Z. Wahrscheinlichkeitstheorie verw. Gebiete 29, 351 – 360 (1974)

Determination of All Additive Quasiarithmetic Mean Codeword Lengths

J. Aczél

1.

Campbell, 1966, has introduced *quasiarithmetic mean codeword lengths* in the following manner.

Let $Y=\{\eta_1, \eta_2, \ldots, \eta_K\}$ be a finite set of messages and let $Q=\{q_1, q_2, \ldots, q_K\}$ be an associated distribution of probabilities, so that the probability of η_k is q_k $(k=1, 2, \ldots, K)$ and

$$\sum_{k=1}^{K} q_k=1; \quad q_k \geqq 0 \quad (k=1, 2, \ldots, K). \tag{1}$$

Suppose that we wish to represent the messages in Y by *codewords*, i.e. by finite sequences of elements of the set $\{0, 1, \ldots, D-1\}$ where $D>1$. There is a uniquely decipherable code (see, e.g., Reza, 1961) which represents η_k by a codeword of length (number of elements) n_k $(k=1, 2, \ldots, K)$ if and only if the set of *positive integer codeword lengths* $N=\{n_1, n_2, \ldots, n_K\}$ satisfies the *Kraft inequality*

$$\sum_{k=1}^{K} D^{-n_k} \leqq 1. \tag{2}$$

Let now $\phi: [1, \infty[\to R$ be a continuous strictly increasing function. It has an inverse ϕ^{-1} which is also continuous and strictly increasing. This defines a *quasiarithmetic mean codeword length*

$$L(Q, N; \phi)=\phi^{-1}\left[\sum_{k=1}^{K} q_k \phi(n_k)\right] \tag{3}$$

for all N satisfying (2). The reason for calling L a mean length is that, for $N=\{n, n, \ldots, n\}$, i.e. when all codewords are of equal length n, then $L(Q, N; \phi)=n$. Moreover, if $\phi(x)=\phi_0(x)=x$ $(x \in [1, \infty[)$, then

$$L(Q, N; \phi)=\sum_{k=1}^{K} q_k n_k, \tag{4}$$

the ordinary or *arithmetic mean codeword length*. Campbell 1965, 1966 has also introduced the *exponential mean codeword length*, for which $\phi(x)=\phi_t(x)=D^{tx}$ $(x \in [1, \infty[; t \neq 0)$,

$$L(Q, N; \phi_t)=\frac{1}{t} \log_D \sum_{k=1}^{K} q_k D^{t n_k}. \tag{5}$$

It is easy to see that $\lim_{t \to 0} L(Q, N; \phi_t)=L(Q, N; \phi_0)$.

Important inequalities are known for the mean codeword lengths (4) and (5) (see, e.g., Reza, 1961; Campbell, 1965; Aczél, 1974, and Section 4 of the present paper). These give essentially the Shannon and Rényi entropies as lower bounds of (4) and (5), respectively, and show also that there exist uniquely decipherable codes for which these mean codeword lengths come within a unit (bit) from their lower bounds. The proof of the latter facts use a translativity property of (4) and (5), the generalizations of which we will examine in Section 3. The inequalities, mentioned above, can also be translated into optimal coding statements with respect to certain cost functions, related to ϕ in (3). This we will see in Section 4, in modification of results by Campbell, partly published (Campbell, 1965; 1966) and partly unpublished.

The question arises, *why* the mean codeword lengths (4) and (5) have been chosen, say, among the *quasiarithmetic* mean codeword lengths (3). In our main result, in Section 2, we will show that the following rather natural *additivity* condition *characterizes* them.

Consider two independent sets of messages $X=\{\xi_1, \xi_2, \ldots, \xi_J\}$ and $Y=\{\eta_1, \eta_2, \ldots, \eta_K\}$ with associated probability distributions $P=\{p_1, p_2, \ldots, p_J\}$ and $Q=\{q_1, q_2, \ldots, q_K\}$. Since X and Y are independent, the probability of the pair (ξ_j, η_k) is $p_j q_k$ $(j=1, 2, \ldots, J; k=1, 2, \ldots, K)$. We denote by PQ the probability distribution $\{p_1 q_1, p_1 q_2, \ldots, p_1 q_K, p_2 q_1, p_2 q_2, \ldots, p_2 q_K, \ldots, p_J q_1, p_J q_2, \ldots, p_J q_K\}$. Let ξ_j be represented by a codeword of length m_j $(j=1, 2, \ldots, J)$ and let η_k be represented by a codeword of length n_k $(k=1, 2, \ldots, K)$. Moreover, suppose that we use the same symbols $\{0, 1, \ldots, D-1\}$ in all these representations. The pair (ξ_j, η_k) may be represented by a codeword of length m_j+n_k $(j=1, 2, \ldots, J;$ $k=1, 2, \ldots, K)$. Let us denote these three distributions of lengths by

$$M=\{m_1, m_2, \ldots, m_J\}, \qquad N=\{n_1, n_2, \ldots, n_K\}$$

and

$$M+N=\{m_1+n_1, m_1+n_2, \ldots, m_1+n_K,$$
$$m_2+n_1, m_2+n_2, \ldots, m_2+n_K, \ldots, m_J+n_1, m_J+n_2, \ldots, m_J+n_K\},$$

respectively. If M and N satisfy the Kraft inequality (2) then so does $M+N$ because

$$\sum_{j=1}^{J} D^{-m_j} \leqq 1 \quad \text{and} \quad \sum_{k=1}^{K} D^{-n_k} \leqq 1 \tag{6}$$

imply

$$\sum_{j=1}^{J} \sum_{k=1}^{K} D^{-(m_j+n_k)} \leqq 1.$$

Thus there exists indeed a uniquely decipherable code with $M+N$ as set of codeword lengths for

$$X \times Y=\{\xi_1\eta_1, \xi_1\eta_2, \ldots, \xi_1\eta_K, \xi_2\eta_1, \xi_2\eta_2, \ldots, \xi_2\eta_K, \ldots, \xi_J\eta_1, \xi_J\eta_2, \ldots, \xi_J\eta_K\}.$$

If L is to be a measure of mean *lengths*, it is natural to require that

$$L(PQ, M+N; \phi)=L(P, M; \phi)+L(Q, N; \phi), \tag{7}$$

i.e.,

$$\phi^{-1}\left[\sum_{j=1}^{J}\sum_{k=1}^{K} p_j q_k \phi(m_j+n_k)\right]=\phi^{-1}\left[\sum_{j=1}^{J} p_j \phi(m_j)\right]+\phi^{-1}\left[\sum_{k=1}^{K} q_k \phi(n_k)\right]. \quad (8)$$

We call the properties (7) or (8) *additivity*. They are supposed for all positive integers m_j and n_k satisfying (6) and for all p_j, q_k $(j=1, 2, \ldots, J;\ k=1, 2, \ldots, K)$ satisfying (1) and

$$\sum_{j=1}^{J} p_j=1; \qquad p_j \geqq 0 \qquad (j=1, 2, \ldots, J). \quad (9)$$

The problem of *finding all additive* (7), *quasiarithmetic mean codeword lengths* (3) has not been solved before (cf. Campbell, 1966; Aczél, 1974). Instead, Campbell, 1966, has generalized the codeword lengths n_k $(k=1, 2, \ldots, K)$ so that they become arbitrary *real* numbers satisfying (2), and has solved (8) in this case. In this paper *we solve the* original *problem, with positive integer codeword lengths.* We *restrict ourselves to* $J=K=2$, thus making the result more general. This has also the advantage that, because of $D \geqq 2$, $m_1 \geqq 1$, $m_2 \geqq 1$, $n_1 \geqq 1$, $n_2 \geqq 1$, (6) *is always satisfied.*

2.

Theorem 1. *The arithmetic and the exponential mean codeword lengths* (4) *and* (5) *are the only quasiarithmetic mean codeword lengths* (3) *which are additive* (7) *with* $J=K=2$ *(for two-place distributions).*

Proof. For $J=K=2$, (7) or (8) can be written as

$$\begin{aligned}&\phi^{-1}[p_1 q_1 \phi(m_1+n_1)+p_1 q_2 \phi(m_1+n_2)+p_2 q_1 \phi(m_2+n_1)+p_2 q_2 \phi(m_2+n_2)]\\ &\quad=\phi^{-1}[p_1 \phi(m_1)+p_2 \phi(m_2)]+\phi^{-1}[q_1 \phi(n_1)+q_2 \phi(n_2)]\end{aligned} \quad (10)$$

where

$$p_1 \geqq 0,\ p_2 \geqq 0,\ p_1+p_2=1,\ q_1 \geqq 0,\ q_2 \geqq 0,\ q_1+q_2=1, \quad (11)$$

m_1, m_2, n_1, n_2 are positive integers.

Put into (10) $m_1=m_2=m$, $q_1=1-q$, $q_2=q$, in order to get

$$\phi^{-1}[(1-q)\phi(n_1+m)+q\phi(n_2+m)]=\phi^{-1}[(1-q)\phi(n_1)+q\phi(n_2)]+m \quad (12)$$

for all

$$q \in [0, 1]; \qquad n_1, n_2, m \text{ positive integers.} \quad (13)$$

We need the following

Lemma. *Let* ϕ, ψ *be continuous, strictly increasing functions defined on* $[1, \infty[$. *The equation*

$$\phi^{-1}[(1-q)\phi(n_1)+q\phi(n_2)]=\psi^{-1}[(1-q)\psi(n_1)+q\psi(n_2)] \quad (14)$$

holds for

$$n_1=1, \qquad n_2 \text{ arbitrary integer greater than } 1, \quad (15)$$

$q \in [0, 1]$ *arbitrary, if and only if there exist constants* $\alpha>0$, β *such that*

$$\psi(x)=\alpha\phi(x)+\beta \qquad \text{for all } x \in [1, \infty[. \quad (16)$$

Proof of the Lemma. The "if" part is obvious. In order to prove the "only if" part, put into (14) $n_1=1$, $n_2>1$. Denote

$$a_1=\phi(n_1)=\phi(1), \qquad a_2=\phi(n_2)-\phi(n_1)>0,$$
$$b_1=\psi(n_1)=\psi(1), \qquad b_2=\psi(n_2)-\psi(n_1)>0.$$

Then (14) goes over into

$$\phi^{-1}(a_2 q+a_1)=\psi^{-1}(b_2 q+b_1) \qquad (q\in[0,1]). \tag{17}$$

Now denote

$$y=b_2 q+b_1$$

and notice (cf. (15)) that y runs through $[\psi(1), \lim_{n\to\infty}\psi(n)[$ when $q\in[0,1]$, $n_2=2,3,\ldots$ (ψ, being increasing, has a finite or infinite limit as $n\to\infty$). So (17) goes over into

$$\psi^{-1}(y)=\phi^{-1}(A_2 y+A_1) \qquad (A_2>0)$$

or

$$\psi(x)=\alpha\phi(x)+\beta \qquad \text{for all } x\in[1,\infty[, \tag{16}$$

where $\alpha=1/A_2=b_2/a_2>0$, q.e.d.

Continuation of the proof of Theorem 1. Denote

$$\psi_m(x)=\phi(x+m) \qquad (x\in[1,\infty[\,; m=1,2,\ldots).$$

Then (12) goes over into

$$\phi^{-1}[(1-q)\phi(n_1)+q\phi(n_2)]=\psi_m^{-1}[(1-q)\psi_m(n_1)+q\psi_m(n_2)]$$

for all

$$q\in[0,1]; \qquad n_1, n_2 \text{ arbitrary integers.}$$

Thus, by the Lemma (the "constants" α, β in (16) will now depend upon m)

$$\phi(x+m)=\psi_m(x)=\alpha(m)\phi(x)+\beta(m) \qquad (x\in[1,\infty[\,; m=1,2,\ldots). \tag{18}$$

We distinguish two cases:

(i) $\alpha(m)\equiv 1$. Put then into (18) $x=n$ ($n=1,2,\ldots$), in order to get

$$\phi(m+n)=\phi(n)+\beta(m) \qquad \text{for all } m,n=1,2,\ldots. \tag{19}$$

Since the left hand side of (19) is symmetric in m and n, the right hand side has to be symmetric too,

$$\phi(n)+\beta(m)=\phi(m)+\beta(n)$$

and thus (put a constant for n) we have

$$\beta(m)=\phi(m)+c \qquad \text{for all } m=1,2,\ldots.$$

This transforms (18) into

$$\phi(x+m)=\phi(x)+\phi(m)+c \qquad (x\in[1,\infty[\,; m=1,2,\ldots). \tag{20}$$

(ii) If there exists an n_0 such that $\alpha(n_0)\neq 1$, then we derive from (18)

$$\phi(x+m+n)=\alpha(n)\phi(x+m)+\beta(n)=\alpha(m)\alpha(n)\phi(x)+\alpha(n)\beta(m)+\beta(n).$$

The left hand side is again symmetric in m and n, so the right hand side has to be symmetric too,

$$\alpha(n)\beta(m)+\beta(n)=\alpha(m)\beta(n)+\beta(m)$$

or, with $n=n_0$ $(\alpha(n_0)\neq 1)$, we have

$$\beta(m)=B[\alpha(m)-1].$$

Putting this into (18) we get

$$\phi(x+m)=\alpha(m)[\phi(x)+B]-B \tag{21}$$

or, with $x=n$ $(n=1, 2, \ldots)$ and again by symmetry,

$$\phi(m+n)+B=\alpha(m)[\phi(n)+B]=\alpha(n)[\phi(m)+B]. \tag{22}$$

By supposition, ϕ is strictly increasing, thus $\phi(n)\not\equiv -B$ and therefore (substitute into (22) $n=n_1$ with $\phi(n_1)\neq -B$)

$$\alpha(m)=a[\phi(m)+B].$$

Putting this into (21), we finally get

$$\phi(x+m)=a\,\phi(x)\,\phi(m)+aB\,\phi(x)+aB\,\phi(m)+aB^2-B. \tag{23}$$

Both (20) and (23) are of the form

$$\phi(x+m)=a\,\phi(x)\,\phi(m)+b\,\phi(x)+b\,\phi(m)+c \tag{24}$$

with

$$a=0,\ b=1 \qquad \text{in the case (i)}, \tag{25}$$

and (since ϕ is not constant on $[2, \infty[$)

$$a\neq 0,\ b=aB,\ c=aB^2-B \qquad \text{in the case (ii)}. \tag{26}$$

So (10) goes over into

$$\begin{aligned}\phi^{-1}(a[p_1\,\phi(m_1)+p_2\,\phi(m_2)][q_1\,\phi(n_1)+q_2\,\phi(n_2)]+b[p_1\,\phi(m_1)+p_2\,\phi(m_2)]\\ +b[q_1\,\phi(n_1)+q_2\,\phi(n_2)]+c)\\ =\phi^{-1}[p_1\,\phi(m_1)+p_2\,\phi(m_2)]+\phi^{-1}[q_1\,\phi(n_1)+q_2\,\phi(n_2)]\end{aligned} \tag{27}$$

with the variables restricted only by (11). If $m_1=n_1=1$ and $m_2, n_2=2, 3, \ldots$, then, as p_2 and q_2 run through $[0, 1]$,

$$u=p_1\,\phi(m_1)+p_2\,\phi(m_2), \qquad v=q_1\,\phi(n_1)+q_2\,\phi(n_2)$$

assume all values in $[\phi(1), \lim_{n\to\infty}\phi(n)[$ (ϕ being increasing, the finite or infinite limit $\lim_{n\to\infty}\phi(n)$ exists). Therefore (27) goes over into

$$\phi^{-1}(auv+bu+bv+c)=\phi^{-1}(u)+\phi^{-1}(v) \qquad \text{for all } u, v\in[\phi(1), \lim_{n\to\infty}\phi(n)[$$

and, with $x=\phi^{-1}(u)$, $y=\phi^{-1}(v)$,

$$\phi(x+y)=a\,\phi(x)\,\phi(y)+b\,\phi(x)+b\,\phi(y)+c \qquad \text{for all } x, y\in[1, \infty[. \tag{28}$$

For the constants in (28) we have one of the two cases (25) or (26). In the case (25), we get that f, defined by

$$f(x)=\phi(x)+c \qquad (x\in[1,\infty[), \tag{29}$$

satisfies the functional equation

$$f(x+y)=f(x)+f(y) \qquad \text{for all } x, y\in[1,\infty[. \tag{30}$$

With ϕ also f is increasing, and so, by Aczél, 1966 and Aczél – Baker – Djoković – Kannappan – Radó, 1971, $f(x)=\gamma x$ $(\gamma>0)$ and

$$\phi(x)=\gamma x+\delta \ (\gamma>0) \qquad \text{for all } x\in[1,\infty[. \tag{31}$$

In the case (26), we get that g defined by

$$g(x)=a[\phi(x)+B] \qquad (x\in[1,\infty[;\ a\neq 0) \tag{32}$$

$[g(m)=\alpha(m);\ m=1,2,\ldots]$ satisfies

$$g(x+y)=g(x)\,g(y) \qquad \text{for all } x, y\in[1,\infty[. \tag{33}$$

From (32) we see that g is strictly monotonic. On the other hand, as (33) shows, if there were an x_0 for which $g(x_0)=0$ then $g(x_0+y)=0$ for all $y\in[1,\infty[$ which would contradict the strict monotonicity of g. Thus g is (strictly monotonic and) nowhere zero and, according to the above references,

$$g(x)=D^{tx} \ (t\neq 0) \qquad \text{for all } x\in[1,\infty[$$

and

$$\phi(x)=\gamma D^{tx}+\delta \ (\gamma t>0) \qquad \text{for all } x\in[1,\infty[. \tag{34}$$

Putting (31) or (34) into (3) we get (4) and (5), respectively, and this concludes the proof of our Theorem 1.

On the other hand, *the functions given by* (31) *and* (34) *satisfy* (8) *for all* $J>1$, $K>1$ [and all m_j, n_k, p_j, q_k $(j=1,2,\ldots,J;\ k=1,2,\ldots,K)$ satisfying (6), (9) and (1)], thus *the arithmetic and exponential means* (4) *and* (5) *are always additive* (7).

3.

The property (12) or its generalization, both called *translativity*,

$$\phi^{-1}\left[\sum_{k=1}^{K} q_k\,\phi(n_k+m)\right]=\phi^{-1}\left[\sum_{k=1}^{K} q_k\,\phi(n_k)\right]+m \tag{35}$$

whenever (1) and (2) are satisfied, is quite important in itself. It serves (cf. Aczél, 1974) to prove certain uniqueness properties of the so called Shannon and Rényi entropies which are the lower bounds of our mean codeword lengths (4) and (5). We will come back to this later briefly. On the other hand, after allowing non-integer codeword lengths, Campbell, 1966 has deduced (31) and (34) from the translativity (12) alone. Thus, in the case of those generalized codeword lengths, the translativity (12) and the additivity (8) are equivalent. This is *not so* anymore *for the proper positive integer codeword lengths, not even* (35) *implies* (8) *or* (10)

[of course, (8) does imply (35)]. We will give, however, the general solution of the translativity equation (12) and we will show that (35) and (12) are equivalent.

If we have (12) for (13), then we can proceed, as in the proof of Theorem 1, till (24) with (25) or (26). From (24) we get then

$$\phi^{-1}(auv_m+bu+bv_m+c)=\phi^{-1}(u)+\phi^{-1}(v_m) \quad \text{for all } u\in[\phi(1), \lim_{n\to\infty}\phi(n)[, \quad \text{but only for all } v_m=\phi(m),\ m=1,2,\ldots. \tag{36}$$

However, (24) and (36) imply (35):

$$\phi^{-1}\left[\sum_{k=1}^{K} q_k\,\phi(n_k+m)\right]=\phi^{-1}\left[a\,\phi(m)\sum_{k=1}^{K} q_k\,\phi(n_k)+b\sum_{k=1}^{K} q_k\,\phi(n_k)+b\,\phi(m)+c\right]$$
$$=\phi^{-1}\left[\sum_{k=1}^{K} q_k\,\phi(n_k)\right]+m.$$

Thus (12) indeed implies (35) and, since (12) is the special case $K=2$ of (35), the equivalence of these two equations is established.

In order to solve (35) or (12) or, equivalently, (24) in the cases (25) and (26), introduce again the functions f and g defined by (29) and (32), respectively. They will satisfy now the functional equations

$$f(x+m)=f(x)+f(m) \qquad (x\in[1,\infty[;\ m=1,2,\ldots) \tag{37}$$

and

$$g(x+m)=g(x)\,g(m) \qquad (x\in[1,\infty[;\ m=1,2,\ldots), \tag{38}$$

respectively. Again ϕ and thus g can be strictly monotonic only if g is nowhere 0 [$g(x_0)=0$ would imply $g(x_0+m)=0$ for all $m=1,2,\ldots$].

It is easy to construct the general continuous strictly increasing solution of (37):

$$f(x)=\begin{cases}\text{arbitrary continuous increasing on } [1,2] \text{ but with } f(2)=2f(1),\\ f(x-k)+kf(1) \quad \text{for } x\in\,]k+1,k+2]\ (k=1,2,\ldots)\end{cases} \tag{39}$$

and the general continuous strictly monotonic (increasing, if $a>0$, decreasing if $a<0$) solution of (38)

$$g(x)=\begin{cases}\text{arbitrary strictly monotonic continuous on } [1,2] \text{ but with } g(2)=g(1)^2,\\ g(x-k)\,g(1)^k \quad \text{for } x\in\,]k+1,k+2]\ (k=1,2,\ldots).\end{cases} \tag{40}$$

So we have proved the following (the "if" part is easily checked).

Theorem 2. *The translativity equations* (12) *and* (35) *are equivalent. A function* ϕ *is continuous, strictly increasing and satisfies* (12) *or* (35) *if, and only if,*

$$\phi(x)=f(x)-c \qquad (x\in[1,\infty[)$$

or

$$\phi(x)=\frac{1}{a}\,g(x)-B \qquad (x\in[1,\infty[)$$

where $a\neq 0$, B, c *are constants and* f *and* g *are given by* (39) *and* (40) (g *increasing if* $a>0$ *and decreasing if* $a<0$).

4.

It is well known (Reza, 1961; Campbell, 1965, 1966; Aczél, 1974) that for all Q and N satisfying (1) and (2), respectively,

$$L(Q, N; \phi_0)=\sum_{k=1}^{K} q_k n_k \geqq -\sum_{k=1}^{K} q_k \log_D q_k, \qquad (0 \log 0 := 0) \tag{41}$$

and, for $t>-1$, $t\neq 0$,

$$L(Q, N; \phi_t)=\frac{1}{t}\log_D \sum_{k=1}^{K} q_k D^{t n_k} \geqq \frac{t+1}{t}\log_D \sum_{k=1}^{K} q_k^{1/(t+1)}, \qquad (0^\alpha := 0). \tag{42}$$

The right hand side of (41) is the *Shannon entropy* while on the right hand side of (42) *Rényi entropies* [of order $1/(t+1)$] stand.

One advantage of allowing non-integer codeword lengths is (Campbell, 1966), that the lower bounds at the right hand sides of (41) and (42) are actually attained. But even if we restrict ourselves to integer codeword lengths, it is easy to prove (Reza, 1961; Campbell, 1965; Aczél, 1974) that

$$L(Q, N^*; \phi_0)=\sum_{k=1}^{K} q_k n_k^* < -\sum_{k=1}^{K} q_k \log_D q_k + 1 \tag{43}$$

if

$$-\log_D q_k \leqq n_k^* < -\log_D q_k + 1 \qquad (k=1, 2, \ldots, K) \tag{44}$$

and, *for all* $t\neq -1$, $t\neq 0$,

$$L(Q, N^*; \phi_t)=\frac{1}{t}\log_D \sum_{k=1}^{K} q_k D^{t n_k^*} < \frac{t+1}{t}\log_D \sum_{k=1}^{K} q_k^{1/(t+1)} + 1, \tag{45}$$

if

$$-\log_D \left(q_k^{1/(t+1)} \Big/ \sum_{i=1}^{K} q_i^{1/(t+1)}\right) \leqq n_k^* < -\log_D \left(q_k^{1/(t+1)} \Big/ \sum_{i=1}^{K} q_i^{1/(t+1)}\right) + 1 \qquad (k=1, 2, \ldots, K). \tag{46}$$

We can get these from the transitivity of (4) and (5).

As to $t=-1$, it is easy to show that

$$\lim_{t\to -1}\left(\frac{t+1}{t}\log_D \sum_{k=1}^{K} q_k^{1/(t+1)}\right) = -\log_D \max(q_1, q_2, \ldots, q_K). \tag{47}$$

(Thus the right hand side of (47) is the *Rényi entropy of order* ∞.) So, by going over to the limit $t\to -1$ in (42), we get

$$L(Q, N; \phi_{-1})=-\log_D \sum_{k=1}^{K} q_k D^{-n_k} \geqq -\log_D \max(q_1, q_2, \ldots, q_K).$$

More generally, Campbell has recently proved (communication by correspondance) that *for all* $t \leqq -1$

$$L(Q, N; \phi_t)=\frac{1}{t}\log_D \sum_{k=1}^{K} q_k D^{t n_k} \geqq \frac{1}{t}\log_D \max(q_1, q_2, \ldots, q_K) \tag{48}$$

while (again for $t \leqq -1$)

$$L(Q, N^*; \phi_t) = \frac{1}{t} \log_D \sum_{k=1}^{K} q_k D^{t n_k^*} < \frac{1}{t} \log_D \max(q_1, q_2, \ldots, q_K) + 1 \tag{49}$$

if

$$n_{k_0}^* = 1, \quad n_k^* \geqq \log_D \frac{D-1}{D(K-1)} \quad (k \neq k_0) \qquad \textit{where } q_{k_0} = \max(q_1, q_2, \ldots, q_K). \tag{50}$$

(All these $\{n_1^*, n_2^*, \ldots, n_k^*\}$ do also satisfy (2).)

On the right hand sides of (43), (45) and (49), $+1$ can be replaced by *arbitrarily small* $+\varepsilon > 0$ if we encode *sequences of independent messages* consecutively.

The minimum or lower bound properties (41), (42) and (48) give interest to the following interpretation of quasiarithmetic mean codeword lengths, cf. Campbell, 1966. The function ϕ in (3) can be understood as *cost function*, $\phi(n)$ being the cost of using a codeword of length n. It is reasonable to suppose that ϕ is (strictly) increasing on the set of positive integers and then it can always be extended to a function strictly increasing and continuous on $[1, \infty[$. This is suitable because then ϕ^{-1} can be applied on more than a denumerable set.

Now the *average cost* of encoding the messages $Y = \{\eta_1, \eta_2, \ldots, \eta_k\}$ (probability distribution $Q = \{q_1, q_2, \ldots, q_k\}$) by a distribution $N = \{n_1, n_2, \ldots, n_k\}$ of codeword lengths is

$$C = \sum_{k=1}^{K} q_k \phi(n_k).$$

A coding problem of some interest is to *minimize the cost* C by an appropriate choice of the distribution N, subject to the costraint (2). Since $L(Q, N; \phi) = \phi^{-1}(C)$ and ϕ^{-1} is (continuous and) strictly increasing, an equivalent problem is to minimize the mean codeword length $L(Q, N; \phi)$.

There are multiplicative and additive constants contained in the cost functions as given by (31) and (34). (They do not influence the mean codeword lengths (4) and (5).) For calculating the average costs it may be advisable to *normalize* them. A possible normalization would assign unit cost to encoding a codeword of length 1 and zero cost in the (idealized) case of a codeword of length 0. Then we still have

$$\tilde{\phi_0}(n) = n \qquad (n = 0, 1, 2, \ldots) \tag{51}$$

but, instead of ϕ_t, we have

$$\tilde{\phi_t}(n) = \frac{D^{tn} - 1}{D^t - 1} \qquad (t \neq 0;\ n = 0, 1, 2, \ldots). \tag{52}$$

(One of the advantages is that $\tilde{\phi_0} = \lim_{t \to 0} \tilde{\phi_t}$ while $\phi_0 \neq \lim_{t \to 0} \phi_t$.) The inequalities (41), (42) and (48) show that *the average costs cannot be less than*

$$-\sum_{k=1}^{K} q_k \log_D q_k \quad (0 \log 0 := 0) \qquad \textit{for } t = 0, \tag{53}$$

$$\frac{\left(\sum_{k=1}^{K} q_k^{1/(t+1)}\right)^{t+1} - 1}{D^t - 1} \qquad \textit{for } t \neq 0,\ t > -1, \tag{54}$$

and

$$\frac{1-\max(q_1, q_2, \ldots, q_K)}{1-D^t} \quad \text{for } t \leqq -1, \tag{55}$$

whenever the cost functions are $\phi_t^{\sim}$, *given by*

$$\phi_0^{\sim}(x)=x \quad \text{and} \quad \phi_t^{\sim}(x)=\frac{D^{tx}-1}{D^t-1} \quad \text{for } t \neq 0 \ (x \in [1, \infty[)$$

[cf. (51), (52)] which, by Theorem 1 and the above, are the normalized forms of the cost functions in all cases of additive mean codeword lengths (8).

The inequalities (44), (46) and (50) show with what N we *get near to the lower bounds* (53), (54), *and* (55) *of the average costs*, respectively.

References

Aczél, J.: Lectures on Functional Equations and Their Applications. New York-London: Academic Press 1966

Aczél, J.: On Shannon's Inequality, Optimal Coding, and Characterizations of Shannon's and Rényi's Entropies. To be published in Symposia Mathematica, Ist. Naz. Alta Mat., Roma. New York: Academic Press 1974

Aczél, J., Baker, J.A., Djoković, D.Ž., Kannappan, Pl., Radó, F.: Extensions of Certain Homomorphisms of Subsemigroups to Homomorphisms of Groups. Aequationes Math. **6**, 263–271 (1971)

Campbell, L.L.: A Coding Theorem and Rényi's Entropy. Information and Control **8**, 423–429 (1965)

Campbell, L.L.: Definition of Entropy by Means of a Coding Problem. Z. Wahrscheinlichkeitstheorie verw. Gebiete **6**, 113–118 (1966)

Reza, F.M.: An Introduction to Information Theory. New York-Toronto-London: McGraw-Hill 1961

J. Aczél
Faculty of Mathematics
University of Waterloo
Waterloo, Ont., Canada

(Received February 27, 1973)