

Source Codes as Random Number Generators¹

Karthik Visweswariah Sanjeev Kulkarni Sergio Verdú

Department of Electrical Engineering, Princeton University, Princeton, NJ, U.S.A. 08544

Abstract — The use of optimal variable-length source codes as optimal random bit generators is investigated. We show in what sense source codes can be considered to be random bit generators.

I. INTRODUCTION

The problem of random number generation involves deterministically transforming a random (non-ideal) source so that the output is a sequence of fair coin flips. Constructive methods to generate random bits from finite order Markov sources have been reported in [2], [3] among others. It was shown in [1] that approximately random bits can be generated from a known random source at its entropy rate. Approximation was considered in [1] in the sense of vanishing variational distance, normalized divergence and other measures. The practically important problem of constructing *universal* optimal random number generators without knowledge of the non-ideal source distributions has remained open.

Since optimal source codes remove all redundancy from a source we would expect them to be prime candidates for random number generation. However, our first result is negative: even in the “favorable” setting of a memoryless source with dyadic distribution, a Huffman code fails to supply almost fair coin flips in the sense of vanishing variational distance. We show in what sense optimal source codes in general and Lempel-Ziv codes in particular are optimal random bit generators. The Lempel-Ziv algorithm has been shown in [4] to be optimal for testing whether or not a source produces unbiased coin flips; the present paper establishes yet another use of the Lempel-Ziv algorithm: optimal universal random bit generation from stationary ergodic sources.

II. DEFINITION

Definition 1 A sequence of deterministic mappings

$$\{\phi_n : A^n \mapsto \{0, 1\}^*\}$$

is a rate- R random bit generator for a source \mathbf{Z} if R is the largest scalar for which there exists a sequence of sets G_n of positive integers such that the following conditions are met:

$$\lim_{n \rightarrow \infty} P(l(\phi_n(Z^n)) \in G_n) = 1 \quad (1)$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{r \in G_n} r P(l(\phi_n(Z^n)) = r) = R \quad (2)$$

$$\limsup_{n \rightarrow \infty} \max_{r \in G_n} \frac{1}{r} D(\phi_n(Z^n) || B^r) = 0 \quad (3)$$

where B^r has the equiprobable distribution on $\{0, 1\}^r$ and Z^n is Z^n restricted to $\{z^n : l(\phi_n(z^n)) = r\}$.

¹This work was supported in part by the National Science Foundation under Grants NYI award IRI-9457645 and NCR 9523805

In the above definition, $l(x)$ denotes the length of a binary string x and $D(P||Q)$ denotes the divergence between distributions P and Q . G_n can be viewed as the set of “good” output lengths which guarantee almost pure randomness in the sense of (3). Thanks to (1), it is not necessary to worry about finding G_n ; furthermore, output lengths outside G_n are not accounted for in the computation of rate.

III. MAIN RESULTS

We can show that optimal source codes are optimal random bit generators without stationarity and ergodicity assumptions on the source. The only assumption we place on the source is that its inf-entropy rate (optimum rate of random number generation according to [1]) is positive.

Theorem 1 *The Shannon source code generates random bits at the maximum possible rate.*

Theorem 2 *The Huffman source code generates random bits at the maximum possible rate.*

Theorem 3 *For a stationary ergodic source, the Lempel-Ziv algorithm generates random bits at the maximum possible rate: the entropy rate of the source.*

The above results are proved by means of two alternative sufficient conditions which can be used to test the applicability of our machinery to optimal source codes other than those mentioned above.

Our next main result shows the strength of our definition of a random bit generator. Let $K(z^n|n)$ denote the Kolmogorov complexity of z^n given its length n .

Theorem 4 *If a code ϕ is a rate- H random bit generator for a stationary ergodic source with entropy rate H then $K(\phi(Z^n)|l(\phi(Z^n)))/l(\phi(Z^n)) \rightarrow 1$ in probability as $n \rightarrow \infty$.*

This result is interesting because if $\lim_{n \rightarrow \infty} K(z^n|n)/n = 1$ then the sequence z will pass all computable statistical tests for randomness [5].

REFERENCES

- [1] S. Vembu, S. Verdú, “Generating Random Bits from an Arbitrary Source: Fundamental limits,” *IEEE Trans. Information Theory*, vol. 41, pp. 1322-1332, September 1995.
- [2] P. Elias, “The efficient construction of an unbiased random sequence,” *Ann. Math. Statist.*, vol. 43, pp. 865-870, 1972.
- [3] Y. Peres, “Iterating von Neumann’s procedure for generating random bits,” *Ann. Statist.*, vol. 20, no. 1, pp. 590-597, 1992.
- [4] J. Ziv, “Compression, Tests for Randomness and Estimating the Statistical Model of an Individual Sequence,” *Sequences: Combinatorics, Compression, Security, and Transmission*, Ed. R.M. Capocelli, Springer Verlag, 1990.
- [5] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, Wiley series in telecommunications, 1991.