability $\frac{1}{2}$ and all strings beginning with 11 are equiprobable. The rest of the strings have zero probability. In the fixed-length case, we generate one random bit—0 if the all-zero string occurs and 1, otherwise.

¹Most results in this paper do not need the finite alphabet assumption. We point out where it is needed.

But if we allow ourselves the freedom to output variable number of bits, we can do much better. In this case we output a null string when the all-zero string occurs. We output $(n-2)$ bits whenever one of the other nonzero probability strings occur. On average, we generate $\frac{n}{2} - 1$ bits in this approach. Conditioned on the number of bits generated, we get equally-likely bits. Notice that the fixed-length scenario corresponds to the worst case, and the variable-length scenario corresponds to the average case.

In the preceding example, we are able to generate exact equiprobable bits. In general, this may not be possible, as we illustrate in [19] with an example.

Now we state the formal definitions.

Definition 1: For any $\epsilon > 0$, R is said to be an ϵ -achievable Intrinsic Randomness (IR) rate for source Z if for any $\gamma > 0$ there exists a sequence of deterministic mappings $\phi_n : A^n \rightarrow \{0, 1\}^r$ such that for all sufficiently large n ,

$$\frac{r}{n} > R - \gamma$$

and

$$\Delta(\phi_n(Z^n), B^r) < \epsilon$$

where B^r is the equiprobable distribution on $\{0, 1\}^r$ and Δ is a measure of the distance between probability distributions to be specified in the sequel.²

Definition 2: If R is ϵ -achievable for all $\epsilon > 0$ then it is called an achievable IR rate.

The maximum achievable intrinsic randomness (MIR) rate of a source Z is denoted by $U_v(Z)$, $U_b(Z)$, and $U_d(Z)$ according to the following three respective choices of the distance measures in Definition 1:

1) $U_v(Z)$: MIR rate according to variational distance:

If P and Q are two probability measures defined on the same measurable space (Ω, \mathcal{F}) then the variational distance between them is given by

$$\begin{aligned} \Delta(P, Q) &= d(P, Q) = \sum_{\omega \in \Omega} |P(\omega) - Q(\omega)| \\ &= 2 \sup_{E \in \mathcal{F}} |P(E) - Q(E)|. \end{aligned}$$

2) $U_b(Z)$: MIR rate according to d-bar distance:

Let P and Q be probability distributions on $\{0, 1\}^r$, and U^r and \tilde{U}^r be the corresponding random variables. The d-bar distance between the two distributions is

$$\Delta(P, Q) = \bar{d}_r(P, Q) = \inf_{P \vee Q} E \left[\frac{1}{r} d_H(U^r, \tilde{U}^r) \right].$$

Here the expectation is with respect to some joint distribution of U^r and \tilde{U}^r and the infimum is over all such joint distributions on $\{0, 1\}^r \times \{0, 1\}^r$ with the first and second marginals P and Q , respectively. $d_H(u^r, \tilde{u}^r)$ refers to the Hamming distance between the two r -bit strings u^r and \tilde{u}^r , i.e., the number of positions in which they differ.

The d-bar distance was introduced by Ornstein in his famous isomorphism paper [11]. Many properties of d-bar distance,

²The first argument in $\Delta(\phi_n(Z^n), B^r)$ denotes the probability distribution of the random variable $\phi_n(Z^n)$.

including the fact that it is less than or equal to half the variational distance, are derived in [13]. This fact implies that $U_v(Z) \leq U_b(Z)$.

3) $U_d(Z)$: MIR rate according to normalized divergence: If P and Q are distributions on $\{0, 1\}^r$, then their normalized divergence is given by

$$\Delta(P, Q) = \frac{1}{r} D(P||Q) = \frac{1}{r} \sum_{a^r \in \{0, 1\}^r} P(a^r) \log \frac{P(a^r)}{Q(a^r)}.$$

One of the main results of this paper is that the maximal intrinsic randomness rate of a discrete source (according to each of the three foregoing distance measures) is equal to the inf-entropy rate of the source, which is defined as follows.

Definition 3: The inf-entropy rate of Z , $\underline{H}(Z)$, is the largest extended real number α , that satisfies for all $\delta > 0$

$$\lim_{n \rightarrow \infty} P_{Z^n} \{ z^n : \frac{1}{n} \log \frac{1}{P_{Z^n}(z^n)} \leq \alpha - \delta \} = 0.$$

The inf-entropy rate and the sup-entropy rate $\bar{H}(Z)$ of a random source were introduced in [1]. They can be referred to as the liminf and limsup in probability, respectively, of the sequence of random variables

$$\left\{ \frac{1}{n} \log \frac{1}{P_{Z^n}(Z^n)} \right\}_{n=1}^{\infty}.$$

If both quantities are equal, then the sequence converges in probability to the entropy rate of the source

$$H(Z) = \lim_{n \rightarrow \infty} \frac{1}{n} E \left[\log \frac{1}{P_{Z^n}(Z^n)} \right]$$

provided the source alphabet is finite, as proved in [1]. It is further shown in [1] that the sup-entropy rate is the minimum achievable fixed-length source encoding rate (see [3] for definitions).

Our main result in the fixed-length case is the following:

Theorem 1: For any discrete source Z

$$U_v(Z) = U_b(Z) = U_d(Z) = \underline{H}(Z).$$

In variable-length random number generation, the pertinent definitions are as follows:

Definition 4: For any $\epsilon > 0$, R is an ϵ -achievable variable-length intrinsic randomness rate for a source Z if for any $\delta > 0$, there exists a sequence of sets I_n of nonnegative integers, a sequence of partitions

$$A^n = \bigcup_{r \in I_n} J_r^{(n)}$$

and a sequence of deterministic mappings

$$\{\phi_{n,r} : J_r^{(n)} \mapsto \{0, 1\}^r\}_{r \in I_n}$$

such that the following conditions are met:

(C1) For all n

$$\frac{1}{n} \sum_{r \in I_n} r P_{Z^n}(J_r^{(n)}) > R - \delta. \quad (1)$$

(C2) For all sufficiently large n

$$\max_{r \in I_n} \Delta(\phi_{n,r}(Z_r^n), B^r) < \epsilon$$

where B^r has the equiprobable distribution on $\{0,1\}^r$ and Z_r^n is Z^n restricted to $J_r^{(n)}$.

The crux of the above definition is to partition the space of source output strings into different sets and output a bit string whose length depends on the particular partition realized. We require that, conditioned on the length of the output bit string, we get almost equiprobable bits. In the above definition we can replace the maximum over $r \in I_n$ with an average taken with respect to the distribution $P_{Z^n}(J_r^{(n)})$, without impacting our results.

Definition 5: If R is an ϵ -achievable VLIR rate for all $\epsilon > 0$, then it is called an achievable VLIR rate. The maximum-achievable variable-length intrinsic randomness (MVLIR) rate of a source Z is denoted by $V_v(Z)$, $V_b(Z)$ and $V_d(Z)$ according to the following respective choices for the distance measure Δ : variational distance, d-bar distance, or normalized divergence.

The main result in the variable-length case is the following theorem:

Theorem 2: For every finite-alphabet source Z

$$V_v(Z) = V_b(Z) = V_d(Z) = \liminf_{n \rightarrow \infty} \frac{1}{n} H(Z^n).$$

III. EXAMPLES

This section is devoted to some examples that illustrate the conditions under which the MIR and MVLIR rates are equal.

A. Stationary Ergodic Sources

According to the Shannon–McMillan theorem, for any stationary ergodic source, the sequence of random variables

$$\left\{ \frac{1}{n} \log \frac{1}{P_{Z^n}(Z^n)} \right\}$$

converges in probability to

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(Z^n).$$

Thus

$$\underline{H}(Z) = \lim_{n \rightarrow \infty} \frac{1}{n} H(Z^n). \quad (2)$$

Convergence in probability of the self-information random variables to a constant is equivalent to saying that the inf- and sup-entropy rates equal that constant. Using Theorems 1 and 2, (2) implies that the MIR rate equals the MVLIR rate for stationary ergodic sources.

B. Stationary Nonergodic Source

We now give an example of a stationary nonergodic source for which the MIR and MVLIR rates are not equal. At the beginning of time nature selects one of two Bernoulli sources with probabilities p and q , respectively, where $q \neq p \neq 1 - q$. It selects the first i.i.d. source with probability $0 < \theta < 1$

and this selection is independent of the two source sequence realizations. The inf-entropy of the joint source is clearly $\min\{h(p), h(q)\}$. On the other hand, the limit of the per-symbol entropy exists and is equal to $\theta h(p) + (1 - \theta)h(q)$. Hence for this source the MIR and MVLIR rates are not equal.

C. Nonstationary Source

Nonstationary nonergodic sources may have identical MIR and MVLIR rates. To see this, consider the following source (which appears in the counterexample to the separation theorem in [10]).

Let J be an infinite set of positive integers given by

$$J = \{i \in N : 2^{2k-1} \leq i < 2^{2k}, \text{ for some } k \in N\}$$

It can be enumerated as

$$J = \{2, 3, 8, 9, 10, 11, 12, 13, 14, 15, 32, 33, \dots, 62, 63, 128, \dots\}.$$

Let Z be a binary memoryless nonstationary source whose distribution P_Z , is given by

$$P_Z(i) = \begin{cases} 1/2, & i \in J \\ 1, & i \notin J. \end{cases}$$

That is, at time $i \in J$ the source is equally likely $\{0,1\}$ and at time $i \notin J$ the source is deterministic.

To evaluate $\underline{H}(Z)$, write

$$\frac{1}{n} \log P_{Z^n}(z^n) = \frac{1}{n} \sum_{i=1}^n \log P_Z(z_i)$$

and observe that $\log P_Z(z_i)$ is deterministic, attaining the value -1 bit for $i \in J$ and 0 for $i \notin J$. (For convenience, the logarithms in this example have base 2.) Thus it is straightforward to verify that

$$\underline{H}(Z) = \liminf_{n \rightarrow \infty} \frac{J(n)}{n} \quad (3)$$

where $J(n)$ stands for the cardinality of the intersection of J with the set $\{1, 2, \dots, n\}$, i.e.

$$J(n) \triangleq |J \cap \{1, 2, \dots, n\}|.$$

The entropy of Z^n is clearly $J(n)$ and hence the liminf of the per-symbol entropy equals

$$\liminf_{n \rightarrow \infty} \frac{J(n)}{n}$$

and (3) implies that the MIR and MVLIR rates are equal (to $1/3$ [10]).

IV. TWO LEMMAS

We prove two lemmas which we call the *aggregation lemma* and *continuity lemma*, respectively. The former is the crux of the direct part and the latter is important for the converse part.

We begin with the aggregation lemma. The basic idea in this is that if we have a probability distribution whose probability values are very small (but which may widely differ from each other), then we can aggregate them into larger clusters of approximately equal probability.

Lemma 1 (Aggregation Lemma): Consider a sequence of random variables $\{Y^n\}_{n=1}^\infty$ where Y^n takes values in A^n . Suppose there exists a sequence of sets $\{J^{(n)} : J^{(n)} \subset A^n\}$ (note that $J^{(n)}$ need not be a cartesian product) that satisfies the following: i)

$$\lim_{n \rightarrow \infty} P_{Y^n}(J^{(n)}) = 1 \quad (4)$$

and ii) there exists $\alpha > 0$ such that for all sufficiently large n and for all $y^n \in J^{(n)}$

$$P_{Y^n}(y^n) \leq 2^{-n\alpha}. \quad (5)$$

Then for every $0 < \theta \leq \frac{2\alpha}{3}$, we can find a sequence of deterministic mappings

$$\{\phi_n : A^n \mapsto \{0, 1\}^r, r = \lfloor n\alpha - n\theta \rfloor\}$$

such that

$$\lim_{n \rightarrow \infty} \Delta(\phi_n(Y^n), B^r) = 0 \quad (6)$$

where B^r has the equiprobable distribution on $\{0, 1\}^r$ and (6) holds in both variational distance and normalized divergence.

Proof: Let us denote $\epsilon_n = 1 - P_{Y^n}(J^{(n)})$. Expression (4) implies that $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Fix some $0 < \theta \leq \frac{2\alpha}{3}$. We will consider n sufficiently large so that (5) is true and $\epsilon_n < \frac{1}{2}$. The second condition is just to make sure that we are not considering sets with too few elements.

Let us first observe that $|J^{(n)}| \geq (1 - \epsilon_n)2^{n\alpha}$ which is a consequence of (5). We will construct a sequence of deterministic mappings

$$\phi_n : A^n \mapsto \{0, 1\}^r, \quad \text{with } r = \lfloor n\alpha - n\theta \rfloor$$

and

$$\Delta(\phi_n(Y^n), B^r) \rightarrow 0 \quad (7)$$

where (7) holds according to both the variational distance and the normalized divergence measures.

We need to *aggregate* the probability masses of P_{Y^n} into 2^r bins, such that the probability of each bin is roughly 2^{-r} . We will be able to do so even though we have not put any conditions on the probabilistic structure of Y beyond (5). The construction is as follows: All but one bin will be filled exclusively with elements of the set $J^{(n)}$. Let $s = n\alpha - \frac{n\theta}{2}$. Order the elements of $J^{(n)}$ arbitrarily and place elements in the first bin $B_n(1)$ until its probability satisfies

$$P_{Y^n}(B_n(1)) \geq 2^{-r} - 2^{-s}. \quad (8)$$

At that point, the construction of $B_n(1)$ is complete. Clearly, due to (5),

$$P_{Y^n}(B_n(1)) < 2^{-r}. \quad (9)$$

To see this, note that if (9) were not true, we could remove an element from the bin and still (8) would hold, resulting in a contradiction.

Continue this procedure with $B_n(2), \dots, B_n(K)$ where $K \leq 2^r - 1$. Note that the probability of each of those bins satisfies (8) and (9) as well. How long can we continue this procedure? We stop either because there are not enough

elements in $J^{(n)}$ to make up a new bin that satisfies (8) or because we hit the limit $K = 2^r - 1$. Due to (9), $J^{(n)}$ will not be exhausted until there are at least $\lfloor (1 - \epsilon_n)2^r \rfloor$ bins, implying $K \geq (1 - \epsilon_n)2^r - 1$.

Finally define

$$\begin{aligned} B_n(K+1) &= A^n - \bigcup_{i=1}^K B_n(i) \\ &= (A^n - J^{(n)}) \cup (J^{(n)} - \bigcup_{i=1}^K B_n(i)). \end{aligned}$$

Now we upper-bound the probability of the bin $B_n(K+1)$. The contribution due to the first part $(A^n - J^{(n)})$ is ϵ_n . The contribution due to the second part can be upper-bounded from (8) as $1 - \epsilon_n - K(2^{-r} - 2^{-s})$ and this quantity, using $K \geq (1 - \epsilon_n)2^r - 1$, can be further upper-bounded by $(1 - \epsilon_n)2^{r-s} + 2^{-r} - 2^{-s}$. Hence

$$\begin{aligned} P_{Y^n}(B_n(K+1)) &\leq \epsilon_n + (1 - \epsilon_n)2^{r-s} + 2^{-r} - 2^{-s} \\ &\leq \epsilon_n + 2 \times 2^{r-s} \end{aligned} \quad (10)$$

where the last inequality follows from the fact that $2^{-r} \leq 2^{r-s}$ which is a consequence of the assumption $\theta \leq \frac{2\alpha}{3}$.

If $K < 2^r - 1$, for the sake of completeness, we define $B_n(K+2), \dots, B_n(2^r)$ to be empty bins.

The deterministic mapping $\phi_n : A^n \mapsto \{0, 1\}^r$ simply assigns a unique string of r bits to each bin.

Now let us evaluate the distance of $\phi_n(Y^n)$ from the equiprobable distribution on $\{0, 1\}^r$. In variational distance we have,

$$\begin{aligned} d(\phi_n(Y^n), B^r) &= \sum_{i=1}^{2^r} |2^{-r} - P_{Y^n}(B_n(i))| \\ &\leq \sum_{i=1}^K |2^{-r} - P_{Y^n}(B_n(i))| \\ &\quad + |2^{-r} - P_{Y^n}(B_n(K+1))| \\ &\quad + \sum_{i=K+2}^{2^r} 2^{-r}. \end{aligned} \quad (11)$$

We bound each of the three terms of (11) separately. The first term is upper-bounded by $K2^{-s}$ using (8) and (9) and this is less than or equal to 2^{r-s} . The second term is upper-bounded by $2^{-r} + \epsilon_n + 2 \times 2^{r-s}$ using (10). The third term is simply $1 - (K+1)2^{-r}$ which is less than ϵ_n .

Putting the three bounds together and taking n sufficiently large, we conclude that

$$\begin{aligned} d(\phi_n(Y^n), B^r) &\leq 3 \times 2^{r-s} + 2^{-r} + 2\epsilon_n \\ &\leq 2^{-\frac{2\theta}{3}} + 2\epsilon_n \end{aligned} \quad (12)$$

which goes to zero with n .

Now the convergence in normalized divergence is proved as follows: We will first lower-bound $\frac{1}{r}H(\phi_n(Y^n))$.

$$\begin{aligned} \frac{1}{r}H(\phi_n(Y^n)) &> \frac{1}{r} \sum_{i=1}^K P_{Y^n}(B_n(i)) \log \frac{1}{P_{Y^n}(B_n(i))} \\ &\geq K(2^{-r} - 2^{-s}) \\ &\geq (1 - \epsilon_n)(1 - 2^{r-s}) - (2^{-r} - 2^{-s}) \end{aligned}$$

where we have used the fact that, for $1 \leq i \leq K$

$$P_{Y^n}(B_n(i)) \log \frac{1}{P_{Y^n}(B_n(i))} \geq r(2^{-r} - 2^{-s})$$

which follows from (8) and (9).

Hence

$$\begin{aligned} \frac{1}{r}D(\phi_n(Y^n)||B^r) &= 1 - \frac{1}{r}H(\phi_n(Y^n)) \\ &\leq \epsilon_n + 2^{r-s} + 2^{-r} - 2^{-s} \\ &\leq \epsilon_n + 2^{-\frac{n\theta}{4}} \end{aligned} \quad (13)$$

which goes to zero with n .

Hence we have

$$\Delta(\phi_n(Y^n), B^r) \leq 2\epsilon_n + 2^{-(n\theta/4)} \quad (14)$$

which is true for all sufficiently large n . This incorporates both (12) and (13).

Remark: We would also be interested in the special case of $\epsilon_n = 0$ for all n . Then the upper bound (14) depends only on θ and is uniform in α provided $\alpha \geq \frac{3\theta}{2}$. This fact will be useful later on.

Now we turn our attention to the next lemma which we call the continuity lemma. This lemma turns out to be the key ingredient of our converse results in the subsequent sections. We state the lemma in a unified way but the proofs for the two distance measures, d-bar distance and normalized divergence, are quite different and are given separately. Once we have the continuity lemma, the proof of the actual converse statements are identical.

Lemma 2 (Continuity Lemma): Let $\{B^r\}_{r=1}^\infty$ be a sequence of random variables such that B^r has the equiprobable distribution on $\{0, 1\}^r$. Let $\{\tilde{B}^r\}_{r=1}^\infty$ be a sequence of random variables also taking values in $\{0, 1\}^r$ that satisfies,

$$\lim_{r \rightarrow \infty} \Delta(\tilde{B}^r, B^r) = 0 \quad (15)$$

where the distance measure Δ is the d-bar distance (resp., the normalized divergence).

Then the process \tilde{B} satisfies $\underline{H}(\tilde{B}) = 1$.

We will first prove the statement for the d-bar distance.

Proof (d-Bar Distance): We will argue by contradiction. Suppose \tilde{B} satisfies (15) and yet there exists a $\lambda > 0$ such that $\underline{H}(\tilde{B}) < 1 - \lambda$. (Since \tilde{B}^r takes values on $\{0, 1\}^r$ it trivially satisfies $\underline{H}(\tilde{B}) \leq 1$.)

Let us first choose a sequence of sets $G_r \subset \{0, 1\}^r$ such that

$$G_r = \left\{ \tilde{b}^r : \frac{1}{r} \log \frac{1}{P_{\tilde{B}^r}(\tilde{b}^r)} \leq 1 - \frac{\lambda}{2} \right\}. \quad (16)$$

By definition of inf-entropy rate, $\{G_r\}$ satisfies

$$P_{\tilde{B}^r}(G_r) \geq \alpha(\lambda) \quad (17)$$

infinitely often in r , for some $\alpha(\lambda) > 0$. We will focus on those r for which (17) is true. From now on we will write α instead of $\alpha(\lambda)$.

The size of G_r satisfies

$$|G_r| \leq 2^{r(1-\frac{\lambda}{2})}. \quad (18)$$

Now we choose a $\delta > 0$ such that

$$\sqrt{\delta} \leq \min \left\{ \sqrt{\frac{\alpha}{2}}, \frac{\lambda}{4}, \frac{1}{16} \right\}.$$

Since the process \tilde{B} converges to B in d-bar distance, let us choose r so large that $\bar{d}_r(\tilde{B}^r, B^r) \leq \delta^2$. Now define a sequence of sets $J_r \subset \{0, 1\}^r \times \{0, 1\}^r$ such that

$$J_r = \{(\tilde{b}^r, b^r) : d_H(\tilde{b}^r, b^r) \leq r\delta\} \quad (19)$$

where d_H refers to the Hamming distance. Clearly, due to the fact that the $\bar{d}_r(\tilde{B}^r, B^r) \leq \delta^2$, there exists a joint distribution $P_{\tilde{B}^r B^r}$ such that

$$P_{\tilde{B}^r B^r}(J_r) \geq 1 - \delta. \quad (20)$$

In the above, we have used the fact that the infimum in the definition of \bar{d} -distance is actually achievable.

Now we define another set $D_r \subset \{0, 1\}^r$ which is a superset of G_r

$$D_r = \{b^r \in \{0, 1\}^r : \exists \tilde{b}^r \in G_r, \text{ such that } d_H(\tilde{b}^r, b^r) \leq r\delta\}. \quad (21)$$

We will show that $P_{B^r}(D_r) \geq \frac{\alpha}{2}$. Note that the set $G_r \times D_r^c$ is a subset of J_r^c , where the superscript c denotes complementation. Hence

$$P_{\tilde{B}^r B^r}(G_r \times D_r^c) \leq P_{\tilde{B}^r B^r}(J_r^c) \leq \delta$$

which implies

$$P_{\tilde{B}^r}(G_r^c) + P_{B^r}(D_r) \geq 1 - \delta \quad (22)$$

and hence from (17)

$$P_{B^r}(D_r) \geq \alpha - \delta \geq \frac{\alpha}{2} \quad (23)$$

as needed.

This immediately gives a lower bound on the size of D_r

$$|D_r| \geq \frac{\alpha}{2} 2^r. \quad (24)$$

From the definition of D_r it is clear that an element of D_r is obtained by flipping the components of some element $\tilde{b}^r \in G_r$ in at most $r\delta$ places. There are

$$\sum_{k \leq r\delta} \binom{r}{k}$$

such elements that arise from each element of G_r . Hence D_r satisfies

$$|D_r| \leq |G_r| \sum_{k \leq r\delta} \binom{r}{k}. \quad (25)$$

The following is true if we take r to be suitably large:

$$\sum_{k \leq r\delta} \binom{r}{k} \leq 2^{3rk(\delta)} \leq 2^{r\sqrt{\delta}}$$

where we have used the fact that

$$h(\delta) = \delta \log \frac{1}{\delta} + (1-\delta) \log \frac{1}{1-\delta} \leq \frac{1}{3}\sqrt{\delta}, \quad \text{for } \sqrt{\delta} \leq \frac{1}{16}.$$

Hence we can say that

$$|D_r| \leq |G_r| 2^{r\sqrt{\delta}} \leq 2^{r(1-\frac{1}{3}+\sqrt{\delta})} \leq 2^{r(1-\frac{1}{4})}. \quad (26)$$

For large enough r , (26) contradicts (24), establishing the result.

Proof (Divergence): We will now prove the continuity lemma for the case of normalized divergence. Assume that

$$\lim_{r \rightarrow \infty} \frac{1}{r} D(\tilde{B}^r || B^r) = 0. \quad (27)$$

We will show that this implies $\underline{H}(\tilde{B}) = 1$. We can rewrite (27) as

$$\lim_{r \rightarrow \infty} \frac{1}{r} H(\tilde{B}^r) = \lim_{r \rightarrow \infty} \sum_{\tilde{b}^r} \frac{1}{r} P_{\tilde{B}^r}(\tilde{b}^r) \log \frac{1}{P_{\tilde{B}^r}(\tilde{b}^r)} = 1.$$

Note that for all r we trivially have $\frac{1}{r} H(\tilde{B}^r) \leq 1$.

Let us fix an arbitrary $\epsilon > 0$. Let

$$E_r = \left\{ \tilde{b}^r : \frac{1}{r} \log \frac{1}{P_{\tilde{B}^r}(\tilde{b}^r)} \leq 1 - \epsilon \right\}.$$

Denote $\beta_r = P_{\tilde{B}^r}(E_r)$. We want to show that β_r goes to zero with increasing r . We will concentrate on those r for which $\beta_r > 0$. We write the normalized entropy of \tilde{B}^r as

$$\begin{aligned} \frac{1}{r} H(\tilde{B}^r) &= \sum_{\tilde{b}^r \in E_r} \frac{1}{r} P_{\tilde{B}^r}(\tilde{b}^r) \log \frac{1}{P_{\tilde{B}^r}(\tilde{b}^r)} \\ &\quad + \sum_{\tilde{b}^r \in E_r^c} \frac{1}{r} P_{\tilde{B}^r}(\tilde{b}^r) \log \frac{1}{P_{\tilde{B}^r}(\tilde{b}^r)}. \end{aligned} \quad (28)$$

The first term of (28) can be easily upper-bounded by $(1-\epsilon)\beta_r$.

The second term in (28) is equal to

$$\begin{aligned} (1-\beta_r) \frac{1}{r} H(\tilde{B}^r | \tilde{B}^r \notin E_r) - \frac{1}{r} \log(1-\beta_r) \\ \leq (1-\beta_r) - \frac{1}{r} \log(1-\beta_r). \end{aligned}$$

Adding the bounds for the two terms, we get an overall upper bound for the normalized entropy as

$$1 - \epsilon\beta_r - \frac{1}{r} \log(1-\beta_r).$$

Now if $\beta_r > \beta > 0$ infinitely often in r , the bound will imply that the normalized entropy is bounded away from 1—which is a contradiction. Hence β_r tends to 0. From the definition of β_r , we conclude that $\underline{H}(\tilde{B}) \geq 1 - \epsilon$. Since the choice of $\epsilon > 0$ is arbitrary, we get the desired result.

Corollary: Under the conditions of Lemma 2, the following holds:

$$\lim_{r \rightarrow \infty} \frac{1}{r} H(\tilde{B}^r) = 1.$$

Proof: For normalized divergence, this is trivially true. For the case of d-bar distance, it is a simple consequence of the fact that for any finite alphabet process \tilde{B} if $\underline{H}(\tilde{B}) = \overline{H}(\tilde{B})$, then $\lim_{r \rightarrow \infty} \frac{1}{r} H(\tilde{B}^r)$ exists and is also equal to the above two quantities ([1, Lemma 1]). Then, noting that $\underline{H}(\tilde{B}) \leq \overline{H}(\tilde{B}) \leq 1$, because the size of the alphabet of \tilde{B} is 2, and the corollary readily follows.

V. FIXED-LENGTH RANDOM NUMBER GENERATION

In this section we prove Theorem 1 which we repeat here for convenience.

Theorem 1: For any discrete source Z

$$U_v(Z) = U_b(Z) = U_d(Z) = \underline{H}(Z).$$

To show this result, we first recall [13] that the variational distance and the d-bar distance satisfy, for all P_{B^r} and $P_{\tilde{B}^r}$.

$$\bar{d}_r(P_{B^r}, P_{\tilde{B}^r}) \leq \frac{1}{2} d(P_{B^r}, P_{\tilde{B}^r}). \quad (29)$$

It follows from (29) that

$$U_v(Z) \leq U_b(Z). \quad (30)$$

Thus it will be enough to show

$$1) \quad U_v(Z) \geq \underline{H}(Z)$$

$$2) \quad U_b(Z) \leq \underline{H}(Z)$$

$$3) \quad U_d(Z) = \underline{H}(Z).$$

Note that the normalized divergence measure neither dominates nor is dominated by the other two distance measures.

A. The Direct Part

In this section we prove the achievability part of Theorem 1.

Lemma 3: Every discrete source Z satisfies

$$a) \quad U_v(Z) \geq \underline{H}(Z) \quad (31)$$

$$b) \quad U_d(Z) \geq \underline{H}(Z). \quad (32)$$

Proof: We might as well assume that $\underline{H}(\mathbf{Z}) > 0$, otherwise there is nothing to prove. Fix $0 < \gamma < \frac{1}{2}\underline{H}(\mathbf{Z})$. Let $r = \lfloor n(\underline{H}(\mathbf{Z}) - \gamma) \rfloor$. We need to aggregate the probability masses of P_{Z^n} into 2^r bins, such that the probability of each bin is roughly 2^{-r} . This is exactly the problem we have dealt with in the aggregation lemma. We denote $\alpha = \underline{H}(\mathbf{Z}) - \frac{\gamma}{2}$ and define the set

$$J^{(n)} = \{z^n \in A^n : P_{Z^n}(z^n) \leq 2^{-n\alpha}\}.$$

By definition of inf-entropy rate, $P_{Z^n}(J^{(n)}) \rightarrow 1$ as $n \rightarrow \infty$. Now we invoke the aggregation lemma to claim that there exists a sequence of deterministic mappings $\phi_n : A^n \rightarrow \{0, 1\}^r$, with

$$r = \lfloor n\alpha - \frac{n\gamma}{2} \rfloor \tag{33}$$

such that

$$\Delta(\phi_n(Z^n), B^r) \rightarrow 0 \tag{34}$$

where (34) holds according to both the variational distance and the normalized divergence measures. Since the choice of γ is arbitrary, it follows that $\underline{H}(\mathbf{Z})$ is an achievable intrinsic randomness rate according to both measures and the proof of the direct part is complete.

B. The Converse Part

In this section we show the converse results for the fixed-length random number generation problem. We will show that $U_b(\mathbf{Z}) \leq \underline{H}(\mathbf{Z})$ and $U_d(\mathbf{Z}) \leq \underline{H}(\mathbf{Z})$. Coupled with the direct part we get the main result

$$U_v(\mathbf{Z}) = U_b(\mathbf{Z}) = U_d(\mathbf{Z}) = \underline{H}(\mathbf{Z}).$$

Let us assume that $\underline{H}(\mathbf{Z}) < \infty$, otherwise, there is nothing to prove. Now we will contradict the converse statement and assume that there exists a $\gamma > 0$ such that $\underline{H}(\mathbf{Z}) + \gamma$ is an achievable IR rate in the sense of \bar{d} -distance (resp., normalized divergence).

This implies that there exists a sequence of deterministic mappings $\{\phi_n : A^n \rightarrow \{0, 1\}^r\}$ such that for all n

$$\frac{r}{n} \geq \underline{H}(\mathbf{Z}) + \frac{\gamma}{2} \tag{35}$$

and

$$\lim_{n \rightarrow \infty} \Delta(\phi_n(Z^n), B^r) = 0$$

where Δ is the \bar{d} -distance (resp., the normalized divergence).

It follows from the continuity lemma proved earlier that the process $\{\bar{B}^r = \phi_n(Z^n)\}$ has inf-entropy rate of 1.

Since $\phi_n : A^n \mapsto \{0, 1\}^r$ is a deterministic function

$$P_{Z^n}(z^n) \leq P_{\bar{B}^r}(\phi_n(z^n)).$$

Thus for all $z^n \in A^n$

$$\begin{aligned} \frac{1}{n} \log \frac{1}{P_{Z^n}(z^n)} &\geq \frac{r}{nr} \log \frac{1}{P_{\bar{B}^r}(\phi_n(z^n))} \\ &\geq \left[\underline{H}(\mathbf{Z}) + \frac{\gamma}{2} \right] \left[\frac{1}{r} \log \frac{1}{P_{\bar{B}^r}(\phi_n(z^n))} \right]. \end{aligned}$$

If, furthermore, z^n belongs to the subset

$$F_n = \left\{ z^n : \frac{1}{r} \log \frac{1}{P_{\bar{B}^r}(\phi_n(z^n))} \geq 1 - \epsilon \right\}$$

with

$$0 < \epsilon < \frac{\gamma}{4(\underline{H}(\mathbf{Z}) + \frac{\gamma}{2})}$$

then

$$\begin{aligned} \frac{1}{n} \log \frac{1}{P_{Z^n}(z^n)} &\geq \left[\underline{H}(\mathbf{Z}) + \frac{\gamma}{2} \right] [1 - \epsilon] \\ &\geq \underline{H}(\mathbf{Z}) + \frac{\gamma}{4}. \end{aligned} \tag{36}$$

Clearly, $P_{Z^n}(F_n) \rightarrow 1$ as $n \rightarrow \infty$ due to the fact that $\underline{H}(\bar{\mathbf{B}}) = 1$. Thus (36) contradicts the definition if $\underline{H}(\mathbf{Z})$, thereby establishing the converse result.

VI. VARIABLE-LENGTH RANDOM NUMBER GENERATION

In this section we handle the variable-length random generation. Note that the proof of our direct part relies on the finite alphabet assumption, unlike the result in the fixed-length case. We repeat Theorem 2 here for convenience.

Theorem 2: For every finite alphabet source \mathbf{Z} :

$$V_r(\mathbf{Z}) = V_b(\mathbf{Z}) = V_d(\mathbf{Z}) = \liminf_{n \rightarrow \infty} \frac{1}{n} H(Z^n).$$

We will split the proof into direct and converse parts:

A. Direct Part

We will show that for arbitrary $\epsilon > 0$ and $\delta > 0$, $H - \delta$ is an ϵ -achievable VLIR rate where

$$H = \liminf_{n \rightarrow \infty} \frac{1}{n} H(Z^n).$$

We might as well assume that $H > 0$, otherwise there is nothing to prove.

The main idea of the proof is to partition A^n into sets such that within a given set we have elements with roughly similar probability masses. We find the total probability of each set in the partition and only keep those sets whose probability is not too low. Within each of these sets, we apply the aggregation lemma to synthesize a close-to-equiprobable distribution.

Now we proceed with the formal proof. Let $\theta > 0$ be such that

$$\theta \leq \min\left\{ \frac{\delta}{20}, \frac{H}{4}, \frac{1}{5} \right\}.$$

From now on we will focus on n large enough that

$$\frac{1}{n} H(Z^n) \geq H - \frac{\delta}{2}.$$

Let us divide the interval $[3\theta, \log |A| + \theta]$ into intervals of length θ :

$$[(\ell + 2)\theta, (\ell + 3)\theta), \quad \ell = 1, 2, \dots, L$$

where

$$L \leq \frac{\log |A|}{\theta} < \infty.$$

Note that L is finite, independent of n , and depends only on θ .

Let us define a partition of A^n as follows:

$$G_\ell^{(n)} = \left\{ z^n : \frac{1}{n} \log \frac{1}{P_{Z^n}(z^n)} \in [(\ell+2)\theta, (\ell+3)\theta] \right\} \quad (37)$$

along with the set

$$E_n = A^n - \bigcup_{\ell=1}^L G_\ell^{(n)}.$$

Now let

$$K_n = \left\{ \ell \in \{1, 2, \dots, L\} : P_{Z^n}(G_\ell^{(n)}) > \frac{1}{L^2} \right\}. \quad (38)$$

Finally, we set

$$G_0^{(n)} = A^n - \bigcup_{\ell \in K_n} G_\ell^{(n)}.$$

Clearly, for all $\ell \in \{1, 2, \dots, L\}$

$$P_{Z^n}(z^n) \leq 2^{-n(\ell+2)\theta}, \quad \text{for all } z^n \in G_\ell^{(n)} \quad (39)$$

and in addition, if $\ell \in K_n$,

$$|G_\ell^{(n)}| \geq \frac{1}{L^2} 2^{n(\ell-2)\theta}. \quad (40)$$

For each $\ell \in K_n$, with $r = \lfloor n\ell\theta \rfloor$, we define the conditional distribution

$$P_{Z_r^n}(z^n) = \begin{cases} \frac{P_{Z^n}(z^n)}{P_{Z^n}(G_\ell^{(n)})}, & \text{if } z^n \in G_\ell^{(n)} \\ 0, & \text{otherwise} \end{cases}$$

where Z_r^n is Z^n restricted to $G_\ell^{(n)}$.

Clearly, for every $\ell \in K_n$

$$P_{Z_r^n}(z^n) \leq L^2 2^{-n(\ell+2)\theta} \leq 2^{-n(\ell+1)\theta} \quad (41)$$

where the last inequality is true if we take n sufficiently large.

The sets $\{G_\ell^{(n)}\}_{\ell \in K_n}$ are precisely the sets for which we will output a nonempty bit string. The length of the bit string is $r = \lfloor n\ell\theta \rfloor$. In order to ensure that the bit string we output is almost equiprobable, we perform the aggregation procedure on each of the sets $G_\ell^{(n)}$ separately. This would mean that we would partition each of the $G_\ell^{(n)}$, $\ell \in K_n$ into 2^r ($r = \lfloor n\ell\theta \rfloor$) bins.

Now recall the upper bound (14) on the distance measures derived in Lemma 1. We apply this bound to each conditional distribution $P_{Z_r^n}$ where $\ell \in K_n$ when n is sufficiently large. In the notation of Lemma 1, $\alpha = (\ell+1)\theta \geq 2\theta$. Using the lemma, we can claim that if n is large enough, there exists a deterministic transformation

$$\phi_{n,r} : G_\ell^{(n)} \mapsto \{0, 1\}^r, \quad r = \lfloor n\ell\theta \rfloor$$

such that

$$\Delta(\phi_{n,r}(Z_r^n), B^r) \leq 2^{-\frac{n\theta}{L}}. \quad (42)$$

(In the present case we can set $\epsilon_n = 0$ in (14).) By the remark following Lemma 1, the upper bound is *uniform* for all values of $\ell \in K_n$. Hence, we can take the maximum over ℓ on the left-hand side. Now letting n go to ∞ , we see that we have satisfied the second condition needed in the definition of achievability.

To be notationally consistent with Definition 1, we set

$$I_n = \{r = \lfloor n\ell\theta \rfloor\}_{\ell \in K_n} \cup \{0\}$$

and we can just reindex the partition according to r instead of ℓ .

Now we need to compute the average number of bits generated by our scheme. We will need the following bounds:

$$S_1 = \frac{1}{n} \sum_{z^n : \frac{1}{n} \log \frac{1}{P_{Z^n}(z^n)} < 3\theta} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)} < 3\theta$$

and

$$S_2 = \frac{1}{n} \sum_{z^n : \frac{1}{n} \log \frac{1}{P_{Z^n}(z^n)} \geq \log |A| + \theta} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)} \leq (\log |A| + \theta) 2^{-n\theta}.$$

Hence

$$\begin{aligned} \frac{1}{n} \sum_{z^n \in E_n} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)} &= S_1 + S_2 \\ &< 3\theta + (\log |A| + \theta) 2^{-n\theta} \\ &\leq 4\theta \end{aligned} \quad (43)$$

where (43) is true for large n .

Let us write

$$\begin{aligned} \frac{1}{n} H(Z^n) &= \frac{1}{n} \sum_{z^n \in A^n} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)} \\ &= \frac{1}{n} \sum_{\ell=1}^L \sum_{z^n \in G_\ell^{(n)}} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)} \\ &\quad + \frac{1}{n} \sum_{z^n \in E_n} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)} \\ &< \sum_{\ell=1}^L P_{Z^n}(G_\ell^{(n)}) (\ell+3)\theta + 4\theta. \end{aligned} \quad (44)$$

Here the inequality (44) is due to (37) and (43).

We further bound the first term on the right-hand side of (44) as

$$\begin{aligned} \sum_{\ell=1}^L P_{Z^n}(G_\ell^{(n)}) (\ell+3)\theta &= \sum_{\ell \in K_n} P_{Z^n}(G_\ell^{(n)}) (\ell+3)\theta \\ &\quad + \sum_{\ell \notin K_n} P_{Z^n}(G_\ell^{(n)}) (\ell+3)\theta \\ &\leq \frac{1}{n} \sum_{\ell \in K_n} (r+1) P_{Z^n}(G_\ell^{(n)}) \\ &\quad + 3\theta + \frac{\theta}{L^2} \sum_{\ell \notin K_n} (\ell+3) \end{aligned} \quad (45)$$

$$\begin{aligned} &\leq \frac{1}{n} \sum_{\ell \in K_n} r P_{Z^n}(G_\ell^{(n)}) + \frac{1}{n} \\ &\quad + 3\theta + \frac{\theta(L+3)(L+4)}{2L^2} \\ &\leq \frac{1}{n} \sum_{\ell \in K_n} r P_{Z^n}(G_\ell^{(n)}) + 6\theta \end{aligned} \quad (46)$$

where (45) is due to $r = \lfloor n\ell\theta \rfloor$ and (46) is true for large n .

Putting together (44) and (46) we get

$$\frac{1}{n} \sum_{t \in K_n} r P_{Z^n}(G_t^{(n)}) \geq \frac{1}{n} H(Z^n) - 10\theta \geq H - \delta$$

as desired. This concludes the proof of the direct part.

B. Converse Part

We show that $V_b(Z) \leq H$ and $V_d(Z) \leq H$. Both proofs invoke the corresponding continuity lemma and so we develop them in parallel, using Δ to denote d-bar distance (resp., normalized divergence).

By way of contradiction, let us assume that $H + 2\delta$ is achievable for some $\delta > 0$. By definition of achievability there exist a sequence of sets I_n of nonnegative integers, a sequence of partitions

$$A^n = \bigcup_{r \in I_n} J_r^{(n)}$$

and a sequence of deterministic mappings

$$\{\phi_{n,r} : J_r^{(n)} \mapsto \{0, 1\}^r\}_{r \in I_n}$$

such that the following conditions are true:

(C1) For all n

$$\frac{1}{n} \sum_{r \in I_n} r P_{Z^n}(J_r^{(n)}) > H + \delta \tag{47}$$

and

(C2) For all sufficiently large n

$$\sup_{r \in I_n} \Delta(\phi_{n,r}(Z_r^n), B^r) < \epsilon$$

where B^r has the equiprobable distribution on $\{0, 1\}^r$ and Z_r^n is Z^n restricted to $J_r^{(n)}$.

We consider the set

$$I'_n = I_n \cap \left\{ \frac{n\delta}{5}, \dots, \lceil n \log |A| \rceil \right\}.$$

Clearly

$$\frac{1}{n} \sum_{r \leq \frac{n\delta}{5}} r P_{Z^n}(J_r^{(n)}) \leq \frac{\delta}{5}. \tag{48}$$

We want to apply the continuity lemma suitably. The trick is to define a "worst" sequence of partitions and apply the continuity lemma to this sequence and then claim that the result holds for any sequence of partitions. To be precise, we define a sequence of conditional distributions as follows: first let

$$r_n = \arg \min_{r \in I'_n} \frac{1}{r} H(\phi_{n,r}(Z_r^n)).$$

Clearly, $r_n \geq \frac{n\delta}{5}$. The sequence of conditional distributions is defined as $\{P_{Z_r^n}\}_{n=1}^\infty$ where $P_{Z_r^n}$ is the distribution of the random variable Z_r^n .

Now (C2) implies that

$$\lim_{n \rightarrow \infty} \Delta(\phi_{n,r_n}(Z_{r_n}^n), B^{r_n}) = 0. \tag{49}$$

Let us consider the sequence of random variables $\{\tilde{B}^{r_n}\}$ whose distributions are defined as

$$P_{\tilde{B}^{r_n}}(\tilde{b}^{r_n}) = \frac{P_{Z^n}(\phi_{n,r_n}(Z_{r_n}^n) = \tilde{b}^{r_n})}{P_{Z^n}(Z^n \in J_{r_n}^{(n)})}.$$

To complete the definition, set $\tilde{B}^k = B^k$ for all positive integers $k \notin \{r_n\}_{n=1}^\infty$. Thus

$$\lim_{k \rightarrow \infty} \Delta(\tilde{B}^k, B^k) = 0.$$

Now we can invoke the corollary to Lemma 2, and apply it to $\{\tilde{B}^k\}_{k=1}^\infty$ and conclude that

$$\lim_{k \rightarrow \infty} \frac{1}{k} H(\tilde{B}^k) = \lim_{n \rightarrow \infty} \frac{1}{r_n} H(\phi_{n,r_n}(Z_{r_n}^n)) = 1. \tag{50}$$

where we have used the fact that if a sequence converges, then any subsequence also converges to the same limit. Therefore, choosing

$$0 < \theta < \frac{\delta}{5H + 4\delta}$$

provided n is chosen large enough

$$\min_{r \in I'_n} \frac{1}{r} H(\phi_{n,r}(Z_r^n)) \geq 1 - \theta. \tag{51}$$

Consider the following string of inequalities:

$$\begin{aligned} \frac{1}{n} H(Z^n) &\geq \frac{1}{n} \sum_{r \in I_n} H(Z_r^n) P_{Z^n}(J_r^{(n)}) \\ &\geq \frac{1}{n} \sum_{r \in I_n} H(\phi_{n,r}(Z_r^n)) P_{Z^n}(J_r^{(n)}) \\ &\geq \frac{1}{n} \sum_{r \in I'_n} r \left[\frac{1}{r} H(\phi_{n,r}(Z_r^n)) \right] P_{Z^n}(J_r^{(n)}) \\ &\geq \frac{1}{n} \sum_{r \in I'_n} r(1 - \theta) P_{Z^n}(J_r^{(n)}) \end{aligned} \tag{52}$$

$$\geq (1 - \theta)(H + \frac{4\delta}{5}) \tag{53}$$

$$\geq H + \frac{3\delta}{5} \tag{54}$$

where (52) follows from (51) and (53) is due to (47) and (48). Since (54) is true for all large n , we obtain a contradiction

of the fact that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} H(Z^n) = H.$$

This proves the converse statement.

REFERENCES

- [1] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752-772, May 1993.
- [2] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147-1157, July 1994.
- [3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.
- [4] J. von Neumann, "Various techniques used in connection with random digits," *Nat. Bur. Stand. Appl. Math. Ser.*, vol. 12, pp. 36-38, 1951. Reprinted in the *Collected Works of von Neumann*, vol. 5.
- [5] P. Elias, "The efficient construction of an unbiased random sequence," *Ann. Math. Statist.*, vol. 43, pp. 865-870, 1972.
- [6] Y. Peres, "Iterating non Neumann's procedure for extracting random bits," *Ann. Statist.*, vol. 20, no. 1, pp. 590-597, 1992.
- [7] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423. Reprinted in book form by the Univ. of Illinois Press, 1949.
- [8] R. M. Gray and L. D. Davisson, Eds., *Ergodic and Information Theory* (benchmark papers in Electrical Engineering and Computer Science). Dowden, Hutchinson and Ross, 1977.
- [9] Ja. G. Sinai, "Weak isomorphism of transformations with an invariant measure," *Sov. Math.-Dokl.*, vol. 3, pp. 1725-1729, 1962 (translated by the Amer. Math. Soc.).
- [10] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 44-54, Jan. 1995.
- [11] D. S. Ornstein, "Bernoulli shifts with the same entropy are isomorphic," *Adv. Math.*, vol. 4, pp. 337-352, 1970.
- [12] ———, "An application of ergodic theory to probability theory," *Ann. Probab.*, vol. 1, pp. 43-65, 1973.
- [13] R. M. Gray, D. L. Neuhoff, and P. C. Shields, "A generalization of Ornstein's distance with applications to information theory," *Ann. Probab.*, vol. III, pp. 515-328, Apr. 1975.
- [14] R. M. Gray and D. S. Ornstein, "Block coding for discrete stationary d-continuous noisy channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 292-306, 1979.
- [15] D. L. Neuhoff and P. C. Shields, "Channel entropy and primitive approximation," *Ann. Probab.*, vol. 10, no. 1, pp. 188-198, 1982.
- [16] D. Zuckerman, "Simulating BPP using a general weak random source," in *Proc. 32nd IEEE Symp. on Foundations of Computer Science*, 1991, pp. 79-89.
- [17] Y. Steinberg and S. Verdú, "Course approximation of source statistics and rate distortion theory," to be published in *IEEE Trans. Inform. Theory*.
- [18] M. Blum, "Independent unbiased coin flips from a correlated biased source: A finite state Markov chain," in *Proc. 25th IEEE Symp. on Foundations of Computer Science*. Silver Spring, MD: IEEE Computer Soc. Press, 1984, pp. 425-433.
- [19] S. Vembu, "Information theory without ergodicity assumptions: Some new results," Ph.D. dissertation, Dept. Elec. Eng., Princeton Univ., Princeton, NJ, Mar. 1994.