Optimal prefix codes for some families of two-dimensional geometric distributions

Frédérique Bassino, Julien Clément, Gadiel Seroussi, Alfredo Viola

Optimal prefix codes for some families of two-dimensional geometric distributions

Frédérique Bassino[1], Julien Clément[2], Gadiel Seroussi[3], and Alfredo Viola[4⋆]

[1] Université de Marne-la-Vallée, France. bassino@univ-mlv.fr.
[2] CNRS UMR 6072, GREYC, Université de Caen, France. Julien.Clement@info.unicaen.fr.
[3] Mathematical Sciences Research Institute, Berkeley, CA, USA. gadiel@msri.org.
[4] Universidad de la República, Montevideo, Uruguay. viola@fing.edu.uy.

Abstract. Lossless compression is studied for pairs of independent integer-valued symbols emitted by a source with a geometric probability distribution of parameter $q \in (0,1)$. Optimal prefix codes are described for $q = 1/2^k$ $(k > 1)$ and $q = 1/\sqrt[k]{2}$ $(k > 0)$. The codes described differ from previously characterized cases related to the geometric distribution in that their corresponding trees are of unbounded width, and in that an infinite set of distinct optimal codes is required to cover any interval $(0, \varepsilon)$, $\varepsilon > 0$, of values of q.

1 Introduction

In 1966, Golomb [1] described optimal prefix codes for some geometric distributions over the nonnegative integers. In [2], these *Golomb codes* were shown to be optimal for *all* geometric distributions, namely, distributions of the form

$$p_i = (1-q)q^i \quad i \geq 0,$$

for some real value of q, $0 < q < 1$, where p_i represents the probability of the implied random variable assuming the value of i. Geometric distributions arise in practice when encoding *run lengths* (Golomb's original motivation in [1]), and in image compression when encoding prediction residuals, which are well-modeled by *two-sided geometric distributions*. Optimal prefix codes for the latter were characterized in [3], based on some (sometimes non-intuitive) variants of Golomb codes. Codes based on the Golomb construction have the practical advantage of allowing the encoding of a symbol i using a simple formula based on the integer value of i, without the need for code tables or other non-trivial memory requirements. This has led to their adoption in many practical applications (see, e.g., [4],[5]).

When dealing with sequences of independent, identically distributed random variables, however, symbol-by-symbol encodings can incur significant redundancy relative to the entropy of the distribution, especially in the low entropy range. One way to ameliorate this problem, while keeping the simplicity and low latency of the encoding and decoding operations, is to consider short blocks of d symbols, and use a prefix code for the blocks. In this paper, we study optimal prefix codes for pairs (blocks of length d=2) of independent, identically and geometrically distributed random variables, namely, distributions on pairs of nonnegative integers (i, j) with

$$\mathrm{Prob}\big((i,j)\big) = p_i\, p_j = (1-q)^2 q^{i+j} \quad i, j \geq 0. \tag{1}$$

⋆ This research is partially supported by project PDT S/C/IF/54/178 2006-2008.

We refer to this distribution as a *two-dimensional geometric distribution (TDGD)*, defined on the alphabet of integer pairs $\mathcal{A} = \{ (i,j) \mid i,j \geq 0 \}$.

Aside from the mentioned practical motivation, the problem is of intrinsic combinatorial interest. It was proven in [6] that, if the entropy $-\sum_{i\geq 0} p_i \log p_i$ of a distribution over the nonnegative integers is finite, optimal (prefix) codes exist and can be obtained, in the limit, from Huffman codes for truncated versions of the alphabet. However, the proof does not give a way for effectively constructing these optimal codes, and in fact, there are few families of distributions over the integers for which an effective construction is known [7, 8]. An algorithmic approach to building optimal codes is presented in [8], which includes geometric distributions and various generalizations. The approach, though, is not applicable to TDGDs, as explicitly noted in [8]. Some fundamental characteristic properties of the families of codes for the one-dimensional case turn out not to hold in the two-dimensional case. Specifically, the codes described in [1] and [3] satisfy the following: (a) for a fixed value of the parameter q, the *width* of the code tree (number of codewords of any one length) is bounded, and (b) there is a value $q = q_0 > 0$ such that all distributions (from the respective family) with $q < q_0$ admit the same optimal prefix code. As we shall see in the sequel (and was also noted with a different terminology in [8]), these properties do not hold for TDGDs.

The remainder of this extended summary is structured as follows. In Section 2 we present some background and notation, and we describe the technique of Gallager and Van Voorhis [2] for constructing optimal prefix codes for infinite alphabets, which we also apply in our constructions. As noted already in [2], most of the work and ingenuity in applying the technique goes into discovering appropriate "guesses" of the basic components on which the construction iterates, and in describing the structure of the resulting codes. That is indeed where most of our effort will be spent in the subsequent sections. In Section 3, we present a construction of optimal codes for two-dimensional geometric distributions with $q = 2^{-k}$ for any integer $k > 1$. We compute the *Kraft polynomials* [9] of the optimal codes, and use them to find expressions for the average codelength, which we apply to estimate the per-symbol redundancy of the codes relative to the entropy rate of geometric distribution. In Section 4, we briefly describe the construction of optimal codes for distributions with $q = 1/\sqrt[k]{2}$ for any positive integer k.

For both of the families of parameters studied in sections 3 and 4, the code trees obtained have only a finite number of non-isomorphic *whole subtrees* (i.e., subtrees consisting of a node and all of its descendants). However, contrary to the previously known results, the tree widths are not bounded, and, in the case $q = 2^{-k}$, there is an infinite sequence of distinct codes as $k \to \infty$, i.e., $q \to 0$, again in contrast with previously characterized cases. We show, however, that there is a *limiting code* as $k \to \infty$, in the sense that there exists an unbounded function $L(k)$ such that all optimal code trees for $k' \geq k$ are identical in their first $L(k)$ levels. Finally, in Section 5 we present some open problems and directions for further research. We note that in the one-dimensional case, it was proved in [2] that the family of codes that are optimal for the values $q = 1/\sqrt[k]{2}$ studied in Section 4 (the same sequence, dense near $q = 1$, mentioned by Golomb in [1]) contains optimal codes for *all* values

of q. This is not true in the two-dimensional case, even when considering also the values $q = 2^{-k}$ of Section 3, which are dense near 0. Thus, one of our main open problems is the complete characterization of a family of optimal codes covering all values of q. Another direction for further research is the extension of the results to blocks of $d > 2$ integer symbols. With a slightly different formalism, presented in [10], the construction of Section 4 extends to arbitrary values of d.

Given the space constraints of this extended summary, most results are presented without proof, and some descriptions are very brief. Complete proofs and descriptions, as well as additional results, will be given in the full version [10].

2 Preliminaries

We are interested in encoding the alphabet $\mathcal{A}$ of integer pairs (i, j), $i, j \geq 0$, using a binary prefix code C. As is well known, such a code can be associated with a rooted binary tree, whose leaves correspond, bijectively, to symbols in $\mathcal{A}$, and where each branch is labeled with a binary digit. The binary codeword assigned to a symbol is "read off" the labels of the branches on the path from the root of the tree to the corresponding leaf. We shall not distinguish between the code C and its associated binary tree, or between alphabet symbols and leaves of the tree. Also, two trees will be considered *equivalent* if for each $\ell \geq 0$, both trees have the same number of leaves at depth ℓ.

We call $s(i, j) = i + j$ the *signature* of $(i, j) \in \mathcal{A}$. For a given signature $f = s(i, j)$, there are $f+1$ pairs with signature f, all with the same probability, $w(f)=(1-q)^2 q^f$, under the distribution (1) on $\mathcal{A}$. Hence, given a prefix code C, symbols of the same signature may be freely permuted without affecting the average code length of C. Thus, for simplicity, we can also regard the correspondence between leaves and symbols as one between leaves and elements of the *multiset*

$$\overline{\mathcal{A}} = \{0, 1, 1, 2, 2, 2, \ldots, \underbrace{f, \ldots, f}_{f+1 \text{ times}}, \ldots\}. \tag{2}$$

In constructing the tree, we do not distinguish between different occurrences of a signature f; for actual encoding, the $f+1$ leaves labeled with f are mapped to the symbols $(0, f), (1, f-1), \ldots, (f, 0)$ in some arbitrary order.

Consider a prefix code C. Let T be a subtree of C, and let $s(x)$ denote the signature associated with a leaf x of T. We define the *weight*, $w(T)$, and *cost*, $c(T)$, of T, respectively, as

$$w(T) = \sum_{x \text{ leaf of } T} w(s(x)), \qquad \text{and} \qquad c(T) = \sum_{x \text{ leaf of } T} \text{depth}(x) w(s(x)),$$

with $w(f) = (1-q)^2 q^f$ for $f \geq 0$. When $T = C$, we have $w(T) = 1$, and $c(T)$ is the average code length of C. Our goal is to find a prefix code C that minimizes this cost.

In deriving the structure and optimality of our prefix codes, we shall rely on the method outlined below, due to Gallager and Van Voorhis [2], and adapted here to our terminology and notation.

– Define a countable sequence of *reduced alphabets* $(\mathcal{S}_f)_{f=-1}^{\infty}$, where $\mathcal{S}_f$ is a multiset containing the signatures $0, 1, \ldots, f$ (with multiplicities as in (2)), and where the signatures strictly greater than f are partitioned into a finite number of nonempty classes we shall refer to as *virtual symbols*, which are also elements of $\mathcal{S}_f$. We naturally associate with each virtual symbol a probability equal to the sum of the probabilities of the signatures it contains.
– Verify that the sequence of reduced alphabets $(\mathcal{S}_f)_{f=-1}^{\infty}$ is compatible with the bottom-up Huffman procedure. This means that after a certain number of merging steps of the Huffman algorithm on the reduced alphabet $\mathcal{S}_f$, one gets $\mathcal{S}_{f'}$ with $f' < f$.[1]
– Apply the Huffman algorithm to the reduced alphabet $\mathcal{S}_{-1}$ (containing no signatures of the original alphabet, only virtual symbols).
– Finally, use a convergence argument that ensures that the sequence of finite codes $(C_n)_{n\geq 1}$ obtained *converges* to an infinite code C. Namely for every $i \geq 1$, codewords of C_n being sorted by non-decreasing length following the lexicographical order, the ith codeword of C_n is eventually constant when n grows, and equal to the ith codeword of C.

This method, utilized also in [3], produces an optimal prefix code for the given probability distribution on the underlying countable alphabet. The difficult part is to guess the structure of the sequence of reduced alphabets.

Quasi-uniform sources. We say that a finite source with probabilities $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{N-1}$ is *quasi-uniform* if either $N \leq 2$ or $\sigma_0 \leq \sigma_{N-2} + \sigma_{N-1}$. An optimal prefix code for a quasi-uniform source of N probabilities consists of $2^{\lceil \log N \rceil} - N$ codewords of length $\lfloor \log N \rfloor$, and $2N - 2^{\lceil \log N \rceil}$ codewords of length $\lceil \log N \rceil$, the shorter codewords being assigned to the symbols with the larger probabilities [2]. We refer to such a code as a *quasi-uniform code*, and denote by $\mathcal{Q}(i, N)$ the codeword it assigns to the ith symbol, which has probability σ_i, $0 \leq i < N$.

3 Family of parameters $q = 2^{-k}$

We introduce some notations for describing the recursive construction of trees with weights associated to their leaves. The notation will be based on grammatical production rules together with scalar multiplication. After assuming that the integer k defining the parameter $q = 2^{-k}$ is fixed, we slightly abuse notation and regard q as a symbolic indeterminate in the production rules. A leaf associated with weight q^f will be denoted $\boxed{q^f}$ (in turn, this weight will be associated with the signature f, the normalizing coefficient $(1-q)^2$ being immaterial to the construction). Given a tree $\mathcal{T}$ and a scalar quantity g, $g\mathcal{T}$ denotes the tree resulting from multiplying the weights of all the leaves of $\mathcal{T}$ by g.

The tree $\mathcal{C}^m$ is defined as the complete tree of depth m, with 2^m leaves labeled $\boxed{q^0}$ (or, equivalently, $\boxed{1}$). Its construction can be described by the following production rules:

$$\mathcal{C}^m \to \begin{array}{c} \bullet \\ \diagup\ \diagdown \\ \mathcal{C}^{m-1}\ \mathcal{C}^{m-1} \end{array}, \quad \mathcal{C}^0 \to \boxed{1}.$$

[1] A way to test Huffman compatibility is to use the *sibling property* [11] that characterizes Huffman trees as the trees whose nodes can be listed in non-increasing order of probability in such way that two sibling nodes are adjacent in the list.

The *infinite* tree (and associated multiset of leaf weights) $\mathcal{L}_q^k$ is defined by the following rules, where k is the fixed integer referred to above:

$$\mathcal{L}_q^0 \to q\mathcal{L}_q^k, \quad \mathcal{L}_q^m \to \begin{array}{c} \bullet \\ \mathcal{L}_q^{m-1}\ \mathcal{C}^{m-1} \end{array}, \text{ for } 0 < m \le k.$$

In words, $\mathcal{L}_q^k$ consists of a complete tree $\mathcal{C}^k$ with 2^k-1 leaves of weight q^0, and with the remaining leaf serving as the root of $q\mathcal{L}_q^k$. Thus, $\mathcal{L}_q^k$ has 2^k-1 leaves of weight q^f at depth $(f+1)k$ for all $f\ge 0$, and no other leaves.

The main result of this section is presented in the following proposition that can be seen as describing, at the same time, the optimal tree, and the sequence of reduced alphabets used in the proof of optimality following the method of [2].

Proposition 1. *Let $q = 2^{-k}$ with $k > 1$. Then, signatures $f \in \overline{\mathcal{A}}$ are distributed in the optimal prefix tree for the TDGD with parameter q according to the following cases:*

1. *Assume $0 \le f < 2^{k-1}$, and write $f = 2^i + j - 1$ with $0 \le j \le 2^i - 1$. Then all signatures f are distributed on two levels in the following way:*

$$q^f \cdot \Big[\underbrace{\boxed{1}\cdots\boxed{1}}_{2^{k-1}-j-1 \text{ times}} \quad \mathcal{R}_f \ \boxed{1} \underbrace{\Big(\begin{array}{c} \bullet \\ \boxed{1}\ \boxed{1} \end{array} \Big)}_{j \text{ times}} \Big]$$

   The multiset $q^f\mathcal{R}_f$ represents a tree containing all the signatures strictly greater than f.
2. *Let $f \ge 2^{k-1}$, and write $f = 2^{k-1} - 1 + \ell(2^k - 1) + j$. Then the signatures f are distributed in the coding tree according to the five cases below. The trees (and associated multisets) $q^f\mathcal{R}_j$ represent a virtual symbol containing all the signatures not contained in the other virtual symbols of types $\mathcal{C}^{k-1}$ and $\mathcal{L}_q^{k-1}$ at the same level.*

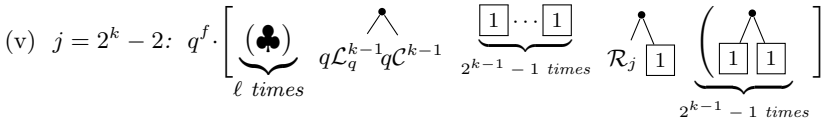
   In the following five cases, (♣) stands for $\Big[\begin{array}{c} \bullet \\ q\mathcal{L}_q^{k-1}\ q\mathcal{C}^{k-1} \end{array} \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^k-1 \text{ times}} \Big]$

   (i) $0 \le j < 2^{k-1} - 2$: $q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell \text{ times}} \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^{k-1}-j-1 \text{times}} \ \mathcal{R}_j \ \boxed{1} \underbrace{\Big(\begin{array}{c} \bullet \\ \boxed{1}\ \boxed{1} \end{array} \Big)}_{j \text{ times}} \Big]$

   (ii) $j = 2^{k-1} - 2$: $q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell \text{ times}} \ \begin{array}{c} \bullet \\ q\mathcal{C}^{k-1}\ \mathcal{R}_j \end{array} \ \underbrace{\Big(\begin{array}{c} \bullet \\ \boxed{1}\ \boxed{1} \end{array} \Big)}_{2^{k-1}-1 \text{ times}} \Big]$

   (iii) $2^{k-1}-2 < j < 2^k - 3$: $q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell \text{ times}} \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^{k-1} \text{ times}} \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^k-2-j \text{ times}} \ \begin{array}{c} \bullet \\ q\mathcal{C}^{k-1}\ \mathcal{R}_j \end{array} \ \underbrace{\Big(\begin{array}{c} \bullet \\ \boxed{1}\ \boxed{1} \end{array} \Big)}_{j-2^{k-1}+1 \text{ times}} \Big]$

   (iv) $j = 2^k - 3$: $q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell \text{ times}} \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^{k-1} \text{ times}} \ \boxed{1} \ \begin{array}{c} \bullet \\ q\mathcal{L}_q^{k-1}\ \mathcal{R}_j \end{array} \ \underbrace{\Big(\begin{array}{c} \bullet \\ \boxed{1}\ \boxed{1} \end{array} \Big)}_{2^{k-1}-2 \text{ times}} \Big]$



The proof (which is omitted here) computes the weights of the signatures and virtual symbols in each case, and verifies that the sibling property holds. It also verifies that by applying the Huffman procedure to the reduced alphabet corresponding to each case, one obtains a configuration corresponding to the previous case, in cyclic fashion (i.e., (v)→(iv)→(iii)→(ii)→(i)→(v)), with the value of ℓ decreasing by one with each cycle, until Case 2(i) is reached with ℓ=0 and j=0, in which case the Huffman merging leads to Case 1 of the proposition.

The construction of the optimal prefix tree stemming from Proposition 1 can be outlined as follows.

1. The first level of the tree (descending directly from the root) is composed of two nodes labeled by $\boxed{1}$ and $\mathcal{R}_0$ respectively (Case 1 with $f = 0$). As long as $f < 2^{k-1}$, $q^{f-1}\mathcal{R}_{f-1}$ is replaced by the subtree associated with the quasi-uniform code for the $f + 1$ symbols of signature f and the virtual symbol $q^f\mathcal{R}_f$ containing all the symbols strictly greater than f.
2. The rest of the tree can be constructed in slices for $f \geq 2^{k-1}$, where each slice contains all the external nodes with signatures $2^{k-1} + \ell(2^k - 1) \leq f < 2^{k-1} + (\ell + 1)(2^k - 1)$ $(\ell \geq 0)$.

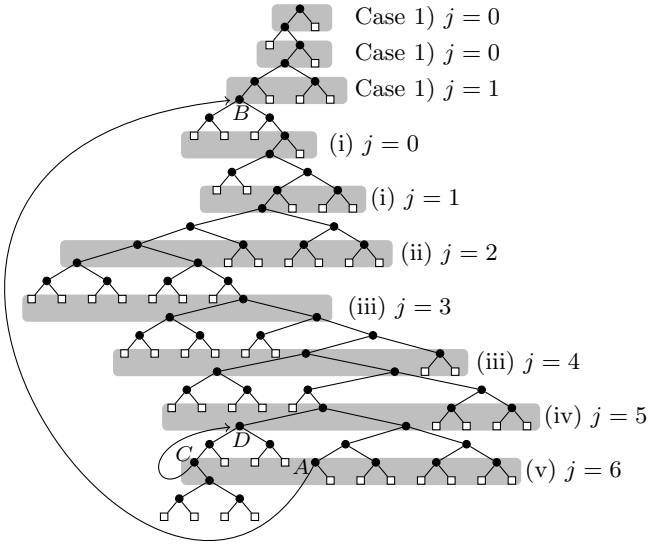


Fig. 1. Optimal prefix code tree for a TDGD with q=1/8.

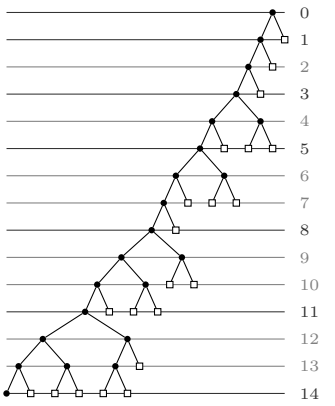


Fig. 2. Top of a limit tree for $q = 2^{-k}$ when $k \to \infty$ (first fourteen levels).

Example. Figure 1 describes the structure of the infinite optimal coding tree when $k = 3$ ($q = 1/8$). The cycling structure of Proposition 1 ((i)→(ii)→(iii)→(iv)→(v)→(i)

when traversing top-down) is represented by the loop-back edge $A \to B$, which indicates that a copy of the tree rooted at B is inserted as a child of A (with labels appropriately shifted). The subtree rooted at D, with the loop-back edge $C \to D$, describes in a concise manner the infinite tree $\mathcal{L}_q^3$. Each time the edge $A \to B$ is traversed, the parameter ℓ (number of patterns (♣) in Proposition 1) increases by one. The first instance of node A occurs for $f = 9$ (that is, Case 2(v) with $j = 6$ and $\ell = 0$). Then, $f = 10$ is described by Case 2(i) with $j = 0$ and $\ell = 1$. The 11 leaves with signature $f = 10$ are obtained as follows. The first four leaves are those shown at the last level of Figure 1, descending from the first recursive occurrence of node B. Three more come from the loop-back edge $C \to D$ (the three leaves two levels down from node D). The remaining four leaves come from the back edge $A \to B$, and are the leaves in the first three levels of the subtree rooted at B.

Notice that it follows from Proposition 1 that the width of the optimal tree for a given value of $q = 2^{-k}$ is unbounded (for example, each of the cases in the proposition has parts that grow monotonically with ℓ, which is unbounded). Also, since, clearly, different values of k lead to different trees $\mathcal{L}_q^k$, it also follows from the proposition that, as $k \to \infty$, we have an infinite sequence of distinct optimal trees. As mentioned in Section 1, the opposite properties hold for the previously characterized cases of distributions derived from the geometric distribution (cf. [8]).

A limit code. The construction of the part of the optimal tree for $1/2^{k'}$ that contains all the symbols $0 \le f < 2^{k-1} - 3$ (Case 1 and Case 2 (i) with $\ell = 0$) is the same for all $k' \ge k$. More precisely, one can build this part of the tree, beginning with $f = 0$ and the quasi-uniform coding tree for $f + 2$ symbols, by recursively replacing f by $f + 1$, as long as $f < 2^k - 3$, and the first leaf of the deepest level of the current tree by the quasi-uniform coding tree for $f + 2$ symbols. For $0 \le f \le 2^k - 4$, the quasi-uniform coding tree with $f + 2$ symbols defines $\lceil \log_2 (f + 2) \rceil$ levels in the optimal tree for 2^{-k}. Collecting these contributions for f from 0 to $2^k - 4$, we verify that the distribution of nodes in the claimed number of levels remains invariant for $k' > k$. Therefore, the distribution of the number of leaves at each level in the optimal trees for $q = 2^{-k}$ progressively "stabilizes" as $k \to \infty$. This observation leads to the following proposition.

Proposition 2. *When $k \to \infty$, the sequence of optimal coding trees for $q = 2^{-k}$ converges to a limit tree that can be constructed, up to tree equivalence, as follows: beginning with the quasi-uniform coding tree for $n = 2$ symbols, recursively replace the first leaf of the last level of the current tree by the quasi-uniform coding tree for $n + 1$ symbols, and increase n.*

Figure 2 shows the first fourteen levels of the limit tree of Proposition 2. Notice that the first eleven levels of the limit tree coincide with those of the tree of Figure 1, up to reordering of nodes at each level. The limit code admits a very simple encoding procedure: given a pair (m, n), with signature $f = m+n$, we write $f = 2^i + j - 1$, with $0 \le j < 2^i$ and $i \ge 0$. We encode (m, n) with a binary codeword xy, where $x = 0^{(i-1)(f+1)+2j+1}$ identifies the path to the root of the quasi-uniform tree that contains all the leaves of signature f, and $y = \mathcal{Q}(m + 1, f + 2)$ (recall that $\mathcal{Q}(m, N)$ is the code for m in a quasi-uniform code on N symbols). A matching decoding

procedure is easily derived. Encoding and decoding procedures for all the codes in this section are presented in [10].

Kraft polynomial Let Σ be an alphabet with a probability distribution $(\mu_i)_{i\in\Sigma}$ and C a prefix code on Σ with codewords lengths $(\ell_i)_{i\in\Sigma}$. The *Kraft polynomial* [9] of the code C is the formal series

$$P(z) = \sum_{i\in\Sigma} \mu_i z^{\ell_i}.$$

The average codelength of C is obtained as $c(C) = z\frac{\partial}{\partial z}P(z)\big|_{z=1}$.

We derive the Kraft polynomial for the optimal codes of Proposition 1. For $0 \leq j < 2^k - 1$, and any integer i, define

$$Q_j(z) = \begin{cases} (2^{k-1}-j-1)z^{k-1} + (2j+1)z^k & \text{if } 0 \leq j < 2^{k-1}-2 \\ 2(2^{k-1}-1)z^k & \text{if } j = 2^{k-1}-2 \\ (3\times 2^{k-1}-j-2)z^{k-1} + 2(j-2^{k-1}+1)z^k & \text{if } 2^{k-1}-2 < j < 2^k-3 \\ (2^{k-1}+1)z^{k-1} + 2(2^{k-1}-2)z^k & \text{if } j = 2^k-3 \\ (2^{k-1}-1)z^{k-1} + (2^k-1)z^k & \text{if } j = 2^k-2 \end{cases}$$

$$\text{and } P_{i,j}(z) = (2^i - j - 1)z^i + (2j+1)z^{i+1}.$$

Then, the Kraft polynomials $P_{2^{-k}}(z)$ for the optimal codes (when $q = 2^{-k}$), and $P_0(z)$ for the limit code, are given by:

$$P_{2^{-k}}(z) = (1-q)^2\Big(\sum_{i=0}^{k-2}\sum_{j=0}^{2^i-1} q^{2^i+j-1}P_{i,j}(z)z^{2^i(i-1)+1+j(i+1)} + \frac{q^{2^{k-1}-1}z^{2^{k-1}(k-2)+1}}{1-(qz^k)^{2^k-1}}\Big(\sum_{j=0}^{2^k-2}(qz^k)^j Q_j(z) + (2^k-1)\frac{1}{z}\frac{(qz^k)^{2^k}}{1-qz^k}\Big)\Big)$$

$$\text{and } P_0(z) = (1-q)^2\sum_{i=0}^{\infty}\sum_{j=0}^{2^i-1} q^{2^i+j-1}P_{i,j}(z)z^{2^i(i-1)+1+j(i+1)}.$$

Redundancy. Figure 3 presents plots of redundancy per integer symbol as a function of q, relative to the entropy rate of the geometric distribution of parameter q, for $q \leq \frac{1}{2}$. Let C_k denote the optimal prefix code for a TDGD with $q = 2^{-k}$. Plots are shown for the Golomb code on single integer symbols, the best code C_k for each q, and the optimal code for each q. Code lengths for the latter were approximated empirically. It is observed in the figure that the family $\{C_k\}$ provides a good approximation to the optimal codes for arbitrary q over the low entropy range. However, we can also observe that optimal codes for some values of q will be strictly outside of the families characterized in this paper.

4 The family of parameters $q = 1/\sqrt[k]{2}$

We introduce the following grammar to define the trees $\mathcal{T}_u^d$ for $d = 0, 1, 2$.

$$\mathcal{T}_u^0 \to \boxed{1}, \quad \mathcal{T}_u^d \to \overset{\bullet}{\wedge}_{u\mathcal{T}_u^d\, u\mathcal{T}_u^{d-1}} \ (d = 1, 2). \tag{3}$$

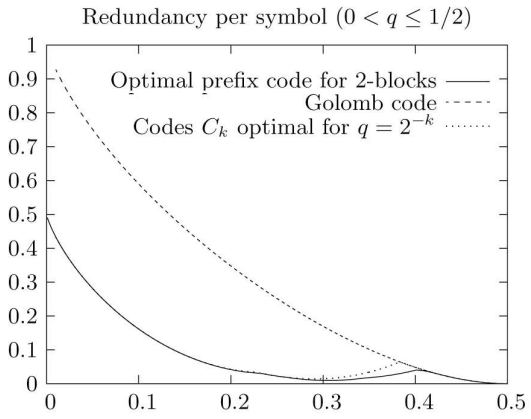


Fig. 3. The redundancy (in bits per symbol) for the optimal prefix code (empirical), the Golomb code and the optimal codes C_k.

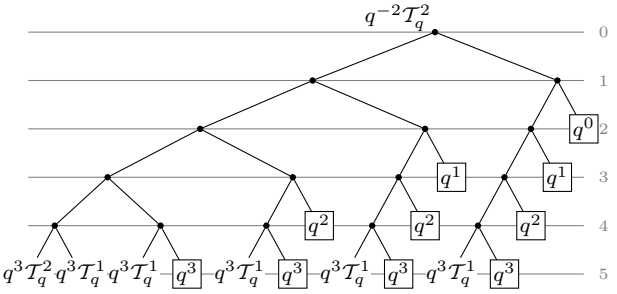


Fig. 4. Top of the tree $q^{-2}\mathcal{T}_q^2$.

The tree $q^r\mathcal{T}_{q^k}^1$ is easily recognized as that of a unary code (Golomb code of order 1), with leaves appropriately weighted. Moreover, the tree $q^r\mathcal{T}_{q^k}^2$ can be recognized as a convolution of two such unary trees, i.e., a unary where each leaf becomes the root of another unary tree (see Figure 4). It follows from this observation, and straightforward symbolic manipulations, that $w(q^r\mathcal{T}_{q^k}^2) = w(\boxed{q^r})\left(\frac{q^k}{1-q^k}\right)^2$. It is important to note that if $q = 1/\sqrt[k]{2}$ then, $w(q^r\mathcal{T}_{q^k}^2) = w(q^r\mathcal{T}_{q^k}^1) = w(q^r\mathcal{T}_{q^k}^0) = w(\boxed{q^r})$. This observation is the basis of the construction and proof in Proposition 3.

We observe also that the tree $q^{-2}\mathcal{T}_q^2$ gives an optimal prefix code for k=1, as can be verified by checking that this coding tree satisfies the sibling property of Huffman trees. The top of this tree is shown in Figure 4. This construction is generalized in Proposition 3 for all $k > 1$. This formalism is explored in [10], and can be generalized to the study of optimal codes on blocks of fixed $d > 2$ integer symbols.

Proposition 3. *Let $q = 1/\sqrt[k]{2}$ with $k \geq 1$. Then, an optimal prefix tree for a TDGD with parameter q can be obtained by the application of the Huffman algorithm to the finite source*

$$\mathcal{S}_T = \{\underbrace{q^{-2k}\mathcal{T}_{q^k}^2}_{1\ time}, \underbrace{q^{-2k+1}\mathcal{T}_{q^k}^2}_{2\ times}, \underbrace{q^{-2k+2}\mathcal{T}_{q^k}^2}_{3\ times} \ldots, \underbrace{q^{-k-1}\mathcal{T}_{q^k}^2}_{k\ times}\} \cup \{\underbrace{q^{-k}\mathcal{T}_{q^k}^2}_{k-1\ times}, \ldots, \underbrace{q^{-3}\mathcal{T}_{q^k}^2}_{2\ times}, \underbrace{q^{-2}\mathcal{T}_{q^k}^2}_{1\ times}\}.$$

It is shown in [10] that a prescribed sequence of pairings of symbols $q^{-i}\mathcal{T}_{q^k}^2$ with $2 \leq i \leq k-1$ leads from the reduced alphabet $\mathcal{S}_T$ to a quasi-uniform source. Efficient coding and decoding algorithms for the codes of Proposition 3 is also presented in [10].

Example. Let $k = 4$, i.e., $q = 1/\sqrt[4]{2}$. The source $\mathcal{S}_T$ has $k^2 = 16$ symbols, leading to a reduced alphabet $\mathcal{S}$ (corresponding to a quasi-uniform source) with 13 symbols. An optimal coding tree for $\mathcal{S}_T$ in this example is given in Figure 5. Circled nodes indicate the pairings leading to the quasi-uniform source $\mathcal{S}$.

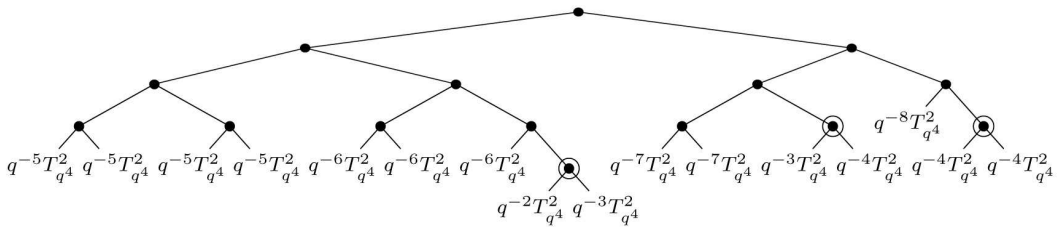


Fig. 5. The top of the tree for $\mathcal{S}_T$ for $q = 1/\sqrt[4]{2}$. Circled nodes are composite nodes of the source $\mathcal{S}$.

5 Conclusion

We characterized optimal prefix codes for two sub-families of two-dimensional geometric distributions, namely, those with parameters $q = 2^{-k}$ with $k>1$, or $q = 1/\sqrt[k]{2}$ with $k > 0$. The codes characterized here are in fact optimal for values of q in a set of positive measure containing the above discrete sequences. However, these codes do not cover the full range of values of q in the interval $(0, 1)$. Characterizing optimal prefix codes for TDGDs over the full interval is the subject of ongoing research. Future work will include also further generalizations to higher dimensions, i.e., blocks of $d > 2$ integer symbols. Of interest also is the derivation of analogous results for blocks of two-sided geometric distributions [3].

References

1. S. W. Golomb, "Run length encodings," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 399-401, 1966.
2. R. G. Gallager and D. C. V. Voorhis, "Optimal source codes for geometrically distributed integer alphabets," *IEEE Trans. Inf. Theory*, pp. 228–230, 1975.
3. N. Merhav, G. Seroussi, and M. J. Weinberger, "Optimal prefix codes for sources with two-sided geometric distributions," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 229–236, 2000.
4. R. F. Rice, "Some practical universal noiseless coding techniques," Tech. Rep. JPL-79-22, Jet Propulsion Laboratory, Pasadena, CA, Mar. 1979.
5. M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Proc.*, vol. 9, pp. 1309–1324, Aug. 2000.
6. V. T. T. Linder and K. Zeger, "Existence of optimal prefix codes for infinite source alphabets," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 2026–2028, 1997.
7. J. Abrahams, "Code and parse trees for lossless source encoding," *Communications in Information and Systems*, vol. 1, no. 2, pp. 113–146, 2001.
8. M. J. Golin and K. K. Ma, "Algorithms for constructing infinite huffman codes," Technical Report HKUST-TCSC-2004-07, HKUST, Hong Kong, China, July 2004.
9. M. G. G. Laidlaw, "The construction of codes for infinite sets," in *Proceedings of SAICSIT 2004*, (Stellenbosch, Western Cape, South Africa), pp. 157–165, 2004.
10. F. Bassino, J. Clément, G. Seroussi, and A. Viola, "Optimal prefix codes for two-dimensional geometric distributions." Preprint, 2006.
11. R. G. Gallager, "Variations on a theme by Huffman," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 668–674, 1978.