Prefix Codes for Power Laws with Countable Support

Michael B. Baer, Member, IEEE

Abstract—In prefix coding over an infinite alphabet, methods that consider specific distributions generally consider those that decline more quickly than a power law (e.g., Golomb coding). Particular power-law distributions, however, model many random variables encountered in practice. For such random variables, compression performance is judged via estimates of expected bits per input symbol. This correspondence introduces a family of prefix codes with an eye towards near-optimal coding of known distributions. Compression performance is precisely estimated for well-known probability distributions using these codes and using previously known prefix codes. One application of these near-optimal codes is an improved representation of rational numbers.

Index Terms—Coding of integers, continued fractions, infinite alphabet, optimal prefix code, power law, rational numbers, search trees, Shannon entropy.

I. INTRODUCTION

Consider discrete power-law distributions, those of the form

\[ p(i) \sim ci^{-\alpha} \]

for constants \( c > 0 \) and \( \alpha > 1 \), where \( p(i) \) is the probability of symbol \( i \), and \( f(i) \sim g(i) \) implies that the ratio of the two functions goes to 1 with increasing \( i \). Such distributions could be either inherently discrete or discretized versions of continuous power-law distributions.

Several researchers in varied fields have, in classic papers ranging from decades to centuries old, observed power-law behavior for various discrete phenomena. These include distribution of wealth [1], [2], town and city populations [2], [3], word frequency [2], [4], [5], numbers of species of a given genus [2], [6], and terms in continued fractions [7], [8]. More recent papers model various Internet phenomena [9]. So active is the topic that several surveys and popular expositions exist, e.g., [9]–[11].

However, there has been relatively little work on lossless compression of symbols obeying such distributions, in spite of a rich literature on prefix coding problems [12]. Exponential-Golomb codes [13] (generalizations of Elias’ \( \gamma \) code [14]) are a good fit for certain power laws [15], [16], leading to their widespread use in compressing video and numerical data [15], [17]. To the author’s knowledge, compression performance of prefix codes [19], [20].

We estimate compression performance for dozens of code/distribution combinations. For fixed codes, these estimates are rigorously shown to be precise.

II. BACKGROUND, FORMALIZATION, AND MOTIVATION

The most common infinite-alphabet codes are codes that are optimal for geometric [23], [24] and geometrically-based [25]–[29] distributions. For geometric distributions, these are known as Golomb codes, and are based on the unary code — ones terminated by a zero, i.e., a code consisting of codewords the form \( \{1^0\} \) for \( j \geq 0 \). In a Golomb code \( G_k \), a unary code prefix precedes a binary code suffix. This binary code is a complete binary code, in that it has \( k \) codewords of the same length or length differing by at most one. For example, the alphabetic complete binary code of size three that is monotonically nonincreasing in length is \( \{0,1,11\} \), so the Golomb code \( G_3 \) is \( \{00,010,011,100,1010,\ldots\} \). If the complete binary code suffix is of constant length, the overall Golomb code is also called a Rice code. Rice codes are used in standards such as JPEG-LS [30]. Codes that exhibit an efficient coding rate for power laws, by contrast, are not known to be optimal (excepting those with finite support and trivial examples for dyadic probability mass functions).

We restrict ourselves to binary codes and assume that the symbols to be coded are positive integers. Thus, an infinite-alphabet source emits symbols drawn from the alphabet \( X = \{1,2,3,\ldots\} \). (Some applications code the alphabet \( X_0 = \{0,1,2,\ldots\} \) or the alphabet \( X_2 = \{0,-1,1,-2,\ldots\} \), but any code of either form can be mapped trivially to a code on \( X \).) Symbol \( i \) has probability \( p(i) > 0 \), forming probability mass function \( P = \{p(i)\} \). The source symbols are coded into binary codewords. The codeword \( c(i) \in \{0,1\}^* \), corresponding to symbol \( i \), has length \( n(i) \in \mathbb{Z}_+ \), thus defining length distribution \( N = \{n(i)\} \). An optimal code is one that minimizes \( \sum_{i \in X} p(i)n(i) \) with the constraint of a corresponding code being uniquely decodable, which one is if and only if the Kraft inequality, \( \sum_{i \in X} 2^{-n(i)} \leq 1 \), is satisfied. We can assume without loss of generality that these codes are prefix codes, that is, codes where there are no two codewords of the form \( c(i) \) and \( c(j) = c(i)x \), where \( c(i)x \) denotes the concatenation of strings \( c(i) \) and (nontrivial) \( x \). (In a similar use of notation, \( 0^k \) and \( 1^k \) denote \( k \)'s, respectively. Note also that we use \( \lg \) to denote \( \log_2 \) and \( \ln \) to denote \( \log_e \), where \( e \) is the base of the natural logarithm.)

One cannot use the Huffman source coding algorithm [31] to find an optimal code, as one can for a finite source alphabet. However, it is sensible that a code over the integers should be monotonic, that is, that \( n(i) \geq n(i+1) \) for all \( i \geq 0 \). An exchange argument easily shows that this is necessary for the code to be optimal given a distribution for which \( p(i) > p(i+1) \) for all \( i \).

Also desirable is for a code to be alphabetic or order preserving; that is, if \( c(i,j) \) is the \( j \)th bit of the \( i \)th codeword, then \( c(i+1,j) < c(i,j) \) only if there is a \( k < j \) such that \( c(i+1,k) \neq c(i,k) \). Alphabetic codes allow the prefix coding tree to be used as a decision tree, which is useful for search problems, as in [32], [33]. It is also useful for implementation of arithmetic coding: Because binary arithmetic coding is much faster than other types of arithmetic coding, a decision tree can reduce an infinite-alphabet source into a binary source for fast arithmetic coding, as in [15]. In addition, order preservation is necessary for the ordered representation of rational numbers as integers in continued fractions [19], [20]; in this correspondence we improve upon these representations.

Any valid monotonic prefix code has a (possibly different) alphabetic prefix code with the same length distribution. For example, the Elias \( \gamma \) code was first presented in a nonalphabetic version, then transformed into alphabetic form (as a decision tree) in [32]. Where there is ambiguity, we will assume use of the alphabetic version of a code.

Another desirable property is one we call “smoothness”:

**Definition:** We call \( N = \{n(i)\} \) \( j \)-smooth if, for every \( i > j \), if \( n(i+1) = n(i) + 2 \), then \( n(i+1) - n(i) \leq 1 \), that is, there are no “jumps” followed by “plateaus”: weakly smooth means that it is \( j \)-smooth for some \( j \). Thus, for any \( j \), a \( j \)-smooth code includes all weakly smooth codes. Similarly, \( 0 \)-smooth (or strongly smooth) codes include all \( j \)-smooth (and thus weakly smooth) codes. Also, we call a
\( P = \{ p(i) \} \) \( j \)-antiunary if, for every \( i > j \), \( p(i) < p(i+1)+p(i+2) \); antiunary means that it is \( j \)-antiunary for some \( j \).

**Observation:** No \( j \)-antiunary distribution has an optimal code which is not \( j \)-smooth. Thus no antiunary distribution has an optimal code which is not weakly smooth.

**Proof:** Suppose a \( j \)-antiunary distribution \( P \) has an optimal code with lengths \( N \) which is not \( j \)-smooth. Then there exists an \( i > j \) such that \( n(i+1) = n(i+2) \) and \( n(i+1) = n(i) + 1 \). Consider \( N' = \{ n'(i) \} \) for which \( n'(k) = n(k) \) except at values \( n'(i) = n(i) + 1 \), \( n'(i+1) = n(i+1) - 1 \), and \( n'(i+2) = n(i+2) - 1 \). Clearly \( N' \) satisfies the Kraft inequality and \( \sum_i p(i)n'(i) < \sum_i p(i)n(i) \), so \( N \) is not optimal.

Every power law is antiunary, but most previously presented codes suitable for power-law distributions are not weakly smooth, so they could not be optimal solutions. The proof shows that, when such codes are applied to antiunary distributions, it is always a simple matter to improve such a code for use with such a distribution.

For many probability distributions, however, there is no guarantee that an optimal code would be computationally tractable, let alone computationally practical for compression applications. We thus judge performance of candidate codes by expected bits per coded symbol rather than by strict optimality. One of the contributions of this correspondence is a comparison of various codes for well-known power-law distributions.

### III. A NEW FAMILY OF CODES FOR INTEGERS

We propose a family of monotonic, alphabetic, computational efficient, 0-smooth codes, starting with the code shown in the center set of columns \( (n_0(\cdot) \) and \( c_0(\cdot) \) of Table I which is defined as

\[
c_0(i) = \begin{cases} 
0b(i-1,3), & i < 4 \\
1c_0(\frac{i-2}{2})0, & i = \{ 4, 6, 8, \ldots \} \\
1c_0(\frac{i+2}{2})1, & i = \{ 5, 7, 9, \ldots \}.
\end{cases}
\]

The term \( b(j, k) \) denotes the \((j+1)\)th codeword of a complete binary code with \( k \) items, which is order-preserving (alphabetic), with the first \( 2^{\lceil \log k \rceil} - k \) items having length \( \log k \) and the last \( 2k - 2^{\lceil \log k \rceil} \) items having length \( \log k \). In this case, that means that \( c_0(1) = 0b(0,3) = 00, c_0(2) = 0b(1,3) = 010, \) and \( c_0(3) = 0b(2,3) = 011 \). Thus, for example, \( c_0(12) = 1c_0(5) = 11c_0(1)10 = 110010 \) is shown to lie in ranges of two possible values for each codeword (or one for the first, which has \( n(1) = 1 \)). The code word lengths of Code 0, \( k = \log_{\frac{2}{2^k}}(i-1) \mod 2^k \), and \( \zeta \) is the Riemann zeta function \( \zeta(s) \) defined and well-approximated as follows:

\[
p_G^K(i) \triangleq -\log \left( 1 - \frac{1}{(i+1)^2} \right) \approx \frac{\log e}{(i+1)^2}
\]

This shows how it is a power law. The Gauss-Kuzmin distribution is the one for which to code when expressing coefficients of continued fractions, as in [19], [38], in which EG0 is proposed for use, and [20], in which Yokoo’s code is proposed. Code \(-1\) is only about 0.008% worse than the (approximated) optimal code, whereas Yokoo’s code is 0.449% worse and the Elias \( \gamma \) code (EG0) is 1.007% worse.

Note also that Code \(-2\) is a good code for the zeta distribution with parameter \( s = 2 \), where the zeta distribution is defined as

\[
p_\zeta(i) \triangleq \frac{1}{i^\zeta(s)}
\]

and \( \zeta \) is the Riemann zeta function \( \zeta(s) \) defined and well-approximated as follows:

\[
p_\zeta(i) = \left\{ \begin{array}{ll}
1^{i-1}, & i \leq -k \\
1^{-k}c_0(i+k), & i > k.
\end{array} \right.
\]

All codes presented here are 0-smooth (strongly smooth), and can be coded and decoded using only additions, subtractions, and shifts such that the total number of operations is proportional to the number of encoded output bits.

### IV. APPLICATION

Table II lists various distributions for which no optimal code is known and estimates, in expected bits per input symbol, of coding performance using several different codes. The entropy and the expected bits per symbol of an optimal code are also estimated. \( H \) denotes the entropy of the distribution \( (H(P) = -\sum_i p(i) \log p(i)) \) and \( N^{*} \) (the expected codeword length of) the optimal code. Golin denotes the best Golin code [35]; Code \( k \) denotes the best of the codes introduced here; \( \pi \) denotes the Levenshtein (\( \text{Levenshtein} \)) code [36]; \( \gamma/\delta/\omega/EGk \) denotes the best of the Elias codes [14] and exponential-Golomb codes [13], which in these examples is always the Elias \( \gamma \) code (EG0); \( Y \) denotes Yokoo’s code for the Gauss-Kuzmin distribution [20]; and \( Gk \) denotes the best Golomb code (with parameter \( k \)) [23]. These codes are defined in the cited papers and the definitions are repeated in the Appendix, which also explains the methods by which the estimations of bits per symbol are calculated. In cases for which there are multiple codes and/or parameters, the best one is chosen, and indicated in superscript. Note that, as in previous papers on these and similar codes [13], [37], the best code is chosen by its empirical performance; there appears to be no simple rule for deciding which code to use.

We show the performance for the overall best fixed code for each distribution in bold in Table II and, if a Golin code is better, this is in italics. Note that Golin codes do well for inputs with rapidly declining probabilities, whereas Yokoo’s code and the codes introduced here have the best results for inverse square probability mass functions. However, Golin codes, in being calculated on the fly, are often impractical, both due to the potential for rounding errors to lead to coding errors and due to the computational complexity of the required floating point divisions.

We find that Code \(-1\) is of particular interest as it happens to be an excellent code for the Gauss-Kuzmin distribution, defined and well-approximated as follows:

\[
p_G^K(i) = -\log \left[ 1 - \frac{1}{(i+1)^2} \right] \approx \frac{\log e}{(i+1)^2}
\]

This shows how it is a power law. The Gauss-Kuzmin distribution is the one for which to code when expressing coefficients of continued fractions, as in [19], [38], in which EG0 is proposed for use, and [20], in which Yokoo’s code is proposed. Code \(-1\) is only about 0.008% worse than the (approximated) optimal code, whereas Yokoo’s code is 0.449% worse and the Elias \( \gamma \) code (EG0) is 1.007% worse.

Note also that Code \(-2\) is a good code for the zeta distribution with parameter \( s = 2 \), where the zeta distribution is defined as

\[
p_\zeta(i) = \frac{1}{i^\zeta(s)}
\]

and \( \zeta \) is the Riemann zeta function \( \zeta(s) \) defined and well-approximated as follows:

\[
p_\zeta(i) = \left\{ \begin{array}{ll}
1^{i-1}, & i \leq -k \\
1^{-k}c_0(i+k), & i > k.
\end{array} \right.
\]
within the allowed ranges. However, we can empirically find better codes, showing that Code -2, although the best simply described code we know of, is about 0.005% worse than an optimal code.

A third distribution family is that of Yule [6] and Simon [2],

$$p_{\rho}^{YS}(i) \equiv \rho B(i, \rho + 1) \quad \left( p_{\rho}^{YS}(i) = \rho \cdot \frac{(i-1)! \cdot \gamma}{\rho + 1} \right)$$

where $B(i, j)$ is the beta function, $\rho > 0$, and the right equation applies for integer $\rho$. Thus, for example, if $\rho = 1$, then $p(i) = 1/(i+1)$. Several statistics, from species population to word frequencies, have been observed to obey a Yule-Simon distribution, most often with parameter $\rho = 1$ [2]. This particular distribution is also related to continued fractions, being the distribution of the first coefficient when the number represented is chosen uniformly over the unit interval $(0, 1)$. For $p_{1}^{YS}$, Yokoo’s code is 0.066% better than Code -1.

The estimates in Table I were calculated based on finite sums and estimates of the remaining infinite sum. For fixed codes and for entropy, these codes are as calculated in the Appendix, and are thus accurate to the precision given. The Golin code was estimated based on the partial code and conditional entropy of the remaining items. Similarly, optimal expected codeword lengths were estimated using an optimal code for the partial sum and the entropy of the remaining items; although not having the same guaranteed accuracy, the results seem to provide accurate estimates based upon the behavior of coding truncated probability distributions of increasing size. In [39], it is shown that sequences of such truncated distributions always have a subsequence converging to the optimal code, providing theoretical justification for the use of this technique. Values that are exactly calculated from infinite sums, rather than estimated, are indicated by the reduced number of figures (for multiples of 0.1) or through ellipses in the case of

$$2.66666 \ldots = \frac{5}{3}, 1.94737 \ldots = \frac{\zeta(1.5)}{\zeta(2.5)}, \text{ and } 1.36843 \ldots = \frac{\zeta(2)}{\zeta(3)}.$$
APPENDIX
Consider all codes and probability distributions that are monotonic and for which we can find $\alpha, \beta, \kappa > 0, \mu, \xi > 0, \tau > 0, \nu > 0, \phi > 0$ such that
\[ n(i) \in [\tau \ln(i + \mu + 1) + \alpha, \nu \ln(i + \mu) + \beta] \]
and
\[ p(i) \in \left[ \frac{\phi}{(i + \kappa)^{\xi + 1}}, \frac{\phi}{(i + \kappa)^{\xi + 1}} \right] \]
for large enough $i \geq i_{\text{min}}$. Then, for $x > i_{\text{min}}$, we have
\[
\sum_{i=x}^{\infty} p(i)n(i) \geq \int_{x}^{\infty} p(i)n(i-1)di \\
\geq \int_{x}^{\infty} \tau \phi \ln(i + \mu) + \alpha \phi \frac{di}{(i + \kappa)^{\xi + 1}} \\
\geq \phi \int_{x}^{\infty} \tau \phi \ln(i + \mu) + \tau \phi \ln(\min(x, \mu)) + \alpha \phi \xi^{-1} + \phi \xi \ln(i + \mu) \ln(\xi + 1), \kappa \xi = \frac{\phi}{\kappa} - \frac{\phi}{(i + \kappa)^{\xi + 1}} \]
where $f_{\text{min}}(x) = \min(\ln(i + \mu) - \ln(i + \mu), x)$, and
\[
\sum_{i=x}^{\infty} p(i)n(i) \leq \int_{x}^{\infty} p(i-1)n(i)di \\
\leq \int_{x}^{\infty} \nu \phi \ln(i + \mu) + \beta \phi \frac{di}{(i + \kappa)^{\xi + 1}} \\
\leq \phi \int_{x}^{\infty} \nu \phi \ln(i - 1) + \nu \phi f_{\text{max}}(x) + \beta \phi \xi^{-1} + \beta \phi \xi \ln(i - 1), \kappa \xi = \frac{\phi}{\kappa} - \frac{\phi}{(i - 1)^{\xi + 1}} \]
where $f_{\text{max}}(x) = \max(\ln(i + \mu) + \ln(i - 1), 0)$, providing upper and lower bounds to average codeword length using code $N = \{n(i)\}$ for probability distribution $P = \{p(i)\}$. Other distributions (such as Golomb codes) and entropy can be bounded similarly.

Such an approach enables us to find estimates with accuracies limited only by the precision of the partial summations (i.e., round-off error). For the probability distributions currently under consideration, we have:

| $P_{\text{GK}}$ | $P_{\text{YG}}$ | $P_{\phi}$ |
|---|---|---|
| $\frac{1}{\xi}$ | $\frac{1}{\phi}$ | $\frac{1}{\kappa}$ |

In order to find bounds for expected codeword lengths, we should first define the codes we are using. Since we only care about codeword lengths, we use code definitions that apply to $\mathcal{X}$ and have the same lengths $N$ as the (equivalent but possibly different) original definitions:

**Elias $\gamma$**
\[
c_{\gamma}(i) = \begin{cases} 
0, & i = 1 \\
1c_{\gamma}(\frac{i}{2}), & i \in \{2, 4, 6, \ldots\} \\
1c_{\gamma}(\frac{i}{2+1}), & i \in \{3, 5, 7, \ldots\} 
\end{cases}
\]

**Elias $\delta$**
\[
c_{\delta}(i) = c_{\gamma}(\lceil \log_2 i \rceil) b(i - 2^{\lceil \log_2 i \rceil} + 1, 2^{\lceil \log_2 i \rceil})
\]

**Elias $\omega$**
\[
c_{\omega}(i) = \begin{cases} 
0, & i = 1 \\
c_{\omega}(\lceil \log_2 i \rceil) b(i - 2^{\lceil \log_2 i \rceil} + 1, 2^{\lceil \log_2 i \rceil}) & i > 1 
\end{cases}
\]

**Golomb $\mu$**
\[
c_{\mu}(i) = \begin{cases} 
0, & i = 1 \\
c_{\omega}(\lceil \log_2 i \rceil) b(i - 2^{\lceil \log_2 i \rceil} + 1, 2^{\lceil \log_2 i \rceil}) & i > 1 
\end{cases}
\]

**EGk $c_{\text{EGk}}(i)$**
\[
c_{\text{EGk}}(i) = \begin{cases} 
0, & i = 1 \\
100, & i = 2 \\
101, & i = 3 \\
10^{q_i} 00b(i - 2^{q_i} + m_i), & i < q_i \\
10^{q_i} 01b(i - q_i, 2^{q_i} - m_i), & i \geq q_i 
\end{cases}
\]

where $c_{\gamma}(i)$ is all but the last bit of $c_{\gamma}$, $g_i \triangleq \lg i$, $m_i \triangleq (2^{q_i} - (-1)^{q_i})/3$, and $q_i \triangleq 2^{q_i} + m_i$. Recall that $b(j, k)$ denotes the $(j + 1)$th codeword of a complete binary code with $k$ items.

For these codes, $\alpha, \beta, \mu, \tau > 0, \nu > 0$ can be
\[
\begin{array}{c|cccc}
\gamma, \text{Yokoo} & \alpha & \beta & \mu & \tau \\
\hline
\text{(i > 1)} & 2 & 2 & -1 & 2\lg e \\
\text{Code k} & 1 & 2 & 2 & 2\lg e \\
\text{(k < 0)} & 0 & 1 & 2 & 2\lg e \\
\hline
\end{array}
\]

where $\alpha_0 = 1 - 2\lg 3$. (Parameters for $\delta$ codes, $\omega$ codes, EGk codes, and Code $k$ for $k > 0$ can be similarly formulated, but these are unused here, as the $\gamma$ code is clearly better for all distributions considered.)

For finding the best code within code families with multiple codes — such as Code $k$, EGk, and Gk (Golomb code $k$, defined in the main text) — partial sums can be used to limit the number of codes tested to a finite number. For example, these codes have $n(1) \rightarrow \infty$ as $k \rightarrow +\infty$, so at some point $p(1)n(1)$ will be too large to consider Code $k$ with parameters $k > k_{\text{max}}$ for some $k_{\text{max}}$. Similarly, as $k \rightarrow -\infty$, the unary portion of the code can be used for the partial sum.

Lacking $\alpha, \beta, \mu, \tau, \nu$, an obvious lower bound for $\sum_{i=1}^{\infty} p(i)\gamma(i)$ is $\sum p(i)n(i)$, but a more accurate bound can be found via entropy bounding with a value of $x$ such that $\sum_{i=1}^{\infty} 2^{-n(i)} = 1 - 2^{-y_x}$ for some $y_x$. For such values, since the code can be assumed without loss of generality to be monotonic, the codewords can be assumed to be all the leaves of a subtree rooted at depth $y_x$. Since any normalized tree is subject to the entropy bound $\sum p(i)n(i) \geq H(P)$, we can normalize to find a useful bound for the overall code. Let us first assign
\[
\sigma_x \triangleq \sum_{i=1}^{x} p(i), \quad H_x \triangleq \frac{\sum_{i=1}^{\infty} p(i)\lg 1/p(i)}{\sum_{i=1}^{\infty} p(i)}
\]
\[H_x^{\text{cond}} \triangleq \sum_{i=1}^{\infty} p(i)\lg 1 - \sigma_x \frac{p(i)}{p(i)} = \lg(1 - \sigma_x) + \frac{H_x}{1 - \sigma_x}
\]
where $H_x$ can be lower-bounded by as previously described. Thus, applying the entropy bound to the normalized subtree,
\[
\sum_{i=x}^{\infty} p(i) \geq (y_x + H_x^{\text{cond}})(1 - \sigma_x)
\]
This is useful for the codes calculated on the fly, e.g., Golin’s codes.

Golin’s original approach, alg$I$, starts by finding the minimum value $k_1$ such that
\[
\sum_{i=1}^{2^{k_1}} p(i) > \frac{3 - \sqrt{5}}{2} = 0.381966 \ldots
\]
and assigning the first $2^{k_1}$ inputs code $0b(i - 1, 2^{k_1})$. The algorithm then normalizes the remaining inputs and finds the minimum value $k_2$ such that
\[
\sum_{i=2^{k_1} + 1}^{2^{k_1} + 2^{k_2}} p(i) > \frac{3 - \sqrt{5}}{2}\text{ where } p_1(i) = \frac{p(i)}{1 - \sum_{j=1}^{k} p(j)}
\]
and assigns the next $2^k$ inputs code $10b(i-1 - 2^h,j)$. Continuing as needed, the algorithm sequentially finds minimum $k_h$ (given $k_1$ through $k_{h−1}$) such that

$$S(k_h, P) \triangleq \sum_{i=1}^{K(h)} K(h−1) p(i) \mid 1 - \sum_{i=1}^{K(h)} p(i) > 3 - \sqrt{5} \over 2$$

where $K(h) \triangleq \sum_{j=1}^{h−1} 2^j$, and assigns code

$$t^{h−1}0b(i - 1 - \sum_{j=1}^{h−1} 2^j)$$

to items $1 + K(h−1) = 1 + \sum_{j=1}^{h−1} 2^j$ through $K(h) = \sum_{j=1}^{h} 2^j$. This top-down approach is quite similar to Shannon–Fano coding [44], a modification of which results in

$$\sum_{i=1}^{H} a(i)\mid 1 - \sum_{i=1}^{H} a(i) > 3 - \sqrt{5} \over 2$$

that is, the group of a power of two that results in the most even division between those grouped and those left ungrouped. (Note that Shannon–Fano codes use the overall “best split” whereas these codes use the best split that groups items together in powers of two.)

REFERENCES

[1] V. Pareto, Cours d’économie politique. Geneva, Switzerland: Droz, 1896.

[2] H. A. Simon, “On a class of skew distribution functions,” Biometrika, vol. 42, no. 3/4, pp. 425–440, 1955.

[3] F. Auerbach, “Das Gesetz der Bevölkerungskonzentration,” Petermanns Geographische Mitteilungen, vol. 59, pp. 74–76, 1913.

[4] J. B. Estoup, Gammes Stenographiques. Paris, France: Institut Stenographique de France, 1916.

[5] G. K. Zipf, “Relative frequency as a determinant of phonetic change,” Harvard Studies in Classical Philology, vol. 40, pp. 1–95, 1929.

[6] G. U. Yule, “A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.,” Philos. Trans. Roy. Soc. London Ser. B, vol. 213, pp. 21–87, 1928.

[7] C. F. Gauss, “Eine Aufgabe der Wahrscheinlichkeitsrechnung,” 1800, available from http://www.gdz.sub.uni-goettingen.de/cgi-bin/digitbgci?PPN225957348

[8] R. O. Kuzmin, “Sur un probleme de Gauss,” in Atti del Congresso Internazionale dei Matematici, vol. 6, Sept. 1928, pp. 83–89.

[9] M. Mitzenmacher, “A brief history of generative models for power law and lognormal distributions,” Internet Mathematics, vol. 1, no. 2, pp. 226–251, 2004.

[10] M. E. J. Newman, “Power laws, Pareto distributions and Zipf’s law,” Contemporary Physics, vol. 46, no. 5, pp. 323–351, Sept. 2005.

[11] N. N. Taleb, The Black Swan: The Impact of the Highly Improbable. New York, NY: Random House, 2007.

[12] J. Abrahams, “Huffman-type codes for infinite source distributions,” IEEE Trans. Inf. Theory, vol. IT-43, no. 6, pp. 2019–2029, June 1997.

[13] D. A. Huffman, “A method for the construction of minimum-redundancy codes,” Proc. IRE, vol. 40, no. 9, pp. 998–1010, Sept. 1952.

[14] J. L. Bentley and A. C. Yao, “An almost optimal solution for unbounded searching,” Inf. Processing Letters, vol. 5, no. 3, pp. 82–87, Aug. 1976.

[15] A. Kato, “Huffman-like optimal codes and search codes for infinite alphabets,” IEEE Trans. Inf. Theory, vol. IT-46, no. 2, pp. 121–135, Mar. 2000.

[16] D. E. Knuth, The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, 3rd ed. Reading, MA: Addison-Wesley, 1997.

[17] D. W. Matula and P. Kornerup, “An ordering preserving finite binary encoding of the rationals,” in Proc. 6th Symposium on Computer Arithmetic, 1983, pp. 201–209.