Minimum Expected Length of Fixed-to-Variable Lossless Compression of Memoryless Sources

Wojciech Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907, U.S.A.
Email: spa@cs.purdue.edu

Sergio Verdú
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, U.S.A.
Email: verdu@princeton.edu

***Abstract*—Conventional wisdom states that the minimum expected length for fixed-to-variable length encoding of an n-block memoryless source with entropy H grows as $nH+O(1)$. However, this performance is obtained under the constraint that the code assigned to the whole n-block is a prefix code. Dropping this unnecessary constraint we show that the minimum expected length grows as**

$$nH - \frac{1}{2}\log n + O(1)$$

unless the source is equiprobable.

I. Introduction

Lossless symbol-by-symbol compressors are required to satisfy the condition of "unique decodability" whereby different input strings are assigned different compressed versions. Since any uniquely decodable code must assign lengths to the various symbols that satisfy Kraft's inequality, while a prefix code is guaranteed to exist with those symbol lengths, it is enough to restrict attention to prefix codes without loss of generality. Achieved by the Huffman code, an exact expression for the minimum average length of a prefix symbol-by-symbol binary code is unknown (cf. [20] for an asymptotic analysis); Shannon [19] showed it to be upper bounded by the entropy (in bits) of the probability distribution of the symbols plus one bit, while Macmillan [15] showed it to be lower bounded by the entropy. However, the paradigm of symbol-by-symbol compression is severely suboptimal even for memoryless sources. For example, they are unable to exploit the redundancy of biased coin flips. The conventional conceptual (not algorithmic) approach to deal with this inefficiency is to partition the source string of length n into blocks of length k and apply the symbol-by-symbol approach at the block level. The resulting average compressed length per source symbol is equal to the entropy of each symbol, $H(X)$, plus at most $1/k$ bits if the source is memoryless, or more generally, equal to the entropy of k consecutive symbols divided by k plus at most $1/k$ bits. Thus, to achieve the best efficiency, we can let $k = n$, apply a Huffman code to the whole n-tuple and the resulting average compressed length behaves as

$$L_n = nH(X) + O(1). \tag{1}$$

where the $O(1)$ term belongs to $[0, 1]$.

As argued in [23], it is possible to attain average compressed length lower than (1). The reason is that it is unnecessary, and in fact wasteful, to impose the prefix condition on a code that operates at the level of the whole file to be compressed. Applying prefix codes to n-block supersymbols is only optimal in terms of the linear growth with n (it attains the entropy rate for stationary ergodic sources); however, as far as sublinear terms, this conventional approach incurs loss of optimality. The optimal fixed-to-variable length code* performs no blocking on the source output; instead the optimal compressor for a length-n output chooses an encoding table that lists all source realizations of length n in decreasing probabilities (breaking ties using a lexicographical ordering on the source symbols) and assigns, starting with the most probable, the binary strings of increasing lengths $\{\emptyset, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \ldots\}$. Note that such a code would not work if applied symbol-by-symbol (or block-by-block) as the decompressor would not be able to recover losslessly the original source string.

Such optimal codes have been previously considered under the rubric of *one-to-one* codes, but because of their misguided standing as non-uniquely decodable *symbol-by-symbol* codes, they have failed to attract much attention.

In the rest of this paper, Section II deals with the nonasymptotic analysis of one-to-one codes. Section III deals with the asymptotic analysis of the minimum average length, L_n^*, of codes for memoryless sources with known distributions. Under the assumption that the the source is not equiprobable, we show that

$$L_n^* = nH(X) - \frac{1}{2}\log_2 n + O(1). \tag{2}$$

for large n.

II. Nonasymptotic analysis of one-to-one codes

Consider a probability distribution P_X on a set of ordered elements $\mathcal{X}$. Define $\pi_X : \mathcal{X} \mapsto \{1, \ldots, |\mathcal{X}|\}$ by $\pi_X(a) < \pi_X(b)$ if $P_X(a) > P_X(b)$ or if $P_X(a) = P_X(b)$ and $a < b$. Thus, $\pi_X(x) = \ell$ if x is the ℓ-th most probable element in

*Optimal code not just in the sense of average length but in the sense that the cumulative distribution function of its length is larger than or equal to that of any other code.

$\mathcal{X}$ according to distribution P_X, with ties broken according to the ordering in $\mathcal{X}$. It is easy to verify that

$$P_X(x)\pi_X(x) \le 1 \tag{3}$$

for all $x \in \mathcal{X}$: if (3) failed to be satisfied for $x_0 \in \mathcal{X}$, there would be at least $\pi_X(x_0)$ masses strictly larger than $1/\pi_X(x_0)$.

The one-to-one code assigns to x the shortest binary string (ties broken with the ordering $0 < 1$) not assigned to any element y with $\pi_X(y) < \pi_X(x)$. Thus, we obtain the simple but important conclusion that the length of the encoding of x is $\lfloor \log_2 \pi_X(x) \rfloor$. Finding an expression for the minimum average length

$$L(X) = \mathbb{E}[\lfloor \log_2 \pi_X(X) \rfloor] \tag{4}$$

as a function of P_X appears to be challenging. For X equiprobable on a set of $M = |\mathcal{X}|$ elements, it can be shown that the average length of the one-to-one code is (cf. [13])

$$\begin{aligned} L(X) & \\ &= \frac{1}{M}\sum_{i=1}^{M} \lfloor \log_2 i \rfloor & (5)\\ &= \lfloor \log_2 M \rfloor + \frac{1}{M}\left(2 + \lfloor \log_2 M \rfloor - 2^{\lfloor \log_2 M \rfloor + 1}\right) & (6) \end{aligned}$$

which simplifies to

$$\frac{1}{M}\sum_{i=1}^{M} \lfloor \log_2 i \rfloor = \frac{(M+1)\log_2(M+1)}{M} - 2 \tag{7}$$

when $M+1$ is a power of 2.

A simple upper bound first noticed in [26] is obtained as

$$\begin{aligned} L(X) &= \mathbb{E}[\lfloor \log_2 \pi_X(X) \rfloor] & (8)\\ &\le \mathbb{E}[\log_2 \pi_X(X)] & (9)\\ &\le \mathbb{E}\left[\log_2 \frac{1}{P_X(X)}\right] & (10)\\ &= H(X) & (11) \end{aligned}$$

where (10) follows from (3). Various lower bounds have been proposed in [1]–[3], [8], [14], [16], [24], [25]. Distilling the main ideas in [1], the following result gives the tightest known bound.

Theorem 1: Define the monotonically increasing function $\psi : \mathbb{R}^+ \mapsto \mathbb{R}^+$ by

$$\psi(x) = x + (1+x)\log_2(1+x) - x\log_2 x \tag{12}$$

Then,

$$\psi^{-1}(H(X)) \le L(X) \tag{13}$$

Proof: For brevity denote $Y = \lfloor \log_2 \pi_X(X) \rfloor$, and $Z = Y + 1$

$$\begin{aligned} H(X) &= H(X|Y) + H(Y) & (14)\\ &\le \mathbb{E}[Y] + H(Y) & (15)\\ &= \mathbb{E}[Y] + H(Z) & (16)\\ &= \mathbb{E}[Y] + \mathbb{E}[Z]h(1/\mathbb{E}[Z]) - D(P_Z\|G_{1/\mathbb{E}[Z]}) & (17)\\ &\le \psi(\mathbb{E}[Y])) & (18) \end{aligned}$$

where

- (14) $\Longleftarrow$ Y is a deterministic function of X;
- (15) $\Longleftarrow$ $H(X|Y=k) \le k$ bits;
- (17) uses the binary entropy function $h(\cdot)$ and the divergence with respect to a geometric (positive) distribution $G_p(k) = p(1-p)^{k-1}$;
- (18) $\Longleftarrow$ $D(\|) \ge 0$.

■

Weakening the bound in (13) by

$$\psi(x) \le x + \log_2(e + ex)$$

and using the upper bound (11) we obtain the bound in [1]:

$$H(X) - \log_2(H(X)+1) - \log_2 e \le \mathbb{E}[\lfloor \log_2 \pi_X(X) \rfloor] \tag{19}$$

Another way of weakening (13) is to use the monotonic increasing nature of $(1+x)\log(1+x) - x\log x$ and (11) to conclude

$$\begin{aligned} L(X) &\ge H(X) - (1+L(X))\log_2(1+L(X)) & (20)\\ &\quad -L(X)\log_2 L(X) & (21)\\ &\ge H(X) - (1+H(X))\log_2(1+H(X)) & (22)\\ &\quad -H(X)\log_2 H(X) \end{aligned}$$

which is the bound found in [2].

III. Asymptotic minimum average length

We assume henceforth that the source is memoryless with distribution P_X. We abbreviate the minimum average length of the encoding of an n-tuple of the source by

$$L_n^* = L(X^n). \tag{23}$$

The minimum average length for a binary memoryless source with bias p has been investigated in great detail (up to $o(1)$ term) in [22]. For fair coin flips ($p = \frac{1}{2}$), the exact result can be obtained from (6) letting $M = 2^n$:

$$L_n^* = n - 2 + 2^{-n}(n+2), \tag{24}$$

in contrast to

$$L_n = n \tag{25}$$

obtained with the Huffman code operating on n-tuples (or single bits).

If $p \ne \frac{1}{2}$, [22] shows that

$$L_n^* = nh(p) - \frac{1}{2}\log_2 n + O(1) \tag{26}$$

and in fact [22] characterizes the $O(1)$ explicitly showing that its behavior depends on whether $\log_2 \frac{1-p}{p}$ is rational.

Our main result is presented next; its proof is outlined in Section IV.

Theorem 2: For a memoryless source with finite alphabet $\mathcal{A}$, the minimum expected length of a lossless binary encoding of X^n is given by

$$L_n^* = \lfloor n \log_2 |\mathcal{A}| \rfloor + o(1). \tag{27}$$

if the source is equiprobable, and by

$$L_n^* = nH(X) - \frac{1}{2}\log_2 n + O(1) \tag{28}$$

if the source is not equiprobable.

IV. Proof of Theorem 2

Expression (27) for non-redundant sources follows from (6). Henceforth, we assume that the source is not equiprobable. We abbreviate $|\mathcal{A}| = m$, denote by $p_1, \dots p_m$ the atoms of P_X such that

$$p_1 \le p_2, \quad \dots \quad p_{m-1} \le p_m,$$

and we denote

$$B_i = \log\frac{p_m}{p_i} \tag{29}$$

for $i = 1, \dots, m-1$. Note that the entropy of P_X can be expressed as

$$H(X) = \log\frac{1}{p_m} + \sum_{i=1}^{m-1} p_i B_i \tag{30}$$

Let $\mathbf{k} = (k_1, \dots, k_m)$ such that $k_1 + \dots + k_m = n$ denote the *type* of an n-string; the probability of each such string is equal to

$$p^{\mathbf{k}} = p_1^{k_1} \cdots p_m^{k_m}. \tag{31}$$

Denote the set of all types of n-strings drawn from an alphabet of m elements by

$$\mathcal{T}_{n,m} = \{(k_1, \dots, k_m) \in \mathbb{N}^m, k_1 + \dots + k_m = n\} \tag{32}$$

We introduce an order among types:

$$\mathbf{l} \preceq \mathbf{k} \quad \text{iff} \quad p^{\mathbf{l}} \ge p^{\mathbf{k}}.$$

and we sort all types from the smallest index (largest probability) to the largest. This can be accomplished by observing that $p^{\mathbf{l}} \ge p^{\mathbf{k}}$ is equivalent to

$$l_1 B_1 + \dots + l_{m-1}B_{m-1} \le k_1 B_1 + \dots + k_{m-1}B_{m-1}. \tag{33}$$

There are

$$\binom{n}{\mathbf{k}} = \binom{n}{k_1, \dots, k_m}$$

sequences of type $\mathbf{k}$ and we list them in lexicographic order. Then the optimum code assigns length $\lfloor \log i \rfloor$ to the ith sequence ($1 \le i \le m^n$) in this list. Denote the number of sequences more probable than or equal to type $\mathbf{k}$ as

$$A_{\mathbf{k}} := \sum_{\mathbf{l} \preceq \mathbf{k}} \binom{n}{\mathbf{l}}. \tag{34}$$

Using somewhat informal but intuitive notation, $\mathbf{k}+1$ and $\mathbf{k}-1$ denote the *next* and *previous* types, respectively, in the sorted list of the elements of $\mathcal{T}_{n,m}$. Clearly, starting from position $A_{\mathbf{k}}$ the next $\binom{n}{\mathbf{k}+1}$ sequences have probability $p^{\mathbf{k}+1}$. Thus the average code length can be computed as follows

$$\begin{aligned} L_n^* &= \sum_{\mathbf{k}\in\mathcal{T}_{n,m}} p^{\mathbf{k}} \sum_{i=A_{\mathbf{k}-1}+1}^{A_{\mathbf{k}}} \lfloor \log i \rfloor \\ &= \sum_{\mathbf{k}\in\mathcal{T}_{n,m}} p^{\mathbf{k}} \sum_{i=1}^{\binom{n}{\mathbf{k}}} \lfloor \log(A_{\mathbf{k}} - i) \rfloor \\ &= \sum_{\mathbf{k}\in\mathcal{T}_{n,m}} p^{\mathbf{k}} \sum_{i=1}^{\binom{n}{\mathbf{k}}} \lfloor \log A_{\mathbf{k}}(1 - i/A_{\mathbf{k}}) \rfloor \\ &= \sum_{\mathbf{k}\in\mathcal{T}_{n,m}} \binom{n}{\mathbf{k}} p^{\mathbf{k}} \log A_{\mathbf{k}} + O(1), \\ &= \log A_{n\mathbf{p}} + O(1), \end{aligned} \tag{35}$$

where (35) follows along the same lines as [9], [12]. Thus we need to evaluate

$$A_{n\mathbf{p}} = \sum_{p^{\mathbf{l}} \ge p^{n\mathbf{p}}} \binom{n}{\mathbf{l}}. \tag{36}$$

Let now

$$l_i = np_i + x_i \tag{37}$$

for $i = 1, \dots, m-1$. Then, by (33) the summation set in (36) can be written as

$$p^{\mathbf{l}} \ge p^{n\mathbf{p}} \;\leftrightarrow\; B_1 x_1 + \dots + B_{m-1}x_{m-1} \le 0. \tag{38}$$

Thus

$$A_{n\mathbf{p}} = \sum_{\mathbf{x}} \binom{n}{n\mathbf{p} + \mathbf{x}} \tag{39}$$

where the summation is over the hyperspace $B_1 x_1 + \dots + B_{m-1}x_{m-1} \le 0$.

The next step is to use Stirling's formula

$$n! = \sqrt{2\pi n} \cdot n^n e^{-n}(1 + O(1/n)) \tag{40}$$

to estimate the summands in (39). A long computation whose details are omitted reveals that

$$\begin{aligned} &\binom{n}{n\mathbf{p} + \mathbf{x}} \\ &= \frac{1}{(2\pi)^{(m-1)/2}} \frac{1}{\sqrt{p_1 \cdots p_m}} \frac{1}{n^{(m-1)/2}} 2^{nH(X)} \\ &\cdot \left(\frac{p_m}{p_1}\right)^{x_1} \cdots \left(\frac{p_m}{p_{m-1}}\right)^{x_{m-1}} \left(1 + O(1/\sqrt{n})\right) \\ &\cdot \exp\left(-\frac{x_1^2}{2np_1} - \dots - \frac{x_{m-1}^2}{2np_{m-1}} - \frac{(x_1 + \dots + x_{m-1})^2}{2np_m}\right) \\ &= \left(1 + O(1/\sqrt{n})\right) C \frac{2^{nH(X)}}{n^{(m-1)/2}} \\ &\cdot \exp\left(B_1 x_1 + \dots + B_{m-1}x_{m-1}\right) \\ &\cdot \exp\left(-\frac{1}{2n}\mathbf{x}^T \mathbf{\Sigma}^{-1}\mathbf{x}\right) \end{aligned} \tag{41}$$

where $\mathbf{\Sigma}$ is an appropriately chosen invertible covariance matrix, and

$$\mathbf{x} = (x_1, \dots, x_{m-1})$$

We are now in the position to evaluate the sum (39). First, we split it into two sums:

- a sum over the $(m-2)$-dimensional hyperplane $B_1x_1 + \cdots + B_{m-1}x_{m-1} = 0$ which we denote as $\mathcal{D}^{m-2}$
- a sum over $B_1x_1 + \cdots + B_{m-1}x_{m-1} < 0$.

Introducing the notation:

$$\mathbf{b}^T = [B_1, \dots, B_{m-1}], \tag{42}$$

(39) together with (41) yields (C in different lines need not be the same constant)

$$\begin{aligned} &A_{n\mathbf{p}} \\ &= \frac{C2^{nH(X)}}{n^{(m-1)/2}} \left(\sum_{\mathbf{b}^T\mathbf{x}=0} \exp\left(-\frac{1}{2n}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right) \right. \\ &+ \left. \sum_{\mathbf{b}^T\mathbf{x}<0} \exp\left(\mathbf{b}^T\mathbf{x} - \frac{1}{2n}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right)\right). \end{aligned} \tag{43}$$

Clearly, the second sum is bounded since it is an exponential sum for $B_1x_1 + \cdots + B_{m-1}x_{m-1} < 0$.

Furthermore, the multidimensional normal distribution integral [10] leads us to conclude that

$$\int_{\mathcal{D}^{m-2}} \exp\left(-\frac{1}{2n}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right) = Cn^{(m-2)/2}. \tag{44}$$

Combining it, and using Euler-Maclaurin formula for replacing discrete sums by integrals, we finally arrive at

$$\begin{aligned} \log A_{n\mathbf{p}} &= \log\left(C\frac{2^{nH(X)}}{n^{(m-1)/2}}n^{(m-2)/2} + O\left(\frac{2^{H(X)}}{n^{(m-1)/2}}\right)\right) \\ &= nH(X) - \frac{1}{2}\log n + O(1) \end{aligned} \tag{45}$$

In view of (35) this completes the proof of Theorem 2.

Example. To illustrate our methodology, we explain it in some details for the case of $m = 3$ symbols with probability $p_1 < p_2 < p_3$. We need to evaluate (with $B_1 = \log(p_3/p_1)$ and $B_2 = \log(p_3/p_2)$) the following

$$A_{np_1,np_2} = \sum_{k_1B_1+k_2B_2 \le np_1B_1+np_2B_2} \binom{n}{k_1, k_2}.$$

As before, we denote $k_1 = np_1 + x$ and $k_2 = np_2 + y$ to arrive at

$$\begin{aligned} \binom{n}{np_1+x, np_2+y} &= \frac{1}{\sqrt{2\pi p_1p_2p_3n}} 2^{nH(\mathbf{p}} \left(\frac{p_3}{p_1}\right)^x \left(\frac{p_3}{p_2}\right)^y \\ &\times \exp\left(-\frac{x^2}{2np_1} - \frac{y^2}{2np_2} - \frac{(x+y)^2}{2np_3}\right)(1 + O(1/\sqrt{n}). \end{aligned}$$

Then (cf. Figure 1)

$$A_{n\mathbf{p}} = \sum_{B_1x+B_2y\le 0} \binom{n}{np_1+x, np_2+y}$$

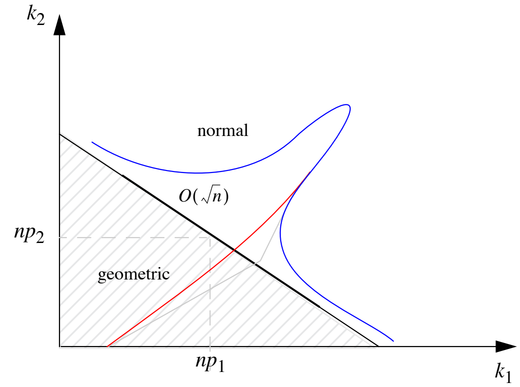

Fig. 1. Illustration for $m = 3$

.

$$\begin{aligned} &\sim \frac{2^{nH(X)}}{n\sqrt{2\pi p_1p_2p_3}} \sum_{B_1x+B_2y=0} \exp\left(-\frac{x^2}{2np_1} - \frac{y^2}{2np_2} - \frac{(x+y)^2}{2np_3}\right) \\ &= O(\sqrt{n})\frac{2^{nH(X)}}{n} = C\frac{2^{nH(X)}}{\sqrt{n}}, \end{aligned}$$

where the last equality follows from the normal approximation on the line $B_1x+B_2y = 0$ (this part contributes $O(\sqrt{n})$), and the first approximation is a consequence of geometric decay of the multinomial coefficient away from the line $B_1x+B_2y = 0$, that is, for $B_1x + B_2y < 0$. This is illustrated in Figure 1.

Acknowledgment

The work of W. Szpankowski was supported in part by the NSF Grants CCF-0513636, DMS-0503742, DMS-0800568, and CCF -0830140, NSA Grant H98230-08-1-0092, and the AFOSR Grant FA8655-08-1-3018. The work of S. Verdú was supported by NSF Grant CCF-0635154.

References

[1] N. Alon and A. Orlitsky, "A Lower Bound on the Expected Length of One-to-One Codes," *IEEE Trans. Information Theory*, 40, 1670-1672, 1994.

[2] C. Blundo and R. de Prisco, "New bounds on the expected length of one-to-one codes," *IEEE Trans. Information Theory,* vol. 42, pp. 246-250, Jan. 1996

[3] J. Cheng and T. K. Huang, "New Bounds on the Expected Length of Optimal One-to-One Codes," Vol. 53, No. 5, pp. 1884-1895, May 2007

[4] M. Drmota, "A Bivariate Asymptotic Expansion of Coefficients of Powers of Generating" Functions, *Europ. J. Combinatorics*, 15, 139-152, 1994.

[5] M. Drmota, H-K. Hwang, and W. Szpankowski, "Precise Average Redundancy of an Idealized Arithmetic Coding," *Proc. Data Compression Conference*, 222-231, Snowbird, 2002.

[6] M. Drmota and R. Tichy, *Sequences, Discrepancies, and Applications*, Springer Verlag, Berlin Heidelberg 1997

[7] M. Drmota and W. Szpankowski, "Precise Minimax Redundancy and Regret," *IEEE Trans. Information Theory*, 50, No. 11, 2004.

[8] J. G. Dunham, "Optimal noiseless coding of random variables, *IEEE Trans. Information Theory,* vol. IT-26, no. 3, p. 345, May 1980.

[9] P. Flajolet, "Singularity Analysis and Asymptotics of Bernoulli Sums," *Theoretical Computer Science*, 215, 371–381, 1999.

[10] B. Gnedenko, *The Theory of Probability and Elements of Statistics*, Chelsea Pub. Company, New York, 1991.

[11] P. Henrici, *Applied and Computational Complex Analysis*, vol. 2, John Wiley & Sons, New York, 1977.

[12] P. Jacquet and W. Szpankowski, "Entropy Computations via Analytic Depoissonization," *IEEE Trans. Information Theory*, 45, 1072-1081, 1999.

[13] D. Knuth, *The Art of Computer Programming: Fundamental Algorithms,* Vol. 1. Addison-Wesley, 1997.

[14] S. K. Leung-Yan-Cheong, T. Cover, "Some Equivalences between Shannon Entropy and Kolmogorov Complexity," *IEEE Trans. Information Theory*, 24, 331-338, 1978.

[15] B. MacMillan, "Two inequalities implied by unique decipherability," *IRE Trans. Information Theory*, vol. 2, pp. 115-116, Dec. 1956.

[16] J. Rissanen, "Tight lower bounds for optimum code length," *IEEE Trans. Information Theory,* vol. IT-28, no. 2, pp. 348349, Mar. 1982.

[17] J. Rissanen, "Universal Coding, Information, Prediction, and Estimation," *IEEE Trans. Information Theory,* vol. IT-30, no. 4, pp. 629636, July 1984.

[18] S. Savari and A. Naheta, "Bounds on the Expected Cost of One-to-One Codes," *2004 IEEE Int.Symp. Information Theory*, p. 92, Chicago, IL, July 2004.

[19] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.

[20] W. Szpankowski, "Asymptotic Average Redundancy of Huffman (and Other) Block Codes," *IEEE Trans. Information Theory*, 46, 2434-2443, 2000.

[21] W. Szpankowski, *Average Case Analysis of Algorithms in Sequences*, John Wiley & Sons, New York, 2000.

[22] W. Szpankowski, "One-to-One Code and Its Anti-Redundancy," *IEEE Trans. Information Theory*, vol. 54, pp. 4762-4766, Oct. 2008.

[23] S. Verdú, "teaching it," XXVIII Shannon Lecture, *2007 IEEE International Symposium on Information Theory*, Nice, France, June 28, 2007. Also: IEEE Information Theory Society Newsletter, Dec. 2007.

[24] E. I. Verriest, "An achievable bound for optimal noiseless coding of a random variable, *IEEE Trans. Information Theory,* vol. IT-32, no. 4, pp. 592-594, July 1986.

[25] C. Weidmann, "New Upper Bounds on the Expected Length of One-to-one Codes," *2002 IEEE International Symposium on Information Theory*, Lausanne, Switzerland, June 30 -July 5, 2002.

[26] A. D. Wyner, "An Upper Bound on the Entropy Series," *Inform. Control*, 20, 176-181, 1972.