

Source Code with Cost as A Nonuniform Random Number Generator

Te Sun Han and Osamu Uchida¹
 Graduate School of Information Systems
 University of Electro-Communications
 Chofugaoka 1-5-1, Chofu, Tokyo, 182-8585 Japan
 e-mail: han@is.uec.ac.jp, o-uchida@hn.is.uec.ac.jp

Abstract — We show that an optimal source code with cost function for code symbols can be regarded as a random number generator generating a random sequence (not necessarily a sequence of fair coin bits) as the target distribution in the sense that the normalized conditional divergence between the distribution of the generated codeword distribution and the target distribution vanishes as the block length tends to infinity.

I. INTRODUCTION

In 1998, Visweswariah *et al.* [1] and Han [2] have independently shown that an optimal variable-length source code can be regarded as a variable-length random number generator in the sense that the normalized divergence distance between the distribution of the generated codeword process and the *uniform* distribution vanishes as the block length tends to infinity.

On the other hand, as is well known, if we impose *unequal costs* on code symbols, it is no longer optimal to use the code which minimizes the average codeword length. Karp [3] has given an algorithm for constructing minimum-redundancy prefix codes with unequal cost symbols. Naturally, there would exist a bias in the frequency of code symbols generated by an optimal source code with cost. Can we then consider the optimal variable-length source code with cost as a variable-length *nonuniform* random number generator? The purpose of this study is to demonstrate that the answer to this question is "yes".

II. VARIABLE-LENGTH SOURCE CODING WITH COST

Let \mathcal{X} be a *countably infinite* source alphabet and \mathcal{Y} be a *finite* code alphabet, respectively. In the sequel all the logarithms are taken to the base $K \equiv |\mathcal{Y}|$, where $|\mathcal{Y}|$ denotes the cardinality of \mathcal{Y} . We denote the set of all non-null finite length sequences taken from \mathcal{Y} by \mathcal{Y}^* . Let us now define a *general source* as an infinite sequence $\mathbf{X} = \{X^n = (X_1^{(n)}, \dots, X_n^{(n)})\}_{n=1}^\infty$ of n -dimensional random variables X^n where each component random variable $X_i^{(n)}$ ($1 \leq i \leq n$) takes values in \mathcal{X} . The class of sources thus defined covers a very wide range of source including all nonstationary and/or nonergodic sources.

Next, we define the *cost function* $c : \mathcal{Y}^* \rightarrow \mathbf{R}^+ \equiv (0, +\infty]$ as follows: First, each symbol $y \in \mathcal{Y}$ is assigned the corresponding cost $c(y)$ such that $0 < c(y) \leq +\infty$ ($\forall y \in \mathcal{Y}$), and then the *additive cost* $c(\mathbf{y})$ of $\mathbf{y} = (y_1, y_2, \dots, y_k) \in \mathcal{Y}^k$ is defined by $c(\mathbf{y}) \equiv \sum_{i=1}^k c(y_i)$.

Definition 1 : R is called an *achievable variable-length source coding cost-rate* for the source \mathbf{X} if there exists a variable-length prefix encoder $\varphi_n : \mathcal{X}^n \rightarrow \mathcal{Y}^*$

given the cost function $c : \mathcal{Y}^* \rightarrow \mathbf{R}^+$ such that $\limsup_{n \rightarrow \infty} \frac{1}{n} E\{c(\varphi_n(X^n))\} \leq R$, and the infimum of R that are achievable variable-length source coding cost-rates is denoted by $R_v^c(\mathbf{X})$, which we call the *infimum achievable variable-length source coding cost-rate*.

Theorem 1 : For any general source \mathbf{X} , we have

$$R_v^c(\mathbf{X}) = \frac{1}{\alpha_c} \limsup_{n \rightarrow \infty} \frac{1}{n} H(X^n),$$

where the *cost capacity* α_c is the positive unique root α of the equation $\sum_{y \in \mathcal{Y}} K^{-\alpha c(y)} = 1$ and $H(X^n) \equiv -\sum_{\mathbf{x} \in \mathcal{X}^n} P_{X^n}(\mathbf{x}) \log P_{X^n}(\mathbf{x})$.

III. SOURCE CODE WITH COST AS A NONUNIFORM RANDOM NUMBER GENERATOR

Given a variable-length prefix encoder $\varphi_n : \mathcal{X}^n \rightarrow \mathcal{Y}^*$, we define $\mathcal{D}_m \equiv \{\mathbf{x} \in \mathcal{X}^n \mid l(\varphi_n(\mathbf{x})) = m\}$ for any positive integer m , where $l(\cdot)$ denotes the length of a string, and we put $\mathcal{J}(\varphi_n) \equiv \{m \mid \Pr\{X^n \in \mathcal{D}_m\} > 0\}$. For any $m \in \mathcal{J}(\varphi_n)$, we define X_m^n as the random variable taking values in \mathcal{D}_m with the distribution given by $P_{X_m^n}(\mathbf{x}) \equiv \frac{P_{X^n}(\mathbf{x})}{\Pr\{X^n \in \mathcal{D}_m\}}$ ($\mathbf{x} \in \mathcal{D}_m$). For any positive integer m , $V^{(m)}$ indicates an i.i.d. sequence of length m . Let us now define the conditional divergence by

$$D(\varphi_n(X^n) \| V^{(I_n)} | I_n) \equiv \sum_{m \in \mathcal{J}(\varphi_n)} \Pr\{I_n = m\} D(\varphi_n(X_m^n) \| V^{(m)})$$

where I_n is the random variable such that $I_n = m$ for $X^n \in \mathcal{D}_m$.

Then, we have the following main theorem.

Theorem 2 : We assume that the entropy rate of the general source \mathbf{X} has the limit $\lim_{n \rightarrow \infty} \frac{1}{n} H(X^n)$. Let $\varphi_n : \mathcal{X}^n \rightarrow \mathcal{Y}^*$ be any *optimal* variable-length prefix encoder in the sense that

$$\lim_{n \rightarrow \infty} \frac{1}{n} E\{c(\varphi_n(X^n))\} = R_v^c(\mathbf{X}).$$

If we define the probability distribution $\mathbf{q}_c = \{q_c(y)\}_{y \in \mathcal{Y}}$ corresponding to the cost function c by $q_c(y) = K^{-\alpha_c c(y)}$ ($y \in \mathcal{Y}$), then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} D(\varphi_n(X^n) \| V^{(I_n)} | I_n) = 0,$$

where $V^{(m)}$ stands for the i.i.d. sequence of length m subject to the distribution \mathbf{q}_c .

REFERENCES

- [1] K. Visweswariah, S. R. Kulkarni and S. Verdú, "Source codes as random number generators," *IEEE Trans. Inform. Theory*, vol. 44, pp.462-471, Mar. 1998.
- [2] T. S. Han, *Information-Spectrum Methods in Information Theory*, Baifukan-Press, Tokyo, 1998 (In Japanese).
- [3] R. M. Karp, "Minimum redundancy coding for the discrete noiseless channel," *IRE Trans. Inform. Theory*, vol. IT-7, pp.27-38, Jan. 1961.

¹O. Uchida is now with the Dept. of Network Engineering, Kanagawa Institute of Technology, Atsugi, Kanagawa, 243-0292 Japan.