I

Identification for Sources

R. Ahlswede, B. Balkenhol, and C. Kleinewächter

1 Introduction

1.1 Pioneering Model

The classical transmission problem deals with the question how many possible messages can we transmit over a noisy channel? Transmission means there is an answer to the question "What is the actual message?" In the identification problem we deal with the question how many possible messages the receiver of a noisy channel can identify? Identification means there is an answer to the question "Is the actual message u?" Here u can be any member of the set of possible messages.

Allowing randomized encoding the optimal code size grows double exponentially in the blocklength and somewhat surprisingly the second order capacity equals Shannon's first order transmission capacity (see [3]).

Thus Shannon's Channel Coding Theorem for Transmission is paralleled by a Channel Coding Theorem for Identification. It seems natural to look for such a parallel for sources, in particular for noiseless coding. This was suggested by Ahlswede in [4].

Let $(\mathcal{U}, P)$ be a source, where $\mathcal{U} = \{1, 2, \ldots, N\}$, $P = (P_1, \ldots, P_N)$, and let $\mathcal{C} = \{c_1, \ldots, c_N\}$ be a binary prefix code (PC) for this source with $\|c_u\|$ as length of c_u. Introduce the RV U with $\text{Prob}\,(U = u) = p_u$ for $u = 1, 2, \ldots, N$ and the RV C with $C = c_u = (c_{u_1}, c_{u_2}, \ldots, c_{u\|c_u\|})$ if $U = u$.

We use the PC for noiseless identification, that is user u wants to know whether the source output equals u, that is, whether C equals c_u or not. He iteratively checks whether $C = (C_1, C_2, \ldots)$ 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?

In order to calculate this quantity we introduce for the binary tree $T_{\mathcal{C}}$, whose leaves are the codewords $c_1, \ldots, c_N$, the sets of leaves $\mathcal{C}_{ik} (1 \leq i \leq N; 1 \leq k)$, where $\mathcal{C}_{ik} = \{c \in \mathcal{C} : c \text{ coincides with } c_i \text{ exactly until the } k\text{'th letter of } c_i\}$. If C takes a value in $\mathcal{C}_{uk}, 0 \leq k \leq \|c_u\| - 1$, the answers are k times "Yes" and 1 time "No". For $C = c_u$ the answers are $\|c_u\|$ times "Yes". Thus

$$L_{\mathcal{C}}(P, u) = \sum_{k=0}^{\|c_u\|-1} P(C \in \mathcal{C}_{uk})(k+1) + \|c_u\| P_u. \ ^{1}$$

[1] Probability distributions and codes depend on N, but are mostly written without an index N.

R. Ahlswede et al. (Eds.): Information Transfer and Combinatorics, LNCS 4123, pp. 51–61, 2006.

For code $\mathcal{C}$ $L_{\mathcal{C}}(P) = \max_{1 \le u \le N} L_{\mathcal{C}}(P, u)$ is the expected number of checkings in the worst case and $L(P) = \min_{\mathcal{C}} L_{\mathcal{C}}(P)$ is this number for a best code.

Analogously, if $\tilde{\mathcal{C}}$ is a randomized coding, $L_{\tilde{\mathcal{C}}}(P, u)$, $L_{\tilde{\mathcal{C}}}(P)$ and $\tilde{L}(P)$ were also introduced in [4].

What are the properties of $L(P)$ and $\tilde{L}(P)$? In analogy to the role of entropy $H(P)$ in Shannon's Noiseless Source Coding Theorem they can be viewed as approximations to a kind of "identification entropy" functional H_I.

Their investigation is left to future research. We quickly report now two simpler pioneering questions and partial answers from [4]. They shed some light on the idea that in contrast to classical entropy H, which takes values between 0 and ∞, the right functional H_I shall have 2 as maximal value.

Let us start with $P_N = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$ and set $f(N) = L(P_N)$.

1. What is $\sup_N f(N)$ or $\lim_{N \to \infty} f(N)$?
   Starting with an identification code for $N = 2^{k-1}$ a new one for 2^k users is constructed by adding for half of all users a 1 as prefix to the codewords and a 0 for the other half. Obviously we are getting an identification code with twice as many codewords in this way. Now user u has to read the first bit. With probability $\frac{1}{2}$ he then stops and with probability $\frac{1}{2}$ he needs only an expected number of $f(2^{k-1})$ many further checkings. Now an optimal identification code is at least as good as the constructed one and we get the recursion
   $$f(2^k) \le 1 + \frac{1}{2} f(2^{k-1}), f(2) = 1$$
   and therefore
   $$f(2^k) \le 2 - 2^{-(k-1)}.$$
   On the other hand it can be verified that $f(9) = 1 + \frac{10}{9} > 2$ and more generally $f(2^k + 1) > 2$.
2. Is $\tilde{L}(P) \le 2$?
   This is the case under the stronger assumption that encoder and decoder have access to a random experiment with unlimited capacity of common randomness (see [5]).
   For $P = (P_1, \ldots, P_N)$, $N \le 2^n$ write $P^{(n)} = (P_1, \ldots, P_N, 0, \ldots, 0)$ with 2^n components. Use a binary regular tree of depth n with leaves $1, 2, \ldots, 2^n$ represented in binary expansions.
   The common random experiment with 2^n outcomes can be used to use 2^n cyclic permutations of $1, 2, \ldots, 2^n$ for 2^n deterministic codes. For each u we get equally often 0 and 1 in its representation and an expected word length $\le 2 - \frac{1}{2^{n-1}} \le 2$. The error probability is 0.

Remark 1. Note that the **same** tree $T_{\mathcal{C}}$ can be used by **all** users in order to answer their question ("Is it me or not?").

1.2 Further Models and Definitions

The model of identification for sources described can be extended (as for channels in the spirit of [4]) to *generalized identification* (GI) as follows.

There is now a set of users $\mathcal{V}$ (not necessarily equal to $\mathcal{U}$), where user $v \in \mathcal{V}$ has a set $\mathcal{U}_v \subset \mathcal{U}$ of source outputs of his interest, that is, he wants to know whether the source output u is in $\mathcal{U}_v$ or not.

Furthermore we speak of *generalized identification with decoding* (GID), if user v not only finds out whether the output is in $\mathcal{U}_v$, but also identifies it if it is in $\mathcal{U}_v$.

Obviously the two models coincide if $|\mathcal{U}_v| = 1$ for $v \in \mathcal{V}$. Also, they specialize to the original model in **1.1**, if $\mathcal{V} = \mathcal{U}$ and $\mathcal{U}_v = \{v\}$ for $v \in \mathcal{U}$.

For our analysis we use the following definition. We denote by $D(x)$ the set of all proper prefixes of $x \in \{0,1\}^*$, i.e.

$$D(x) \triangleq \{y \in \{0,1\}^* : y \text{ is prefix of } x \text{ and } \|y\| < \|x\|\}. \tag{1.1}$$

e stands for the empty word in $\{0,1\}^*$. For a set $A \subset \{0,1\}^*$ we extend this notion to

$$D(A) \triangleq \bigcup_{x \in A} D(x). \tag{1.2}$$

$\{0,1\}^*$ can be viewed as a binary, regular infinite tree with root e. A code $\mathcal{C}$ corresponds to the subtree $T_{\mathcal{C}}$ with root e and leaves $c_1, \ldots, c_N$.

In the sequel we use a specific example of a code for illustrations of concepts and ideas.

Example 1. Let $\mathcal{C}$ be the set of all words of length 3. Notice that $D(010) = \{e, 0, 01\}$ and $D(\{001, 010\}) = \{e, 0, 00, 01\}$.

The set $\mathcal{C}_v = \{c_u : u \in \mathcal{U}_v\}$ is a code for user v. For GID its codewords have to be uniquely decodable by user v in order to identify the source output. For this he uses the set of stop sequences

$$\mathcal{S}_v = \big\{y_1 \ldots y_k : y_1 \ldots y_{k-1} \in D(\mathcal{C}_v) \text{ and } y_1 \ldots y_k \notin D(\mathcal{C}_v)\big\}. \tag{1.3}$$

By definition of D $\mathcal{C}_v$ is contained in $\mathcal{S}_v$. We can also write

$$\mathcal{S}_v = \big\{xy : x \in \{0,1\}^*, y \in \{0,1\} \text{ with } x \in D(\mathcal{C}_v) \text{ and } xy \notin D(\mathcal{C}_v)\big\}. \tag{1.4}$$

(For $k = 1$ $y_1 \ldots y_{k-1}$ describes the empty word e or the root of the code tree which is element of each set $D(\mathcal{C}_v)$.)

Example 2. For the code of Example 1 we have for $\mathcal{C}_v = \{010\}$ $\mathcal{S}_v = \{1, 00, 011, 010\}$ and we have for $\mathcal{C}_v = \{001, 010\}$ $\mathcal{S}_v = \{1, 000, 001, 010, 011\}$.

With the families of sets of stop sequences $\mathcal{S}_v$ we derive first in Section 2 general lower bounds on the number of checkings for both models. In Section 3 we consider a uniform source and show that $\lim_{N \to \infty} f(N) = 2$. Then, in Section 4, we derive bounds on the maximal individual (average) identification length, which is introduced in Section 2 C.

Finally, in Section 5, we introduce an *average identification* length for the case $\mathcal{V} = \mathcal{U}$, $\mathcal{U}_v = \{v\}$ for $v \in \mathcal{V}$ and derive asymptotic results.

2 A Probabilistic Tool for Generalized Identification

General supposition. We consider here prefix codes $\mathcal{C}$, which satisfy the Kraft inequality with equality, that is,

$$\sum_{u\in\mathcal{U}} 2^{-\|c_u\|} = 1. \tag{2.1}$$

We call them saturated, because they cannot be enlarged.

A. GID

For all $x \in \{0,1\}^*$ let $q_{\mathcal{C}}(P,x) = \begin{cases} 0, & \text{if } x \notin D(\mathcal{C}) \cup \mathcal{C} \\ P_u, & \text{if } x = c_u \\ q_{\mathcal{C}}(P,x0) + q_{\mathcal{C}}(P,x1), & \text{if } x \in D(\mathcal{C}). \end{cases}$

The general supposition implies that for any set of stopping sequences $\mathcal{S}_v$ we have $\mathcal{S}_v \subset D(\mathcal{C}) \cup \mathcal{C}$ and the probability for user v to stop in $x \in \mathcal{S}_v$ equals $q_c(P,x)$. After stopping in x user v has read $\|x\|$ many bits. Therefore the average identification length of user v is

$$L_{\mathcal{C}}(P,v) = \sum_{x\in\mathcal{S}_v} q_{\mathcal{C}}(P,x)\|x\|. \tag{2.2}$$

By definition of $q_{\mathcal{C}}$ we get

$$L_{\mathcal{C}}(P,v) = \sum_{x\in D(\mathcal{C}_v)} q_{\mathcal{C}}(P,x). \tag{2.3}$$

By construction $\mathcal{S}_v$ forms a prefix code. Each codeword has to be uniquely decoded by user v. Furthermore the probabilities $q_{\mathcal{C}}(P,x)$, $x \in \mathcal{S}_v$, define a probability distribution on $\mathcal{S}_v$ by

$$P_{\mathcal{C},v}(x) \triangleq q_{\mathcal{C}}(P,x) \text{ for all } x \in \mathcal{S}_v. \tag{2.4}$$

By the Noiseless Coding Theorem $L_{\mathcal{C}}(P,v)$ can be lower bounded by the entropy $H(P_{\mathcal{C},v})$. More directly, using the grouping axiom we get

$$H(P_{\mathcal{C},v}) = \sum_{x\in D(\mathcal{C}_v)} q_{\mathcal{C}}(P,x) h\left(\frac{q_{\mathcal{C}}(P,x1)}{q_{\mathcal{C}}(P,x)}\right), \tag{2.5}$$

where h is the binary entropy function, and thus

$$L_{\mathcal{C}}(P,v) - H(P_{\mathcal{C},v}) = \sum_{x\in D(\mathcal{C}_v)} q_{\mathcal{C}}(P,x)\left(1 - h\left(\frac{q_{\mathcal{C}}(P,x1)}{q_{\mathcal{C}}(P,x)}\right)\right). \tag{2.6}$$

Suppose $P_u > 0$ for all $1 \le u \le N$, then

$$q_{\mathcal{C}}(P,x) > 0 \text{ and with } \left(\frac{q_{\mathcal{C}}(P,x1)}{q_{\mathcal{C}}(P,x)}\right) \le 1 \text{ for all } x \in D(\mathcal{C})$$

it follows under the general supposition (2.1) for every user $v \in \mathcal{V}$ the average identification length satisfies

Theorem 1

$$L_{\mathcal{C}}(P,v) \geq H(P_{\mathcal{C},v}) \text{ with "=" iff } \frac{q_{\mathcal{C}}(P,x1)}{q_{\mathcal{C}}(P,x)} = \frac{1}{2} \text{ for all } x \in D(\mathcal{C}_v). \tag{2.7}$$

Since P is fixed we write now $L_{\mathcal{C}}(v)$ for $L_{\mathcal{C}}(P,v)$.

B. GI

Suppose we have a node x and a user v with the properties

(a) all codewords having x as prefix are all elements of $\mathcal{C}_v$ or (b) they are all not in $\mathcal{C}_v$.

In this case user v can stop in x and decide whether v occurred or not. By construction of the stop sequences $\mathcal{S}_v$ in (1.3) only case (a) can occur. Therefore we have to start the following algorithm to generate modified sets $\mathcal{S}_v$.

1. If $\mathcal{C}_v$ contains two codewords different only in the last position, say $x_1 \ldots x_k 0$ and $x_1 \ldots x_k 1$ then
   (a) remove these two codewords from $\mathcal{C}_v$ and insert $x_1 \ldots x_k$. This new codeword has the probability $q_{\mathcal{C}}(P, x_1 \ldots x_k)$.
   (b) repeat step 1. Else continue with 2.
2. With the modified sets $\mathcal{C}_v$ construct the sets $\mathcal{S}_v$ as defined in (1.3).

The definition of $L_{\mathcal{C}}(P,v)$, $P_{\mathcal{C},v}$ and $H(P_{\mathcal{C},v})$ are as in (2.2), (2.4) and (2.5). Also the formulas (2.6) and (2.7) hold.

Example 3. Let $\mathcal{C}_v = \{000, 001, 010\}$. After step 1 of the algorithm we get $\mathcal{C}_v = \{00, 010\}$. With step 2 we define $D(\mathcal{C}_v) = \{\varnothing, 0, 01\}$ and $\mathcal{S}_v = \{1, 00, 010, 011\}$.

C. Maximal individual (expected) identification length $L(P)$

For a given probability distribution P and a given code $\mathcal{C}$ user v has uniquely to decode the codewords in $\mathcal{C}_v$.

Using (2.7) we can lower bound $L(P)$ as follows:

(i) Take the set of pairs $\mathcal{M} = \{(\mathcal{C}_v, v) : L(P) = L_{\mathcal{C}}(P,v)\}$.
(ii) Define

$$H_{\max}(P) = \max_{(\mathcal{C}_v,v)\in\mathcal{M}} H(P_{\mathcal{C},v}).$$

Then

$$L(P) \geq H_{\max}(P).$$

Remark 2. Note that

1.

$$\sum_{x\in D(\mathcal{C})} q_{\mathcal{C}}(P,x) = \sum_{u=1}^{N} P_u \|c_u\|.$$

2. Using the grouping axiom it holds

$$\sum_{x \in D(\mathcal{C})} q_{\mathcal{C}}(P,x) h\left(\frac{q_{\mathcal{C}}(P,x1)}{q_{\mathcal{C}}(P,x)}\right) = H(P)$$

for all codes $\mathcal{C}$.

3. If for each code $\mathcal{C}$ there exists a set $\mathcal{C}_v$ (in case B after modification) such that $D(\mathcal{C}_v) = D(\mathcal{C})$, then $L(P) = \sum_{u=1}^{N} P_u \|c_u\|$ where the code $\mathcal{C}$ is the Huffman–code for the probability distribution P.

Example 4. Suppose that $|\mathcal{V}| = \binom{N}{K}$, $K \geq \frac{N}{2}$, and $\{\mathcal{U}_v : v \in \mathcal{V}\} = \binom{[N]}{K}$.

1. In case A there exists for each code $\mathcal{C}$ a set $\mathcal{C}_v$ such that $D(\mathcal{C}_v) = D(\mathcal{C})$.
2. In case B with $K = \frac{N}{2}$ there exists for each code $\mathcal{C}$ a set $\mathcal{C}_v$ such that $D(\mathcal{C}_v) = D(\mathcal{C})$.
3. In case B if $K = N$ and thus $\mathcal{V} = \{v_1\}$,$\mathcal{U}_{v_1} = [N]$, then after modifying $\mathcal{C}_{v_1}$ the set $D(\mathcal{C}_{v_1})$ contains only the root of the tree which means the user v_1 has to read nothing from the received codeword (because he knows already the answer).

Remark 3. Example 4 is motivated by K–identification for channels!

3 The Uniform Distribution

Now we return to the original model of 1.1 with $\mathcal{V} = \mathcal{U}$ and $\mathcal{C}_v = \{c_v\}$ for each $v \in \mathcal{V}$. Let $P = (\frac{1}{N}, \ldots, \frac{1}{N})$. We construct a prefix code $\mathcal{C}$ in the following way. In each node (starting at the root) we split the number of remaining codewords in proportion as close as possible to $(\frac{1}{2}, \frac{1}{2})$.

1. Suppose $N = 2^k$. By construction our code $\mathcal{C}$ contains all binary sequences of length k. It follows that

$$q_{\mathcal{C}}(P,x) = \frac{1}{N}\frac{N}{2^{\|x\|}} = 2^{-\|x\|} \tag{3.1}$$

and by (2.3)

$$L_{\mathcal{C}}(P) = \sum_{x \in D(\mathcal{C}_v)} q_{\mathcal{C}}(P,x) = \sum_{i=0}^{k-1} 2^{-i} = 2 - 2^{-k+1} = 2 - \frac{2}{N}. \tag{3.2}$$

2. Suppose $2^{k-1} < N < 2^k$. By construction the remaining code contains only the codeword lengths $k-1$ and k.

   By (2.3) we add the weights ($q_{\mathcal{C}}(P,x)$) of all nodes of a path from the root to a codeword (leave). Therefore in the worst case, N is odd and we have to add the larger weight.

At the root we split $(\frac{N-1}{2}, \frac{N-1}{2}+1)$. Now we split again the larger one and in the worst case this number is again odd. It follows in general that

$$q_{\mathcal{C}}(P,x) \leq \frac{1}{N}\left(\frac{N-1}{2^{\|x\|}}+1\right). \tag{3.3}$$

Therefore

$$\begin{aligned} L_{\mathcal{C}}(P) &\leq \sum_{i=0}^{k-1}\frac{1}{N}\left(\frac{N-1}{2^i}+1\right) = \sum_{i=0}^{k-1}2^{-i} - \frac{1}{N}\sum_{i=0}^{k-1}2^{-i} + \frac{1}{N}\sum_{i=0}^{k-1}1 \\ &= 2 - \frac{1}{N} - \frac{2}{N} + \frac{2}{N^2} + \frac{k}{N} = 2 + \frac{k-3}{N} + \frac{2}{N^2}. \end{aligned} \tag{3.4}$$

With $k = \lceil log_2(N) \rceil$ it follows

Theorem 2. *For* $P = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$

$$\lim_{N\to\infty} L_{\mathcal{C}}(P) = 2. \tag{3.5}$$

4 Bounds on $L(P)$ for General $P = (P_1, \ldots, P_N)$

A. An upper bound

We will now give an inductive construction for identification codes to derive an upper bound on $L(P)$. Let $P = (P_1, \ldots, P_N)$ be the probability distribution. W.l.o.g. we can assume that $P_i \geq P_j$ for all $i < j$. For $N = 2$ of course we assign 0 and 1 as codewords. Now let $N > 2$. We have to consider two cases:

1. $P_1 \geq 1/2$. In this case we assign 0 as codeword to message 1. We set $P_i'' = \frac{P_i}{\sum_{j=2}^{N} P_j}$ for $i = 2, \ldots, N$. By induction we can construct a code for the probability distribution $P'' = (P_2'', \ldots, P_N'')$ and messages 2 to N get the corresponding codewords for P'' but prefixed with a 1.
2. $P_1 < 1/2$. Choose ℓ such that $\delta_\ell = |\frac{1}{2} - \sum_{i=1}^{\ell} P_i|$ is minimal. Set $P_i' = \frac{P_i}{\sum_{j=1}^{\ell} P_j}$ for $i = 1, \ldots, \ell$ and $P_i'' = \frac{P_i}{\sum_{j=\ell+1}^{N} P_j}$ for $i = \ell+1, \ldots, N$. Analogous to the first case we construct codes for the distributions $P' = (P_1', \ldots, P_\ell')$ (called the *left side*) and $P'' = (P_{\ell+1}'', \ldots, P_N'')$ (called the *right side*). We get the code for P by prefixing the codewords from the left side with 0 and the codewords from the right side with 1.

Trivially this procedure yields a prefix code.

Theorem 3. *Let* $N \in \mathbb{N}$ *and let* $P = (P_1, \ldots, P_N)$. *The previous construction yields a prefix code with* $L(P) \leq 3$.

Proof. The case $N = 2$ is trivial. Now let $N \geq 3$.

Case 1. $P_1 \geq 1/2$: In this case we have $L(P) \leq 1 + \max\left\{P_1, L(P'')\sum_{i=2}^{N} P_i\right\}$, where $L(P'')$ denotes the corresponding maximal identification length for prob-

ability distribution P''. If the maximum is assumed for P_1 we have $L(P) \leq 2$, otherwise we get by induction $L(P) < 1 + 3 \cdot 1/2 < 3$.

Case 2. $P_1 < 1/2$ for $i = 1, \ldots, N$: In this case we have

$$L(P) \leq 1 + \max \left\{ L(P') \cdot \sum_{i=1}^{\ell} P_i, \quad L(P'') \cdot \sum_{i=\ell+1}^{N} P_i \right\}.$$

Choose ℓ' such that $\sum_{i=1}^{\ell'} P_i \leq 1/2 < \sum_{i=1}^{\ell'+1} P_i$. Obviously either $\ell = \ell'$ or $\ell = \ell' + 1$.

Subcase: $\ell = \ell'$. Suppose the maximum is assumed on the left side. Then without changing the maximal identification length we can construct a new probability distribution $P''' = (P'''_1, \ldots, P'''_{\ell+1})$ by $P'''_1 = \sum_{i=\ell+1}^{N} P_i$ and $P'''_i = P_{i-1}$ for $2 \leq i \leq \ell + 1$. Since $P'''_1 \geq 1/2$ we are back in case 1. If the maximum is assumed on the right side then let $P'''_1 = \sum_{i=1}^{\ell} P_i$ and $P'''_i = P_{i+\ell-1}$ for all $2 \leq i \leq n-\ell+1$. Notice that in this case $P'''_1 \geq 1/3$ (because $P'''_1 \geq 1/2 - P'''_2/2 \geq 1/2 - P'''_1/2$). Thus by induction $L(P''') \leq 1 + 3 \cdot 2/3 \leq 3$.

Subcase: $\ell = \ell' + 1$. If the maximum is on the right side we set $P'''_1 = \sum_{i=1}^{\ell} P_i \geq 1/2$, $P'''_i = P_{i+\ell-1}$ for $2 \leq i \leq n - \ell + 1$ and we are again back in case 1. Now suppose the maximum is taken on the left side. Since $\sum_{i=1}^{\ell} P_i - 1/2 \leq 1/2 - \sum_{i=1}^{\ell'} P_i$ it follows that $\delta_\ell \leq P_\ell/2$. Because $P_{\ell'} \leq (2\ell')^{-1}$ we have $\delta_\ell \leq (4\ell')^{-1} = (4(\ell-1))^{-1}$. Also note that $\ell \geq 2$. The case $\ell = 2$ is again trivial. Now let $\ell > 2$. Then $L(P) < 3 \cdot (1/2 + \frac{1}{4(\ell-1)}) \leq 3 \cdot (1/2 + 1/8) < 3$.

5 An Average Identification Length

We consider here the case where not only the source outputs but also the users occur at random. Thus in addition to the source $(\mathcal{U}, P)$ and RV U, we are given $(\mathcal{V}, Q)$, $\mathcal{V} \equiv \mathcal{U}$, with RV V independent of U and defined by $\text{Prob}\,(V = v) = Q_v$ for $v \in \mathcal{V}$. The source encoder knows the value u of U, but not that of V, which chooses the user v with probability Q_v. Again let $\mathcal{C} = \{c_1, \ldots, c_N\}$ be a binary prefix code and let $L_{\mathcal{C}}(P, u)$ be the expected number of checkings on code $\mathcal{C}$ for user u. Instead of $L_{\mathcal{C}}(P) = \max_{u \in \mathcal{U}} L_{\mathcal{C}}(P, u)$, the maximal number of expected checkings for a user, we consider now the average number of expected checkings

$$L_{\mathcal{C}}(P, Q) = \sum_{v \in \mathcal{V}} Q_v L_{\mathcal{C}}(P, v) \tag{5.1}$$

and the average number of expected checkings for a best code

$$L(P, Q) = \min_{\mathcal{C}} L_{\mathcal{C}}(P, Q). \tag{5.2}$$

(The models GI and GID can also be considered.)

We also call $L(P, Q)$ the average identification length. $L_{\mathcal{C}}(P, Q)$ can be calculated by the formula

$$L_{\mathcal{C}}(P, Q) = \sum_{x \in D(\mathcal{C})} q_{\mathcal{C}}(Q, x) q_{\mathcal{C}}(P, x). \tag{5.3}$$

In the same way as (5.3) we get the conditional entropy

$$H_{\mathcal{C}}(P\|Q) = \sum_{x\in D(\mathcal{C})} q_{\mathcal{C}}(Q,x)q_{\mathcal{C}}(P,x)h\left(\frac{q_{\mathcal{C}}(P,x1)}{q_{\mathcal{C}}(P,x)}\right). \tag{5.4}$$

5.1 Q Is the Uniform Distribution on $\mathcal{V} = \mathcal{U}$

We begin with $|\mathcal{U}| = N = 2^k$, choose $\mathcal{C} = \{0,1\}^k$ and note that

$$\sum_{\substack{x\in D(\mathcal{C})\\ \|x\|=i}} q_{\mathcal{C}}(P,x) = 1 \text{ for all } 0 \le i \le k. \tag{5.5}$$

By (3.1) for all $x \in \{0,1\}^*$ with $\|x\| \le k$

$$q_{\mathcal{C}}(Q,x) = 2^{-\|x\|} \tag{5.6}$$

and thus by (5.3) and then by (5.5)

$$L_{\mathcal{C}}(P,Q) = \sum_{i=0}^{k-1} \sum_{\substack{x\in D(\mathcal{C})\\ \|x\|=i}} 2^{-i} q_{\mathcal{C}}(P,x) \tag{5.7}$$

$$= \sum_{i=0}^{k-1} 2^{-i} = 2 - 2^{-k+1} = 2 - \frac{2}{N}. \tag{5.8}$$

We continue with the case $2^{k-1} < N < 2^k$ and construct the code $\mathcal{C}$ again as in Section 3. By (3.3)

$$q_{\mathcal{C}}(Q,x) \le \frac{1}{N}\left(\frac{N-1}{2^{\|x\|}} + 1\right). \tag{5.9}$$

Therefore

$$\begin{aligned} L_{\mathcal{C}}(P,Q) &= \sum_{x\in D(\mathcal{C})} q_{\mathcal{C}}(Q,x)q_{\mathcal{C}}(P,x) \le \frac{1}{N}\sum_{x\in D(\mathcal{C})}(\frac{N-1}{2^{\|x\|}}+1)q_{\mathcal{C}}(P,x) \\ &= \frac{1}{N}\sum_{i=0}^{k-1}(\frac{N-1}{2^i}+1)\sum_{\substack{x\in D(\mathcal{C})\\ \|x\|=i}} q_{\mathcal{C}}(P,x) \le \frac{1}{N}\sum_{i=0}^{k-1}(\frac{N-1}{2^i}+1)\cdot 1 \\ &= 2 + \frac{k-3}{N} + \frac{2}{N^2} \text{ (see (3.4))}. \end{aligned} \tag{5.10}$$

With $k = \lceil log_2(N) \rceil$ it follows that

Theorem 4. *Let $N \in \mathbb{N}$ and $P = (P_1, \ldots, P_N)$, then for $Q = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$*

$$\lim_{N\to\infty} L_{\mathcal{C}}(P,Q) = 2. \tag{5.11}$$

Example 4 with average identification length for a uniform Q^*
We get now

$$L_{\mathcal{C}}(P,Q^*) = \sum_{x\in D(\mathcal{C})} \frac{|\{v : x \in D(\mathcal{C}_v)\}|}{|\mathcal{V}|} q_{\mathcal{C}}(P,x) \tag{5.12}$$

and for the entropy in (5.4)

$$H_{\mathcal{C}}(P\|Q^*) = \sum_{x\in D(\mathcal{C})} \frac{|\{v : x \in D(\mathcal{C}_v)\}|}{|\mathcal{V}|} q_{\mathcal{C}}(P,x) h\left(\frac{q_{\mathcal{C}}(P,x1)}{q_{\mathcal{C}}(P,x)}\right). \tag{5.13}$$

Furthermore let $\mathcal{C}_0$ be the set of all codes $\mathcal{C}$ with $L_{\mathcal{C}}(P,Q^*) = L(P,Q^*)$. We define

$$H(P\|Q^*) = \max_{\mathcal{C}\in\mathcal{C}_0} H_{\mathcal{C}}(P\|Q^*). \tag{5.14}$$

Then

$$L(P,Q^*) \geq H(P\|Q^*). \tag{5.15}$$

Case $N = 2^n$: We choose $\mathcal{C} = \{0,1\}^n$ and calculate $\frac{|\{v:x\in D(\mathcal{C}_v)\}|}{|\mathcal{V}|}$. Notice that for any $x \in D(\mathcal{C})$ we have $2^{n-\|x\|}$ many codewords with x as prefix.

Order this set. There are $\binom{N-1}{K-1}$ $(K-1)$–element subsets of $\mathcal{C}$ containing the first codeword in this set. Now we take the second codeword and $K-1$ others, but not the first. In this case we get $\binom{N-2}{K-1}$ further sets and so on.

Therefore $|\{v : x \in D(\mathcal{C}_v)\}| = \sum_{j=1}^{2^{n-\|x\|}} \binom{2^n-j}{K-1}$ and (5.14) yields

$$\begin{aligned} L_{\mathcal{C}}(P,Q^*) &= \frac{1}{\binom{N}{K}} \sum_{x\in D(\mathcal{C})} \sum_{j=1}^{2^{n-\|x\|}} \binom{2^n-j}{K-1} q_{\mathcal{C}}(P,x) \\ &= \frac{1}{\binom{2^n}{K}} \sum_{i=0}^{n-1} \left(\sum_{j=1}^{2^{n-i}} \binom{2^n-j}{K-1}\right) \left(\sum_{\substack{x\in D(\mathcal{C}) \\ \|x\|=i}} q_{\mathcal{C}}(P,x)\right) \\ &= \frac{1}{\binom{2^n}{K}} \sum_{i=0}^{n-1} \left(\sum_{j=1}^{2^{n-i}} \binom{2^n-j}{K-1}\right) \text{ (by (5.5))}. \end{aligned} \tag{5.17}$$

Lets abbreviate this quantity as $g(n,K)$. Its asymptotic behavior remains to be analyzed.

Exact values are

$g(n,1) = 2 - \frac{2}{2^n}$, $g(n,2) = \frac{2}{3}\frac{5\cdot 2^{-n}-9+4\cdot 2^n}{2^n-1}$

$g(n,3) = -\frac{2}{7}\frac{49\cdot 2^n-70+32\cdot 2^{-n}-11\cdot 4^n}{(2^n-1)(2^n-2)}$, $g(n,4) = \frac{4}{105}\frac{-2220+908\cdot 2^{-n}-705\cdot 4^n+1925\cdot 2^n+92\cdot 8^n}{(2^n-1)(2^n-2)(2^n-3)}$

We calculated the limits ($n \to \infty$)

K	1	2	3	4	5	6	7	8	9
$\lim_{n\to\infty} g(n,K)$	2	$\frac{8}{3}$	$\frac{22}{7}$	$\frac{368}{105}$	$\frac{2470}{651}$	$\frac{7880}{1953}$	$\frac{150266}{35433}$	$\frac{13315424}{3011805}$	$\frac{2350261538}{513010785}$

This indicates that $\sup_K \lim_{n\to\infty} g(n,K) = \infty$.

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, Preprint 97–118, SFB 343 "Diskrete Strukturen in der Mathematik", Universität Bielefeld, 1997; General theory of information transfer:updated, General Theory of Information Transfer and Combinatorics, a Special Issue of Discrete Applied Mathematics, to appear.
5. R. Ahlswede and I. Csiszár, Common randomness in Information Theory and Cryptography, Part II: CR capacity, IEEE Trans. Inf. Theory, Vol. 44, No. 1, 55–62, 1998.
6. C.C. Campbell, Definition of entropy by means of a coding problem, Z. Wahrscheinlichkeitstheorie u. verw. Geb., 113–119, 1966.