1999

Asymptotic Average Redundancy of Huffman (and Shannon-Fano) Block Codes

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
99-021

Szpankowski, Wojciech, "Asymptotic Average Redundancy of Huffman (and Shannon-Fano) Block Codes" (1999). *Department of Computer Science Technical Reports*. Paper 1452.
<https://docs.lib.purdue.edu/cstech/1452>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.
Please contact epubs@purdue.edu for additional information.

**SYMPTOTIC AVERAGE REDUNDANCY OF HUFFMANN
(AND SHANNON-FANO) BLOCK CODES**

Wojciech Spankowski

**Department of Computer Science
Purdue University
West Lafayette, IN 47907**

**CSD TR #99-021
July 1999**

Asymptotic Average Redundancy of Huffman (and Shannon–Fano) Block Codes

July 14, 1999

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

Abstract

We study the redundancy of Huffman code (which, incidentally, is as old as the author of this paper). It has been known from the inception of this code that in the worst case Huffman code redundancy — defined as the excess of the code length over the optimal (ideal) code length — is not more than one. Over more than forty years, insightful, elegant and useful constructions have been set up to determine tighter bounds on the redundancy. One should mention here Gallager's upper bound of $p_1 + 0.086$ (p_1 is the probability of the most likely symbol), and tighter bounds due to Capocelli, de Santis and others. However, to the best of our knowledge no precise asymptotic results have been reported in literature thus far. We consider here a memoryless binary source generating a sequence of length n distributed as *binomial*(n, p) with p being the probability of emitting 0. Based on recent result of Stubbley, we prove that for $p \neq 1/2$ the average redundancy $\bar{R}_n$ of the Huffman code as $n \rightarrow \infty$ becomes

$$\bar{R}_n \sim \begin{cases} \frac{3}{2} - \frac{1}{\ln 2} = 0.057304\dots & \alpha = \log_2(1-p)/p \text{ irrational} \\ \frac{3}{2} - \frac{1}{M} (\langle \beta M n \rangle - \frac{1}{2}) - \frac{1}{M(1-2^{-1/M})} 2^{-(n\beta M)/M} & \alpha = \frac{N}{M} \text{ rational} \end{cases}$$

where M, N are integers such that $\gcd(N, M) = 1$, $\langle x \rangle = x - [x]$ is the fractional part of x , and $\beta = -\log_2(1-p)$. The appearance of the fractal-like function $\langle \beta M n \rangle$ explains the erratic behavior of the Huffman redundancy, and its “resistance” to succumb to a precise analysis. In fact, from the above we also can recover the Gallager's upper bound. As a side result, we prove that the average redundancy of the Shannon–Fano code is

$$\bar{R}_n \sim \begin{cases} \frac{1}{2} & \alpha = \log_2(1-p)/p \text{ irrational} \\ \frac{1}{2} - \frac{1}{M} (\langle M n \beta \rangle - \frac{1}{2}) & \alpha = \frac{N}{M} \text{ rational} \end{cases}$$

as $n \rightarrow \infty$. These findings are obtained through analytic methods such as Fourier analysis and theory of distribution of sequences modulo 1.

Index Terms — Huffman code, Shannon–Fano code, average redundancy, Fourier analysis, distribution of sequences, Weyl's criterion.

*This research was supported in part by NSF Grants NCR-9415491 and C-CR-9804760.

1 Introduction

Since the appearance more than 45 years ago of Huffman's classical paper [12] on optimal variable length source coding, Huffman coding still remains one of the most familiar topics in information theory (cf. [1, 2, 3, 11, 17, 25, 26]). Recall that Huffman code is an iterative algorithm built over the associated Huffman tree, in which the two nodes with lowest weights are combined into a new node with a weight that is the sum of the weights of its two children. Such a construction is not unique and there are cases of Huffman codes that lead to a longer code word than the corresponding Shannon–Fano code (cf. [4]). Fortunately, with a simple modification to the Huffman algorithm, it is possible to construct a Huffman code that does have the average code words that are less than or equal to the Shannon–Fano code so that the longest code words are as short as possible (cf. [25]). In this paper, we deal with such modified Huffman codes.

While Huffman has already known that the average code length is asymptotically equal to the entropy of the source, the asymptotic performance of the code is still not fully understood. Here, we concentrate on the average redundancy of the Huffman code, and for the first time we present asymptotically precise results for block Huffman coding sequence binomially distributed. Before we describe our findings, we briefly discuss the redundancy rate problem for lossless coding.

Recent years have seen a resurgence of interest in redundancy rates of lossless coding; (cf. [5, 6, 9, 11, 13, 14, 16, 19, 20, 21, 22, 23, 24, 25, 27, 28, 30, 31]). 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 $C_n : \mathcal{A}^n \rightarrow \{0, 1\}^*$ 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 . Information-theoretic quantities are expressed in binary logarithms written $\lg := \log_2$. We also write $\log := \ln$.

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

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

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

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

as $n \rightarrow \infty$. In this paper, we investigate the average redundancy $\bar{R}_n^H := \bar{R}_n(\mathcal{H})$ of the Huffman code $\mathcal{H}$, and the average redundancy $\bar{R}_n^{SF} := \bar{R}_n(\mathcal{SF})$ of the Shannon-Fano code $\mathcal{SF}$.

To place our findings in a larger perspective, we briefly review some known *precise* (e.g., second-order asymptotics) results:

- If $\mathcal{M}$ is i.i.d. or the class of Markov chains, or more generally the process belongs to a finitely parametrizable class of dimension K , then Rissanen [19] established

$$\bar{R}_n(\mathcal{M}) \sim R_n^*(\mathcal{M}) \sim \frac{K}{2} \lg n.$$

It was also found in [28] that the next term of $\bar{R}_n(\mathcal{S})$ and of $R_n^*(\mathcal{S})$ is $O(1)$, while in [27] a full asymptotic expansion for $R_n^*(\mathcal{S})$ for memoryless sources over an m -ary alphabet was established, namely

$$\begin{aligned} R_n^*(\mathcal{M}) &= \frac{m-1}{2} \lg \left(\frac{n}{2} \right) + \lg \left(\frac{\sqrt{\pi}}{\Gamma(m/2)} \right) + \cdots + \frac{\Gamma(m/2)m}{3\Gamma(m/2-1/2)} \cdot \frac{\sqrt{2}}{\sqrt{n}} \\ &+ \left(\frac{3+m(m-2)(2m+1)}{36} - \frac{\Gamma^2(m/2)m^2}{9\Gamma^2(m/2-1/2)} \right) \cdot \frac{1}{n} + O\left(\frac{1}{n^{3/2}}\right) \end{aligned}$$

where $\Gamma(x)$ is the Euler gamma function.

- Csiszár and Shields [5] have studied order r Markov renewal sequences in which a 1 is inserted every $T_0, T_1, \dots$ of 0's, where $\{T_i\}$ is either an i.i.d. or Markov renewal or r -order Markov renewal process. We denote such sources as $\mathcal{R}_r$. The authors of [5] proved that $\bar{R}_n(\mathcal{R}_r) = R^*(\mathcal{R}_r) = \Theta(n^{(r+1)/(r+2)})$ for $r = 1, 2, \dots$ which reduces to $\Theta(\sqrt{n})$ when $r = 0$. Recently, Flajolet and Szpankowski [9] established a precise asymptotic expansion of $R_n^*(\mathcal{R}_0)$ for the renewal processes, namely

$$R_n^*(\mathcal{R}_0) = \frac{2}{\log 2} \sqrt{cn} - \frac{5}{8} \lg n + \frac{1}{2} \lg \log n + O(1)$$

where $c = \frac{\pi^2}{6} - 1 \approx 0.645$.

- Louchard and Szpankowski [16], Savari [20], Wyner [30], and Jacquet and Szpankowski [13] proved that Lempel-Ziv codes in the class of i.i.d. and Markov processes have either rate $\Theta(n/\log n)$ for LZ'78 or $\Theta(n \log \log n / \log n)$ for LZ'77 code. The bound $\Theta(n/\log n)$

cannot be improved to an asymptotic equivalence due to some “wobbles” appearing in this formula. More precisely, for a binary alphabet with 0’s generated with probability p and 1’s with probability $q = 1 - p$, the authors of [16] showed that

$$\bar{R}_n(\mathcal{LZ}) = H \left(2 - \gamma - \frac{1}{2H} h_2 + \omega - \delta(n) \right) \frac{n}{\log n} + O \left(\frac{n \log \log n}{\log^2 n} \right)$$

where $H = -p \log p - q \log q > 0$ is the entropy rate, $\gamma = 0.577 \dots$ is the Euler constant, $h_2 = p \log^2 p + q \log^2 q$, and

$$\omega = - \sum_{k=1}^{\infty} \frac{p^{k+1} \log p + q^{k+1} \log q}{1 - p^{k+1} - q^{k+1}}.$$

The function $\delta(x)$ fluctuates with mean zero and a tiny amplitude when $\log p / \log q$ is rational (the amplitude of $\delta(x)$ is smaller than 10^{-6} for the unbiased memoryless source) but satisfies $\lim_{x \rightarrow \infty} \delta(x) = 0$ otherwise.

- Savari and Gallager [22] and Savari [21] analyzed Tunstall’s variable-to-fixed codes for memoryless and Markovian sources. For memoryless binary source, it was proved that

$$\bar{R}_n(\mathcal{T}) \sim - \frac{H \lg H + 0.5 h_2}{\lg n}$$

provided $B = \frac{\log p}{\log q}$ is irrational, where h_2 is defined as above. The authors of [21, 22] do not address the case when B is rational, but one expects some fluctuation in this case. We observe that B is irrational for uncountable number of p , and it is rational for only countable number of p .

To the best of our knowledge, no asymptotic results have been reported in literature on the average redundancy of Huffman codes. However, many elegant, insightful and useful lower and upper bounds on $\bar{R}_n^H = \bar{R}_n(\mathcal{H})$ are known. Gallager [11] proved that $\bar{R}_n(\mathcal{H}) \leq p_1 + \lg(2(\log e)/e) \approx p_1 + 0.086$ where p_1 is the probability of the most likely symbol. This bound was further improved by Capocelli and de Santis [2, 3], Manstetten [17], Stubble [25, 26] and others (cf. [1]). To understand the level of difficulty in establishing such asymptotics, we use the tight bounds of Stubble [25] and plot in Figure 1 the average redundancy $\bar{R}_n^H$ for the Huffman block code of length n binomially distributed with p denoting the probability of emitting 0 and $q = 1 - p$ representing the probability of generating 1. In Figure 1(a) we consider $\alpha = \lg(1 - p)/p$ irrational while in Figure 1(b) α is rational. The distinct characteristics of these two curves are clearly visible. The function in Figure 1(a) seems to converge to a constant (≈ 0.05) for large n , while the curve in Figure 1(b) is quite erratic (with the maximum close to Gallager’s upper bound 0.086!). We shall theoretically justify this behavior in Theorem 1

Figure 1: The average redundancy of Huffman codes versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

of Section 2 (cf. abstract for a brief summary of our main theoretical finding regarding the Huffman code redundancy).

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. Actually, one does not need to look too far since the simplest code, that of Shannon-Fano, exhibits the same kind of behavior. In this case, the average redundancy $\bar{R}_n^{SF}$ can be computed as

$$\bar{R}_n^{SF} = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \left(\lceil -\log_2(p^k (1-p)^{n-k}) \rceil + \log_2(p^k (1-p)^{n-k}) \right).$$

In Figure 2 we plot $\bar{R}_n^{SF}$ for $\alpha = \lg(1-p)/p$ irrational and rational. Again, the curves exhibit different characteristics. In Theorem 2 of Section 2 we provide a theoretical justification for this behavior. Our main result concerning the redundancy of Shannon-Fano codes is also reported in the abstract.

The rest of the paper is organized as follows. In the next section we present our main results. We prove them in the last two sections. In Section 3 we discuss the redundancy of the Shannon-Fano code proving Theorem 2, while in Section 4 we derive the formula for the Huffman code redundancy showing Theorem 1. We establish our results by analytic methods.

Figure 2: The average redundancy of Shannon-Fano codes versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

In particular, we apply Fourier series (cf. [33]) and theory of sequences *uniformly distributed modulo 1* (cf. [8, 15]). These techniques seem to have other applications in information theory. For example, we expect to use them in a forthcoming paper on the analysis of the Context-Tree Weighting scheme [29].

2 Main Results

We start with a brief description of Stubble's [25] tight bounds on the Huffman redundancy for binomially distributed blocks of length n . Let p denote the probability of generating a 0 and $q = 1 - p$ denote the probability of emitting a 1. Throughout, we assume that $p \neq \frac{1}{2}$, however, to simplify some derivations we set $p < \frac{1}{2}$. Certainly, this does not restrict the generality of the analysis. We also write $p(k) = p^k(1-p)^{n-k} = p^k q^{n-k}$.

Let $\mathbf{E}[L_n]$ denote the average length of the (modified) Huffman code which can be written as

$$\mathbf{E}[L_n] = \sum_{k=0}^n \binom{n}{k} p(k) L(k),$$

where

$$L(k) = \frac{1}{\binom{n}{k}} \sum_{j \in S_k} l_j$$

with S_k representing the set of all inputs having probability $p(k)$, and l_j being the length of the j th code. By Gallager's sibling property, we know that code lengths in S_k are either equal to $l(k)$ or $l(k) + 1$ for some integer $l(k)$. If n_k denotes the number of code words in S_k that are equal to $l(k) + 1$, then

$$L(k) = l(k) + \frac{n_k}{\binom{n}{k}}.$$

Thus, the average redundancy $\bar{R}_n^H$ of the Huffman code is

$$\bar{R}_n^H = \sum_{k=0}^n \binom{n}{k} p(k) [L(k) + \lg p(k)] \quad (1)$$

with $L(k)$ as above, and $\lg := \log_2$.

In order to find $l(k)$ and n_k , which are necessary to evaluate (1), one observes that the Huffman code minimizes $\bar{R}_n^H = \mathbf{E}[L_n] - nH$ subject to $\sum_{k=0}^n \binom{n}{k} 2^{-L(k)} \leq 1$. Here, $H = -p \lg p - q \lg q$ is the entropy rate. This can be viewed as a simple optimization problem. Using Lagrangian multipliers one can solve it and prove that $L(k) = \lg p(k)$ if no integer constraints are imposed. If we set $L(k) = l(k) + n_k / \binom{n}{k}$, then one gets

$$\begin{aligned} l(k) &= \lfloor \lg p(k) \rfloor, \\ n_k &= \binom{n}{k} [-\lg p(k) - l(k)]. \end{aligned}$$

In the above, we still did not impose integer constraints on n_k .

Stubley [25] analyzed carefully the above problem and concluded that setting

$$n_k = \left\lceil 2 \binom{n}{k} (1 - p(k) 2^{l(k)}) \right\rceil$$

solves the problem, with the Kraft inequality being satisfied. After algebraic manipulations, he was led to the following asymptotic formula

$$\bar{R}_n^H \sim \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} f_k + 2 \sum_{k=0}^{n-1} \binom{n}{k} p^k q^{n-k} (1 - 2^{f_k})$$

where

$$f_k = \lg p(k) + \lfloor -\lg p(k) \rfloor.$$

We rewrite it in another form. Let

$$\begin{aligned} \alpha &= \log_2 \left(\frac{1-p}{p} \right), \\ \beta &= \log_2 \left(\frac{1}{1-p} \right). \end{aligned}$$

Then, we obtain after some additional algebra,

$$\bar{R}_n^H = 2 - \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle - 2 \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle} + O(\rho^n), \quad (2)$$

where $\rho < 1$ and $\langle x \rangle = x - \lfloor x \rfloor$ being the fractional part of x .

We are now in a position to present our main result concerning the asymptotic behavior of $\bar{R}_n^H$ as $n \rightarrow \infty$. The proof can be found in Section 4.

Theorem 1 *Consider the Huffman block code of length n binomially(n, p) distributed over a binary alphabet. Then, for $p < \frac{1}{2}$ as $n \rightarrow \infty$*

$$\bar{R}_n^H = \begin{cases} \frac{3}{2} - \frac{1}{\log 2} + o(1) \approx 0.057304 & \alpha \text{ irrational} \\ \frac{3}{2} - \frac{1}{M} \left(\langle \beta M n \rangle - \frac{1}{2} \right) - \frac{1}{M(1-2^{-1/M})} 2^{-\langle n \beta M \rangle / M} + O(\rho^n) & \alpha = \frac{N}{M} \end{cases} \quad (3)$$

where N, M are integers such that $\gcd(N, M) = 1$ and $\rho < 1$.

Remark. An extension of Theorem 1 to multi-alphabet is possible. Let us consider a V -ary alphabet $\mathcal{A} = \{1, 2, \dots, V\}$. Based on Stubble's 1998 paper [26] we conclude the following expression for the Huffman redundancy in this case

$$\begin{aligned} \bar{R}_n^H &= 2 - \sum_{k_1 + \dots + k_V = n} \binom{n}{k_1, \dots, k_V} p_1^{k_1} \dots p_V^{k_V} \langle \alpha_1 k_1 + \dots + \alpha_{V-1} k_{V-1} + \beta n \rangle \\ &\quad - 2 \sum_{k_1 + \dots + k_V = n} \binom{n}{k_1, \dots, k_V} p_1^{k_1} \dots p_V^{k_V} 2^{-\langle \alpha_1 k_1 + \dots + \alpha_{V-1} k_{V-1} + \beta n \rangle} + O(\rho^n), \end{aligned}$$

where $\alpha_i = \lg(p_V/p_i)$ for $1 \leq i \leq V-1$ and $\beta = \lg(1/p_V)$. Using the same arguments as in the proof of Theorem 1, we observe that if one of α_i is irrational, then the formula for $\bar{R}_n^H$ from Theorem 1 remains valid, while if all α_i are rationals, then one derives an expression similar to the one in Theorem 1. $\square$

Theorem 1, however, does not really provide an insight into $\bar{R}_n^H$ behavior; in particular, one wonders why it exhibits two different characteristics for α irrational and rational. Furthermore, a comparison between the asymptotic formula of Theorem 1 and the numerical results of Figure 1 is desirable. With respect to the first question, we have to go back to formula (2) and inspect the term $\langle \alpha k + \beta n \rangle$ as $k, n \rightarrow \infty$. From theory of sequences uniformly distributed modulo 1 (cf. [8, 15]) it is known that $\langle \alpha k + \beta n \rangle$ "fills" up the interval $[0, 1)$ "uniformly" for α irrational as $n \rightarrow \infty$ (i.e., it can be viewed as the uniform distribution over $[0, 1)$). A different behavior occurs when α is rational. In this case, the sequence $\langle \alpha k + \beta n \rangle$ is periodic, only

certain points in the interval $[0, 1]$ are selected, and this fact is reflected in the final formula on the redundancy.

Comparing our theoretical finding with the numerical result of Figure 1, we observe a good agreement. Indeed, in the irrational case, the redundancy curve in Figure 1(a) converges to a constant that is approximately equal to 0.056 which coincides with the result of Theorem 1. In the rational case, we observe that the redundancy swings from almost zero to about 0.086. To see it more precisely, and in fact to recover Gallager's upper bound, we set $x = \langle Mn\beta \rangle$. We first observe that $M = 1$ maximizes $\bar{R}_n^H$, and then

$$\bar{R}_n^H(x) = 2 - x - 2^{-x+1}.$$

This leads to

$$\max_{0 \leq x < 1} \bar{R}_n^H = 1 - \frac{1 + \log \log 2}{\log 2} = \lg(2(\lg e)/e) = 0.08607 \dots,$$

which is the Gallager upper bound (since the most likely probability $p_1 = O(1/\sqrt{n})$ in this case). We formulate it as a corollary.

Corollary 1 *The maximum value of the average Huffman redundancy is*

$$\max\{\bar{R}_n^H\} = 1 - \frac{1 + \log \log 2}{\log 2} = \lg(2(\lg e)/e) = 0.08607 \dots,$$

as $n \rightarrow \infty$.

As mentioned in the introduction, the redundancy of the Huffman code is related to the redundancy of the **Shannon–Fano** code that we discuss now. In fact, we shall see that the Huffman redundancy $\bar{R}_n^H$ can be expressed in terms of the average Shannon–Fano redundancy $\bar{R}_n^{SF}$ as

$$\bar{R}_n^H = 1 + \bar{R}_n^{SF} - 2 \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle}. \quad (4)$$

Indeed, the Shannon–Fano code assigns length $\lceil -\lg p(k) \rceil$ to a word of probability $p(k) = p^k q^{n-k}$. Therefore, the average Shannon–Fano redundancy is

$$\begin{aligned} \bar{R}_n^{SF} &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} (\lceil -\lg p(k) \rceil + \lg p(k)) \\ &= 1 + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} (\lg p(k) + \lceil -p(k) \rceil) \\ &= 1 - \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle \end{aligned} \quad (5)$$

where, as before, $\alpha = \lg((1-p)/p)$ and $\beta = \lg(1/(1-p))$.

In Section 3 we prove the following asymptotic performance of the Shannon–Fano code. (In passing we should point out that numerical results of Figure 2 coincide with the theoretical findings of Theorem 2.)

Theorem 2 *Consider the Shannon–Fano block code of length n binomially (n, p) distributed over a binary alphabet. Then, for $p < \frac{1}{2}$ as $n \rightarrow \infty$*

$$\bar{R}_n = \begin{cases} \frac{1}{2} + o(1) & \alpha \text{ irrational} \\ \frac{1}{2} - \frac{1}{M} \left(\langle Mn\beta \rangle - \frac{1}{2} \right) + O(\rho^n) & \alpha = \frac{N}{M}, \quad \gcd(N, M) = 1 \end{cases} \quad (6)$$

where $\rho < 1$.

Remark. The rates of convergence in Theorem 1 and Theorem 2 for the irrational case are not specified. This problem is actually quite subtle, and there is no one simple bound for the rate of convergence for all α . However, for almost all irrational α the $o(1)$ term in Theorem 2 can be replaced by $O\left(\frac{\log^{1+\delta} n}{\sqrt{n}}\right)$ for some $\delta > 0$. More precisely, M. Drmota proved that for functions $f : [0, 1] \rightarrow \mathbb{R}$ with bounded variations $V(f)$ we have

$$\left| \lim_{n \rightarrow \infty} \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f(\langle \alpha k + \beta n \rangle) - \int_0^1 f(t) dt \right| \leq V(f) \cdot \inf_{\eta > 0} \left(e^{-C\eta^2} + \eta \max_{\eta\sqrt{n} \leq N \leq 2\eta\sqrt{n}} D_N(\alpha k) \right)$$

where $D_N(\alpha k)$ is the discrepancy of the sequence $\langle \alpha k \rangle_{k=1}^N$. The discrepancy of the Weyl sequence $\langle \alpha k \rangle$ has been extensively studied and is well understood (cf. [8, 15]). $\square$

3 Derivation of the Shannon–Fano Code Redundancy

We derive our main results in the reverse order we presented them. First, we deal with the redundancy of the Shannon–Fano code proving Theorem 2, and later we extend the analysis to the Huffman code redundancy. This is a natural order of showing our results since according to (4) $\bar{R}_n^{SF}$ is part of $\bar{R}_n^H$.

We are interested in the asymptotics of $\bar{R}_n^{SF}$ given by (5). Our main tool is Fourier series analysis [33]. In particular, we shall use the following result for $x \in \mathbb{R}$

$$\begin{aligned} \langle x \rangle &= \frac{1}{2} - \sum_{m=1}^{\infty} \frac{\sin 2\pi m x}{m\pi} \\ &= \frac{1}{2} - \sum_{m \in \mathbb{Z} - \{0\}} c_m e^{2\pi i m x}, \quad c_m = -\frac{i}{2\pi m}, \end{aligned} \quad (7)$$

where $\mathbb{Z}$ is the set of all integers. Hereafter, we shall write $\sum_{m \neq 0} := \sum_{m \in \mathbb{Z} - \{0\}}$.

3.1 Irrational Case

We first deal with the case when α is irrational. Using (7) in (5) we obtain

$$\begin{aligned}\bar{R}_n^{SF} &= \frac{1}{2} + \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \sum_{m \neq 0} c_m e^{2\pi i m(\alpha k + \beta n)} \\ &= \frac{1}{2} + \sum_{m \neq 0} c_m e^{2\pi i m \beta n} \left(p e^{2\pi i m \alpha} + q \right)^n.\end{aligned}$$

Our goal now is to prove that the last sum in the above is $o(1)$ when α is irrational.

While we could establish the above fact directly, we rather apply theory of *sequences distributed modulo 1* that is quite useful in our situation and finds other applications in information theory. We start with the following definition.

Definition 1 (B-u.d. mod 1) *A sequence $x_n \in \mathbb{R}$ is said to be Bernoulli uniformly distributed modulo 1 (in short: B-u.d. mod 1) if for fixed $0 < p < 1$*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \chi_I(\langle x_k \rangle) = \lambda(I) \quad (8)$$

holds for every interval $I \subset \mathbb{R}$, where $\chi_I(x_n)$ is the characteristic function of I (i.e., it equals to 1 if $x_n \in I$ and 0 otherwise) and $\lambda(I)$ is the Lebesgue measure of I .

The following result summarizes the main property of B-u.d. modulo 1 sequences that we need in the analysis.

Theorem 3 *Let $0 < p < 1$ be a fixed real number and suppose that the sequence x_n is B-uniformly distributed modulo 1. Then for every Riemann integrable function $f : [0, 1] \rightarrow \mathbb{R}$ we have*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f(\langle x_k + y \rangle) = \int_0^1 f(t) dt, \quad (9)$$

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

Proof. The proof is standard, however, details are quite tedious. For the reader convenience, we present a detailed proof proposed by M. Drmota in Appendix A. Here we only sketch the main idea. We first proof (9) for characteristic functions $\chi_I(x_k)$. This follows from Definition 1. Then, we approximate f by a step function (i.e., a combination of characteristic functions) and use the definition of the Riemann integral to bound the integral from below and above. We show that when $n \rightarrow \infty$ these bounds coincide with the left-hand side of (9). ■

To use Theorem 3 effectively, one needs a simple criterion to verify whether a sequence is B-u.d. mod 1. Such a criterion, fortunately, exists and it is basically due to Weyl. Before we

formulate it, we first observe that we can relax the condition of Theorem 3 to functions f that are continuous with period 1.

Theorem 4 (Weyl's Criterion) *A sequence x_n is B-u.d. mod 1 if and only if*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{2\pi i m x_k} = 0 \quad (10)$$

holds for all $m \in \mathbb{Z} - \{0\}$.

Proof. The proof again is standard and the reader is referred to textbooks such as [8, 15]. Basically, it is based on the fact that by Weierstrass's *approximation theorem* every Riemann integrable function f of period 1 can be uniformly approximated by a trigonometric polynomial (i.e., a finite combination of functions of the type $e^{2\pi i m x}$). Some details can be found in Appendix A. ■

Now, we are in position to finish our derivation. We first prove that $\langle \alpha k \rangle$ is B-u.d. mod 1. Indeed, by Weyl's criterion

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{2\pi i m (k\alpha)} = \lim_{n \rightarrow \infty} (pe^{2\pi i m \alpha} + q)^n = 0$$

provided α is irrational. Hence, by Theorem 3, with $f(t) = t$ and $y = \beta n$, we immediately obtain

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle = \int_0^1 t dt = \frac{1}{2}.$$

This proves the irrational part of Theorem 2.

3.2 Rational Case

Now, we turn our attention to the case when α is rational. We assume $\alpha = M/N$ where M, N are integers such that $\gcd(N, M) = 1$. Let $p_{n,k} = \binom{n}{k} p^k q^{n-k}$ and we denote by S_n the sum in (5). We proceed as follows

$$\begin{aligned} S_n &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \left\langle k \frac{N}{M} + \beta n \right\rangle = \sum_{k=0}^n \binom{n}{k} p_{n,k} \left\langle k \frac{N}{M} + \beta n \right\rangle \\ &= \sum_{\ell=0}^{M-1} \sum_{m: k=\ell+mM \leq n} p_{n,k} \left\langle \ell \frac{N}{M} + N + \beta n \right\rangle \\ &= \sum_{\ell=0}^{M-1} \sum_{m: k=\ell+mM \leq n} p_{n,k} \left\langle \frac{\ell}{M} + \beta n \right\rangle \\ &= \sum_{\ell=0}^{M-1} \left\langle \frac{\ell}{M} + \beta n \right\rangle \sum_{m: k=\ell+mM \leq n} p_{n,k}. \end{aligned} \quad (11)$$

To evaluate the last sum we need the following simple lemma. It asserts that if one picks every M th term of the binomial distribution, then the total probability of this sample is “well” approximated by $1/M$.

Lemma 1 *For fixed $\ell \leq M$ and M , there exist $\rho < 1$ such that*

$$\sum_{m: k=\ell+mM \leq n} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{M} + O(\rho^n). \quad (12)$$

Proof. Let $\omega_k = e^{2\pi i k/M}$ for $k = 0, 1, \dots, M-1$ be the M th root of unity. It is well known that

$$\frac{1}{M} \sum_{k=0}^{M-1} \omega_k^n = \begin{cases} 1 & \text{if } M|n \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

where $M|n$ means that M divides n . In view of this, we can write

$$\begin{aligned} \sum_{m: k=\ell+mM \leq n} \binom{n}{k} p^k q^{n-k} &= \frac{1 + (p\omega_1 + q)^{n-\ell} + \dots + (p\omega_{M-1} + q)^{n-\ell}}{M} \\ &= \frac{1}{M} + O(\rho^n), \end{aligned}$$

since $|(p\omega_r + q)| = p^2 + q^2 + 2pq \cos(2\pi r/M) < 1$ for $r \neq 0$. This proves the lemma. ■

From now on, we only deal with the sum $S_n = \frac{1}{M} \sum_{\ell=0}^{M-1} \left\langle \frac{\ell}{M} + \beta n \right\rangle$ ignoring the error term $O(\rho^n)$. We use again the Fourier series (7) and (13) to obtain

$$\begin{aligned} S_n &= \frac{1}{M} \sum_{\ell=0}^{M-1} \left(\frac{1}{2} - \sum_{m \neq 0} c_m e^{2\pi i m(\ell/M + \beta n)} \right) \\ &= \frac{1}{2} - \sum_{m \neq 0} c_m e^{2\pi i m \beta n} \frac{1}{M} \sum_{\ell=0}^{M-1} e^{2\pi i m \ell/M} \\ &\stackrel{(13)}{=} \frac{1}{2} - \frac{1}{M} \sum_{m=kM \neq 0} c_{kM} e^{2\pi i kM \beta n} \\ &\stackrel{(7)}{=} \frac{1}{2} - \frac{1}{M} \left(\frac{1}{2} - \langle \beta n M \rangle \right). \end{aligned}$$

This proves the rational part of (6), and completes the proof of Theorem 2.

4 Derivation of the Huffman Code Redundancy

We establish here Theorem 1, that is, we shall analyze the Huffman redundancy given by (2). Due to (4), we must only deal with the asymptotics of the following sum

$$T_n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle}.$$

As in the previous section, we shall apply Fourier series, but this time we need the following for some $a > 0$

$$2^{-\langle x \rangle / a} = C_0(a) + \sum_{m \neq 0} C_m(a) e^{2\pi i m x}, \quad (14)$$

where

$$C_0(a) = \frac{a}{\log 2} (1 - 2^{-1/a}), \quad (15)$$

$$C_m(a) = \frac{a}{2\pi i m a + \log 2} (1 - 2^{-1/a}), \quad m \neq 0. \quad (16)$$

4.1 Irrational Case

We first consider α irrational. To analyze T_n we can directly applied Theorem 3 with $f(t) = 2^{-t}$ and $y = \beta n$. This leads to

$$\lim_{n \rightarrow \infty} T_n = \int_0^1 2^{-t} dt = \frac{1}{2 \log 2}$$

which proves Theorem 1 for α irrational.

4.2 Rational Case

Next, we consider α rational. This turns out to be a more intricate case. First of all, applying Lemma 1 we can re-write T_n when $\alpha = N/M$ ($\gcd(N, M) = 1$) as

$$T_n = \frac{1}{M} \sum_{\ell=0}^{M-1} 2^{-\langle \frac{\ell}{M} + \beta n \rangle} + O(\rho^n)$$

for some $\rho < 1$. Ignoring the error term, we proceed as follows

$$\begin{aligned} T_n &= C_0(1) + \frac{1}{M} \sum_{m \neq 0} C_m(1) \sum_{\ell=0}^{M-1} \exp \left(2\pi i m \left(\frac{\ell}{M} + n\beta \right) \right) \\ &= C_0(1) + \sum_{m \neq 0} C_m(1) e^{2\pi i m n \beta} \frac{1}{M} \sum_{\ell=0}^{M-1} \left(e^{2\pi i \ell / M} \right)^m \\ &\stackrel{(13)}{=} C_0(1) + \sum_{m=kM \neq 0} C_{kM}(1) e^{2\pi i k n \beta M} \\ &= \frac{1}{2 \log 2} + \frac{1}{M} \sum_{k \neq 0} \frac{1}{2(2\pi i k + \log 2/M)} e^{2\pi i k n \beta M} \\ &= \frac{1}{2 \log 2} + \frac{1}{2(1 - 2^{-1/M})} \sum_{k \neq 0} \frac{(1 - 2^{-1/M})}{(2\pi i k + \log 2/M)} e^{2\pi i k (n\beta M)} \\ &\stackrel{(14)}{=} \frac{1}{2 \log 2} + \frac{1}{2(1 - 2^{-1/M})} \sum_{k \neq 0} C_k(M) e^{2\pi i k (n\beta M)} \\ &= \frac{1}{2M(1 - 2^{-1/M})} 2^{-\langle n\beta M \rangle / M}. \end{aligned}$$

This establishes the rational part of (3), and completes the proof of Theorem 1.

In passing, we should point out that our proofs of the rational cases for the Shannon-Fano code and the Huffman code can be somewhat simplified by the following theorem (suggested by M. Drmota). Its proof follows the footsteps of our analyses above, so we omit it here.

Theorem 5 *Let $0 < p < 1$ be a fixed real number and suppose that $\alpha = \frac{N}{M}$ is a rational number with $\gcd(N, M) = 1$. Then, for every bounded function $f : [0, 1] \rightarrow \mathbb{R}$ we have*

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f(\langle k\alpha + y \rangle) = \frac{1}{M} \sum_{l=0}^{M-1} f\left(\frac{l}{M} + \frac{\langle My \rangle}{M}\right) + O(\rho^n)$$

uniformly for all $y \in \mathbb{R}$ and some $\rho < 1$.

Appendix A: Drmota's Proof of Theorem 3

Let $\mathcal{L}$ be the linear space of all (Lebesgue) integrable functions $f : [0, 1]^k \rightarrow \mathbb{C}$ satisfying

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathbb{R}} \left| \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f(\langle x_k + y \rangle) - \int_0^1 f(t) dt \right| = 0. \quad (\text{A.1})$$

Of course, $\mathcal{L}$ contains all constant functions, i.e. $\mathcal{L} \neq \emptyset$. In a next step, we prove that $\mathcal{L}$ contains all trigonometric polynomials. Since $\mathcal{L}$ is a linear space, it suffices to consider the functions $f(x) = e^{2\pi i m x}$ for m being a non-zero integer. This follows from Weyl's criterion (cf. Theorem 4 since

$$\left| \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{2\pi i m (x_k + y)} \right| = \left| \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{2\pi i m x_k} \right| \rightarrow 0.$$

Next we prove the following closure property of $\mathcal{L}$.

Suppose $f : [0, 1] \rightarrow \mathbb{R}$ is an integrable function such that for every $\varepsilon > 0$ there exist functions $g_1, g_2 \in \mathcal{L}$ with $g_1 \leq f \leq g_2$ and

$$\int_0^1 (g_2(t) - g_1(t)) dt < \varepsilon. \quad (\text{A.2})$$

then $f \in \mathcal{L}$, too.

For this purpose, let $m_n(f, y)$ denote the positive linear functionals

$$m_n(f, y) := \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f(\langle x_k + y \rangle)$$

and $m(f)$ the positive linear functional

$$m(f) := \int_0^1 f(t) dt.$$

By $m_n(g_1, y) \leq m_n(f, y) \leq m_n(g_2, y)$ and $m(g_1) \leq m(f) \leq m(g_2)$ we immediately get

$$\begin{aligned} m(g_1) &= \liminf_{N \rightarrow \infty} m_n(g_1, y) \leq \liminf_{N \rightarrow \infty} m_n(f, y) \\ &\leq \limsup_{N \rightarrow \infty} m_n(f, y) \leq \limsup_{N \rightarrow \infty} m_n(g_2, y) \\ &= m(g_2) \end{aligned}$$

which implies

$$|m(f) - \liminf_{N \rightarrow \infty} m_n(f, y)| < \varepsilon$$

and

$$|m(f) - \limsup_{N \rightarrow \infty} m_n(f, y)| < \varepsilon$$

for every $\varepsilon > 0$. Thus

$$\lim_{N \rightarrow \infty} \sup_{y \in \mathbb{R}} |m_n(f, y) - m(f)| = 0,$$

which means that $f \in \mathcal{L}$, too.

Next, suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is a continuous function. Then by the Weierstrass approximation theorem, for every $\varepsilon > 0$ there exist two trigonometric polynomials g_1, g_2 with $g_1 \leq f \leq g_2$ and

$$\int_0^1 (g_2(t) - g_1(t)) dt < \varepsilon.$$

Hence all continuous functions $f : [0, 1] \rightarrow \mathbb{R}$ are contained in $\mathcal{L}$.

Now, it is easy to derive that every characteristic function $f = \chi_I$ of any interval $I \subseteq [0, 1]$ is in $\mathcal{L}$, too. It is obvious, that for every $\varepsilon > 0$ there exist two continuous functions g_1, g_2 with $g_1 \leq \chi_I \leq g_2$ and

$$\int_0^1 (g_2(t) - g_1(t)) dt < \varepsilon.$$

This also implies that all step functions (i.e. finite linear combinations of characteristic functions) are in $\mathcal{L}$.

Finally, if $f : [0, 1] \rightarrow \mathbb{R}$ is a Riemann integrable function, then by definition for every $\varepsilon > 0$ there exist two step functions g_1, g_2 with $g_1 \leq f \leq g_2$ and

$$\int_0^1 (g_2(t) - g_1(t)) dt < \varepsilon.$$

If $f : [0, 1] \rightarrow \mathbb{C}$ is a Riemann integrable function then we consider the real and imaginary part separately and conclude that all Riemann integrable function $f : [0, 1] \rightarrow \mathbb{C}$ are contained in $\mathcal{L}$.

ACKNOWLEDGEMENT

This work would be impossible without many happy occurrences. John Kieffer asked me about two years ago if “analytic information theory” can handle the Huffman code redundancy. When working with Philippe Jacquet on the entropy computations via analytic methods, and with H-K. Hwang on the redundancy of CTW code, I understood that the theory of *distribution of sequences* is relevant, and I did some initial work on the Shannon–Fano code. Then, happily enough, Julia Abrahams, after my talk at Princeton during the *Fourth Seminar on Analysis of Algorithms*, sent Stubbley’s paper [25] that prompted me to look again at the Huffman redundancy problem. Michael Drmota helped me with many technical questions concerning uniform distribution of sequences, supplied a detailed proof of Theorem 3, and provided many insightful comments. Svante Janson had additional comments during my talk at the *Fifth Seminar on Analysis of Algorithms*, Barcelona, June 1999. Finally, P. Stubbley generously shared his unpublished paper [26]. To all of them I thank whole-heartly.

References

- [1] J. Abrahams, “Code and Parse Trees for Lossless Source Encoding, *Proc. of Compression and Complexity of SEQUENCE’97*, Positano, IEEE Press, 145–171, 1998.
- [2] R. Capocelli and A. de Santis, Tight Upper bounds on the Redundancy of Huffman Codes, *IEEE Trans. Information Theory*, 35, 1084–1091, 1989.
- [3] R. Capocelli and A. de Santis, New Bounds on the Redundancy of Huffman Codes, *IEEE Trans. Information Theory*, 37, 1095–1104, 1991.
- [4] T. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York 1991.
- [5] I. Csiszár and P. Shields, Redundancy Rates for Renewal and Other Processes, *IEEE Trans. Information Theory*, 42, 2065–2072, 1996.
- [6] A. Dembo and I. Kontoyiannis, The Asymptotics of Waiting Times Between Stationary Processes, Allowing Distortion, *Ann. Appl. Probab.*, 9, 413–429, 1999.
- [7] M. Drmota, Private Communications.
- [8] M. Drmota and R. Tichy, *Sequences, Discrepancies, and Applications*, Springer Verlag, Berlin Heidelberg 1997
- [9] P. Flajolet and W. Szpankowski, Analytic Variations on Redundancy Rates of Renewal Processes, INRIA Research Report RR-335, 1998.
- [10] R. Gallager, *Information Theory and Reliable Communications*, New York, Wiley 1968.
- [11] R. Gallager, Variations on the Theme by Huffman, *IEEE Trans. Information Theory*, 24, 668–674, 1978.

- [12] D. Huffman, A Method for the Construction of Minimum Redundancy Codes, *proc. IRE*, 40, 1089-1101, 1952.
- [13] P. Jacquet and W. Szpankowski, Asymptotic Behavior of the Lempel-Ziv Parsing Scheme and Digital Search Trees, *Theoretical Computer Science*, 144, 161-197, 1995.
- [14] I. Kontoyiannis, Pointwise Redundancy in Lossy Data Compression and Universal Data Compression, preprint 1999.
- [15] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. John Wiley, New York 1974.
- [16] G. Louchard and W. Szpankowski, On the Average Redundancy Rate of the Lempel-Ziv Code, *IEEE Trans. Information Theory*, 43, 2-8, 1997.
- [17] D. Manstetten, Tight Upper Bounds on the Redundancy of Huffman Codes, *IEEE Trans. Information Theory*, 38, 220-22, 1980.
- [18] F. Olver, *Asymptotics and Special Functions*, Academic Press, New York 1974.
- [19] J. Rissanen, Complexity of Strings in the Class of Markov Sources, *IEEE Trans. Information Theory*, 30, 526-532, 1984.
- [20] S. Savari, Redundancy of the Lempel-Ziv Incremental Parsing Rule, *IEEE Trans. Information Theory*, 43, 9-21, 1997.
- [21] S. Savari, Variable-to-Fixed Length Codes for Predictable Sources, *Proc. Data Compression Conference*, 481-490, Snowbird 1998.
- [22] S. Savari and R. Gallager, Generalized Tunstall Codes for Sources with Memory, *IEEE Trans. Information Theory*, 43, 658-668, 1997.
- [23] P. Shields, Universal Redundancy Rates Do Not Exist, *IEEE Trans. Information Theory*, 39, 520-524, 1993.
- [24] Y. Shtarkov, Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, 23, 175-186, 1987.
- [25] P. Stubbley, On the Redundancy of Optimum Fixed-to-Variable Length Codes, *Proc. Data Compression Conference*, 90-97, Snowbird 1994.
- [26] P. Stubbley, The Redundancy of Optimum Codes for Multinomially Distributed Sources, preprint 1998.
- [27] W. Szpankowski, On Asymptotics of Certain Recurrences Arising in Universal Coding, *Problems of Information Transmission*, 34, No.2, 55-61, 1998.
- [28] M. Weinberger, N. Merhav, and M. Feder, Optimal Sequential Probability Assignments for Individual Sequences, *IEEE Trans. Information Theory*, 40, 384-396, 1994.
- [29] F.M. Willems, Y. Shtarkov and T. Tjalkens, The Context-Tree Weighting Method: Basic Properties, *IEEE Trans. Information Theory*, 41, 653-664, 1995.

- [30] A. J. Wyner, The Redundancy and Distribution of the Phrase Lengths of the Fixed-Database Lempel-Ziv Algorithm, *IEEE Trans. Information Theory*, 43, 1439–1465, 1997.
- [31] E-H. Yang and J. Kieffer, A Simple Technique for Bounding the Pointwise Redundancy of the 1978 Lempel-Ziv Algorithm, *Proc. Data Compression Conference*, Snowbird 1999.
- [32] J. Ziv, Back from Infinity: A Constrained Resources Approach to Information Theory, *IEEE Information Theory Society Newsletter*, 48, 30-33, 1998.
- [33] A. Zygmund, *Trigonometric Series*, Cambridge University Press, New York, 1959.