# A One-to-One Code and Its Anti-redundancy*

Wojciech Szpankowski†
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

**Abstract**

One-to-one codes are "one shot" codes that assign a distinct codeword to source symbols and are not necessarily prefix codes (more generally, uniquely decodable). Interestingly, as Wyner proved in 1972, for such codes the average code length can be *smaller* than the source entropy. By how much? We call this difference the *anti-redundancy*. Various authors over the years have shown that the anti-redundancy can be as big as minus the logarithm of the source entropy. However, to the best of our knowledge precise estimates do not exist. In this note, we consider a block code of length $n$ generated for a binary memoryless source, and prove that the average anti-redundancy is

$$-\frac{1}{2}\log_2 n + C + F(n) + o(1)$$

where $C$ is a constant and either $F(n) = 0$ if $\log_2(1-p)/p$ is irrational (where $p$ is the probability of generating a "0") or otherwise $F(n)$ is a fluctuating function as the code length increases. This relatively simple finding requires a combination of analytic tools such as precise evaluation of Bernoulli sums, the saddle point method, and theory of distribution of sequences modulo 1.

*Index Terms* — Prefix codes, one-to-one codes, average redundancy, Bernoulli sums, saddle point method, distribution of sequences modulo 1.

*Preliminary version was presented at *2005 International Symposium on Information Theory*, Adelaide, 2005.

†This work was supported in part by the NSF Grants CCF-0513636, DMS-0503742, and DMS-0800568, NIH Grant R01 GM068959-01, NSA Grant H98230-08-1-0092, and the AFOSR Grant FA8655-08-1-3018.

## 1 Introduction

Traditionally, source coding deals with prefix (or more generally, uniquely decodable) codes that are injections from an alphabet $\mathcal{A}$ into the binary strings $\{0,1\}^*$. Already in 1948 Shannon observed that for such codes the average code length cannot be smaller than the entropy of the source. A simple proof of this fact may go as follows: Let $K = \sum_x 2^{-L(x)} \le 1$ (by Kraft's inequality for prefix codes). Then, using standard notations (e.g., $L(x)$ stands for the code length of the source symbol $x$, $P(x)$ is the probability of $x$, and $H(X) = -\sum_{x\in\mathcal{A}} P(x)\log_2 P(x)$ is the source entropy), we have

$$\mathbf{E}[L(X)] - H(X) = \sum_{x\in\mathcal{A}} P(x)L(x) + \sum_{x\in\mathcal{A}} P(x)\log_2 P(x) = \sum_{x\in\mathcal{A}} P(x)\log_2 \frac{P(x_1^n)}{2^{-L(x)}/K} - \log K \ge 0$$

since the divergence cannot be negative.

The next natural step is to ask by how much the average code length exceeds the entropy. This is called the *average redundancy* which is nonnegative for prefix codes (i.e., codes satisfying Kraft's inequality), as shown above. Over the last twenty years a substantial literature was built to address this problem (e.g., see [7] for some recent developments).

Occasionally, encodings are not necessarily prefix free. In *one-to-one codes* a distinct codeword is assigned to each source symbol and unique decodability is not required. Such codes are usually one shot codes and there is one designated an"end of message" symbol that is distinct from all other (source) symbols. Wyner [17] in 1972 proved that the average code length $L$ is actually smaller than the source $X$ entropy $H(X)$. A lower bound for the average code length of such codes was first established in [13] and then improved by Alon and Orlitsky [1] who proved that

$$L \ge H(X) - \log(H(X)+1) - \log e. \tag{1}$$

Some recent results on one-to-one codes are reported in [4, 14].

As with prefix codes, one can study the difference between the average code length and the entropy. For one-to-one codes we shall call this difference the *anti-redundancy*. Thus the anti-redundancy is defined as

$$\bar{R} = L - H(X)$$

and from [1] we conclude that $\bar{R} = \Omega(-\log H(X))$. A question arises whether this lower bound is a universal one for a class of sources. Alon and Orlitsky [1] showed that the lower bound is achievable for the geometric distribution. In this note we consider a block one-to-one code for a binary memoryless source over $\{0,1\}^n$ and analyze precisely the average anti-redundancy $\bar{R}_n$ showing that the bound in [1] is not tight for such sources.

Let us briefly discuss our main findings. We consider a source sequence $X_1^n = X_1 \dots X_n$ generated by a binary memoryless source with $p$ being the probability of generating a "0". We assume $p \le 1-p := q$ and order all probabilities $p^k q^{n-k}$ in a nondecreasing fashion assigning a codeword length $\lfloor \log_2 j \rfloor$ to the $j$th message where $1 \le j \le 2^n$. Observe that for every $1 \le k \le n$ there are $\binom{n}{k}$ messages of the same probability that we order randomly. Our goal is to estimate the average code length $L_n$ defined precisely in (3) below. We shall prove that for $p < 1/2$

$$L_n = nh(p) - \frac{1}{2}\log_2 n + C + F(n) + o(1)$$

where $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the entropy rate, $C$ is an explicitly computable constant, and $F(n) \equiv 0$ when $\log_2(1-p)/p$ is irrational and $F(n)$ is a fluctuating function of $n$ when $\log_2(1-p)/p$ is rational. Interestingly enough, fluctuations appear only in the third order term of the asymptotic expansion, while for the Huffman and arithmetic codes the fluctuations contribute already to the second term [6, 15].

To obtain our main result we need a battery of analytic techniques. Namely, a formula to deal with sums of floor functions, asymptotics for the Bernoulli sums, the saddle point method, and the theory of distribution of sequences modulo 1. The interested reader is referred to [10, 16] for detailed accounts on these methods. In the next section we present our main results that are proved in Section 3.

## 2 Main Results

We consider a binary memoryless source $X$ over the binary alphabet $\mathcal{A} = \{0, 1\}$ generating a sequence $x_1^n = x_1, \ldots, x_n \in \mathcal{A}^n$. Then $P(x_1^n) = p^k q^{n-k}$, where $k$ is the number of 0s in $x_1^n$ and throughout this paper we shall assume that $p \le q$. We now list all $2^n$ probabilities in a nonincreasing order

$$q^n \left(\frac{p}{q}\right)^0 \ge q^n \left(\frac{p}{q}\right)^1 \ge \cdots \ge q^n \left(\frac{p}{q}\right)^n. \tag{2}$$

Let us assign consecutive natural numbers $j$ ($1 \le j \le 2^n$) to each probability on the list of $P(x_1^n)$. Clearly, there are $\binom{n}{k}$ equal probabilities $p^k q^{n-k}$. Define

$$A_k = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k}, \qquad A_{-1} = 0.$$

Starting from the position $A_{k-1} + 1$ of the list (2), the next $\binom{n}{k}$ probabilities are the same and equal to $p^k q^{n-k}$.

For each $j = A_{k-1} + i$, $1 \le i \le \binom{n}{k}$, we now assign the codelength

$$\lfloor \log_2(j) \rfloor = \lfloor \log_2(A_{k-1} + i) \rfloor$$

to the $j$th binary string. Thus the average code length is

$$\begin{aligned} L_n &= \sum_{k=0}^{n} p^k q^{n-k} \sum_{j=A_{k-1}+1}^{A_k} \lfloor \log_2(j) \rfloor \qquad (3) \\ &= \sum_{k=0}^{n} p^k q^{n-k} \sum_{i=1}^{\binom{n}{k}} \lfloor \log_2(A_{k-1} + i) \rfloor. \end{aligned}$$

Our goal is to estimate $L_n$ asymptotically for large $n$.

Let us first simplify the above formula for $L_n$. We need to handle the inner sum that contains the floor function. To evaluate this sum we apply the following identity (cf. Knuth [12] Ex. 1.2.4-42)

$$\sum_{j=1}^{N} a_j = N a_N - \sum_{j=1}^{N-1} j(a_{j+1} - a_j)$$

for any sequence $a_j$. Using this, we easily find an explicit formula for the inner sum of (3), namely

$$\begin{aligned} S_{n,k} &= \sum_{j=1}^{\binom{n}{k}} \lfloor \log_2(A_{k-1}+j) \rfloor = \binom{n}{k} \lfloor \log_2 A_k \rfloor - (2^{\lfloor \log_2(A_k) \rfloor + 1} - 2^{\lceil \log_2(A_{k-1}+2) \rceil}) \\ &+ (A_{k-1}+1)(1 + \lfloor \log_2(A_k) \rfloor - \lceil \log_2(A_{k-1}+2) \rceil). \end{aligned}$$

After some algebra, using $\lfloor x \rfloor = x - \langle x \rangle$ and $\lceil x \rceil = x + \langle -x \rangle$ where $\langle x \rangle = x - \lfloor x \rfloor$ is the fractional part of $x$, we finally reduce the formula for $L_n$ to the following

$$\begin{aligned} L_n &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \lfloor \log_2 A_k \rfloor & (4) \\ &- 2 \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \log_2 A_k \rangle} & (5) \\ &+ \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \frac{1 + A_{k-1}}{\binom{n}{k}} \left(1 + \log_2\left(\frac{A_k}{A_{k-1}+2}\right) - \langle -\log_2(A_{k-1}+2) \rangle - \langle \log_2 A_k \rangle \right) & (6) \\ &- \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \frac{A_{k-1}}{\binom{n}{k}} \left(2^{-\langle \log_2 A_k \rangle + 1} - 2^{\langle -\log_2(A_{k-1}+2) \rangle}\right) & (7) \\ &+ 2 \sum_{k=0}^{n} p^k q^{n-k} 2^{\langle -\log_2(A_{k-1}+2) \rangle}. & (8) \end{aligned}$$

In the next section, we evaluate asymptotically sums (4)– (8) leading to our main result of this paper.

**Theorem 1** *Consider a binary memoryless source and the one-to-one block code described above. Then for* $p < \frac{1}{2}$

$$\begin{aligned} L_n &= nh(p) - \frac{1}{2}\log_2 n - \frac{3+\ln(2)}{2\ln(2)} + \log_2 \frac{1-p}{1-2p} \frac{1}{\sqrt{2\pi p(1-p)}} + \frac{p}{1-2p} \log_2\left(\frac{2(1-p)}{p}\right) \\ &+ F(n) + o(1) & (9) \end{aligned}$$

*where* $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ *is the entropy rate,* $\alpha = \log_2(1-p)/p$, $\beta = \log_2(1/(1-p)$ *and* $F(n) = 0$ *if* $\log_2 \frac{1-p}{p}$ *is irrational. If* $\log_2 \frac{1-p}{p} = N/M$ *for some integers* $M, N$ *such that* $\gcd(N, M) = 1$, *then*

$$F(n) = -\frac{1-p}{1-2p} H_M(n\beta)[x] - \frac{p}{1-2p} H_M(n\beta - \alpha)[-x] - \frac{2(1-3p)}{1-2p} H_M(n\beta)[2^{-x}] + \frac{p}{1-2p} H_M(n\beta-\alpha)[2^x]$$

*where*

$$H_M(y)[f] := \frac{1}{M\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left( \left\langle M \left( y - \log_2 \left( \frac{1-2p}{1-p} \sqrt{2\pi pqn} \right) - \frac{x^2}{2\ln 2} \right) \right\rangle - \int_0^1 f(t)dt \right) dx$$

*for some Riemann function* $f$*.*
*For* $p = \frac{1}{2}$*, we have*

$$L_n = nh(1/2) - 2 + 2^{-n}(n+2)$$

*for every* $n \geq 1$*.*

In view of Theorem 1, we again see that asymptotic behavior of the redundancy or anti-redundancy depends on the rationality/irrationality of $\log_2(1-p)/p$ (cf. [6, 7, 15]). In Figure 1 we plot the "constant part" $L_n - nh(p) + 0.5\log_2(n)$ versus $n$. We observe change of "mode" from a "converging mode" to a "fluctuating mode", when switching from $\alpha = \log_2(1-p)/p$ irrational (cf. Fig. 1(a)) to rational (cf. Fig. 1(b)). This phenomenon was already observed in [7, 15] for Huffman and Shannon codes.

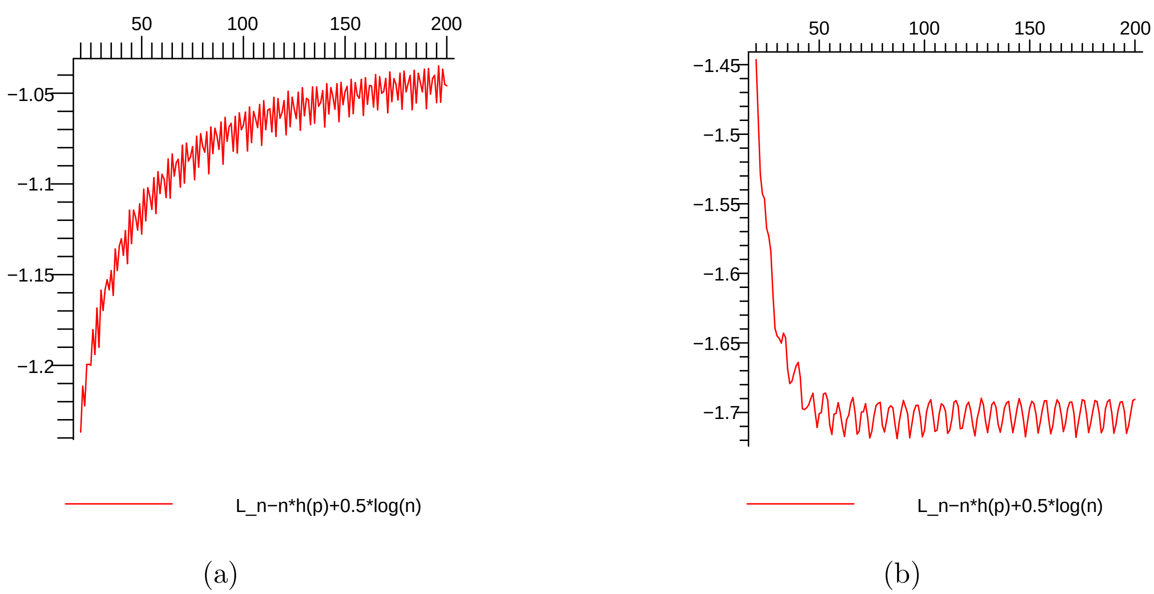


Figure 1: Plots of $L_n - nh(p) + 0.5\log(n)$ (y-axis) versus $n$ (x-axis) for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

Finally, one may conclude from the lower bound proved in [1] that the leading term of $\bar{R}_n$ is $-\log_2 n$ (e.g., the bound is achievable for the geometric distribution). In this paper we prove that for memoryless sources (i.e., the binomial distribution) the average anti-redundancy is asymptotically equal to $-\frac{1}{2}\log_2 n$ showing that the lower bound of [1] is not tight in this case.

Furthermore, our findings indicate that the minimax anti-redundancy for a class of binary memoryless sources is $-\frac{1}{2}\log_2 n + O(1)$.

## 3 Analysis

In this section we analyze asymptotically the four terms of $L_n$ as presented in (4)–(8). Throughout we write $\log := \log_2$. We start with (4) which we split as follows

$$\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\lfloor \log_2 A_k\rfloor = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\log_2 A_k - \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\langle \log_2 A_k\rangle,$$

and define

$$a_n = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\log_2 A_k, \tag{10}$$

$$b_n = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\langle \log_2 A_k\rangle. \tag{11}$$

The first sum, and most sums discussed here, falls under the so called *Bernoulli sum* paradigm (discussed in depth in [9, 11]) defined as

$$B_n = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}f(k)$$

where $f(k)$ is a suitable function. In general, at least for poly-log functions $f$

$$B_n \sim f(\lfloor np\rfloor),$$

however, a more sophisticated analysis (i.e., singularity analysis or analytic depoissonization) is required to find second order asymptotic terms, as we aspire here. For example, in [9, 11] it is shown that

$$\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\log k = \log(pn) + \frac{p-1}{2pn} - \frac{p^2-6p+5}{12p^2n^2} + O(n^{-3}).$$

The second sum $b_n$ requires a different approach that falls under the *Bernoulli distributed sequences modulo 1* scenario. The reader is referred to [5, 8, 6, 15] for a detailed discussion.

In order to evaluate these sums we first need to estimate asymptotically $A_k$ around $k = np$ which is presented in the next lemma. Throughout the paper we shall use small positive constants $\delta > 0$ and $\varepsilon > 0$ that can change from line to line.

**Lemma 1** *For large $n$ and $p < 1/2$*

$$A_{np} = \frac{1-p}{1-2p}\frac{1}{\sqrt{2\pi np(1-p)}}2^{nh(p)}\left(1+O(n^{-1/2})\right) \tag{12}$$

*where $h(p)$ is the binary entropy. More precisely, for an $\varepsilon > 0$ and $k = np + \Theta(n^{1/2+\varepsilon})$ we have*

$$A_k = \frac{1-p}{1-2p}\frac{1}{\sqrt{2\pi np(1-p)}}\left(\frac{1-p}{p}\right)^k \frac{1}{(1-p)^n}\exp\left(-\frac{(k-np)^2}{2p(1-p)n}\right)\left(1+O(n^{-\delta})\right) \quad (13)$$

*for some $\delta > 0$.*

**Proof**. We use the saddle point method [16]. Let's first define the generating function of $A_k$, that is,

$$A_n(z) = \sum_{k=0}^{n} A_k z^k = \frac{(1+z)^n - 2^n z^{n+1}}{1-z}.$$

Thus by Cauchy's formula [16]

$$\begin{aligned} A_k &= \frac{1}{2\pi i}\oint \frac{(1+z)^n - 2^n z^{n+1}}{1-z}\frac{dz}{z^{k+1}} \\ &= \frac{1}{2\pi i}\oint \frac{1}{1-z} 2^{n\log(1+z)-(k+1)\log z} dz. \end{aligned}$$

Define $H(z) = n\log(1+z) - (k+1)\log z$. The saddle point $z_0$ solves $H'(z_0) = 0$, and one finds $z_0 = (k+1)/(n-k+1) = p/(1-p)$ and $H''(z_0) = q^3/p$. Thus by the saddle point method

$$A_k = \frac{1}{1-z_0}\frac{1}{\sqrt{2\pi n H''(z_0)}} 2^{nH(z_0)}(1+O(n^{-1/2})).$$

This proves (12). In a similar manner, as shown in [5], we establish (13). ∎

We also need to approximate the binomial distribution around the mean. We shall use the following well known lemma that is a simple consequence of Stirling's approximation.

**Lemma 2** *Let $p_n(k) = \binom{n}{k} p^k q^{n-k}$ where $q = 1-p$ be the binomial distribution. Then for $|k - pn| \le n^{1/2+\varepsilon}$ we have*

$$p_n(k) = \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{(k-pn)^2}{2p(1-p)n}\right) + O(n^{-\delta}) \quad (14)$$

*uniformly as $n \to \infty$. Furthermore*

$$\sum_{|k-np| > \sqrt{p(1-p)} n^{1/2+\varepsilon}} p_n(k) < 2n^{-\varepsilon} e^{-n^{2\varepsilon}/2} \quad (15)$$

*for large $n$.*

**Proof**. These are standard estimates that can be found in textbooks; formula (14) is the local limit theorem for the binomial distribution [2, 5], and (15) is the tail of the binomial distribution (cf. Corollary 1.4 in [2]). ∎

**Remark**. Observe that by (12) and (13) we have

$$\frac{A_k}{\binom{n}{k}} = \frac{1-p}{1-2p} + O(n^{-\delta}) \quad (16)$$

$$\frac{A_{k-1}}{\binom{n}{k}} = \frac{p}{1-2p} + O(n^{-\delta}) \quad (17)$$

for $k = np + O(n^{1/2+\varepsilon})$ since by Stirling's approximation

$$\binom{n}{np} = \frac{1}{\sqrt{2\pi np(1-p)}} 2^{nh(p)} \left(1 + O(n^{-1/2})\right).$$

Now we are in a position to estimate $a_n$ and $b_n$. Observe that based on (15) of Lemma 2 we can restrict the sum to $|k - pn| \le n^{1/2+\varepsilon}$. In fact, by Lemma 1 we have

$$\log A_k = \log A_{np} + \alpha(k - np) - \frac{(k-np)^2}{2pqn \ln 2} + O(n^{-\delta}). \quad (18)$$

Using Lemma 2 we arrive at

$$a_n = \log A_{np} - \frac{1}{2 \ln 2} + O(n^{-\delta})$$

after applying (12).

Now, we deal with the second sum (11), namely

$$b_n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \langle \log_2 A_k \rangle$$

which falls under the Bernoulli distributed sequences modulo 1 methodology, as discussed in [8, 6, 15]. This technique was already used in [6, 15] to estimate the redundancy of the Huffman code and arithmetic codes. Observe first that from (13) we find for $|k - pn| \le n^{1/2+\varepsilon}$

$$\log A_k = \alpha k + n\beta - \log_2 \omega\sqrt{n} - \frac{(k-np)^2}{2pqn \ln 2} + O(n^{-\delta})$$

where $\omega = (1-2p)\sqrt{2\pi pq}/(1-p)$. In order to estimate $b_n$ we need to understand the asymptotics of the following sum

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left\langle \alpha k + n\beta - \log_2 \omega\sqrt{n} - \frac{(k-np)^2}{2pqn \ln 2} \right\rangle.$$

The asymptotics of the above sum depend upon rationality or irrationality of $\alpha$ as proved in [6] (cf. also [8, 15]). In fact, the next lemma follows directly from the analysis of [6].

**Lemma 3** *Let $0 < p < 1$ be a fixed real number and $f : [0, 1] \to \mathbf{R}$ be a Riemann integrable function.*

(i) *If $\alpha$ is irrational, then*

$$\lim_{n\to\infty}\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}f\left(\left\langle k\alpha+y-(k-np)^2/(2pqn\ln 2)\right\rangle\right)=\int_0^1 f(t)\,dt, \qquad (19)$$

*where the convergence is uniform for all shifts $y\in\mathbf{R}$.*

(ii) *Suppose that $\alpha=\frac{N}{M}$ is a rational number with integers $N,M$ such that* $\gcd(N,M)=1$*. Then uniformly for all $y\in\mathbf{R}$*

$$\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}f\left(\left\langle k\alpha+y-(k-np)^2/(2pqn\ln 2)\right\rangle\right)=\int_0^1 f(t)\,dt+G_M(y) \qquad (20)$$

*where*

$$G_M(y)[f]:=\frac{1}{M}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}\left(\left\langle M\left(y-\frac{x^2}{2\ln 2}\right)\right\rangle-\int_0^1 f(t)\,dt\right)dx$$

*is a periodic function with period $\frac{1}{M}$.*

Using this lemma we immediately show that for $\alpha$ irrational

$$b_n=\frac{1}{2}+o(1),$$

while for $\alpha=N/M$ we have

$$b_n=\frac{1}{2}+G_M\left(\beta n-\log_2\omega\sqrt{n}\right)[x]=\frac{1}{2}+H_M(n\beta)[x]+o(1),$$

where $H_M(y)[f]$ is defined in Theorem 1.

Now we consider term (6) which we split into two terms

$$\begin{aligned}
c_n &= \sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k}\frac{1+A_{k-1}}{\binom{n}{k}}\left(1+\log_2(1+\binom{n}{k}A_{k-1}^{-1})\right)+o(1),\\
d_n &= \sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k}\frac{1+A_{k-1}}{\binom{n}{k}}\left(\langle-\log_2 A_{k-1}+2\rangle+\langle\log_2 A_k\rangle\right).
\end{aligned}$$

By (16)-(17), and since by Lemma 2 the sum for $c_n$ is negligible for $|k-np|>n^{1/2+\varepsilon}$, we immediately obtain

$$c_n=\frac{p}{1-2p}+\frac{p}{1-2p}\log_2\frac{1-p}{p}+o(1).$$

Finally, by Lemma 3 and the above we conclude that

$$d_n=\frac{p}{1-2p}\left(H_M(n\beta)[x]+H_M(n\beta-\alpha)[-x]\right)+o(1).$$

for $\alpha$ rational, and $d_n=o(1)$ for $\alpha$ irrational.

Thus, to complete the proof of Theorem 1 we need to evaluate (7) which we recall below

$$e_n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \frac{A_{k-1}}{\binom{n}{k}} \left( 2^{-\langle \log_2 A_k \rangle + 1} - 2^{\langle -\log_2 (A_{k-1}+2) \rangle} \right).$$

This sum can be estimated using Lemmas 2 and 3. In particular,

$$\begin{aligned}
\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \frac{A_{k-1}}{\binom{n}{k}} 2^{-\langle \log_2 A_k \rangle} &= \sum_{|k-np| \le n^{1/2+\varepsilon}} \binom{n}{k} p^k q^{n-k} \frac{A_{k-1}}{\binom{n}{k}} 2^{-\langle \log_2 A_k \rangle} + o(1) \\
&= \frac{p}{1-2p} \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \log_2 A_k \rangle} + o(1) \\
&= \frac{p}{1-2p} \frac{1}{2 \ln 2} + \frac{p}{1-2p} H_M(n\beta)[2^{-x}] + o(1)
\end{aligned}$$

where the last expression follows from

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \log_2 A_k \rangle} = \frac{1}{2 \ln 2} + H_M(n\beta)[2^{-x}] + o(1)$$

by Lemma 3 as in [15]. This also provides asymptotics of the term like (5) and (7), that is

$$e_n = \frac{p}{1-2p} \left( 2H_M(n\beta)[2^{-x}] + H_M(n\beta - \alpha)[2^{x}] \right) + o(1)$$

for $\alpha$ rational, and $e_n = o(1)$ for $\alpha$ irrational. Finally, one easily see that the last sum (8) is $o(1)$. This completes the proof of Theorem 1 since the case $p = 0.5$ is trivial.

## Acknowledgment

The author thanks Prof. M. Drmota (TU Wien) for many useful discussions and Prof. Sergio Verdú for his encouragements.

## 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] B. Bollobás, *Random Graphs*, Academic Press, London 1985.

[3] T. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York 1991.

[4] J Cheng, T K Huang, C Weidmann, New Bounds on the Expected Length of Optimal One-to-One Codes, *IEEE Trans. Information Theory*, 53, 1884 - 1895, 2007.

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

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

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

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

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

[10] P. Flajolet and B. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2008.

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

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

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

[14] S. Savari and A. Naheta, Bounds on the Expected Cost of One-to-One Codes, *2004 ISIT*, p. 92, Chicago, 2004.

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

[16] W. Szpankowski, *Average Case Analysis of Algorithms in Sequences*, John Wiley & Sons, New York, 2001.

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

**Wojciech Szpankowski** is a a Professor of Computer Science and (by courtesy) Electrical and Computer Engineering at Purdue University where he teaches and conducts research in analysis of algorithms, information theory, bioinformatics, analytic combinatorics, random structures, and stability problems of distributed systems. He held several Visiting Professor/Scholar positions, including McGill University, INRIA, France, Stanford, Hewlett-Packard Labs, Universite de Versailles, and University of Canterburry, New Zealand. He is a Fellow of IEEE, the Erskine Fellow, and the Humboldt Fellow. In 2001 he published the book "Average Case Analysis of Algorithms on Sequences", John Wiley & Sons, 2001. He has been a guest editor and an editor of technical journals, including THEORETICAL COMPUTER SCIENCE, the ACM TRANSACTION ON ALGORITHMS, the IEEE TRANSACTIONS ON INFORMATION THEORY, FOUNDATION AND TRENDS IN COMMUNICATIONS AND INFORMATION THEORY, and COMBINATORICS, PROBABILITY, AND COMPUTING. He chaired a number of conferences, including *Analysis of Algorithms*, Gdansk and Berkeley, the *NSF Workshop on Information Theory and Computer Science Interface*, Chicago, and the workshop *Information Beyond Shannon*, Orlando and Venice.