# SGN-2306 Solution to Exercise Set 5

**Problem 1** (*The optimality of Golomb codes for the geometric distribution when Golomb condition is satisfied*)

Let's consider the geometric distribution: $P(i) = (1-\theta)\theta^i$, $\theta \in (0,1)$, $i = 0, 1, 2, \cdots$. Suppose that the Golomb condition is satisfied: it exists $\ell \in \mathcal{Z}$ with property

$$\theta^\ell = 1/2 \tag{1}$$

a) What are the values of parameter $\theta$ for which $\ell \in \{1,2,3\}$.

b) Verify that the probability of symbol $n+\ell$ is half the probability of symbol $n$ for any positive integer $n$. So, we expect that the codeword for the symbol $n+\ell$ to be 1 bit longer than the codeword for the symbol $n$. Conclude that there should be $\ell$ codewords for each possible wordlength, except for the shortest wordlengths.

c) Consider the smallest positive integer $k$ with property $2^k \geq 2\ell$. Prove that for any $D \geq k$, the maximum number of nodes at depth $D$ in a binary tree is larger than $\ell$. What about the depth $D = k-1$ ?

d) The Golomb codes for the original geometric distribution: exactly $\ell$ words of every word length $\geq k$ and $2^{k-1} - \ell$ words of length $k-1$. Prove that the Golomb codes verify with equality the Kraft inequality.

*Hint:*

$$\frac{\ell}{2^k} + \frac{\ell}{2^{k+1}} + \frac{\ell}{2^{k+2}} + \cdots = \frac{\ell}{2^{k-1}}$$

**Solution inspired by the solution of Vesa Peltonen and Gergely Korodi:**

a) It's given that: $\theta^l = \frac{1}{2}$ and $l \in \{1,2,3\}$

So $\theta^{\{1,2,3\}} = 1/2 \Rightarrow \theta = \{\frac{1}{2} = 0.5, \frac{1}{\sqrt{2}} \approx 0.7071, \frac{1}{\sqrt[3]{2}} \approx 0.7937\}$.

b)

$$\frac{P(n+\ell)}{P(n)} = \frac{(1-\theta)\theta^{n+\ell}}{(1-\theta)\theta^n} = \frac{\theta^{n+\ell}}{\theta^n} = \theta^l = \frac{1}{2}$$

which implies

$$-\log_2 P(n+\ell) = -\log_2 P(n) + 1$$

Now the conclusion is direct from the fact that $n$ can take any non-negative value.

c) At depth $D$ in a binary tree, the maximum number of nodes is: $N = 2^D$

$$D \geq k \Rightarrow 2^D \geq 2^k \geq 2\ell > \ell$$

When $D = k - 1$, the equality can occur.

d)

$$(2^{k-1} - \ell)\frac{1}{2^{k-1}} + \frac{\ell}{2^k} + \frac{\ell}{2^{k+1}} + \frac{\ell}{2^{k+2}} + \cdots = 1 - \frac{\ell}{2^{k-1}} + \frac{\ell}{2^{k-1}} = 1$$

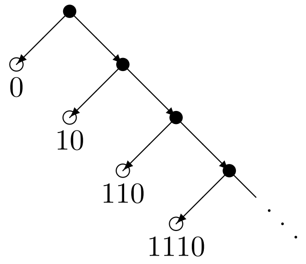
**Problem 2** (*The optimality of Golomb codes for the geometric distribution when Golomb condition is satisfied - continuation*)

a) For $\ell$ chosen as in **Problem 1**, draw the binary trees for Golomb codes when $\ell \in \{1, 2, 3\}$.

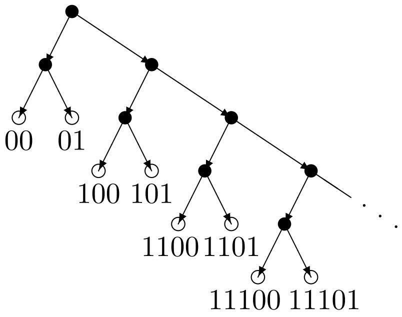
b) Use these trees to explain that the Golomb code with parameter $\ell$ for a nonnegative integer $u$ is obtained by concatenating the binary representation of $r = u \bmod \ell$ written with $\lfloor \log_2 \ell \rfloor$ bits if $r < 2^{\lceil \log_2 \ell \rceil} - \ell$, or $\lceil \log_2 \ell \rceil$ bits otherwise, and the unary representation of $c = \lfloor u/\ell \rfloor$ written with $c + 1$ bits.

**Solution by Gergely Korodi:**

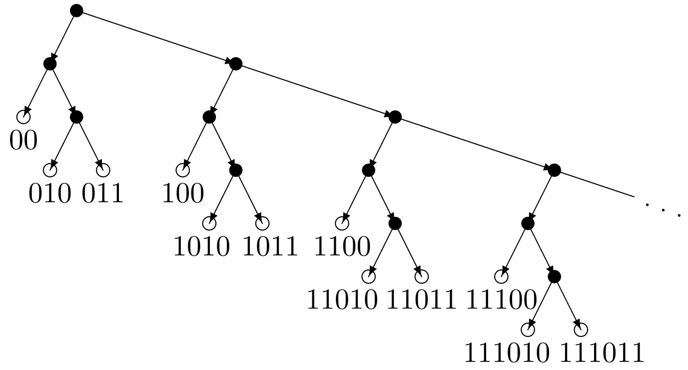
*Part a)* If $l = 1$, then $k = 1$ and the tree is:



If $l = 2$, then $k = 2$ and the tree is:



Finally, if $l = 3$, then $k = 3$ and the tree is:



*Part b)*

The rule presented in the problem is a bit informal. As is, it is not ready for a straight implementation. But using the information from Problem 1 d) it is easy to produce the complete rule. Here it is:

Let $c$ be $\lfloor \frac{u}{l} \rfloor$, let $r$ be $u \bmod l$, $q$ be $2^{\lceil \log l \rceil} - l$, $p$ be $q + \lfloor \frac{1}{2}(r - q) \rfloor$. Now, to get the Golomb code of $u$ first write $c$ number of 1-s followed by a 0. Then if $r < q$ then append to it the binary representation of $r$ in $\lfloor \log l \rfloor$ bits and you're ready. Otherwise ($r \geq q$) append $p$ in $\lfloor \log l \rfloor$ bits, then terminate the codeword with either a 0 or a 1 depending on whether $r - q$ is even or odd.

This rule conforms with the definition found in Problem 1 d).

**Problem 3** (*m-reduced source, Gallager-van Voorhis condition*)

Let's take an arbitrary $\theta \in (0, 1)$.

a) Prove there is a unique positive integer $\ell$ which satisfies the Gallager-van Voorhis condition:

$$\theta^{\ell} + \theta^{\ell+1} \leq 1 < \theta^{\ell} + \theta^{\ell-1} \tag{2}$$

b) For $\ell$ determined at **point a)** and for any integer $m \geq 0$ define an *m-reduced source* as a source with $m + 1 + \ell$ letters with the following probabilities:

$$P_m(i) \quad = \quad \begin{cases} (1-\theta)\theta^i, & 0 \leq i \leq m \\ \frac{(1-\theta)\theta^i}{1-\theta^{\ell}}, & m < i \leq m + \ell \end{cases}$$

Verify that the probabilities for the *m-reduced source* sum to 1 and each of the final $\ell$ probabilities can be regarded as the accumulation of the probabilities of all the integers greater than $m$ in the original geometric source which are in the same equivalence class modulo $\ell$.

*Hint:*

$$\frac{(1-\theta)\theta^i}{1-\theta^{\ell}} = \sum_{j=0}^{\infty} (1-\theta)\theta^{i+j\ell}$$

c) Use (2) to prove that the symbols $m$ and $m + \ell$ have the smallest probabilities. Verify that merging these two least likely letters we obtain an $m - 1$-*reduced source.*

*Hint:* The probabilities of the first $m+1$ letters are monotonically decreasing, as are the probabilities of the last $\ell$ letters. Use (2) to show that the $(m+\ell)$-th letter is no more probable than the $(m-1)$-th and the $(m+\ell-1)$-th letter is more probable than the $m$-th letter.

**Solution by Gergely Korodi:**

*Part a)* Let's solve the inequality to find out if it has indeed one and only one solution:

$$\theta^l(1+\theta) \le 1 < \theta^{l-1}(1+\theta)$$

$$\theta^l \le \frac{1}{(1+\theta)} < \theta^{l-1}$$

$$l \ge \frac{1}{\log\theta} \cdot \log\frac{1}{1+\theta} > l-1$$

$$l \ge -\frac{\log 1+\theta}{\log\theta} > l-1$$

This indeed has one unique solution, namely $l = \left\lceil -\frac{\log 1+\theta}{\log\theta} \right\rceil$.

*Part b)*

$$\sum_{i=0}^{m}(1-\theta)\theta^i + \sum_{i=m+1}^{m+l}\frac{(1-\theta)\theta^i}{1-\theta^l} =$$

$$(1-\theta)\frac{\theta^{m+1}-1}{\theta-1} + \frac{(1-\theta)\theta^{m+1}}{1-\theta^l}\cdot\frac{\theta^l-1}{\theta-1} = 1-\theta^{m+1}+\theta^{m+1} = 1.$$

$$(1-\theta)\theta^{m+k} + (1-\theta)\theta^{m+l+k} + (1-\theta)\theta^{m+2l+k} + \dots = \frac{(1-\theta)\theta^{m+k}}{1-\theta^l}.$$

*Part c)*

$(1-\theta), (1-\theta)\theta, (1-\theta)\theta^2, \dots$ is monotonically decreasing. So is $\frac{(1-\theta)\theta^{m+1}}{1-\theta^l}, \frac{(1-\theta)\theta^{m+2}}{1-\theta^l}, \dots$. Now, since $\theta^{m-1}(\theta^l+\theta^{l+1}) \le \theta^{m-1}$, so $\theta^{m+l} \le \theta^{m-1} - \theta^{m+l-1}$ and thus

$$\frac{(1-\theta)\theta^{m+l}}{1-\theta^l} \le (1-\theta)\theta^{m-1}.$$

Likewise, since $\theta^m(\theta^l+\theta^{l-1}) > \theta^m$, so $\theta^{m+l-1} > \theta^m - \theta^{m+l}$ and thus

$$\frac{(1-\theta)\theta^{m+l-1}}{1-\theta^l} > (1-\theta)\theta^m.$$

$$(1-\theta)\theta^m + \frac{(1-\theta)\theta^{m+l}}{1-\theta^l} = \frac{(1-\theta)(\theta^m-\theta^{m+l}+\theta^{m+l})}{1-\theta^l} = \frac{(1-\theta)\theta^m}{1-\theta^l}$$

which means that

$$P_{m-1}(i) = \begin{cases} (1-\theta)\theta^i & \text{if } 0 \le i \le m-1 \\ \frac{(1-\theta)\theta^i}{1-\theta^l} & \text{if } m-1 < i \le m+l-1 \end{cases}$$

**Problem 4** (The optimality of Golomb codes, Gallager-van Voorhis condition)

a) By iterating the merging process from **Problem 3, point c)**, we are obtaining at the final step, the $-1$-*reduced source*:

$$P_{-1}(i) \;=\; \frac{(1-\theta)\theta^i}{1-\theta^\ell}, \quad 0 \le i \le \ell - 1 \tag{3}$$

Verify that for the $-1$-*reduced source* the sum of the two least likely symbols exceeds the probability of the most likely and conclude the lengths of the optimal codewords for (3) differ by at most 1.

b) Show that if we choose for the symbols $\{0, 1, \ldots, 2^{\lceil \log_2 \ell \rceil} - \ell - 1\}$ codewords with length $\lfloor \log_2 \ell \rfloor$ and for the symbols $\{2^{\lceil \log_2 \ell \rceil} - \ell, \ldots, \ell - 1\}$ codewords with length $\lceil \log_2 \ell \rceil$, then the Kraft inequality is satisfied with equality.

c) Taking into account that $m$ is arbitrary chosen, consider $m \to \infty$ and conclude the Golomb codes (**Problem 2 b)**) are optimal for any geometric source which obeys (2) and not only in the most restrictive condition (1).

**Solution by Gergely Korodi:**

*Part a)*

$$\frac{(1-\theta)\theta^{l-1}}{1-\theta^l} + \frac{(1-\theta)\theta^{l-2}}{1-\theta^l} = \frac{(1-\theta)(\theta^{l-1}+\theta^{l-2})}{1-\theta^l} > \frac{(1-\theta)(\theta^{l-1}+\theta^{l})}{1-\theta^l} > \frac{1-\theta}{1-\theta^l}.$$

Suppose that there are two symbols "u" and "v", whose code lengths differ at least by 2: $l(u) \le l(v) - 2$. If "w" denotes the sibling of "v", then the node from which they directly descend has their cumulative probability $P(v) + P(w) > P(u)$, but this node is at least one level below the level of symbol "u", which is contradiction.

*Part b)*

$$\sum_{i=0}^{2^{\lceil \log l \rceil}-l-1} 2^{-\lfloor \log l \rfloor} + \sum_{i=2^{\lceil \log l \rceil}-l}^{l-1} 2^{-\lceil \log l \rceil} =$$

$$(2^{\lceil \log l \rceil} - l) \cdot 2^{-\lfloor \log l \rfloor} + (l - 1 - 2^{\lceil \log l \rceil} + l + 1) \cdot 2^{-\lceil \log l \rceil} =$$

$$\frac{2^{\lceil \log l \rceil} - l}{2^{\lfloor \log l \rfloor}} + \frac{2l}{2^{\lceil \log l \rceil}} - 1 = \begin{cases} l = 2^k & \Rightarrow \frac{2^k - 2^k}{2^k} + \frac{2 \cdot 2^k}{2^k} - 1 & = 1 \\ 2^k < l < 2^{k+1} & \Rightarrow \frac{2^{k+1} - l}{2^k} + \frac{2l}{2^{k+1}} - 1 & = 1 \end{cases}$$

*Part c)*

**Solution**: (*R.G. Gallager, D.C. Van Voorhis, Optimal source codes for geometrically distributed integer alphabets, IEEE Tr. on Information Theory, March 1975*)

Let $\overline{n}$ be the infimum of the expected codeword length over all uniquely decipherable codes for the considered OSG distribution. Let $\overline{n}_G$ be the expected codeword length for the Golomb code, and let $\overline{n}_m$ be the expected codeword length for the optimal code for an *m-reduced source*. Clearly $\overline{n} \leq \overline{n}_G$; also $\overline{n}_m \leq \overline{n}$, for all $m$, since given uniquely decipherable code for the original source, we can obtain a code for the *m-reduced source* by using the same codewords for $i \leq m$, and the shortest remaining codewords for $m+1 \leq i \leq m+\ell$. Finally $\overline{n}_m$ is increasing with $m$, and has a limit as $m \to \infty$, and thus $\lim_{m\to\infty} \overline{n}_m \leq \overline{n}$. This limit is just $\overline{n}_G$, and thus the Golomb code is optimum.

**Problem 5** (A practical approach on using Golomb codes)

In practical encoders the Golomb codes are used in the particular hypothesis that $\ell$ is a power of 2. Generate the codewords and the trees for $\ell = 7$ and $\ell = 8$ and explain the implementation advantages when $\ell = 8$. In the general case, compute the code length when the unsigned integer $u$ is encoded with Golomb codes for which $\ell = 2^k, k \in \mathcal{Z}$.

**Solution by Gergely Korodi:**

Table 1 contains the code words for $l = 7$ and $l = 8$.

The advantage of choosing $l$ to be the power of 2 is that (using the notations of Solution 2 b)) $q$ is always 0 in this case, so the rule for obtaining the code word for the integer $u$ simplifies to this ($l = 2^k$):

- Output $c = \left\lfloor \frac{u}{l} \right\rfloor$ "1"-s followed by a single "0"
- Output the binary representation of $r = u \bmod l$ in $k$ bits.

This is certainly faster and easier than the general rule. The length of the codeword of $u$ is $c + 1 + k = \left\lfloor \frac{u}{l} \right\rfloor + 1 + k$ bits.

**Problem 6** (*Why limited-length Golomb codes*)

The length of the Golomb-Rice code with parameter $k$ for the non-negative integer $y$ is given by:

$$\left\lfloor \frac{y}{2^k} \right\rfloor + k + 1$$

Compute the code length when $y = 255$ and $k = 1$. Write a short comment on the result taking into account the binary representation for $y$ is 8 bits-long.

**Solution by Gergely Korodi:**

*Part a)* If $y = 255$ and $k = 1$, then $y$ will be encoded in 129 bits. This shows the effect of the blind nature of Golomb codes: if the predictor we use is very good (and hopefully it is) then the errors will be very small, resulting in a small $k$. $k = 1$ corresponds to the case when every symbol's probability is half of that of its predecessor, which is by no means uncommon if a good predictor is being used. However, we have to take into account that

| $u$ | $l = 7$ | $l = 8$ |
|---|---|---|
| 0 | 000 | 0000 |
| 1 | 0010 | 0001 |
| 2 | 0011 | 0010 |
| 3 | 0100 | 0011 |
| 4 | 0101 | 0100 |
| 5 | 0110 | 0101 |
| 6 | 0111 | 0110 |
| 7 | 1000 | 0111 |
| 8 | 10010 | 10000 |
| 9 | 10011 | 10001 |
| 10 | 10100 | 10010 |
| 11 | 10101 | 10011 |
| 12 | 10110 | 10100 |
| 13 | 10111 | 10101 |
| 14 | 11000 | 10110 |
| 15 | 110010 | 10111 |
| 16 | 110011 | 110000 |
| 17 | 110100 | 110001 |
| 18 | 110101 | 110010 |
| 19 | 110110 | 110011 |
| 20 | 110111 | 110100 |
| 21 | 111000 | 110101 |
| 22 | 1110010 | 110110 |
| 23 | 1110011 | 110111 |
| 24 | 1110100 | 1110000 |
| 25 | 1110101 | 1110001 |

Table 1: Golomb codes for $l = 7$ and $l = 8$

there are always exceptions, noise[1] whose context baffle the predictor so that it gives a totally false value. As our example shows, such values will be encoded using very many bits, resulting in a very bad code.

Despite that Golomb codes with a suitable parameter are optimal for a geometric distribution, and despite the prediction errors can be modelled by a geometric distribution *quite well*, the resulting codewords have serious shortcomings. This is because of the "*quite well*": high values are associated with *extremely* low probabilities, however, if they occur only once, that would give them a significantly larger probability than they theoretically deserve. In our example, 255 is associated with the probability $2^{-256}$, which is so small that it theoretically should never occur in any practical amount of data. However, there are always unpredictable pixels of quite a considerable amount, so the presence of some large prediction errors are likely anyway. Because the geometric distribution gives these a very low probability, the Golomb procedure gives a very long codeword.

One possible workaround of this phenomenon is that we realize that above a certain value the prediction errors are no longer geometric, rather linear. For example, a value of 100 is such a large error that it can be considered a noise just as well as 200, and since we should not make any difference between their predicted probabilities. So let's do this: designate a limit $0 \ll v \ll 256$ at and above which we make this distinction. The associated probability to $v$ is $\sum_{i=v}^{\infty}(1-\theta)\theta^i$. In our example, where $k = 1$, $\theta = 1/2$ so this probability is $\frac{1}{2^v}$. Let's associate a single codeword to this probability, which we can regard as an escape code in the followings. Every time we encounter a prediction error that is at least $v$, we send this escape code, followed by the error's character code in clear—that is, in 8 bits. In the example, if we set $v = 10$, than the longest code for an error will be only 18 (escape + character), while the good predictions are still compressed by the original Golomb codes very well.

---

[1] Here under the term "noise" I mean pixels whose value cannot be predicted well from their contexts. This of course is not the same as the usual meaning of noise in Signal Processing.