"These approximations are valid for any positive value of t so long as n is sufficiently large."

A Note on the Computation of the Hamming Bound

by
Efren F. Abaya, Ph.D.*

ABSTRACT

This paper derives a simple approximation to a binomial sum occurring in the Hamming Bound. The approximation is easy to compute and quite accurate, even for modest values of the parameters involved. The approximation permits easy calculation of the minimum number of parity bits required for t error correction.

INTRODUCTION

In coding theory, a lower bound on the number of parity bits r necessary to correct up to t errors in a binary block code of length n (where n is greater than t) is given by the so-called Hamming Bound [1]

$$2^r \geq \sum_{k=\phi}^{t} C(n, k) \tag{1}$$

which involves the combinations C (n, k) of n objects taken k at a time. The sum above is quite difficult to calculate even for modest values of n and t.

The purpose of this short paper is to show that the sum on the right hand side of the inequality can be approximated quite well by an expression that is much simpler to compute. Specifically

$$\sum_{k=\phi}^{t} C(n, k) \approx n^t/t! \tag{2}$$

In this approximation, the ratio of the two values approaches unity as n becomes large. Moreover, the minimum number of parity bits r which is necessary for t error correction is approximated by

$$r \approx t[\log_2(n)] - \log_2 (t!) \tag{3}$$

within $\pm$ 1 of the true value.

These approximations are valid for any positive value of t so long as n is sufficiently large.

*Professor of Electrical Engineering, U.P. College of Engineering, Diliman, Quezon City, Philippines.

Shannon's Second Coding Theorem (as applied to a binary channel) states that it is possible to transmit at a rate arbitrarily close to (but strictly less than) the channel capacity with an arbitrarily low (but positive) probability of error. This is made possible by very long error correcting block codes. It is difficult to appreciate the largeness of the numbers involved without calculating a few of them. It is, however, a tremendous computational effort to calculate (1) for values of t which may range in the hundreds, and values of n in the hundreds of thousands. The approximation in (3) gives a relatively simple way to estimate these numbers.

Even for modest values of n (say, a few hundred to a few thousand) and small values of t which are encountered in practice, equation (3) gives a quick and easily accessible estimate of the number of parity bits required. The equation also permits a quick assessment of the overhead involved in correcting t errors through the ratio

$$\frac{r}{n} \approx t\left(\frac{\log_2 n}{n}\right)$$

which exhibits the well-known increasing efficiency of long block codes.

DEVELOPMENT

We first derive an upper bound to the sum in (1) by dominating it with a geometric sum. Since

$$C(n,k) = \frac{n!}{(n-k)!\,k!} \leq \frac{n^k}{k!}$$

it follows that

$$\begin{aligned} \sum_{k=\phi}^{t} C(n,k) &\leq \sum_{k=\phi}^{t} n^k/k! \\ &\leq \frac{n^t}{t!} \sum_{k=\phi}^{t} (t/n)^k \end{aligned} \tag{4}$$

Taking the sum of the geometric series in (4) gives a tight upper bound. For our purposes, however, a further simplification is useful. Since t is less than or equal to n, we have

$$\begin{aligned} \sum_{k=\phi}^{t} (t/n)^k &\leq 1 + \sum_{k=1}^{t} (t/n) \\ &= 1 + t^2/n \end{aligned}$$

This yields the following upper bound valid so long as t is less than or equal to n:

$$\sum_{k=\phi}^{t} C(n,k) \leq \frac{n^t}{t!}\left(1 + \frac{t^2}{n}\right) \tag{5}$$

Next, we develop a lower bound to the sum in (1). Note that

$$\sum_{k=\phi}^{t} C(n,k) \cdot t!$$

can be taken as a polynomial of degree t. The coefficients starting from the highest degree term of the first ten polynomials are listed in Table 1.

In the Appendix, it is shown that the first three terms of the polynomial are as follows:

$$\sum_{k=\phi}^{t} C(n,k) \,.\, t! = n^t - \frac{t(t-3)}{2} n^{t-1} + \left(\frac{t^4}{8} - \frac{11t^3}{12} + \frac{23t^2}{8} - \frac{25t}{12}\right) n^{t-2} + \ldots.$$

Table 1. Coefficients of $\sum_{k=\phi}^{t} C(n,k) \cdot t!$

t	Coefficients*								
1	1	1							
2	1	1	2						
3	1	0	5	6					
4	1	−2	11	14	24				
5	1	−5	25	5	94	120			
6	1	−9	55	−75	304	444	720		
7	1	−14	112	−350	1,099	364	3,828	5,040	
8	1	−20	210	−1064	3,969	−4,340	15,980	25,584	40,320
9	1	−27	366	−2646	12,873	−31,563	79,064	34,236	270,576	362,880

*The leftmost column corresponds to the highest degree term n^t.

The coefficient of the third term is non-negative for any positive value of t. Therefore, so long as n is large enough

$$\sum_{k=\phi}^{t} C(n,k) \geq \frac{n^t}{t!}\left(1 - \frac{t(t-3)}{2n}\right)$$

Finally,

$$\sum_{k=\phi}^{t} C(n,k) \geq \frac{n^t}{t!}\left(1 - \frac{t^2}{2n}\right) \tag{6}$$

APPROXIMATIONS

From the upper bound (5) and lower bound (6), we have

$$\left| \sum_{k=\phi}^{t} C(n,k) - \frac{n^t}{t!} \right| \leq \frac{n^t}{t!}\left(\frac{t^2}{n}\right) \tag{7}$$

Therefore, if n is sufficiently larger than t, the right hand side below is a simple approximation to the sum in the Hamming bound:

$$\sum_{k=\phi}^{t} C(n,k) \approx n^t/t! \tag{8}$$

From (7) the relative error of the approximation is o (t^2/n) which is small if n is sufficiently large.

The minimum number r of parity bits is given by the logarithm (base 2) of the sum in 1. We can derive a useful expression that gives r directly by manipulating (5) and (6) to give

$$\log_2\left(1 - \frac{t^2}{2n}\right) \leq \log_2 \sum_{k=\phi}^{t} C(n,k) - \log_2 \frac{n^t}{t!} \leq \log_2\left(1 + \frac{t^2}{n}\right) \tag{9}$$

This suggests that if n is greater than t^2 then approximately

$$r \approx t \cdot \log_2 n - \log_2 t! \tag{10}$$

with a difference of not more than one bit.

If t is large, Stirling's Approximation can be used for the last term. Therefore, these approximations are easy to calculate.

REFERENCE

[1] RICHARD W. HAMMING, *Coding And Information Theory,* N. J.: Prentice Hall, Inc., 1980.

APPENDIX

Two properties of permutations and a sum involving combinations are proven below.

The permutation P (n, k) of n objects taken k at a time is equal to $n!/(n-k)!$. For a fixed k, this expression can be viewed as a polynomial of degree k in the variable n. For example,

$P(n, 1) = n$

$P(n, 2) = n^2 - n$

$P(n, 3) = n^3 - 3n^2 + 2n$

$P(n, 4) = n^4 - 6n^3 + 11n^2 - 6n$

$P(n, 5) = n^5 - 10n^4 + 35n^3 + 24n$

Lemma 1

The first three terms of P(n, k) are given by

$$P(n, k) = n^k - \frac{k(k-1)}{2} n^{k-1} + \left(\frac{k^4}{8} - \frac{5k^3}{12} + \frac{3k^2}{8} - \frac{1k}{12}\right) n^{k-2} - \ldots$$

Proof. The proof proceeds by induction. The property can easily be verified for k = 1, 2 and 3. Now assume that the property is true for k−1. From the recursive property of permutations

$$P(n, k) = (n-k+1) \cdot P(n, k-1)$$

a straightforward calculation yields the first three terms given above. This completes the proof.

Lemma 2

For t less than or equal to n,

$$\sum_{k=\phi}^{t} C(n, k) \cdot t! = n^t - \frac{t(t-3)}{2} n^{t-1} + \left(\frac{t^4}{8} - \frac{11t^3}{12} + \frac{23t^2}{8} - \frac{25t}{12}\right) n^{t-2} + \ldots.$$

Proof. The proof proceeds by induction. The given equation can be easily checked for t = 1 and 2. Now assume that the property is true for t−1. Observe that

$$\sum_{k=\phi}^{t} C(n, k) \cdot t! = t \sum_{k=\phi}^{t-1} C(n, k) \cdot (t-1)! + P(n, t)$$

Invoking Lemma 1 and calculating the right hand side yields the first three terms claimed above. This completes the proof.