Optimal prefix codes for pairs of geometrically-distributed random variables

Frédérique Bassino
Institut Gaspard-Monge
Université de Marne-la-Vallée, France
bassino@univ-mlv.fr

Julien Clément
GREYC-CNRS
Université de Caen, France
clement@info.unicaen.fr

Gadiel Seroussi
MSRI
Berkeley, CA, USA
gadiel@msri.org

Alfredo Viola
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, $0 < q < 1$. Optimal prefix codes are described for $q = 1/2^k$ ($k > 1$) and $q = 1/\sqrt[k]{2}$ ($k > 0$). These codes retain some of the low-complexity and low-latency advantage of symbol by symbol coding of geometric distributions, which is widely used in practice, while improving on the inherent redundancy of the approach. From a combinatorial standpoint, 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.**

I. 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

$$\text{Prob}(i) = p_i = (1-q)q^i\,, \quad i \geq 0,$$

for some real-valued parameter q, $0 < q < 1$. 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 (cf. [4],[5]). For computational reasons, these applications often use a sub-family of the class of optimal codes as a good approximation for the whole class over the full range of q (e.g., "power-of-two" Golomb codes in [4] and [5]; see also [6]).

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. One way to mitigate this problem, while keeping the simplicity and low latency of the encoding and decoding operations, is to consider short blocks of $d{>}1$ 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

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

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 [7] (see also [8]) 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 general way for effectively constructing optimal codes, and in fact, there are few families of distributions over countable alphabets for which an effective construction is known [9][10]. An algorithmic approach to building optimal codes is presented in [10], which covers geometric distributions and various generalizations. The approach, though, is not applicable to TDGDs, as explicitly noted in [10]. 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 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 predicted with a different terminology in [10], optimal codes for TDGDs will not satisfy these properties.

The remainder of this extended summary is structured as follows. In Section II 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 III, we present a construction of optimal codes for TDGDs with $q = 2^{-k}$ for any integer $k > 1$.

We compute the *Kraft functions* (called *Kraft polynomials* in [11]) of the optimal codes, and use them to compute the average code lengths, which we apply to estimate the per-symbol redundancy of the codes relative to the entropy rate of the geometric distribution. In Section IV, we describe the construction of optimal codes for distributions with $q = 1/\sqrt[k]{2}$ for any positive integer k. We also summarize a study of the redundancy rates for both families of characterized codes over the full interval $0 < q < 1$. The study suggests that these codes provide a good approximation to the full class of optimal codes for TDGDs (the full characterization of which remains an open problem) over the entire range of q.

For both families of parameters studied in sections III and IV, 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., $q \to 0$. We show, however, that there exists 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 V we present some open problems and directions for further research. 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 [12].

II. 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 usual, we associate C with a rooted (infinite) 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, $w(f){=}(1-q)^2q^f$, 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 fixed 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

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

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.

- Define a countable sequence of finite *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 referred 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 $\mathcal{S}_{-1}$.

While the sequence of reduced alphabets $\mathcal{S}_{f'}$ can be seen as evolving "bottom-up," the infinite prefix code C constructed results from a "top-down" sequence of corresponding finite prefix codes $C_{f'}$. One shows that the sequence of codes $(C_{f'})_{f' \geq 1}$ *converges* to the infinite code C, in the sense that for every $i \geq 1$, with codewords of $C_{f'}$ consistently sorted, the ith codeword of $C_{f'}$ is eventually constant when f' grows, and equal to the ith codeword of C. A corresponding convergence argument on the sequence of average code lengths then establishes the optimality of C.

This method was successfully applied to characterize infinite optimal prefix codes in [2] and [3]. 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 corresponding to the more probable symbols [2]. We refer to such a code as a *quasi-uniform code* (or *tree*), denote it by Q_N, and denote by $\mathcal{Q}(i, N)$ the codeword it assigns to the symbol corresponding to probability σ_i, $0{\leq}i{<}N$.

Tree concatenation. We define the *concatenation* of two coding trees T_1 and T_2, denoted $T_1 \bullet T_2$, as the tree constructed by taking T_1, and making each of its leaves the root of a copy

[1] A way to test Huffman compatibility is to use the *sibling property* [13] 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.

of T_2. For example, the Golomb code of order $m \geq 1$ can be seen as the concatenation $G_m{=}Q_m \bullet G_1$, where G_1 is a *unary tree* consisting of a root whose children are a leaf (say, on the branch labeled '1'), and, recursively, a copy of G_1.

III. The family of parameters $q = 2^{-k}$

We introduce some notations, based on grammatical production rules together with scalar multiplication, for describing the recursive construction of trees with weights associated to their leaves. After assuming that the integer k defining $q = 2^{-k}$ 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.

We denote by $\mathcal{C}^m$ 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:

$$\mathcal{C}^0 \to \boxed{1}, \quad \mathcal{C}^m \to \overset{\bullet}{\diagup\!\diagdown}\!\left(\mathcal{C}^{m-1}\ \mathcal{C}^{m-1}\right).$$

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:

$$\mathcal{L}_q^0 \to q\mathcal{L}_q^k\,, \quad \mathcal{L}_q^m \to \overset{\bullet}{\diagup\!\diagdown}\!\left(\mathcal{L}_q^{m-1}\ \mathcal{C}^{m-1}\right) \quad \text{for } 0 < m \leq k.$$

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 $f{\geq}0$, and no other leaves.

The main result of this section is presented in the following proposition, where we describe the layers of the optimal prefix tree for a TDGD with parameter $q = 2^{-k}$, $k > 1$ (the case $k{=}1$ is covered in Section IV). The proposition 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 an optimal prefix tree for the TDGD with parameter q according to the following cases:

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

$$q^f \cdot \Big[\ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^i\,-\,j\,-\,1\text{ times}}\ \ \overset{\bullet}{\diagup\!\diagdown}\!\left(\mathcal{R}_f\ \boxed{1}\right)\ \ \underbrace{\overset{\bullet}{\diagup\!\diagdown}\!\left(\boxed{1}\ \boxed{1}\right)}_{j\text{ times}}\ \Big]$$

The multiset $q^f\mathcal{R}_f$ represents a tree containing all the signatures strictly greater than f.

2) Let $f \geq 2^{k-1}$, and write $f = 2^{k-1} - 1 + \ell(2^k - 1) + j$. Then the signatures f are distributed in the optimal coding tree according to the five cases below. The trees (and associated multisets) $q^f R_j$ 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. Also, for succinctness, the symbol $(\clubsuit)$ stands for $\Big[\ q\mathcal{L}_q^k\ \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^k\,-\,1\text{ times}}\Big]$.

(i) $0 \leq j < 2^{k-1}{-}2$:

$$q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell\text{ times}}\ \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^{k-1}\,-\,j\,-\,1\text{times}}\ \ \overset{\bullet}{\diagup\!\diagdown}\!\left(\mathcal{R}_j\ \boxed{1}\right)\ \ \underbrace{\overset{\bullet}{\diagup\!\diagdown}\!\left(\boxed{1}\ \boxed{1}\right)}_{j\text{ times}}\Big]$$

(ii) $j = 2^{k-1} - 2$:

$$q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell\text{ times}}\ \ \overset{\bullet}{\diagup\!\diagdown}\!\left(q\mathcal{C}^{k-1}\ \mathcal{R}_j\right)\ \ \underbrace{\overset{\bullet}{\diagup\!\diagdown}\!\left(\boxed{1}\ \boxed{1}\right)}_{2^{k-1}-1\text{ times}}\Big]$$

(iii) $2^{k-1}{-}2 < j < 2^k{-}3$:

$$q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell\text{ times}}\ \ \underbrace{\boxed{1}\cdots\boxed{1}}_{3\cdot 2^{k-1}-2-j\text{ times}}\ \ \overset{\bullet}{\diagup\!\diagdown}\!\left(q\mathcal{C}^{k-1}\ \mathcal{R}_j\right)\ \ \underbrace{\overset{\bullet}{\diagup\!\diagdown}\!\left(\boxed{1}\ \boxed{1}\right)}_{j-2^{k-1}+1\text{ times}}\Big]$$

(iv)] $j = 2^k{-}3$:

$$q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell\text{ times}}\ \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^{k-1}+1\text{ times}}\ \ \overset{\bullet}{\diagup\!\diagdown}\!\left(q\mathcal{L}_q^{k-1}\ \mathcal{R}_j\right)\ \ \underbrace{\overset{\bullet}{\diagup\!\diagdown}\!\left(\boxed{1}\ \boxed{1}\right)}_{2^{k-1}-2\text{ times}}\Big]$$

(v) $j = 2^k{-}2$:

$$q^f \cdot \Big[\underbrace{(\clubsuit)}_{\ell\text{ times}}\ \ q\mathcal{L}_q^k\ \ \underbrace{\boxed{1}\cdots\boxed{1}}_{2^{k-1}-1\text{ times}}\ \ \overset{\bullet}{\diagup\!\diagdown}\!\left(\mathcal{R}_j\ \boxed{1}\right)\ \ \underbrace{\overset{\bullet}{\diagup\!\diagdown}\!\left(\boxed{1}\ \boxed{1}\right)}_{2^{k-1}-1\text{ times}}\Big]$$

□

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 applying the Huffman procedure to the reduced alphabet corresponding to each case, one obtains a configuration corresponding to the previous case, in cyclic fashion, namely, (v)→(iv)→(iii)→(ii)→(i)→(v) (bottom-up). The value of ℓ decreases by one with each cycle, until Case 2(i) is reached with $\ell{=}0$ and $j = 0$, in which case the Huffman merging leads to Case 1 of the proposition.

The construction of the optimal prefix tree derived from Proposition 1 can be outlined as follows (top-down):

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 signatures strictly greater than f.
2) The rest of the tree is constructed in cycles, one for each value of $\ell \geq 0$, generating all the leaves with signatures f, $2^{k-1}{-}1{+}\ell(2^k{-}1) \leq f < 2^{k-1}{-}1{+}(\ell+1)(2^k{-}1)$. Within each cycle, the construction follows the top-down sequence of sub-cases (i)→(ii)→(iii)→(iv)→(v) of Case 2 of Proposition 1.

Example. Figure 1 describes the structure of the infinite optimal coding tree for $q = 1/8$ ($k{=}3$). The loop-back edges

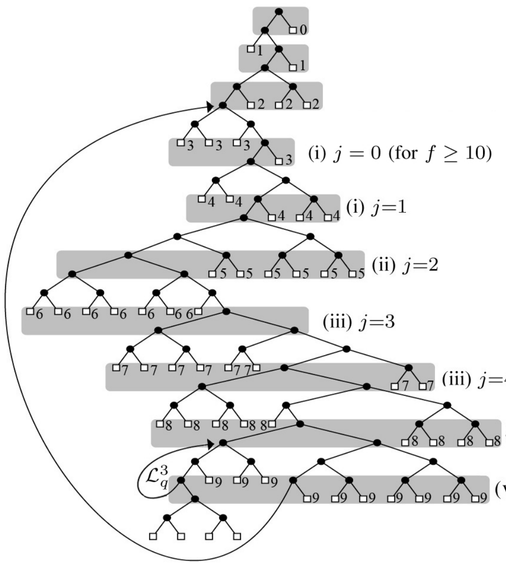

Fig. 1. Optimal prefix code tree for a TDGD with q=1/8, with leaf signatures noted for $f \leq 9$.

in the figure indicate that a copy of the tree rooted at the target node is inserted as a child of the originating node (with labels appropriately shifted). The cycling structure of Proposition 1 ((i)→(ii)→(iii)→(iv)→(v)→(i) when traversing top-down) is represented by the outer loop-back edge—each traversal of the edge represents an increase of the parameter ℓ of the proposition by one. The inner loop-back edge helps describe in a concise manner the infinite tree $\mathcal{L}_q^3$, which is rooted at the node originating the edge. Signatures $0 \leq f \leq 3$ are generated by Case 1 of Proposition 1. Signatures $4 \leq f \leq 9$ correspond to the subcases of Case 2 indicated in the figure, with ℓ=0, and the value of j also indicated. Signature f=10 starts a new cycle, with $\ell = 1$; four leaves with this signature are shown at the deepest level in the figure, three are picked up after traversing the inner loop-back edge, and the remaining four after traversing the outer loop-back edge.

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, different values of k lead to different trees $\mathcal{L}_q^k$, so it also follows from the proposition that there is an infinite sequence of distinct optimal trees as $k \to \infty$. The opposite properties hold for the previously characterized cases related to the geometric distribution (cf. [10]).

The infinite sequence of optimal codes obtained when $k \to \infty$ ($q \to 0$) stabilizes in the limit, as stated in the following proposition, which follows from Proposition 1 (the fact is also mentioned in [14, Ch. 5]).

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 equivalence, as follows: start with Q_n for n=2, recursively replace the first leaf of the deepest level of the current tree by Q_{n+1}, and increase n. □

Figure 2 shows the first fourteen levels of the limit tree of

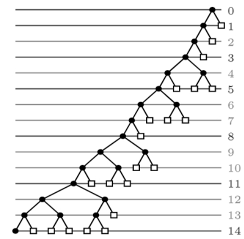

Fig. 2. Top of the limit tree for $q = 2^{-k}$ when $k \to \infty$.

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 $0 \leq j < 2^i$ and $i \geq 0$. 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 $y = \mathcal{Q}(m+1, f+2)$. A matching decoding procedure is easily derived. Encoding and decoding procedures for all the codes in this section are presented in [12][14].

Kraft functions. Let Σ be a countable alphabet, $(\mu_i)_{i \in \Sigma}$ a distribution on Σ, and C a prefix code on Σ with codeword lengths $(\ell_i)_{i \in \Sigma}$. The *Kraft function* [11] of C is defined by the the formal series

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

The average code length of C is then $c(C) = z \frac{\partial}{\partial z} P(z)\big|_{z=1}$. Explicit expressions for the Kraft functions of the optimal codes of Proposition 1 can be derived from the proposition using standard generating function tools, and are presented fully in [12]. As an example, the Kraft function of the optimal code for $q = 1/4$ is given by

$$P_{\frac{1}{4}}(z) = \frac{9}{16}\left(z + \frac{qz}{1-q^3z^6}\left(2z^2 + qz^2\left(3z + qz^2\left(z + 3z^2 + \frac{3qz^3}{1-qz^2}\right)\right)\right)\right)$$

Kraft functions were applied to obtain the average code lengths of the optimal codes of Proposition 1. These lengths were used, in turn, in the redundancy computations summarized in Section IV.

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

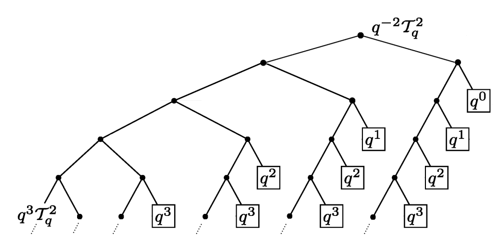

Fig. 3. The tree $q^{-2}\mathcal{T}_q^2$.

The following grammar defines the trees $\mathcal{T}_u^d$ for $d = 0, 1, 2$, and an indeterminate u.

$$\mathcal{T}_q^0 \to \boxed{1}, \quad \mathcal{T}_q^d \to \overset{\wedge}{q\mathcal{T}_q^d \;\; q\mathcal{T}_q^{d-1}} \; (d = 1, 2). \tag{3}$$

The tree $q^r \mathcal{T}^1_{q^k}$ is easily recognized as G_1 (unary tree), with leaves appropriately weighted. Moreover, the tree underlying $q^r \mathcal{T}^2_{q^k}$ is recognized as $G_1 \bullet G_1$ (see Figure 3). This tree is optimal for the (dyadic) TDGD with $q = 1/2$, since G_1 has redundancy zero for the corresponding geometric distribution, and, thus, the same is true for $G_1 \bullet G_1$ on pairs.

It follows from the foregoing discussion, and straightforward symbolic manipulations, that $w(q^r \mathcal{T}^2_{q^k}) = w(\boxed{q^r}) \left(\frac{q^k}{1-q^k}\right)^2$. It is important to note that if $q = 1/\sqrt[k]{2}$ then, $w(q^r \mathcal{T}^2_{q^k}) = w(q^r \mathcal{T}^1_{q^k}) = w(q^r \mathcal{T}^0_{q^k}) = w(\boxed{q^r})$. This observation is the basis of the construction and proof of Proposition 3 below.

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

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

□

It is shown in [12] that a prescribed sequence of pairings of symbols $q^{-i}\mathcal{T}^2_{q^k}$ 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 D_k of Proposition 3 is also presented in [12]. Notice that the concatenation $G_1 \bullet G_1$ plays, for the D_k, the role that G_1 plays for Golomb codes (with a more complex structure of k^2 leaves, defined by $\mathcal{S}_T$, at the "head" of the tree). This observation is explored further in [12], and generalizes to optimal codes on blocks of $d{>}2$ symbols.

Redundancy. Figure 4 presents plots of redundancy per integer symbol as a function of q, for (A) $0 < q < 0.5$, and (B) $0.5 \leq q < 1$, relative to the entropy rate of the geometric distribution of parameter q. Let C_k denote the optimal prefix code for a TDGD with $q = 2^{-k}$, and D_k the optimal code for $q = 1/\sqrt[k]{2}$, $k \geq 1$. Plots are shown for the Golomb code on single integer symbols, the best code C_k or D_k for each q (with $C_1 = D_1$), and the optimal code for each q. Code lengths for the latter were approximated empirically. It is observed in the figure that the families $\{C_k\}$ and $\{D_k\}$ provide good approximations to the optimal codes for all q. However, it is also observed that optimal codes for some values of q will be strictly outside of the families characterized in this paper.

V. Conclusion

Optimal prefix codes were presented for two sub-families of TDGDs, namely, those with parameters $q = 2^{-k}$ with $k{>}1$, or $q = 1/\sqrt[k]{2}$ with $k > 0$. The two families provide good approximations to, but do not contain, all optimal prefix codes for TDGDs with parameters in the full interval $0{<}q{<}1$. 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].

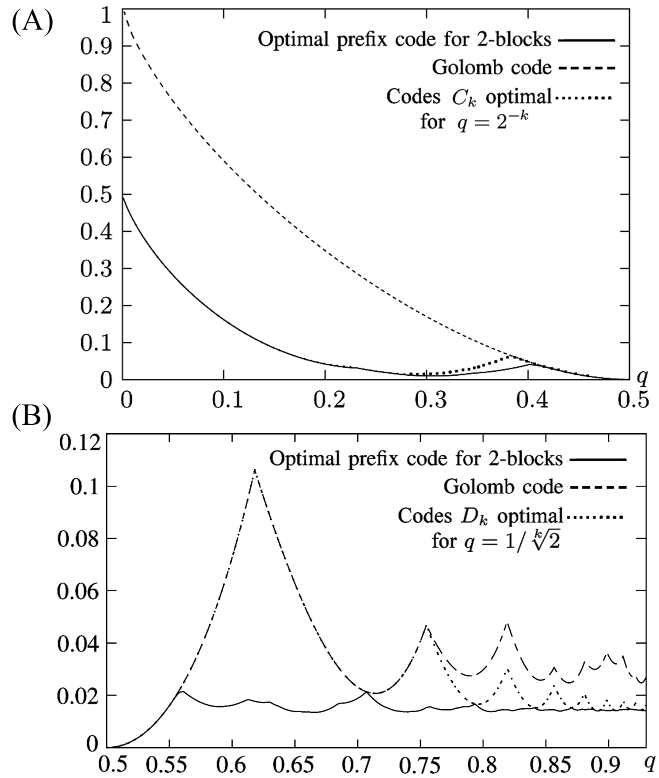

Fig. 4. Redundancy (bits/integer symbol) for the optimal code (empirical), the Golomb code, and the best code C_k or D_k for (A) $0{<}q{<}\frac{1}{2}$, (B) $\frac{1}{2}{\leq}q{<}1$.

References

[1] S. W. Golomb, "Run length encodings," *IEEE Trans. Inf. Theory*, vol. IT-12, pp. 399–401, 1966.

[2] R. G. Gallager and D. C. V. Voorhis, "Optimal source codes for geometrically distributed integer alphabets," *IEEE Trans. Inf. Theory*, vol. IT-21, 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, pp. 229–236, 2000.

[4] R. F. Rice, "Some practical universal noiseless coding techniques," Tech. Rep. JPL-79-22, JPL, Pasadena, CA, 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, 2000.

[6] G. Seroussi and M. J. Weinberger, "On adaptive strategies for an extended family of Golomb-type codes," in *Proc. DCC'97*, (Snowbird, UT), pp. 131–140, 1997.

[7] V. T. T. Linder and K. Zeger, "Existence of optimal prefix codes for infinite source alphabets," *IEEE Trans. Inf. Theory*, vol. 43, pp. 2026–2028, 1997.

[8] A. Kato, T. S. Han, and H. Nagaoka, "Huffman coding with an infinite alphabet," *IEEE Trans. Inf. Theory*, vol. 42, pp. 977–984, 1996.

[9] J. Abrahams, "Code and parse trees for lossless source encoding," *Commun. Inf. Syst.*, vol. 1, pp. 113–146, 2001.

[10] 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.

[11] 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.

[12] F. Bassino, J. Clément, G. Seroussi, and A. Viola, "Optimal prefix codes for pairs of geometrically-distributed random variables." In preparation.

[13] R. G. Gallager, "Variations on a theme by Huffman," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 668–674, 1978.

[14] M. B. Baer, *Coding for General Penalties*. PhD thesis, Stanford University, 2003.