Precise Average Redundancy of an Idealized Arithmetic Coding

Michael Drmota	Hsien-Kuei Hwang	Wojciech Szpankowski*
Institut für Geometrie	Institue of Statistical Science	Dept. of Computer Science
TU Wien	Academia Sinica	Purdue University
A-1040 Wien	11529 Taipei	W. Lafayette, IN 47907
Austria	Taiwan	U.S.A.

Abstract

Redundancy is defined as the excess of the code length over the optimal (ideal) code length. We study the *average* redundancy of an *idealized arithmetic coding* (for memoryless sources with unknown distributions) in which the Krichevsky and Trofimov estimator is followed by the Shannon–Fano code. We shall ignore here important practical implementation issues such as finite precisions and finite buffer sizes. In fact, our idealized arithmetic code can be viewed as an adaptive infinite precision implementation of arithmetic encoder that resembles Elias coding. However, we provide very precise results for the average redundancy that takes into account integer–length constraints. These findings are obtained by analytic methods of analysis of algorithms such as theory of distribution of sequences modulo 1 and Fourier series. These estimates can be used to study the average redundancy of codes for tree sources, and ultimately the context-tree weighting algorithms.

1 Introduction

Recent years have seen a resurgence of interest in redundancy rates of lossless coding (cf. [1, 9, 10, 12, 13, 14]). The redundancy rate problem for a class of sources corresponds to determining how much the actual code length exceeds the optimal (ideal) code length. We define a code $C_n : \mathcal{A}^n \to \{0,1\}^*$ as a mapping from the set $\mathcal{A}^n$ of all sequences of length n over the alphabet $\mathcal{A}$ to the set $\{0,1\}^*$ of binary sequences. We write X_1^n to denote the random variable representing a message of length n. Given a probabilistic source model, we let $P(x_1^n)$ be the probability of the message $x_1^n \in \mathcal{A}^n$. Given a code C_n, we let $L(C_n, x_1^n)$ be the code length for x_1^n. Throughout we shall write log for the binary logarithm.

From Shannon's works we know that the entropy $H_n(P) = -\sum_{x_1^n} P(x_1^n) \log P(x_1^n)$ is an absolute lower bound on the expected code length. Hence $-\log P(x_1^n)$ can be viewed as the "ideal" code length. The *pointwise redundancy* $R_n(C_n, P; x_1^n)$ and the *average redundancy* $\overline{R}_n(C_n, P)$ are defined as

$$\begin{aligned} R_n(C_n, P; x_1^n) &= L(C_n, x_1^n) + \log P(x_1^n), \\ \overline{R}_n(C_n) &= \mathbf{E}_{X_1^n}[R_n(C_n, P; X_1^n)] = \mathbf{E}_{X_1^n}[L(C_n, X_1^n)] - H_n(P), \end{aligned}$$

where the underlying probability measure P represents a particular source model and $\mathbf{E}$ denotes the expectation. Another natural measure of code performance is the *maximal* redundancy defined as $R^*(C_n, P) = \max_{x_1^n}\{R_n(C_n, P; x_1^n)\}$. The *redundancy rate problem*

*This research was supported in part by NSF Grants NCR-9415491 and C-CR-9804760, and contract 1419991431A from sponsors of CERIAS at Purdue.

consists in determining for a class $\mathcal{S}$ of source models the growth rate of

$$R_n^*(\mathcal{S}) = \min_{C_n} \max_{P\in\mathcal{S}} \{R_n^*(C_n, P)\}, \tag{1}$$

$$\overline{R}_n(\mathcal{S}) = \min_{C_n} \max_{P\in\mathcal{S}} \{\overline{R}_n(C_n, P)\}. \tag{2}$$

In this paper, we investigate the average redundancy of arithmetic coding [6] for memoryless sources with *unknown* parameters. Here, we analyze an idealized arithmetic coding in which finite precision and finite buffer sizes are not taken into account. Following [16] we assume that the idealized arithmetic encoding consists of the Krichevsky and Trofimov estimator followed by the Shannon–Fano code (however, using our recent results [14] we could replace the Shannon–Fano code with the Huffman code at the cost of significant complication of the analysis). Arithmetic coding is one of the most popular entropy encoding that virtually appears in every multimedia compression scheme (cf. [5]). It has been known (cf. [16]) that the average redundancy of arithmetic encoding is $O(\log n)$ for source strings of length n, however, to the best of our knowledge no precise estimates are available (cf. [11] for similar results). Here, we present precise asymptotics for the average redundancy $\overline{R}_n^{AC}$ of arithmetic coding and the Krichevsky and Trofimov estimator (KT-estimator) [7]. As a consequence, we can estimate the average redundancy $\overline{R}_n^T$ of codes for tree sources [16] and the context-tree weighting algorithm CTW proposed by Willems, Shtarkov, and Tjalkens [16]. The evaluation of redundancy of latter codes is our ultimate goal, but in this conference version we will not elaborate on these issues.

We now briefly summarize our results. For a sequence x_1^n generated by a memoryless source with unknown parameter θ (i.e., $P(x_1^n) = \theta^k(1-\theta)^{n-k}$), the average redundancy of arithmetic coding (that applies Shannon-Fano code on the top of the KT-estimator) is asymptotically (as $n \to \infty$) equal to (cf. Theorem 1)

$$\overline{R}_n^{AC} = \frac{1}{2}\log n - \frac{1}{2}\log\frac{\pi e}{2} + 2 - E_n + O(n^{-1/2}),$$

where E_n exhibits an "erratic" behavior that depends whether $\log\frac{1-\theta}{\theta}$ is rational or irrational. A graph of E_n is shown in Figure 1. Actually, $E_n \approx \frac{1}{2}$, however, the exact behavior is much more complicated (cf. Theorem 2). We observe that the leading term of $\overline{R}_n^{AC}$ is optimal (cf. [10]) while the constant term is not.

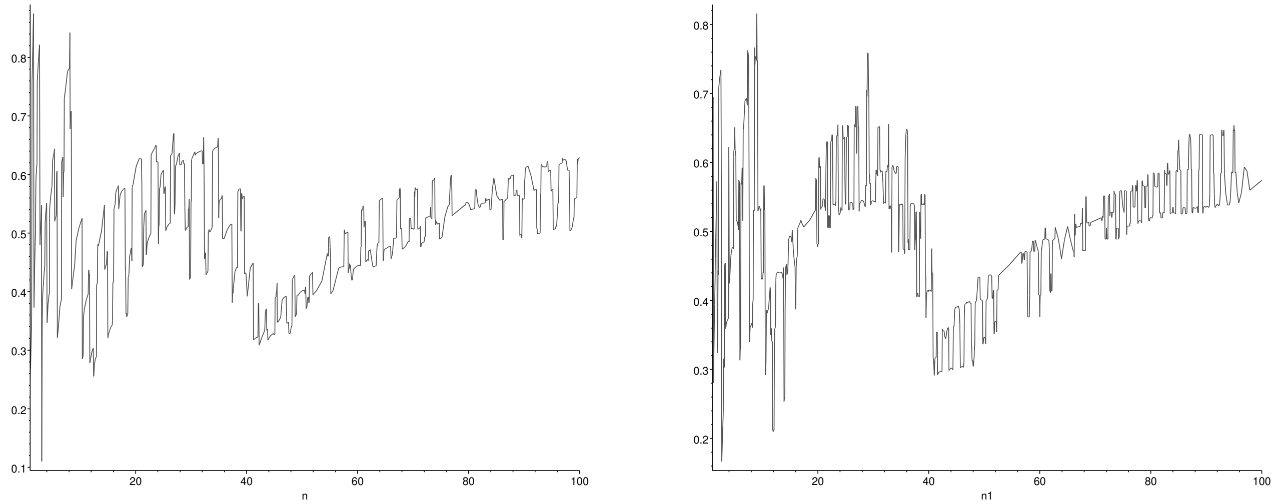
As a simple consequence of the above result, one can obtain the average redundancy of arithmetic coding for tree sources $\mathcal{S}$ (see [16] for a definition) as follows

$$\overline{R}_n^T = \frac{S}{2}\log n + \frac{1}{2}\sum_{1\le j\le S}\log p_j - \frac{S}{2}\log\frac{\pi e}{2} + 2S - E_n' + O(n^{-1/2}),$$

where $S = |\mathcal{S}|$, p_j is the probability of the jth suffix occurrence, (where the j suffix belongs to $\mathcal{S}$), and E_n' is the erratic part of the redundancy (again $E_n' \approx \frac{S}{2}$).

The erratic behavior of the redundancy seems to be a rule rather than an exception. We have already observed this in the redundancy of the Lempel-Ziv code and the Tunstall code (cf. [9, 12]). Actually, one does not need to look too far since the simplest code, that of Shannon-Fano, exhibits the same kind of behavior. The average redundancy $\overline{R}_n^{SF}$ of the Shannon-Fano code for a *known* memoryless source (i.e., for $P(x_1^n) = \theta^k(1-\theta)^{n-k}$ with known θ) can be computed as

$$\overline{R}_n^{SF} = 1 + \sum_{k=0}^{n}\binom{n}{k}\theta^k(1-\theta)^{n-k}\left(\lceil -\log(\theta^k(1-\theta)^{n-k})\rceil + \log(\theta^k(1-\theta)^{n-k})\right).$$



(a) (b)

Figure 1: The "erratic" part, E_n, of the average redundancy of the KT estimator versus n for: (a) irrational $\theta = 1/\pi$; (b) rational $\theta = 1/2$.

In [14] Szpankowski proved that $\overline{R}_n^{SF}$ for large n behaves as follows

$$\overline{R}_n^{SF} = \begin{cases} \frac{3}{2} + o(1) & \alpha = \log(1-\theta)/\theta \quad \text{irrational} \\ \\ \frac{3}{2} - \frac{1}{M}\left(\langle Mn\beta\rangle - \frac{1}{2}\right) + O(\rho^n) & \alpha = \frac{N}{M} \quad \text{rational} \end{cases}$$

where $\beta = -\log(1-\theta)$, the integers M, N are such that $\gcd(N, M) = 1$, and $\langle x \rangle = x - \lfloor x \rfloor$ is the fractional part of x. The same type of behavior is exhibited in the Huffman code redundancy as shown in [14].

2 Main Results

In this section we formulate precisely our results focusing here on the average redundancy of the KT-estimator. Below, we provide only a sketch of the proof delaying details to the next section.

Let x_1^n be a binary sequence of length n generated by a memoryless(θ) source with k "1" and $n-k$ "0", that is, $P(x_1^n) = \theta^k(1-\theta)^{n-k}$. It is assumed that θ is *unknown*. Therefore, to estimate the probability $P(x_1^n)$ we shall use the KT estimator [7, 16] defined as

$$P_e(k, n-k) := \frac{\Gamma(k+1/2)\Gamma(n-k+1/2)}{\pi\Gamma(n)}.$$

To generate an arithmetic encoding, we apply the Shannon-Fano code (cf. [2, 6]) for the probability distribution $P_e(k, n-k)$. That is, the code length L_n is $L_n = \lceil -\log P_e(k, n-k)\rceil + 1$. The average redundancy of the arithmetic coding therefore becomes

$$\overline{R}_n^{AC} = 1 + \sum_{k=0}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \left(\lceil -\log P_e(k, n-k) \rceil + \log \theta^k (1-\theta)^{n-k} \right).$$

Using $\lceil -x \rceil = -x + 1 - \langle -x \rangle$, where $\langle x \rangle$ is the fractional part of x, we reduce the above to the following

$$\overline{R}_n^{AC} = 2 + \sum_{k=0}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \log \frac{\theta^k (1-\theta)^{n-k}}{P_e(k, n-k)} - E_n,$$

where

$$E_n = \sum_{k=0}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \langle -\log P_e(k, n-k) \rangle.$$

Our main result is formulated next.

Theorem 1 *Consider arithmetic coding over memoryless(θ) source. Then*

$$\overline{R}_n^{AC} = \frac{1}{2} \log n - \frac{1}{2} \log \frac{\pi e}{2} + 2 - E_n + O(n^{-1/2}) \tag{3}$$

where E_n behavior depends whether $\gamma = \log \frac{1-\theta}{\theta}$ is rational or not, that is:

(i) *If $\gamma = \log \frac{1-\theta}{\theta}$ is* **rational**, *i.e. $\gamma = \frac{N}{M}$ for some positive integers M, N with* $\gcd(M, N) =$ 1, *then*

$$E_n = \frac{1}{2} + G_M \left(-\log(1-\theta)n + \frac{1}{2} \log \frac{\pi n}{2} \right) + o(1) \tag{4}$$

as $n \to \infty$, where

$$G_M(y) := \frac{1}{M} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left(\left\langle M \left(y - \frac{x^2}{2} \right) \right\rangle - \frac{1}{2} \right) dx$$

is a periodic function with period $\frac{1}{M}$ and maximum $\max |G_M| \le \frac{1}{2M}$.

(ii) *If $\gamma = \log \frac{1-\theta}{\theta}$ is* **irrational**, *then*

$$E_n = \frac{1}{2} + o(1) \tag{5}$$

as $n \to \infty$.

Sketch of Proof. Here we only sketch how to estimate the main part of $\overline{R}_n^{AC}$ delaying the derivation of E_n the next section. In the derivation of E_n, we shall use discrepancy theory and uniformly distributed sequences modulo 1 (cf. [3, 8, 15]).

Our proof first approximates the binomial distribution by its Gauss density, and then estimates the sum by the Gaussian integral, coupling with large deviations of the binomial distribution. By Stirling's formula, we have

$$\log \frac{\theta^k (1-\theta)^{n-k}}{P_e(k, n-k)} = \frac{1}{2} \log n + \frac{1}{2} \log \frac{\pi}{2} - \frac{x^2}{2} + O((|x| + |x|^3) n^{-1/2}),$$

for $k = \theta n + x\sqrt{\theta(1-\theta)n}$ and $x = o(n^{1/6})$. Note that the left-hand side is bounded above by $\frac{1}{2} \log n + 1/2$ for $n \ge 2$ and $k \ne 0, n$. This follows easily from the identity

$$\Gamma(n + 1/2) = \frac{(2n)! \sqrt{\pi}}{4^n \, n!} \qquad (n \ge 0),$$

and the inequalities

$$\sqrt{2\pi n}(n/e)^n \le n! \le e^{1/12}\sqrt{2\pi n}(n/e)^n, \qquad (n \ge 1).$$

On the other hand, by using the local limit theorem

$$\binom{n}{k}\theta^k(1-\theta)^{n-k} = \frac{e^{-x^2/2}}{\sqrt{2\pi\theta(1-\theta)n}}\left(1 + O((1+|x|^3)n^{-1/2})\right), \tag{6}$$

uniformly for $x = o(n^{1/6})$, we deduce that

$$\overline{R}_n^{AC} - E_n = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\left(\frac{1}{2}\log n + \frac{1}{2}\log\frac{\pi}{2} - \frac{x^2}{2}\right)\,dx + O(n^{-1/2}),$$

A straightforward evaluation of the integral leads to (3). The error term can be further refined by expanding more terms in the above, but this will not be used. ■

Remark: Note that it is an easy exercise to derive the Fourier expansion for $G_M(y)$. Its mean value vanishes. Therefore, also in the rational case, E_n varies around $\frac{1}{2}$. However, the essential difference between the rational and the irrational case is that in the rational case E_n never converges (cf. Figure 1). (If $-\log(1-\theta)$ is irrational then the sequence $x_n = -\log(1-\theta)\,n + \frac{1}{2}\log\frac{\pi n}{2}$ is uniformly distributed modulo 1, and if $-\log(1-\theta)$ is rational then x_n is not uniformly distributed modulo 1 but dense in the unit interval.)

3 Derivation of E_n

Our goal is to estimate

$$E_n := \sum_{0\le k\le n} p_n(k)\langle -\log P_e(k, n-k)\rangle,$$

where

$$p_n(k) = p_n(k,\theta) := \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

and

$$P_e(a,b) := \frac{\Gamma(a+\frac{1}{2})\Gamma(b+\frac{1}{2})}{\pi\Gamma(a,b)}.$$

The main result, formulated below, is a consequence of applying analytic tools such as theory of distribution of sequences modulo 1 and Fourier series, as already advocated in [14]. The interested reader is referred to [3, 8, 15].

Theorem 2 (i) *If $\gamma = \log\frac{1-\theta}{\theta}$ is rational, i.e. $\gamma = \frac{N}{M}$ for some positive integers M, N with* $\gcd(M,N) = 1$, *then (4) holds, that is,*

$$E_n = \frac{1}{2} + G_M\left(-\log(1-\theta)n + \frac{1}{2}\log\frac{\pi n}{2}\right) + o(1)$$

as $n \to \infty$, where

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

is a periodic function with period $\frac{1}{M}$ *and maximum* $\max|G_M| \leq \frac{1}{2M}$.

(ii) *If* $\gamma = \log\frac{1-\theta}{\theta}$ *is irrational then*

$$E_n = \frac{1}{2} + o(1) \tag{7}$$

as $n \to \infty$.

We start with the following lemma.

Lemma 1 *Set* $\alpha := -\theta\log\theta - (1-\theta)\log(1-\theta)$. *Then if* $|k - \theta n| \leq n^{7/12}$ *we have*

$$-\log P_e(k, n-k) = \alpha n + \frac{1}{2}\log\frac{\pi n}{2} + \gamma(k-\theta n) - \frac{1}{2\theta(1-\theta)}\frac{(k-\theta n)^2}{n} + \mathcal{O}\left(n^{-1/4}\right)$$

uniformly as $n \to \infty$.

Proof. Stirling's approximation formula. ■

Lemma 2 *If* $|k - \theta n| \leq n^{7/12}$ *we have*

$$p_n(k,\theta) = \frac{1}{\sqrt{2\pi\theta(1-\theta)n}}\exp\left(-\frac{(k-\theta n)^2}{2\theta(1-\theta)n}\right) + \mathcal{O}(n^{-3/4})$$

uniformly as $n \to \infty$.

Proof. Stirling's approximation formula. ■

Lemma 3 *Suppose that* M, N *are positive integers with* $\gcd(M,N) = 1$. *Then for every real number* x *we have*

$$\frac{1}{M}\sum_{v=0}^{M-1}\left(\left\langle x + v\frac{N}{M}\right\rangle - \frac{1}{2}\right) = \frac{1}{M}\left(\langle xM\rangle - \frac{1}{2}\right).$$

Proof. First observe that $\gcd(M,N) = 1$ ensures that the numbers $\langle N : 0 \leq v \leq M-1\rangle$ represent a complete set of residue classes modulo M. Therefore, the sum of interest does not change if we replace N by 1. Next, it is clear that the function of interest

$$f(x) := \frac{1}{M}\sum_{v=0}^{M-1}\left(\left\langle x + \frac{v}{M}\right\rangle - \frac{1}{2}\right)$$

is periodic with period $\frac{1}{M}$. Hence, it suffices to consider $f(x)$ for $0 \leq x < \frac{1}{M}$. In this range we can calculate $f(x)$ by

$$f(x) = \frac{1}{M}\sum_{v=0}^{M-1}\left(x + \frac{v}{M} - \frac{1}{2}\right) = \left(x - \frac{1}{2M}\right).$$

Finally, since $f(x) = f\left(\frac{\langle xM\rangle}{M}\right)$ and $0 \leq \frac{\langle xM\rangle}{M} < \frac{1}{M}$ we directly obtain the proposed representation for $f(x)$. ■

Now, we are in the position to **prove part (i)** of Theorem 2. In a first step we concentrate on k with $|k - \theta n| \leq n^{7/12}$ and subdivide those k into residue classes modulo

M. Afterwards we use Lemma 1 and 2 to approximate the sum by an Gaussian-like integral. Finally, we apply Lemma 3 to simplify the resulting integral:

$$
\begin{aligned}
E_n - \frac{1}{2} &= \sum_{|k-\theta n| \le n^{7/12}} p_n(k) \left(\langle -\log P_e(k, n-k) \rangle - \frac{1}{2} \right) + o(1) \\
&= \sum_{|k-\theta n| \le n^{7/12}} \frac{1}{\sqrt{2\pi\theta(1-\theta)n}} \exp\left(-\frac{(k-\theta n)^2}{2\theta(1-\theta)n} \right) \\
&\qquad \left(\left\langle \alpha n + \frac{1}{2}\log\frac{\pi n}{2} + \frac{M}{N}(k-\theta n) - \frac{1}{2\theta(1-\theta)} \frac{(k-\theta n)^2}{n} + \mathcal{O}\left(n^{-1/4}\right) \right\rangle - \frac{1}{2} \right) + o(1) \\
&= \sum_{v=0}^{M-1} \sum_{k \equiv v \bmod M} \frac{1}{\sqrt{2\pi\theta(1-\theta)n}} \exp\left(-\frac{(k-\theta n)^2}{2\theta(1-\theta)n} \right) \\
&\qquad \left(\left\langle (\alpha - \gamma\theta) n + \frac{1}{2}\log\frac{\pi n}{2} + \frac{M}{N}v - \frac{1}{2\theta(1-\theta)} \frac{(k-\theta n)^2}{n} + \mathcal{O}\left(n^{-1/4}\right) \right\rangle - \frac{1}{2} \right) + o(1) \\
&= \sum_{v=0}^{M-1} \sum_{u \in \mathbf{Z}} \frac{1}{\sqrt{2\pi\theta(1-\theta)n}} \exp\left(-\frac{M^2u^2}{2\theta(1-\theta)n} \right) \\
&\qquad \left(\left\langle (\alpha - \gamma\theta) n + \frac{1}{2}\log\frac{\pi n}{2} + \frac{M}{N}v - \frac{1}{2\theta(1-\theta)} \frac{M^2u^2}{n} + \mathcal{O}\left(n^{-1/4}\right) \right\rangle - \frac{1}{2} \right) + o(1) \\
&= \sum_{v=0}^{M-1} \frac{1}{\sqrt{2\pi\theta(1-\theta)}} \int_{-\infty}^{\infty} \exp\left(-\frac{M^2x^2}{2\theta(1-\theta)} \right) \\
&\qquad \left(\left\langle (\alpha - \gamma\theta) n + \frac{1}{2}\log\frac{\pi n}{2} + \frac{M}{N}v - \frac{M^2}{2\theta(1-\theta)} x^2 \right\rangle - \frac{1}{2} \right) dx + o(1) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \frac{1}{M} \sum_{v=0}^{M-1} \left(\left\langle (\alpha - \gamma\theta) n + \frac{1}{2}\log\frac{\pi n}{2} + \frac{M}{N}v - \frac{x^2}{2} \right\rangle - \frac{1}{2} \right) dx + o(1) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \frac{1}{M} \left(\left\langle M \left((\alpha - \gamma\theta) n + \frac{1}{2}\log\frac{\pi n}{2} - \frac{x^2}{2} \right) \right\rangle - \frac{1}{2} \right) dx + o(1).
\end{aligned}
$$

This proves (4).

We now concentrate on **proving part (ii)** of Theorem 2, that is, (7). We have to introduce some notation.

Definition 1 *Let $(x_k)_{k \ge 1}$ be a sequence of real numbers. The* **discrepancy** $D_N(x_k)$ *is defined by*

$$
D_N(x_k) := \sup_{0 \le y \le 1} \left| \frac{|\{n \le N : \langle x_k \rangle \in [0, y)\}|}{N} - y \right|
$$

and the **uniform discrepancy** $\tilde{D}_N(x_k)$ *by*

$$
\tilde{D}_N(x_k) := \sup_{\kappa \ge 0} \sup_{0 \le y \le 1} \left| \frac{|\{k \le N : \langle x_{k+\kappa} \rangle \in [0, y)\}|}{N} - y \right|.
$$

Furthermore, (x_n) is said to be **uniformly distributed modulo 1** *if* $\lim_{N \to \infty} D_N(x_k) = 0$ *and (x_k) is said to be* **well distributed modulo 1** *if* $\lim_{N \to \infty} \tilde{D}_N(x_k) = 0$.

In particular we will apply this concept for the Weyl sequence $x_k = \gamma k$.

Lemma 4 *Suppose that γ is an irrational real number. Then the sequence $x_k := \gamma k$ is well distributed modulo 1.*

Proof. Weyl criterion (cf. [3, 15]). ∎

We will also apply two standard tools of the theory of uniformly distributed sequences (cf. [3]).

Lemma 5 (Koksma inequality) *Suppose that $f : [0,1] \to \mathbf{R}$ is of bounded variation $V(f) = \int_0^1 d|f|$. Then, for all real sequences (x_k)*

$$\left| \frac{1}{N} \sum_{k=1}^{N} f(\langle x_k \rangle) - \int_0^1 f(t)\, dt \right| \leq V(f)\, D_N(x_k).$$

Lemma 6 *Suppose that (x_k) and (y_k) are real sequences. Then*

$$|D_N(x_k) - D_N(y_k)| \leq \operatorname{dist}\left(\{x_1, \ldots, x_N\}, \{y_1, \ldots, y_N\}\right),$$

where

$$\operatorname{dist}\left(\{x_1, \ldots, x_N\}, \{y_1, \ldots, y_N\}\right) := \min_{\pi \in S_N} \max_{1 \leq k \leq N} |\langle x_k - y_{\pi(k)} \rangle|,$$

and S_N denotes the set of all permutations π of the numbers $\{1, 2, \ldots, N\}$.

Furthermore we will use the following technical estimate.

Lemma 7 *For a real sequence (x_k) set $\delta_N := \sup_{N' \geq N} \tilde{D}_{N'}(x_k)$. Then the discrepancy of the sequence $y_k := x_k + \frac{k^2}{M}$ (with $M \geq 1$) can be estimated by*

$$D_N(y_k) \leq \inf_{L \geq 1} \left(2\delta_{N/(2L)} + \frac{N^2}{ML} \right).$$

Proof. For $0 \leq l < L$ set

$$\mathcal{N}_l := \{k \geq 1 : lN^2 \leq k^2 L \leq (l+1)N^2\}$$

and $N_l := |\mathcal{N}_l|$. Obviously, we have $N_0 \geq N_1 \geq \cdots \geq N_{L-1} \sim \frac{N}{2L}$. Since

$$\operatorname{dist}\left(\left\{ x_k + \frac{l}{L}\frac{N^2}{M} : k \in \mathcal{N}_l \right\}, \left\{ x_k + \frac{n^2}{M} : k \in \mathcal{N}_l \right\} \right) \leq \frac{N^2}{MN}$$

we obtain by Lemma 6

$$\left| D_{N_l}\left(\left\{ x_k + \frac{l}{L}\frac{N^2}{M} : k \in \mathcal{N}_l \right\} \right) - D_{N_l}\left(\left\{ x_k + \frac{k^2}{M} : k \in \mathcal{N}_l \right\} \right) \right| \leq \frac{N^2}{MN}.$$

Hence

$$\begin{aligned}
N\, D_N\left(x_k + \frac{k^2}{M} \right) &\leq \sum_{l=0}^{L-1} N_l\, D_{N_l}\left(\left\{ x_k + \frac{k^2}{M} : k \in \mathcal{N}_l \right\} \right) \\
&\leq \sum_{l=0}^{L-1} N_l \left(D_{N_l}\left(\left\{ x_k + \frac{l}{L}\frac{N^2}{M} : k \in \mathcal{N}_l \right\} \right) + \frac{N^2}{MN} \right) \\
&\leq \sum_{l=0}^{L-1} N_l \left(2\delta_{N_l} + \frac{N^2}{MN} \right) \leq N \left(2\delta_{N_{L-1}} + \frac{N^2}{MN} \right)
\end{aligned}$$

proves the lemma. ∎

Corollary 1 *Suppose that (x_k) is a well distributed sequence modulo 1. Then there exists a monotonely decreasing sequence (ε_N) with $\lim_{N\to\infty} \varepsilon_N = 0$ such that for all $N \le n^{7/12}$*

$$D_N\left(x_k + c_1\frac{k^2}{n} + \mathcal{O}(n^{-1/4})\right) \le \varepsilon_N. \tag{8}$$

Proof. By choosing $L = \lfloor\sqrt{N}\rfloor$ in Lemma 7 and by Lemma 6 we have

$$D_N\left(x_k + c_1\frac{k^2}{n} + \mathcal{O}(n^{-1/4})\right) \le 2\delta_{\lfloor\sqrt{N}/2\rfloor} + N^{-3/14} + \mathcal{O}(N^{-3/7})$$

Thus, we can choose

$$\varepsilon_N := 2\delta_{\lfloor\sqrt{N}/2\rfloor} + N^{-3/14} + C_1 N^{-3/7}$$

for some constant $C_1 > 0$. By construction we have $0 < \varepsilon_{N+1} < \varepsilon_N$ and $\lim_{N\to\infty} \varepsilon_N = 0$. ∎

Now we can complete the proof of Theorem 2. Set

$$T_1(n) := \sum_{0\le k<\lfloor\theta n\rfloor} p_n(k)\left(\langle -\log P_e(k, n-k)\rangle - \frac{1}{2}\right)$$

and

$$T_2(n) := \sum_{\lfloor\theta n\rfloor\le k\le n} p_n(k)\left(\langle -\log P_e(k, n-k)\rangle - \frac{1}{2}\right).$$

We show that $T_2(n) = o(1)$ as $n \to \infty$. (Of course, in the same fashion it is possible to prove $T_1(n) = o(1)$ which then completes the proof of Theorem 2.)

First of all we note that $\sum_{\lfloor\theta n\rfloor + n^{7/12}\le k\le n} p_n(k) = o(1)$ as $n \to \infty$. Thus, it suffices to consider the sum

$$\tilde{T}_2(n) := \sum_{\lfloor\theta n\rfloor\le k\le\lfloor\theta n\rfloor + n^{7/12}} p_n(k)\left(\langle -\log P_e(k, n-k)\rangle - \frac{1}{2}\right).$$

We further note that for $\lfloor\theta n\rfloor < k \le \lfloor\theta n\rfloor + n^{7/12}$ we have $p_n(k) > p_n(k+1)$, and that Lemma 5 (applied to the function $f(x) = \langle x+\delta\rangle - \frac{1}{2}$) and (8) imply that for $\lfloor\theta n\rfloor \le N \le \lfloor\theta n\rfloor + n^{7/12}$

$$\begin{aligned}\left|\sum_{\lfloor\theta n\rfloor\le k\le\lfloor\theta n\rfloor+N}\left(\langle -\log P_e(k, n-k)\rangle - \frac{1}{2}\right)\right| &\le N\, D_N\left(\gamma k + c_1\frac{k^2}{n} + \mathcal{O}(n^{-1/4})\right)\\ &\le N\,\varepsilon_N.\end{aligned}$$

Finally, partial summation (cf. [15]) yields

$$\begin{aligned}|\tilde{T}_2(n)| &= \left|\sum_{\lfloor\theta n\rfloor\le k\le\lfloor\theta n\rfloor+n^{7/12}} p_n(k)\left(\langle -\log P_e(k, n-k)\rangle - \frac{1}{2}\right)\right|\\ &\le p_n(\lfloor\theta n + n^{7/12}\rfloor)\lfloor n^{7/12}\rfloor\varepsilon_{\lfloor n^{7/12}\rfloor} + \sum_{\lfloor\theta n\rfloor\le k\le\lfloor\theta n\rfloor+n^{7/12}} (p_n(k) - p_n(k+1))(k - \lfloor\theta n\rfloor)\varepsilon_{k-\lfloor\theta n\rfloor}\\ &\le \varepsilon_{\lfloor n^{1/4}\rfloor}\left(p_n(\lfloor\theta n + n^{7/12}\rfloor)\lfloor n^{7/12}\rfloor + \sum_{\lfloor\theta n\rfloor\le k\le\lfloor\theta n\rfloor+n^{7/12}} (p_n(k) - p_n(k+1))(k - \lfloor\theta n\rfloor)\right)\end{aligned}$$

$$
\begin{aligned}
& + \sum_{\lfloor\theta n\rfloor \le k \le \lfloor\theta n\rfloor + n^{1/4}} (p_n(k) - p_n(k+1))(k - \lfloor\theta n\rfloor) \\
\le \;& \varepsilon_{\lfloor n^{1/4}\rfloor} \sum_{\lfloor\theta n\rfloor \le k \le \lfloor\theta n\rfloor + n^{7/12}} p_n(k) + n^{1/4} p_n(\lfloor\theta n\rfloor) \le \varepsilon_{\lfloor n^{1/4}\rfloor} + \mathcal{O}(n^{-1/4}) = o(1),
\end{aligned}
$$

which proves that $T_2(n) = o(1)$ as $n \to \infty$. As already mentioned exactly the same reasoning works for $T_1(n)$, too, and shows that $T_1(n) = o(1)$ as $n \to \infty$. ■

References

[1] A. Barron, J. Rissanen, and B. Yu, The Minimum Description Length Principle in Coding and Modeling, *IEEE Trans. Information Theory*, 44, 2743-2760, 1998.

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

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

[4] M. Drmota and W. Szpankowski, Generalized Shannon Code Minimizes the Maximal Redundancy, *Proc. LATIN'02*, Cancun, Mexico, 2002.

[5] J. Gibson, T. Berger, T. Lookabaugh, R. Baker *Multimedia Compression: Applications & Standards*, Morgan Kaufmann Publishers 1998.

[6] P. Howard and J. Vitter, Analysis of Arithmetic Coding for Data Compression, Brown University, Department of Computer Science, *Proc. Data Compression Conference*, 3–12, Snowbird 1991.

[7] R. Krichevsky and V. Trofimov, The Performance of Universal Coding, *IEEE Trans. Information Theory*, 27, 199-207, 1981.

[8] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. John Wiley & Sons, New York 1974.

[9] G. Louchard and W. Szpankowski, On the Average Redundancy Rate of the Lempel-Ziv Code, *IEEE Trans. Information Theory*, 43, 2–8, 1997.

[10] J. Rissanen, Complexity of Strings in the Class of Markov Sources, *IEEE Trans. Information Theory*, 30, 526–532, 1984.

[11] B. Ryabko, and A.N. Fionov, An Efficient Method for Adaptive Arithmetic Coding of Sources with Large Alphabets, *Problems of Information Transmission*, 35, 95–108, 1999.

[12] S. Savari, Redundancy of the Lempel-Ziv Incremental Parsing Rule, *IEEE Trans. Information Theory*, 43, 9–21, 1997.

[13] Y. Shtarkov, Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, 23, 175–186, 1987.

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

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

[16] F.M. Willems, Y. Shtarkov and T. Tjalkens, The Context–Tree Weighting Method: Basic Properties, *IEEE Trans. Information Theory*, 41, 653–664, 1995.

[17] J. Ziv, Back from Infinity: A Constrained Resources Approach to Information Theory, *IEEE Information Theory Society Newsletter*, 48, 30-33, 1998.