Table 5. Comparison of Var(p) and Var(p)

$n = 500$

Case	Var ($\tilde{p}$)	Var ($\hat{p}$)	$\frac{\text{Var}(\tilde{p})}{\text{Var}(\hat{p})}$
1	0.006979	0.003643	1.92
2	0.003321	0.0008726	3.81
3	0.002651	0.00009919	26.74
4	0.001773	0.0004761	3.72
5	0.002551	0.003193	0.80
6	0.003192	0.003985	0.80

Using (9) the variances oi the estimators using suboptimum quantiles were computed for the six cases and are given in Table 5. Var($\hat{p}$) is also shown again for comparison.

As might be expected, the results for the six cases vary. However, only for Case 3 is the result quite poor, quantitatively speaking, and this is due to the fact that opt. $s = 0.0496$ is smaller than all four of the orders of the suboptimum quantiles. Nevertheless, even for this case, since the standard deviation of $\tilde{p}$ is 0.0515, there is considerable probability that a given p will at least provide a basis for the qualitative conclusion that p is small. It should also be noted that Var($\tilde{p}$), in common with Var($\hat{p}$), Var($\hat{p}_1$) and Var($\hat{p}_2$), is inversely proportional to n. Thus, for sufficiently large n, even in Case 3 the suboptimum estimator of p will also be quantitatively significant. And large sample sizes are not uncommon in space experiments.

E. Instantaneously Synchronizable Block Code Dictionaries

J. J. Stiffler

1. Introduction

Let D(N) be a block code dictionary consisting of N-symbol words defined over an r-symbol alphabet. The code is said to be instantaneously synchronizable if the knowledge of any consecutive $2N$-1 symbols in a sequence of code words is sufficient to determine synchronization (i.e., to establish which symbol of each word is the initial symbol). Comma-free codes (Ref. 19) are instantaneously synchronizable as well as prefix codes (Ref. 20). These and other constructions will be discussed, and the number of words in the dictionaries for a given N will be compared for the various techniques.

2. Several Constructions

Comma-free codes have the property that no N-tuple consisting of the last k symbols of one code word followed by the first N-k symbols of another (not necessarily different) word can be a code word for any k, $1 \leqslant k \leqslant N - 1$. Such code dictionaries are clearly instantaneously synchronizable. They have two disadvantages, however. First, it is necessary to observe the symbols in groups of N and to determine if they are code words or not. This can be time consuming if N is large. Second, the encoding and decoding procedures for comma-free codes are generally quite complex.

The first of these disadvantages is effectively nonexistent in a more restricted class of codes: the prefix codes. All words in a prefix code dictionary begin with the same m-symbol prefix and are so constrained that, regardless of the word sequence, this prefix can occur *only* at the beginning of a code word. Thus, when the m-symbol prefix is observed it may immediately be concluded that the initial code word symbols have been found.

Another construction which results in an even more restricted class of codes, but one for which both of the above objectives to comma-free codes are to some extent overcome, is the following: An m-symbol prefix is reserved as a marker or "comma" as before, but now the constraint length is shortened from N to some smaller integer n. The word length N is taken to be the number of symbols separating the initial symbols of two successive occurrences of the comma. Thus, if the comma is transmitted after every k^{th} n-tuple, $N = m + kn$. The n-tuples must be so constrained that no m-tuple formed from any concatenation of allowable n-tuples or from any n-tuple and the comma itself can be the comma. Since the n-tuples can be selected independently, the encoding and decoding processes can be kept within reasonable bounds even for large values of N.

Some special cases of this last construction are: (1) $m = 1$; one symbol must be set aside for a comma and not used in any of the code words. There are $(r-1)^n$ possible n-tuples which may be used. The encoding is extremely simple in this case. (2) $m = n$; one n-tuple is used as a comma. The number of n-tuples which may be used as code words has been determined in Ref. 21. When n is even it is possible to use

$$r^n\left(1 - \frac{r}{r-1}\frac{1}{r^{n/2}} + \frac{1}{(r-1)r^n}\right) \tag{1}$$

such n-tuples, and when n is odd, the number becomes

$$r^n\left(1 - \frac{2r-1}{r-1_r}\frac{1}{r^{\frac{n+1}{2}}} + \frac{1}{(r-1)r^n}\right) \tag{2}$$

(3) $m = 2n$; two consecutive n-tuples are used to define the comma. If the two-n-tuples are taken to be

$$a\,a \cdots\cdots\cdots\cdot a\,a$$

and

$$a\,a \cdots\cdots\cdots\cdot a\,b$$

respectively, where a and b are any two of the r symbols of the alphabet, it is evident that only one n-tuple need be excluded from the set used to transmit data; viz the n-tuple $aa \cdots\cdots\, a$. This is apparent when it is observed that if the $2n$-tuple

$$\underbrace{a\,a \cdots\cdot a}_{n\text{-symbols}} \quad \underbrace{a\,a \cdots\cdot a\,b}_{n\text{-symbols}}$$

is to be found within the data sequence, then the last ℓ symbols of one n-tuple, followed by a second n-tuple which is in turn followed by the first n-ℓ symbols of a third n-tuple must result in this sequence of symbols. But either $1 \leqslant \ell \leqslant n-1$, in which case the middle n-tuple must be the prohibited $aa \cdots\cdot a$, or else $\ell = 0$, and this n-tuple must occur initially. Hence, $r^n - 1$ n-tuples can be used to form the data portion of the code words. The encoding and decoding procedures here are nearly as simple as those encountered when a single symbol is used as a comma. It is only necessary to exclude the n-tuple $aa \cdots\cdot a$ from the data sequence (and then only when it occurs as a subword, not as an overlap of two such subwords). When this is done, the n-tuples may be transmitted directly.

3. Comparison of Dictionary Sizes for Large N

All instantaneously synchronizable codes clearly must have the property that if an N-tuple is in the dictionary then none of its cyclic permutations can be in the dictionary. If an N-tuple has a periodicity less than N it is a cyclic permutation of itself and cannot be in the dictionary. Thus no instantaneously synchronizable dictionary can have more than

$$W(N, r) = \frac{1}{N}\sum_{d/\mathrm{n}} \mu(d)\; r^{n/d} \tag{3}$$

words where $\mu(d)$ is the Mobius function defined as follows: Let $d = p^{\alpha}_{1}{}_1 p^{\alpha}_{2}{}_2 \cdots\, p^{\alpha}_{l}l$ be the factorization of d into prime powers. Then

$$\mu(d) = \begin{cases} 1 & \text{if } d = 1 \\ (-1)^l & \text{if } \alpha_1 = \alpha_2 = \cdots = \alpha_l = 1 \\ 0 & \text{if } \alpha_i > 1 \text{ for any } i \end{cases} \tag{4}$$

(cf Ref. 19). It is known that comma-free codes can be constructed achieving this bound for odd lengths N. When N is even comma-free codes containing at least $2/e\ r^N/N$ words can be constructed (Ref. 22).

It is shown in Ref. 20 that the number of words $P(N, r)$ in a prefix dictionary is for $r = 2$ and asymptotically with N,

$$\frac{1}{e}\frac{2^N}{N} \geqslant P(N, 2) \geqslant \frac{1}{2\log_2 e}\frac{2^N}{N}. \tag{5}$$

This result can be generalized, for arbitrary r, to establish that

$$\frac{1}{e}\frac{r^N}{N} \geqslant P(N, r) \geqslant \frac{1}{(r-1)\,r^{\frac{1}{r-1}}\log_r e}\frac{r^N}{N}. \tag{6}$$

Asymptotic bounds on the number of words $C(N, r)$ in large comma dictionaries are also readily obtained. When $m = 1$, the number of words is $C_1(N, r) = (r-1)^{N-1}$ regardless of the value of n. For purposes of comparison this may be written

$$C_1(N, r) = \frac{N\left(1 - \frac{1}{r}\right)^N}{r-1}\frac{r^N}{N}. \tag{7}$$

When $m = n$ we have, for even n,

$$C_n(N, r) = \left[r^n\left(1 - \frac{r}{r-1}\frac{1}{r^{n/2}} + \frac{1}{(r-1)r^n}\right)\right]^{\frac{N}{n}-1} \tag{8}$$

since $N = (k+1)n$. Thus,

$$C_n(N, r) = \frac{r^N}{r^n}\left[1 - \frac{r}{r-1}\frac{1}{r^{n/2}}\left(1 - \frac{1}{r^{n/2+1}}\right)\right]^{\frac{N}{n}-1}. \tag{9}$$

We define a new function of N, $a = a(N)$ by the relationship

$$n\, r^{n/2} = N/a. \tag{10}$$

Then,

$$C_n(N,r) = \frac{r^N}{r^n}\left[1 - \frac{r}{r-1}\frac{an}{N}\left(1 - \frac{an}{rN}\right)\right]^{N/n-1} \quad (11)$$

which for large N yields

$$C_n(N,r) \backsim \frac{r^N}{r^n}\,exp\left\{-\frac{r}{r-1}a\right\} =$$

$$r^N\,exp\left\{-\left(n\log_e r + \frac{r}{r-1}\frac{N}{nr^{n/2}}\right)\right\}. \quad (12)$$

It is easily verified that

$$exp\left\{-\left(n\log_e r + \frac{r}{r-1}\frac{N}{nr^{r/2}}\right)\right\}$$

monotonically increases with n for

$$\log_e r < \frac{r}{r-1}\frac{N}{nr^{n/2}}\left(\frac{1}{n} + \frac{1}{2}\log_e r\right) \quad (13)$$

and is a monotonically decreasing function of n otherwise. Hence, a maximum occurs when

$$\log_e r = \frac{r}{r-1}\,a\left(\frac{1}{n} + \frac{1}{2}\log_e r\right) \sim \frac{r}{r-1}\frac{a}{2}\log_e r. \quad (14)$$

Thus, for the dictionary size to be maximum

$$a = \frac{N}{nr^{n/2}} \approx \frac{2\,(r-1)}{r}$$

and, for large n, is essentially constant. To maximize the size of the dictionary, then, for a given constraint length n it is required that

$$N = 2(r-1)n\,r^{(n/2-1)}. \quad (15)$$

Since

$$r^{n/2} < nr^{n/2} < r^{n/2(1+\varepsilon)} \quad (16)$$

for sufficiently large n, and for any $\epsilon > 0$, it follows that, asymptotically, when N satisfies Eq. (15),

$$\left(\frac{rN}{2(r-1)}\right)^{\frac{2}{1+\varepsilon}} < r^n < \left(\frac{rN}{2(r-1)}\right)^2 \quad (17)$$

and from Eq. (12),

$$C_n\left[N = 2\,(r-1)\,nr^{(n/2-1)}, r\right] \sim \left[\frac{2(r-1)}{er}\right]^2\frac{r^N}{N^2}. \quad (18)$$

When n is odd a similar argument establishes that

$$C_n\left[N = \frac{2(r-1)}{2r-1}\,nr^{\frac{n+1}{2}}, r\right] \backsim r\left[\frac{2(r-1)}{e(2r-1)}\right]^2\frac{r^N}{N^2}. \quad (19)$$

Finally, when $m = 2n$,

$$C_{2n}(N,r) = (r^n - 1)^{\frac{N}{n}-2}$$

$$= \frac{r^N}{r^{2n}}\left(1 - \frac{1}{r^n}\right)^{\frac{n}{N}-2}. \quad (20)$$

Defining

$$nr^n = \frac{N}{a}, \quad (21)$$

we find that

$$C_{2n}(N,r) \sim \frac{r^N}{r^{2n}}\,e^{-a} = r^N\,exp\left[-\left(\frac{N}{nr^n} + 2n\log_e r\right)\right] \quad (22)$$

The expression $N/(nr)^n + 2n \log_e r$ is minimum when a $\approx$ 2. Further, since for sufficiently large n,

$$r^n < \frac{N}{2} < r^{n(1+\varepsilon)} \quad (23)$$

when $N = 2nr^n$, and for any $\epsilon > 0$, it follows that

$$\left(\frac{N}{2}\right)^{\frac{1}{1+\varepsilon}} < r^n < \frac{N}{2} \quad (24)$$

and

$$C_{2n}(N = 2nr^n, r) \sim \frac{4}{e^2}\frac{r^N}{N^2}. \quad (25)$$

It should be emphasized that these asymptotic results [Eqs. (18), (19) and (25)] are not valid for every value of N, but rather only when N is the proper multiple of nr^n. Thus, for arbitrarily large values of N, the code dictionary may be much smaller than r^N/N^2. Eq. (25) holds for large integer values of n and Eq. (18) for large even integers n. Eq. (19), on the other hand, is valid only for certain odd values of n, since

$$N = \frac{2(r-1)}{2\,r-1}\,nr^{\frac{n+1}{2}}$$

must also be an integer.

In conclusion, the comma codes discussed here are always less efficient than prefix codes and comma-free codes, the former having, asymptotically with the word length N, only $1/N^{th}$ as many code words as the latter. Nevertheless, the simplicity of encoding and decoding using these codes may more than make up for the defect.

References

1. Moore, E. F., *Sequential Machines, Selected Papers,* Addison-Wesley, Reading, Mass., 1964.
2. Golomb, S. W., and Welch, L. R., *Non-Linear Shift Register Sequences,* Technical Memorandum No. 20-149, Jet Propulsion Laboratory, Pasadena, California, 1957.
3. Easterling, M., "A Skin-Tracking Radar Experiment Involving the COURIER Satellite," Trans. *IRE, Space Electronics and Telemetry,* Vol. SET-8 1962, Pp. 76–84.
4. Bose, R. C., and Ray-Chaudhuri, D. K., "On a class of Error-correcting Binary Group Codes," *Information and Control* 3, 1960, pp. 68–79.
5. Bose, R. C., and Ray-Chaudhuri, D. K., "Further results on Error-correcting Binary Group Codes," *Information and Control* 3, 1960, pp. 279–290.
6. Chien, R. T., "Cyclic Decoding Procedures for Bose-Chaudhuri-Hocquenghem Codes," *IEEE Profession Group on Information Theory, 1964,* pp. 357–362.
7. Peterson, W. W., *Error-Correcting Codes,* John Wiley and Sons, New York, 1961, p. 176.
8. Massey, J. L., private communication, 1965.
9. Solomon, G., SPS 37-29, Vol. 4, Jet Propulsion Laboratory, Pasadena, 1965, pp. 296–298. See also Fredricksen, H. N. Ibid, p. 299.
10. Forney, D., private communication, 1965.
11. Daykin, G. E., "Distribution of Bordered Persymmetric Matrices in a Finite Field," *Journal für die reine und angewandte Mathematik* 203, 1960, pp. 47–54.
12. Berlekamp, E. R., *The Enumeration of Matrices by Rank,* Bell Telephone Laboratories, Murray Hill, N. J., 1963.
13. Berlekamp, E. R., "The Distribution of Matrices Resulting from Newton's Identities in a Field of Characteristic Two," in *Notes on System Theory, VII,* ERL, University of California, Berkeley, 1965.
14. Trench, W. F., "An Algorithm for the Inversion of Finite Toeplitz Matrices," *Journal of the Society of Industrial and Applied Mathematics,* Vol. 12, 1965, pp. 515–522.
15. *Quantile System,* Space Program Summaries No. 37-32, Vol. 3, Jet Propulsion Laboratory, Pasadena (to be published).
16. Eisenberger, I., and Posner, E. C., *Systematic Statistics Used for Data Compression in Space Telemetry,* Technical Report No. 32-510, Jet Propulsion Laboratory, Pasadena, Oct. 1, 1963.
17. Eisenberger, I., *Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles,* I, Technical Report, Jet Propulsion Laboratory, Pasadena (to be published).
18. Eisenberger, I., *Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles,* II, Technical Report, Jet Propulsion Laboratory, Pasadena (to be published).

References (Cont'd)

19. Cramér, H., *Mathematical Methods of Statistics,* Princeton University Press, Princeton, N. J., 1946, pp. 367–370.

20. Golomb, S. W., Gordon, B., and Welch, L. R., "Comma-Free Codes," *Canadian Journal of Mathematics,* Vol. 10, No. 2, 1958, pp. 202–209.

21. Gilbert, E. N., "Synchronization of Binary Messages," IRE Transactions on Information Theory, IT-6, 1960.

22. Kendall, W. B., "Optimum Synchronizing Words for Fixed Word-Length Code Dictionaries" (to be published).

23. Jiggs, B. H., "Recent Results in Comma-Free Codes," *Canadian Journal of Mathematics.*