V
Identification Entropy

R. Ahlswede

Abstract. Shannon (1948) has shown that a source $(\mathcal{U}, P, U)$ with output U satisfying Prob $(U = u) = P_u$, can be encoded in a prefix code $\mathcal{C} = \{c_u : u \in \mathcal{U}\} \subset \{0,1\}^*$ such that for the entropy

$$H(P) = \sum_{u \in \mathcal{U}} -p_u \log p_u \leq \sum p_u ||c_u|| \leq H(P) + 1,$$

where $||c_u||$ is the length of c_u.

We use a prefix code $\mathcal{C}$ for another purpose, namely noiseless identification, that is every user who wants to know whether a u $(u \in \mathcal{U})$ of his interest is the actual source output or not can consider the RV C with $C = c_u = (c_{u_1}, \dots, c_{u||c_u||})$ and check whether $C = (C_1, C_2, \dots)$ coincides with c_u in the first, second etc. letter and stop when the first different letter occurs or when $C = c_u$. Let $L_\mathcal{C}(P, u)$ be the expected number of checkings, if code $\mathcal{C}$ is used.

Our discovery is an identification entropy, namely the function

$$H_I(P) = 2\left(1 - \sum_{u \in \mathcal{U}} P_u^2\right).$$

We prove that $L_\mathcal{C}(P, P) = \sum\limits_{u \in \mathcal{U}} P_u \; L_\mathcal{C}(P, u) \geq H_I(P)$ and thus also that

$$L(P) = \min_{\mathcal{C}} \max_{u \in \mathcal{U}} L_\mathcal{C}(P, u) \geq H_I(P)$$

and related upper bounds, which demonstrate the operational significance of identification entropy in noiseless source coding similar as Shannon entropy does in noiseless data compression.

Also other averages such as $\bar{L}_\mathcal{C}(P) = \frac{1}{|\mathcal{U}|} \sum\limits_{u \in \mathcal{U}} L_\mathcal{C}(P, u)$ are discussed in particular for Huffman codes where classically equivalent Huffman codes may now be different.

We also show that prefix codes, where the codewords correspond to the leaves in a regular binary tree, are universally good for this average.

1 Introduction

Shannon's Channel Coding Theorem for Transmission [1] is paralleled by a Channel Coding Theorem for Identification [3]. In [4] we introduced noiseless source coding for identification and suggested the study of several performance measures.

R. Ahlswede et al. (Eds.): Information Transfer and Combinatorics, LNCS 4123, pp. 595–613, 2006.

Interesting observations were made already for uniform sources $P^N = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)$, for which the worst case expected number of checkings $L(P^N)$ is approximately 2. Actually in [5] it is shown that $\lim_{N\to\infty} L(P^N) = 2$.

Recall that in channel coding going from transmission to identification leads from an exponentially growing number of manageable messages to double exponentially many. Now in source coding roughly speaking the range of average code lengths for data compression is the interval $[0,\infty)$ and it is $[0,2)$ for an average expected length of optimal identification procedures. Note that no randomization has to be used here.

A discovery of the present paper is an identification entropy, namely the functional

$$H_I(P) = 2\left(1 - \sum_{u=1}^{N} P_u^2\right) \tag{1.1}$$

for the source $(\mathcal{U}, P)$, where $\mathcal{U} = \{1, 2, \dots, N\}$ and $P = (P_1, \dots, P_N)$ is a probability distribution.

Its operational significance in identification source coding is similar to that of classical entropy $H(P)$ in noiseless coding of data: it serves as a good lower bound.

Beyond being continuous in P it has three basic properties.

I. Concavity

For $p = (p_1, \dots, p_N)$, $q = (q_1, \dots, q_N)$ and $0 \le \alpha \le 1$

$$H_I(\alpha p + (1-\alpha)q) \ge \alpha H_I(p) + (1-\alpha) H_I(q).$$

This is equivalent with

$$\sum_{i=1}^{N} (\alpha p_i + (1-\alpha) q_i)^2 = \sum_{i=1}^{N} \alpha^2 p_i^2 + (1-\alpha)^2 q_i^2 + \sum_{i\ne j} \alpha(1-\alpha) p_i q_j \le \sum_{i=1}^{N} \alpha p_i^2 + (1-\alpha) q_i^2$$

or with

$$\alpha(1-\alpha) \sum_{i=1}^{N} p_i^2 + q_i^2 \ge \alpha(1-\alpha) \sum_{i\ne j} p_i q_j,$$

which holds, because $\sum_{i=1}^{N} (p_i - q_i)^2 \ge 0$.

II. Symmetry

For a permutation $\Pi : \{1, 2, \dots, N\} \to \{1, 2, \dots, N\}$ and $\Pi P = (P_{1\Pi}, \dots, P_{N\Pi})$

$$H_I(P) = H_I(\Pi P).$$

III. Grouping identity

For a partition $(\mathcal{U}_1, \mathcal{U}_2)$ of $\mathcal{U} = \{1, 2, \dots, N\}$, $Q_i = \sum_{u\in\mathcal{U}_i} P_u$ and $P_u^{(i)} = \frac{P_u}{Q_i}$ for $u \in \mathcal{U}_i (i = 1, 2)$

$$H_I(P) = Q_1^2 H_I(P^{(1)}) + Q_2^2 H_I(P^{(2)}) + H_I(Q), \text{ where } Q = (Q_1, Q_2).$$

Indeed,

$$Q_1^2 2\left(1-\sum_{j\in\mathcal{U}_1}\frac{P_j^2}{Q_1^2}\right)+Q_2^2 2\left(1-\sum_{j\in\mathcal{U}_2}\frac{P_j^2}{Q_2^2}\right)+2(1-Q_1^2-Q_2^2)$$
$$=2Q_1^2-2\sum_{j\in\mathcal{U}_1}P_j^2+2Q_2^2-2\sum_{j\in\mathcal{U}_2}P_j^2+2-2Q_1^2-2Q_2^2$$
$$=2\left(1-\sum_{j=1}^{N}P_j^2\right).$$

Obviously, $0 \leq H_I(P)$ with equality exactly if $P_i = 1$ for some i and by concavity $H_I(P) \leq 2\left(1-\frac{1}{N}\right)$ with equality for the uniform distribution.

Remark. Another important property of $H_I(P)$ is Schur concavity.

2 Noiseless Identification for Sources and Basic Concept of Performance

For the source $(\mathcal{U}, P)$ let $\mathcal{C} = \{c_1, \dots, c_N\}$ be a binary prefix code (PC) with $||c_u||$ as length of c_u. Introduce the RV U with $\text{Prob}(U = u) = P_u$ for $u \in \mathcal{U}$ and the RV C with $C = c_u = (c_{u1}, c_{u2}, \dots, c_{u||c_u||})$ if $U = u$. We use the PC for noiseless identification, that is a user interested in u wants to know whether the source output equals u, that is, whether C equals c_u or not. He iteratively checks whether $C = (C_1, C_2, \dots)$ coincides with c_u in the first, second etc. letter and stops when the first different letter occurs or when $C = c_u$. What is the expected number $L_{\mathcal{C}}(P, u)$ of checkings?

Related quantities are

$$L_{\mathcal{C}}(P) = \max_{1\leq u\leq N} L_{\mathcal{C}}(P, u), \tag{2.1}$$

that is, the expected number of checkings for a person in the worst case, if code $\mathcal{C}$ is used,

$$L(P) = \min_{\mathcal{C}} L_{\mathcal{C}}(P), \tag{2.2}$$

the expected number of checkings in the worst case for a best code, and finally, if users are chosen by a RV V independent of U and defined by $\text{Prob}(V = v) = Q_v$ for $v \in \mathcal{V} = \mathcal{U}$, (see [5], Section 5) we consider

$$L_{\mathcal{C}}(P, Q) = \sum_{v\in\mathcal{U}} Q_v L_{\mathcal{C}}(P, v) \tag{2.3}$$

the average number of expected checkings, if code $\mathcal{C}$ is used, and also

$$L(P, Q) = \min_{\mathcal{C}} L_{\mathcal{C}}(P, Q) \tag{2.4}$$

the average number of expected checkings for a best code.

A natural special case is the mean number of expected checkings

$$\bar{L}_{\mathcal{C}}(P) = \sum_{u=1}^{N} \frac{1}{N} L_{\mathcal{C}}(P, u), \tag{2.5}$$

which equals $L_{\mathcal{C}}(P, Q)$ for $Q = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)$, and

$$\bar{L}(P) = \min_{\mathcal{C}} \bar{L}_{\mathcal{C}}(P). \tag{2.6}$$

Another special case of some "intuitive appeal" is the case $Q = P$. Here we write

$$L(P, P) = \min_{\mathcal{C}} L_{\mathcal{C}}(P, P). \tag{2.7}$$

It is known that Huffman codes minimize the expected code length for PC.

This is not the case for $L(P)$ and the other quantities in identification (see Example 3 below). It was noticed already in [4], [5] that a construction of code trees balancing probabilities like in the Shannon-Fano code is often better. **In fact Theorem 3 of [5] establishes that $L(P) < 3$ for every $P = (P_1, \dots, P_N)$!**

Still it is also interesting to see how well Huffman codes do with respect to identification, because of their classical optimality property. This can be put into the following

Problem: Determine the region of simultaneously achievable pairs $(L_{\mathcal{C}}(P), \sum_u P_u ||c_u||)$ for (classical) transmission and identification coding, where the $\mathcal{C}$'s are PC. In particular, what are extremal pairs? We begin here with first observations.

3 Examples for Huffman Codes

We start with the uniform distribution

$$P^N = (P_1, \dots, P_N) = \left(\frac{1}{N}, \dots, \frac{1}{N}\right), 2^n \le N < 2^{n+1}.$$

Then $2^{n+1} - N$ codewords have the length n and the other $2N - 2^{n+1}$ codewords have the length $n + 1$ in any Huffman code. We call the $N - 2^n$ nodes of length n of the code tree, which are extended up to the length $n + 1$ **extended nodes**.

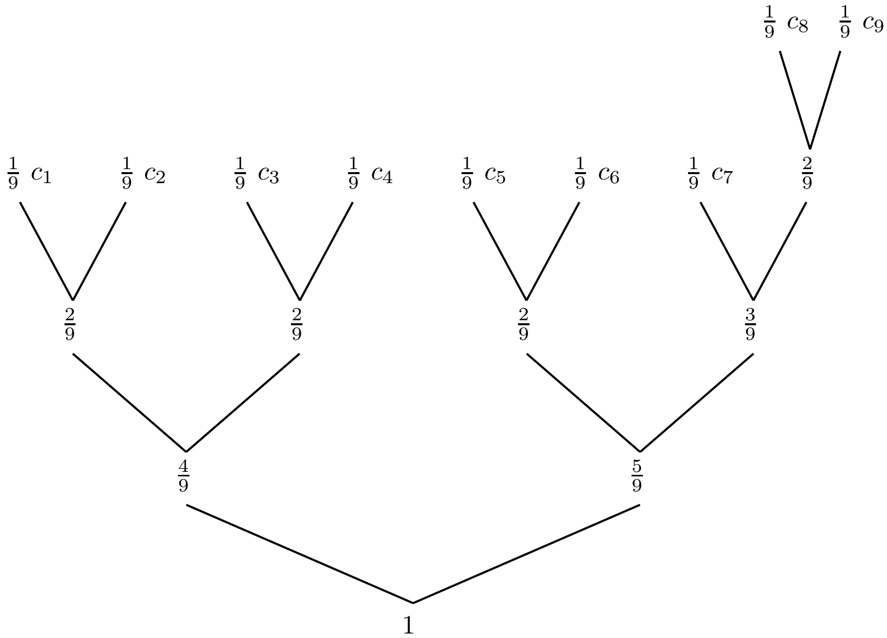
All Huffman codes for this uniform distribution differ only by the positions of the $N - 2^n$ extended nodes in the set of 2^n nodes of length n.

The average codeword length (for data compression) does not depend on the choice of the extended nodes.

However, the choice influences the performance criteria for identification!

Clearly there are $\binom{2^n}{N-2^n}$ Huffman codes for our source.

Example 1. $N = 9$, $\mathcal{U} = \{1, 2, \ldots, 9\}$, $P_1 = \cdots = P_9 = \frac{1}{9}$.



Here $L_{\mathcal{C}}(P) \approx 2.111$, $L_{\mathcal{C}}(P,P) \approx 1.815$ because

$$L_{\mathcal{C}}(P) = L_{\mathcal{C}}(c_8) = \frac{4}{9} \cdot 1 + \frac{2}{9} \cdot 2 + \frac{1}{9} \cdot 3 + \frac{2}{9} \cdot 4 = 2\frac{1}{9}$$

$$L_{\mathcal{C}}(c_9) = L_{\mathcal{C}}(c_8), L_{\mathcal{C}}(c_7) = 1\frac{8}{9}, L_{\mathcal{C}}(c_5) = L_{\mathcal{C}}(c_6) = 1\frac{7}{9},$$

$$L_{\mathcal{C}}(c_1) = L_{\mathcal{C}}(c_2) = L_{\mathcal{C}}(c_3) = L_{\mathcal{C}}(c_4) = 1\frac{6}{9}$$

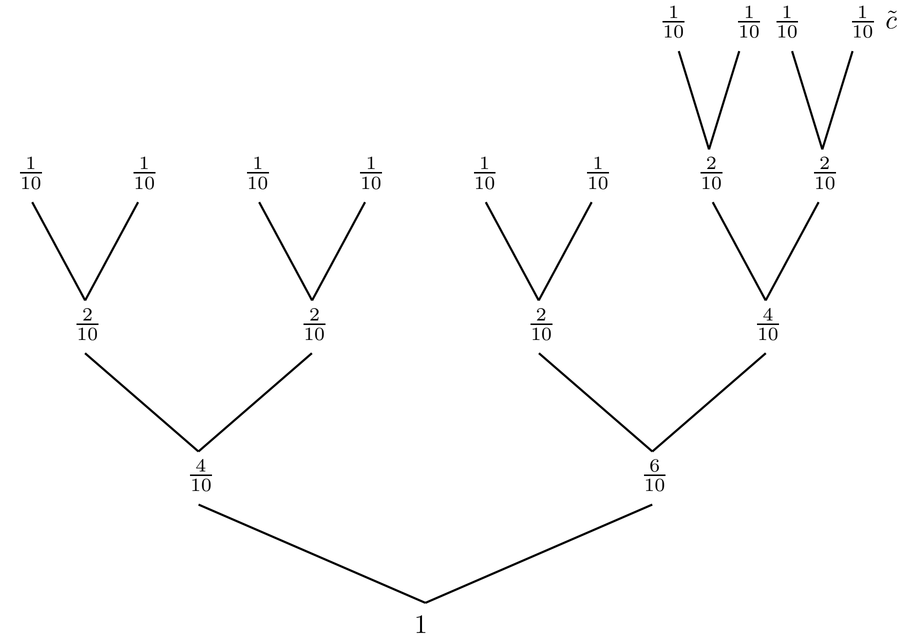
and therefore

$$L_{\mathcal{C}}(P,P) = \frac{1}{9}\left[1\frac{6}{9} \cdot 4 + 1\frac{7}{9} \cdot 2 + 1\frac{8}{9} \cdot 1 + 2\frac{1}{9} \cdot 2\right] = 1\frac{22}{27} = \bar{L}_{\mathcal{C}},$$

because P is uniform and the $\binom{2^3}{9-2^3} = 8$ Huffman codes are equivalent for identification.

Remark. Notice that Shannon's data compression gives
$H(P) + 1 = \log 9 + 1 > \sum_{u=1}^{9} P_u ||c_u|| = \frac{1}{9} 3 \cdot 7 + \frac{1}{9} 4 \cdot 2 = 3\frac{2}{9} \geq H(P) = \log 9.$

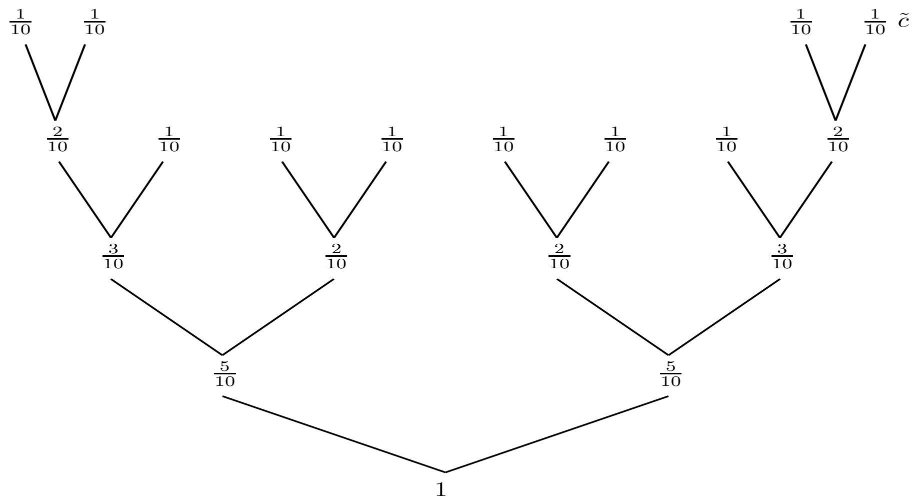
Example 2. $N = 10$. There are $\binom{2^3}{10-2^3} = 28$ Huffman codes.
The 4 worst Huffman codes are maximally unbalanced.



Here $L_{\mathcal{C}}(P) = 2.2$ and $L_{\mathcal{C}}(P,P) = 1.880$, because

$$L_{\mathcal{C}}(P) = 1 + 0.6 + 0.4 + 0.2 = 2.2$$
$$L_{\mathcal{C}}(P,P) = \frac{1}{10}[1.6 \cdot 4 + 1.8 \cdot 2 + 2.2 \cdot 4] = 1.880.$$

One of the 16 best Huffman codes



Here $L_{\mathcal{C}}(P) = 2.0$ and $L_{\mathcal{C}}(P,P) = 1.840$ because

$$L_{\mathcal{C}}(P) = L_{\mathcal{C}}(\tilde{c}) = 1 + 0.5 + 0.3 + 0.2 = 2.000$$
$$L_{\mathcal{C}}(P,P) = \frac{1}{5}(1.7 \cdot 2 + 1.8 \cdot 1 + 2.0 \cdot 2) = 1.840$$

Table 1. The best identification performances of Huffman codes for the uniform distribution

N	8	9	10	11	12	13	14	15
$L_{\mathcal{C}}(P)$	1.750	2.111	2.000	2.000	1.917	2.000	1.929	1.933
$L_{\mathcal{C}}(P,P)$	1.750	1.815	1.840	1.860	1.861	1.876	1.878	1.880

Actually $\lim_{N\to\infty} L_{\mathcal{C}}(P^N) = 2$, but bad values occur for $N = 2^k + 1$ like $N = 9$ (see [5]).

One should prove that a best Huffman code for identification for the uniform distribution is best for the worst case and also for the mean.

However, for non-uniform sources generally Huffman codes are not best.

Example 3. Let $N = 4$, $P(1) = 0.49$, $P(2) = 0.25$, $P(3) = 0.25$, $P(4) = 0.01$. Then for the Huffman code $||c_1|| = 1$, $||c_2|| = 2$, $||c_3|| = ||c_4|| = 3$ and thus $L_{\mathcal{C}}(P) = 1 + 0.51 + 0.26 = 1.77$, $L_{\mathcal{C}}(P,P) = 0.49 \cdot 1 + 0.25 \cdot 1.51 + 0.26 \cdot 1.77 = 1.3277$, and $\bar{L}_{\mathcal{C}}(P) = \frac{1}{4}(1 + 1.51 + 2 \cdot 1.77) = 1.5125$.

However, if we use $\mathcal{C}' = \{00, 10, 11, 01\}$ for $\{1, \ldots, 4\}$ (4 is on the branch together with 1), then $L_{\mathcal{C}'}(P,u) = 1.5$ for $u = 1, 2, \ldots, 4$ and all three criteria give the same value 1.500 better than $L_{\mathcal{C}}(P) = 1.77$ and $\bar{L}_{\mathcal{C}}(P) = 1.5125$.

But notice that $L_{\mathcal{C}}(P,P) < L_{\mathcal{C}'}(P,P)$!

4 An Identification Code Universally Good for All P on $\mathcal{U} = \{1, 2, \ldots, N\}$

Theorem 1. *Let $P = (P_1, \ldots, P_N)$ and let $k = \min\{\ell : 2^\ell \geq N\}$, then the regular binary tree of depth k defines a PC $\{c_1, \ldots, c_{2^k}\}$, where the codewords correspond to the leaves. To this code $\mathcal{C}_k$ corresponds the subcode $\mathcal{C}_N = \{c_i : c_i \in \mathcal{C}_k, 1 \leq i \leq N\}$ with*

$$2\left(1 - \frac{1}{N}\right) \leq 2\left(1 - \frac{1}{2^k}\right) \leq \bar{L}_{\mathcal{C}_N}(P) \leq 2\left(2 - \frac{1}{N}\right) \tag{4.1}$$

and equality holds for $N = 2^k$ on the left sides.

Proof. By definition,

$$\bar{L}_{\mathcal{C}_N}(P) = \frac{1}{N}\sum_{u=1}^{N} L_{\mathcal{C}_N}(P,u) \tag{4.2}$$

and abbreviating $L_{\mathcal{C}_N}(P,u)$ as $L(u)$ for $u = 1,\dots,N$ and setting $L(u) = 0$ for $u = N+1,\dots,2^k$ we calculate with $P_u \triangleq 0$ for $u = N+1,\dots,2^k$

$$
\begin{aligned}
\sum_{u=1}^{2^k} L(u) =& \big[(P_1+\dots+P_{2^k})2^k\big] \\
&+ \big[(P_1+\dots+P_{2^{k-1}})2^{k-1} + (P_{2^{k-1}+1}+\dots+P_{2^k})2^{k-1}\big] \\
&+ \big[(P_1+\dots+P_{2^{k-2}})2^{k-2} + (P_{2^{k-2}+1}+\dots+P_{2^{k-1}})2^{k-2} \\
&+ (P_{2^{k-1}+1}+\dots+P_{2^{k-1}+2^{k-2}})2^{k-2} \\
&+ (P_{2^{k-1}+2^{k-2}+1}+\dots+P_{2^k})2^{k-2}\big] \\
&+ \qquad \dots \\
&. \\
&. \\
&. \\
&+ \big[(P_1+P_2)2 + (P_3+P_4)2 + \dots + (P_{2^k-1}+P_{2^k})2\big] \\
=& 2^k + 2^{k-1} + \dots + 2 = 2(2^k-1)
\end{aligned}
$$

and therefore

$$\sum_{u=1}^{2^k} \frac{1}{2^k} L(u) = 2\left(1-\frac{1}{2^k}\right). \tag{4.3}$$

Now

$$2\left(1-\frac{1}{N}\right) \le 2\left(1-\frac{1}{2^k}\right) = \sum_{u=1}^{2^k}\frac{1}{2^k}L(u) \le \sum_{u=1}^{N}\frac{1}{N}L(u) =$$

$$\frac{2^k}{N}\sum_{u=1}^{2^k}\frac{1}{2^k}L(u) = \frac{2^k}{N}2\left(1-\frac{1}{2^k}\right) \le 2\left(2-\frac{1}{N}\right),$$

which gives the result by (4.2). Notice that for $N = 2^k$, a power of 2, by (4.3)

$$\bar{L}_{\mathcal{C}_N}(P) = 2\left(1-\frac{1}{N}\right). \tag{4.4}$$

Remark. The upper bound in (4.1) is rough and can be improved significantly.

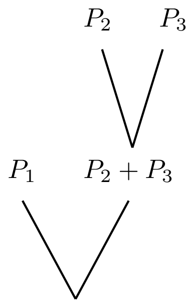
5 Identification Entropy $H_I(P)$ and Its Role as Lower Bound

Recall from the Introduction that

$$H_I(P) = 2\left(1-\sum_{u=1}^{N} P_u^2\right) \text{ for } P = (P_1 \dots P_N). \tag{5.1}$$

We begin with a small source

Example 4. Let $N = 3$. W.l.o.g. an optimal code $\mathcal{C}$ has the structure



Claim.

$$\bar{L}_{\mathcal{C}}(P) = \sum_{u=1}^{3} \frac{1}{3} L_{\mathcal{C}}(P,u) \geq 2\left(1 - \sum_{u=1}^{3} P_u^2\right) = H_I(P).$$

Proof. Set $L(u) = L_{\mathcal{C}}(P,u)$. $\sum\limits_{u=1}^{3} L(u) = 3(P_1 + P_2 + P_3) + 2(P_2 + P_3)$. This is smallest, if $P_1 \geq P_2 \geq P_3$ and thus $L(1) \leq L(2) = L(3)$. Therefore $\sum\limits_{u=1}^{3} P_u L(u) \leq \frac{1}{3} \sum\limits_{u=1}^{3} L(u)$. Clearly $L(1) = 1$, $L(2) = L(3) = 1 + P_2 + P_3$ and $\sum\limits_{u=1}^{3} P_u L(u) = P_1 + P_2 + P_3 + (P_2 + P_3)^2$.
This does not change if $P_2 + P_3$ is constant. So we can assume $P = P_2 = P_3$ and $1 - 2P = P_1$ and obtain

$$\sum_{u=1}^{3} P_u L(u) = 1 + 4P^2.$$

On the other hand

$$2\left(1 - \sum_{u=1}^{3} P_u^2\right) \leq 2\left(1 - P_1^2 - 2\left(\frac{P_2 + P_3}{2}\right)^2\right), \tag{5.2}$$

because $P_2^2 + P_3^2 \geq \frac{(P_2+P_3)^2}{2}$.
Therefore it suffices to show that

$$\begin{aligned} 1 + 4P^2 &\geq 2\big(1 - (1 - 2P)^2 - 2P^2\big) \\ &= 2(4P - 4P^2 - 2P^2) \\ &= 2(4P - 6P^2) = 8P - 12P^2. \end{aligned}$$

Or that $1 + 16P^2 - 8P = (1 - 4P)^2 \geq 0$.

We are now prepared for the first main result for $L(P,P)$.

Central in our derivations are proofs by induction based on **decomposition formulas for trees**.

Starting from the root a binary tree $\mathcal{T}$ goes via 0 to the subtree $\mathcal{T}_0$ and via 1 to the subtree $\mathcal{T}_1$ with sets of leaves $\mathcal{U}_0$ and $\mathcal{U}_1$, respectively. A code $\mathcal{C}$ for $(\mathcal{U}, P)$ can be viewed as a tree $\mathcal{T}$, where $\mathcal{U}_i$ corresponds to the set of codewords $\mathcal{C}_i$, $\mathcal{U}_0 \cup \mathcal{U}_1 = \mathcal{U}$.

The leaves are labelled so that $\mathcal{U}_0 = \{1, 2, \dots, N_0\}$ and $\mathcal{U}_1 = \{N_0+1, \dots, N_0+N_1\}$, $N_0 + N_1 = N$. Using probabilities

$$Q_i = \sum_{u \in \mathcal{U}_i} P_u, \qquad i = 0, 1$$

we can give the decomposition in

Lemma 1. For a code $\mathcal{C}$ for $(\mathcal{U}, P^N)$

$$\begin{aligned}
&L_{\mathcal{C}}((P_1, \dots, P_N), (P_1, \dots, P_N)) \\
&= 1 + L_{\mathcal{C}_0}\left(\left(\frac{P_1}{Q_0}, \dots, \frac{P_{N_0}}{Q_0}\right), \left(\frac{P_1}{Q_0}, \dots, \frac{P_{N_0}}{Q_0}\right)\right) Q_0^2 \\
&\quad + L_{\mathcal{C}_1}\left(\left(\frac{P_{N_0+1}}{Q_1}, \dots, \frac{P_{N_0+N_1}}{Q_1}\right), \left(\frac{P_{N_0+1}}{Q_1}, \dots, \frac{P_{N_0+N_1}}{Q_1}\right)\right) Q_1^2.
\end{aligned}$$

This readily yields

Theorem 2. *For every source* $(\mathcal{U}, P^N)$

$$3 > L(P^N) \geq L(P^N, P^N) \geq H_I(P^N).$$

Proof. The bound $3 > L(P^N)$ restates Theorem 3 of [5]. For $N = 2$ and any $\mathcal{C}$ $L_{\mathcal{C}}(P^2, P^2) \geq P_1 + P_2 = 1$, but

$$H_I(P^2) = 2(1 - P_1^2 - (1 - P_1)^2) = 2(2P_1 - 2P_1^2) = 4P_1(1 - P_1) \leq 1. \tag{5.3}$$

This is the induction beginning.

For the induction step use for any code $\mathcal{C}$ the decomposition formula in Lemma 1 and of course the desired inequality for N_0 and N_1 as induction hypothesis.

$$\begin{aligned}
&L_{\mathcal{C}}((P_1, \dots, P_N), (P_1, \dots, P_N)) \\
&\geq 1 + 2\left(1 - \sum_{u \in \mathcal{U}_0} \left(\frac{P_u}{Q_0}\right)^2\right) Q_0^2 + 2\left(1 - \sum_{u \in \mathcal{U}_1} \left(\frac{P_u}{Q_1}\right)^2\right) Q_1^2 \\
&\geq H_I(Q) + Q_0^2 H_I(P^{(0)}) + Q_1^2 H_I(P^{(1)}) = H_I(P^N),
\end{aligned}$$

where $Q = (Q_0, Q_1)$, $1 \geq H(Q)$, $P^{(i)} = \left(\frac{P_u}{Q_i}\right)_{u \in \mathcal{U}_i}$, and the grouping identity is used for the equality. This holds for every $\mathcal{C}$ and therefore also for $\min\limits_{\mathcal{C}} L_{\mathcal{C}}(P^N)$. $\square$

6 On Properties of $\bar{L}(P^N)$

Clearly for $P^N = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)$ $\bar{L}(P^N) = L(P^N, P^N)$ and Theorem 2 gives therefore also the lower bound

$$\bar{L}(P^N) \geq H_I(P^N) = 2\left(1 - \frac{1}{N}\right), \tag{6.1}$$

which holds by Theorem 1 only for the Huffman code, but then for all distributions.

We shall see later in Example 6 that $H_I(P^N)$ is not a lower bound for general distributions P^N! Here we mean non-pathological cases, that is, not those where the inequality fails because $\bar{L}(P)$ (and also $L(P,P)$) is not continuous in P, but $H_I(P)$ is, like in the following case.

Example 5. Let $N = 2^k + 1$, $P(1) = 1 - \varepsilon$, $P(u) = \frac{\varepsilon}{2^k}$ for $u \neq 1$, $P^{(\varepsilon)} = \left(1 - \varepsilon, \frac{\varepsilon}{2^k}, \dots, \frac{\varepsilon}{2^k}\right)$, then

$$\bar{L}(P^{(\varepsilon)}) = 1 + \varepsilon 2\left(1 - \frac{1}{2^k}\right) \tag{6.2}$$

and $\lim_{\varepsilon \to 0} \bar{L}(P^{(\varepsilon)}) = 1$ whereas $\lim_{\varepsilon \to 0} H_I(P^{(\varepsilon)}) = \lim_{\varepsilon \to 0} \left(2\left(1 - (1-\varepsilon)^2 - \left(\frac{\varepsilon}{2^k}\right)^2 2^k\right)\right) = 0.$

However, such a discontinuity occurs also in noiseless coding by Shannon.

The same discontinuity occurs for $L(P^{(\varepsilon)}, P^{(\varepsilon)})$.

Furthermore, for $N = 2$ $\quad P^{(\varepsilon)} = (1 - \varepsilon, \varepsilon)$, $\bar{L}(P^{(\varepsilon)}) = 1 \quad L(P^{(\varepsilon)}, P^{(\varepsilon)}) = 1$ and $H_I(P^{(\varepsilon)}) = 2(1 - \varepsilon^2 - (1-\varepsilon)^2) = 0$ for $\varepsilon = 0$.

However, $\max_{\varepsilon} H_I(P^{(\varepsilon)}) = \max_{\varepsilon} 2(-2\varepsilon^2 + 2\varepsilon) = 1$ (for $\varepsilon = \frac{1}{2}$). Does this have any significance?

There is a second decomposition formula, which gives useful lower bounds on $\bar{L}_{\mathcal{C}}(P^N)$ for codes $\mathcal{C}$ with corresponding subcodes $\mathcal{C}_0, \mathcal{C}_1$ with uniform distributions.

Lemma 2. *For a code $\mathcal{C}$ for $(\mathcal{U}, P^N)$ and corresponding tree $\mathcal{T}$ let*

$$T_{\mathcal{T}}(P^N) = \sum_{u \in \mathcal{U}} L(u).$$

Then (in analogous notation)

$$T_{\mathcal{T}}(P^N) = N_0 + N_1 + T_{\mathcal{T}_0}(P^{(0)})Q_0 + T_{\mathcal{T}_1}(P^{(1)})Q_1.$$

However, identification entropy is not a lower bound for $\bar{L}(P^N)$. We strive now for the worst deviation by using Lemma 2 and by starting with $\mathcal{C}$, whose parts $\mathcal{C}_0, \mathcal{C}_1$ satisfy the entropy inequality.

Then inductively

$$T_{\mathcal{T}}(P^N) \geq N + 2\left(1 - \sum_{u \in \mathcal{U}_0} \left(\frac{P_u}{Q_0}\right)^2\right) N_0 Q_0 + 2\left(1 - \sum_{u \in \mathcal{U}_1} \left(\frac{P_u}{Q_1}\right)^2\right) N_1 Q_1 \quad (6.3)$$

and

$$\frac{T_{\mathcal{T}}(P^N)}{N} \geq 1 + \sum_{i=0}^{1} 2\left(1 - \sum_{u \in \mathcal{U}_i} \left(\frac{P_u}{Q_i}\right)^2\right) \frac{N_i Q_i}{N} \triangleq A, \text{ say.}$$

We want to show that for

$$2\left(1 - \sum_{u \in \mathcal{U}} P_u^2\right) \triangleq B, \text{ say,}$$

$$A - B \geq 0. \quad (6.4)$$

We write

$$\begin{aligned} A - B &= \left[-1 + 2\sum_{i=0}^{1} \frac{N_i Q_i}{N}\right] + 2\left[\sum_{u \in \mathcal{U}} P_u^2 - \sum_{i=0}^{1} \sum_{u \in \mathcal{U}_i} \left(\frac{P_u}{Q_i}\right)^2 \frac{N_i Q_i}{N}\right] \\ &= C + D, \text{ say.} \end{aligned} \quad (6.5)$$

C and D are functions of P^N and the partition $(\mathcal{U}_0, \mathcal{U}_1)$, which determine the Q_i's and N_i's. The minimum of this function can be analysed without reference to codes. Therefore we write here the partitions as $(\mathcal{U}_1, \mathcal{U}_2)$, $C = C(P^N, \mathcal{U}_1, \mathcal{U}_2)$ and $D = D(P^N, \mathcal{U}_1, \mathcal{U}_2)$. We want to show that

$$\min_{P^N, (\mathcal{U}_1, \mathcal{U}_2)} C(P^N, \mathcal{U}_1, \mathcal{U}_2) + D(P^N, \mathcal{U}_1, \mathcal{U}_2) \geq 0. \quad (6.6)$$

A first idea

Recall that the proof of (5.3) used

$$2Q_0^2 + 2Q_1^2 - 1 \geq 0. \quad (6.7)$$

Now if $Q_i = \frac{N_i}{N}$ $(i = 0, 1)$, then by (6.7)

$$A - B = \left[-1 + 2\sum_{i=0}^{1} \frac{N_i^2}{N^2}\right] + 2\left[\sum_{u \in \mathcal{U}} P_u^2 - \sum_{u \in \mathcal{U}} P_u^2\right] \geq 0.$$

A goal could be now to achieve $Q_i \sim \frac{N_i}{N}$ by rearrangement not increasing $A - B$, because in case of equality $Q_i = \frac{N_i}{N}$ that does it.

This leads to a nice **problem of balancing a partition** $(\mathcal{U}_1, \mathcal{U}_2)$ of $\mathcal{U}$. More precisely for $P^N = (P_1, \ldots, P_N)$

$$\varepsilon(P^N) = \min_{\phi \neq \mathcal{U}_1 \subset \mathcal{U}} \left| \sum_{u \in \mathcal{U}_1} P_u - \frac{|\mathcal{U}_1|}{N} \right|.$$

Then clearly for an optimal $\mathcal{U}_1$

$$Q_1 = \frac{|\mathcal{U}_1|}{N} \pm \varepsilon(P^N) \quad \text{and} \quad Q_2 = \frac{N - |\mathcal{U}_1|}{N} \mp \varepsilon(P^N).$$

Furthermore, one comes to a question of some independent interest. What is

$$\max_{P^N} \varepsilon(P^N) = \max_{P^N} \min_{\phi \neq \mathcal{U}_1 \subset \mathcal{U}} \left| \sum_{u \in \mathcal{U}_1} P_u - \frac{|\mathcal{U}_1|}{N} \right| ?$$

One can also go from sets $\mathcal{U}_1$ to distributions $\mathcal{R}$ on $\mathcal{U}$ and get, perhaps, a smoother problem in the spirit of game theory.

However, we follow another approach here.

A rearrangement

We have seen that for $Q_i = \frac{N_i}{N}$ $D = 0$ and $C \geq 0$ by (6.7). Also, there is "air" up to 1 in C, if $\frac{N_i}{N}$ is away from $\frac{1}{2}$. Actually, we have

$$C = -\left(\frac{N_1}{N} + \frac{N_2}{N}\right)^2 + 2\left(\frac{N_1}{N}\right)^2 + 2\left(\frac{N_2}{N}\right)^2 = \left(\frac{N_1}{N} - \frac{N_2}{N}\right)^2. \tag{6.8}$$

Now if we choose for $N = 2m$ even $N_1 = N_2 = m$, then the air is out here, $C = 0$, but it should enter the second term D in (6.5).

Let us check this case first. Label the probabilities $P_1 \geq P_2 \geq \cdots \geq P_N$ and define $\mathcal{U}_1 = \left\{1, 2, \ldots, \frac{N}{2}\right\}$, $\mathcal{U}_2 = \left\{\frac{N}{2} + 1, \ldots, N\right\}$. Thus obviously

$$Q_1 = \sum_{u \in \mathcal{U}_1} P_u \geq Q_2 = \sum_{u \in \mathcal{U}_2} P_u$$

and

$$D = 2\left(\sum_{u \in \mathcal{U}} P_u^2 - \sum_{i=1}^{2} \frac{1}{2Q_i} \sum_{u \in \mathcal{U}_i} P_u^2\right).$$

Write $Q = Q_1$, $1 - Q = Q_2$. We have to show

$$\sum_{u \in \mathcal{U}_1} P_u^2 \left(1 - \frac{1}{(2Q)^2}\right) \geq \sum_{u \in \mathcal{U}_2} P_u^2 \left(\frac{1}{(2Q_2)^2} - 1\right)$$

or

$$\sum_{u\in\mathcal{U}_1} P_u^2 \frac{(2Q)^2-1}{(2Q)^2} \geq \sum_{u\in\mathcal{U}_2} P_u^2 \left(\frac{1-(2(1-Q))^2}{(2(1-Q))^2}\right). \tag{6.9}$$

At first we decrease the left hand side by replacing $P_1,\dots,P_{\frac{N}{2}}$ all by $\frac{2Q}{N}$. This works because $\sum P_i^2$ is Schur-concave and $P_1 \geq \dots \geq P_{\frac{N}{2}}$, $\frac{2Q}{N} = \frac{2(P_1+\dots+P_{\frac{N}{2}})}{N} \geq P_{\frac{N}{2}+1}$, because $\frac{2Q}{N} \geq P_{\frac{N}{2}} \geq P_{\frac{N}{2}+1}$. Thus it suffices to show that

$$\frac{N}{2}\left(\frac{2Q}{N}\right)^2 \frac{(2Q)^2-1}{(2Q)^2} \geq \sum_{u\in\mathcal{U}_2} P_u^2 \frac{1-(2(1-Q))^2}{(2(1-Q))^2} \tag{6.10}$$

or that

$$\frac{1}{2N} \geq \sum_{u\in\mathcal{U}_2} P_u^2 \frac{1-(2(1-Q))^2}{(2(1-Q))^2((2Q)^2-1)}. \tag{6.11}$$

Secondly we increase now the right hand side by replacing $P_{\frac{N}{2}+1},\dots,P_N$ all by their maximal possible values $\left(\frac{2Q}{N},\frac{2Q}{N},\dots,\frac{2Q}{N},q\right) = (q_1,q_2,\dots,q_t,q_{t+1})$, where $q_i = \frac{2Q}{N}$ for $i=1,\dots,t$, $q_{t+1}=q$ and $t\cdot\frac{2Q}{N}+q = 1-Q$, $t = \left\lfloor \frac{(1-Q)N}{2Q}\right\rfloor$, $q < \frac{2Q}{N}$.

Thus it suffices to show that

$$\frac{1}{2N} \geq \left(\left\lfloor\frac{(1-Q)N}{2Q}\right\rfloor \cdot \left(\frac{2Q}{N}\right)^2 + q^2\right) \frac{1-(2(1-Q))^2}{(2(1-Q))^2((2Q)^2-1)}. \tag{6.12}$$

Now we inspect the easier case $q=0$. Thus we have $N=2m$ and equal probabilities $P_i = \frac{1}{m+t}$ for $i=1,\dots,m+t=m$, say for which (6.12) goes wrong! We arrived at a very simple counterexample.

Example 6. In fact, simply for $P_M^N = \left(\frac{1}{M},\dots,\frac{1}{M},0,0,0\right)$ $\lim_{N\to\infty} \bar{L}(P_M^N) = 0$, whereas

$$H_I(P_M^N) = 2\left(1-\frac{1}{M}\right) \text{ for } N \geq M.$$

Notice that here

$$\sup_{N,M} |\bar{L}(P_M^N) - H_I(P_M^N)| = 2. \tag{6.13}$$

This leads to the

Problem 1. Is $\sup_P |\bar{L}(P) - H_I(P)| = 2$? which is solved in the next section.

7 Upper Bounds on $\bar{L}(P^N)$

We know from Theorem 1 that

$$\bar{L}(P^{2^k}) \leq 2\left(1-\frac{1}{2^k}\right) \tag{7.1}$$

and come to the

Problem 2. Is $\bar{L}(P^N) \leq 2\left(1 - \frac{1}{2^k}\right)$ for $N \leq 2^k$?

This is the case, if the answer to the next question is positive.

Problem 3. Is $\bar{L}\left(\left(\frac{1}{N}, \dots, \frac{1}{N}\right)\right)$ monotone increasing in N?

In case the inequality in Problem 2 does not hold then it should with a very small deviation. Presently we have the following result, which together with (6.13) settles Problem 1.

Theorem 3. *For $P^N = (P_1, \dots, P_N)$*

$$\bar{L}(P^N) \leq 2\left(1 - \frac{1}{N^2}\right).$$

Proof. (The induction beginning $\bar{L}(P^2) = 1 \leq 2\left(1 - \frac{1}{4}\right)$ holds.) Define now $\mathcal{U}_1 = \left\{1, 2, \dots, \left\lfloor \frac{N}{2} \right\rfloor\right\}$, $\mathcal{U}_2 = \left\{\left\lfloor \frac{N}{2} \right\rfloor + 1, \dots, N\right\}$ and Q_1, Q_2 as before. Again by the decomposition formula of Lemma 2 and induction hypothesis

$$T(P^N) \leq N + 2\left(1 - \frac{1}{\left\lfloor \frac{N}{2} \right\rfloor^2}\right) Q_1 \left\lfloor \frac{N}{2} \right\rfloor + 2\left(1 - \frac{1}{\left\lceil \frac{N}{2} \right\rceil^2}\right) Q_2 \cdot \left\lceil \frac{N}{2} \right\rceil$$

and

$$\bar{L}(P^N) = \frac{1}{N} T(P^N) \leq 1 + \frac{2\left\lfloor \frac{N}{2} \right\rfloor Q_1 + 2\left\lceil \frac{N}{2} \right\rceil Q_2}{N} - \frac{2}{\left\lfloor \frac{N}{2} \right\rfloor} \cdot \frac{Q_1}{N} - \frac{2Q_2}{\left\lceil \frac{N}{2} \right\rceil N} \tag{7.2}$$

Case N even: $\bar{L}(P^N) \leq 1 + Q_1 + Q_2 - \left(\frac{4}{N^2} Q_1 + \frac{4}{N^2} Q_2\right) = 2 - \frac{4}{N^2} = 2\left(1 - \frac{2}{N^2}\right) \leq 2\left(1 - \frac{1}{N^2}\right)$

Case N odd: $\bar{L}(P^N) \leq 1 + \frac{N-1}{N} Q_1 + \frac{N+1}{N} Q_2 - 4\left(\frac{Q_1}{(N-1)N} + \frac{Q_2}{(N+1)N}\right) \leq 1 + 1 + \frac{Q_2 - Q_1}{N} - \frac{4}{(N+1)N}$

Choosing the $\left\lceil \frac{N}{2} \right\rceil$ smallest probabilities in $\mathcal{U}_2$ (after proper labelling) we get for $N \geq 3$

$$\bar{L}(P^N) \leq 1 + 1 + \frac{1}{N \cdot N} - \frac{4}{(N+1)N} = 2 + \frac{1 - 3N}{(N+1)N^2} \leq 2 - \frac{2}{N^2} = 2\left(1 - \frac{1}{N^2}\right),$$

because $1 - 3N \leq -2N - 2$ for $N \geq 3$.

8 The Skeleton

Assume that all individual probabilities are powers of $\frac{1}{2}$

$$P_u = \frac{1}{2^{\ell_u}}, \quad u \in \mathcal{U}. \tag{8.1}$$

Define then $k = k(P^N) = \max_{u \in \mathcal{U}} \ell_u$.

Since $\sum\limits_{u\in\mathcal{U}} \frac{1}{2^{\ell_u}} = 1$ by Kraft's theorem there is a PC with codeword lengths

$$||c_u|| = \ell_u. \tag{8.2}$$

Notice that we can put the probability $\frac{1}{2^k}$ at all leaves in the binary regular tree and that therefore

$$L(u) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{2^3}3 + \cdots + \frac{1}{2^t}t + \cdots + \frac{2}{2^{\ell_u}}. \tag{8.3}$$

For the calculation we use

Lemma 3. *Consider the polynomials* $G(x) = \sum\limits_{t=1}^{r} t\cdot x^t + rx^r$ *and* $f(x) = \sum\limits_{t=1}^{r} x^t$, *then*

$$G(x) = x\ f'(x) + r\ x^r = \frac{(r+1)x^{r+1}(x-1) - x^{r+2} + x}{(x-1)^2} + r\ x^r.$$

Proof. Using the summation formula for a geometric series

$$f(x) = \frac{x^{r+1}-1}{x-1} - 1$$

$$f'(x) = \sum_{t=1}^{r} t\ x^{t-1} = \frac{(r+1)x^r(x-1) - x^{r+1} + 1}{(x-1)^2}.$$

This gives the formula for G.

Therefore for $x = \frac{1}{2}$

$$G\left(\frac{1}{2}\right) = -(r+1)\left(\frac{1}{2}\right)^r - \left(\frac{1}{2}\right)^r + 2 + r\left(\frac{1}{2}\right)^r$$

$$= -\frac{1}{2^{r-1}} + 2$$

and since $L(u) = G\left(\frac{1}{2}\right)$ for $r = \ell_u$

$$L(u) = 2\left(1 - \frac{1}{2^{\ell_u}}\right) = 2\left(1 - \frac{1}{2^{\log\frac{1}{P_u}}}\right)$$

$$= 2(1-P_u). \tag{8.4}$$

Therefore

$$L(P^N, P^N) \leq \sum_u P_u(2(1-P_u)) = H_I(P^N) \tag{8.5}$$

and by Theorem 2

$$L(P^N, P^N) = H_I(P^N). \tag{8.6}$$

Theorem 4. [1] *For* $P^N = (2^{-\ell_1}, \dots, 2^{-\ell_N})$ *with 2-powers as probabilities*

$$L(P^N, P^N) = H_I(P^N).$$

This result shows that identification entropy is a right measure for identification source coding. For Shannon's data compression we get for this source $\sum_u p_u ||c_u|| = \sum_u p_u \ell_u = -\sum_u p_u \log p_u = H(P^N)$, again an identity.

For general sources the minimal average length deviates there from $H(P^N)$, but by not more than 1.

Presently we also have to accept some deviation from the identity.

We give now a first (crude) approximation. Let

$$2^{k-1} < N \leq 2^k \tag{8.7}$$

and assume that the probabilities are sums of powers of $\frac{1}{2}$ with exponents not exceeding k

$$P_u = \sum_{j=1}^{\alpha(u)} \frac{1}{2^{\ell_{uj}}}, \ell_{u1} \leq \ell_{u2} \leq \dots \leq \ell_{u\alpha(u)} \leq k. \tag{8.8}$$

We now use the **idea of splitting object $\boldsymbol{u}$ into objects $\boldsymbol{u1, \dots, u\alpha(u)}$**. (8.9)

Since

$$\sum_{u,j} \frac{1}{2^{\ell_{uj}}} = 1 \tag{8.10}$$

again we have a PC with codewords c_{uj} $(u \in \mathcal{U}, j = 1, \dots, \alpha(u))$ and a regular tree of depth k with probabilities $\frac{1}{2^k}$ on all leaves.

Person u can find out whether u occurred, he can do this (and more) by finding out whether $u1$ occurred, then whether $u2$ occurred, etc. until $u\alpha(u)$. Here

$$L(us) = 2\left(1 - \frac{1}{2^{\ell_{us}}}\right) \tag{8.11}$$

and

$$\sum_{u,s} L(us) P_{us} = 2\left(1 - \sum_{u,s} \frac{1}{2^{\ell_{us}}} \cdot \frac{1}{2^{\ell_{us}}}\right) = 2\left(1 - \sum_u \left(\sum_{s=1}^{\alpha(u)} P_{us}^2\right)\right). \tag{8.12}$$

On the other hand, being interested only in the original objects this is to be compared with $H_I(P^N) = 2\left(1 - \sum_u \left(\sum_s P_{us}\right)^2\right)$, which is smaller.

[1] In a forthcoming paper "An interpretation of identification entropy" the author and Ning Cai show that $L_{\mathcal{C}}(P,Q)^2 \leq L_{\mathcal{C}}(P,P) L_{\mathcal{C}}(Q,Q)$ and that for a block code $\mathcal{C}$ $\min_{P \text{ on } \mathcal{U}} L_{\mathcal{C}}(P,P) = L_{\mathcal{C}}(R,R)$, where R is the uniform distribution on $\mathcal{U}$! Therefore $\bar{L}_{\mathcal{C}}(P) \leq L_{\mathcal{C}}(P,P)$ for a block code $\mathcal{C}$.

However, we get

$$\left(\sum_s P_{us}\right)^2 = \sum_s P_{us}^2 + \sum_{s \neq s'} P_{us} P_{us'} \leq 2 \sum_s P_{us}^2$$

and therefore

Theorem 5

$$L(P^N, P^N) \leq 2\left(1 - \sum_u \left(\sum_{s=1}^{\alpha(u)} P_{us}^2\right)\right) \leq 2\left(1 - \frac{1}{2}\sum_u P_u^2\right). \tag{8.13}$$

For $P_u = \frac{1}{N}(u \in \mathcal{U})$ this gives the upper bound $2\left(1 - \frac{1}{2N}\right)$, which is better than the bound in Theorem 3 for uniform distributions.

Finally we derive

Corollary

$$L(P^N, P^N) \leq H_I(P^N) + \max_{1 \leq u \leq N} P_u.$$

It shows the lower bound of $L(P^n, P^N)$ by $H_I(P^N)$ and this upper bound are close.

Indeed, we can write the upper bound

$$2\left(1 - \frac{1}{2}\sum_{u=1}^N P_u^2\right) \text{ as } H_I(P^N) + \sum_{u=1}^N P_u^2$$

and for $P = \max_{1 \leq u \leq N} P_u$, let the positive integer t be such that $1 - tp = p' < p$. Then by Schur concavity of $\sum\limits_{u=1}^N P_u^2$ we get $\sum\limits_{u=1}^N P_u^2 \leq t \cdot p^2 + p'^2$, which does not exceed $p(tp + p') = p$.

Remark. In its form the bound is tight, because for $P^2 = (p, 1-p)$

$$L(P^2, P^2) = 1 \text{ and } \lim_{p \to 1} H_I(P^2) + p = 1.$$

Remark. Concerning $\bar{L}(P^N)$ (see footnote) for $N = 2$ the bound $2\left(1 - \frac{1}{4}\right) = \frac{3}{2}$ is better than $H_I(P^2) + \max\limits_u P_u$ for $P^2 = \left(\frac{2}{3}, \frac{1}{3}\right)$, where we get $2(2p_1 - 2p_1^2) + p_1 = p_1(5 - 4p_1) = \frac{2}{3}\left(5 - \frac{8}{3}\right) = \frac{14}{9} > \frac{3}{2}$.

9 Directions for Research

A. Study

$$L(P, R) \quad \text{for} \quad P_1 \geq P_2 \geq \cdots \geq P_N \quad \text{and} \quad R_1 \geq R_1 \geq \cdots \geq R_N.$$

B. Our results can be extended to q-ary alphabets, for which then identification entropy has the form

$$H_{I,q}(P) = \frac{q}{q-1}\left(1 - \sum_{i=1}^{N} P_i^2\right).^2$$

C. So far we have considered prefix-free codes. One also can study
   a. fix-free codes
   b. uniquely decipherable codes
D. Instead of the number of checkings one can consider other cost measures like the αth power of the number of checkings and look for corresponding entropy measures.
E. The analysis on universal coding can be refined.
F. In [5] first steps were taken towards source coding for K-identification. This should be continued with a reflection on entropy and also towards GTIT.
G. **Grand ideas:** Other data structures
   a. Identification source coding with **parallelism**: there are N identical code-trees, each person uses his own, but informs others
   b. Identification source coding with **simultaneity**: $m (m = 1, 2, \dots, N)$ persons use simultaneously the same tree.
H. It was shown in [5] that $L(P^N) \leq 3$ for all P^N. Therefore there is a universal constant $A = \sup_{P^N} L(P^N)$. It should be estimated!
I. We know that for $\lambda \in (0, 1)$ there is a subset $\mathcal{U}$ of cardinality $\exp\{f(\lambda)H(P)\}$ with probability at least λ for $f(\lambda) = (1-\lambda)^{-1}$ and $\lim_{\lambda \to 0} f(\lambda) = 1$.
   Is there such a result for $H_I(P)$?
   It is very remarkable that in our world of source coding the classical range of entropy $[0, \infty)$ is replaced by $[0, 2)$ – singular, dual, plural – there is some appeal to this range.

References

1. C.E. Shannon, A mathematical theory of communication, Bell Syst. Techn. J. 27, 379-423, 623-656, 1948.
2. D.A. Huffman, A method for the construction of minimum redundancy codes, Proc. IRE 40, 1098-1101, 1952.
3. R. Ahlswede and G. Dueck, Identification via channels, IEEE Trans. Inf. Theory, Vol. 35, No. 1, 15-29, 1989.
4. R. Ahlswede, General theory of information transfer: updated, General Theory of Information Transfer and Combinatorics, a Special issue of Discrete Applied Mathematics.
5. R. Ahlswede, B. Balkenhol, and C. Kleinewächter, Identification for sources, this volume.

[2] In the forthcoming paper mentioned in 1. the coding theoretic meanings of the two factors $\frac{q}{q-1}$ and $\left(1 - \sum_{i=1}^{N} P_i^2\right)$ are also explained.