19 Communication theory

19.1 Random codes (5 units)

Background material for this project is given in the Part II course Coding and Cryptography.

The (binary) Hamming space $\{0,1\}^n$ consists of all possible n-tuples of 0s and 1s. We define a *code* C of length n and size r to be a subset C of the Hamming space $\{0,1\}^n$ with r elements. The *Hamming distance* between two elements $\mathbf{x} = (x_i)$, $\mathbf{y} = (y_i)$ of $\{0,1\}^n$ is the number of places in which $\mathbf{x}$, $\mathbf{y}$ differ. The *minimum distance* of a code C is the minimum Hamming distance $d(C)$ between distinct elements of C. The *information rate* of C is $\frac{1}{n}\log_2 r$. We define the *error-control rate* to be $(d-1)/n$ (note: elsewhere it is often defined as d/n).

In this project we investigate how high the information rate of a randomly-generated code can be, subject to constraints on its error-control rate. In the first study, any randomly chosen code may be considered, whereas in the second study only randomly generated linear codes are allowed. In each study we try two approaches: specify the information rate of the random code and see what error-control rate can be achieved, or specify the error-control rate and see what information rate can be achieved.

Question 1 Write a procedure to find the minimum distance of a code. Use your procedure to write a program which generates random codes of length n and size r and then computes the minimum distance.

Run your program several times with various values of n and r, for each choice finding the best (i.e., largest) d that you can.

Question 2 Now generate codes of length n and minimum distance d by starting with an initial code vector, say $(0,\ldots,0)$ and randomly generating further vectors, adding a new vector to the code if it has distance at least d from all the vectors already in the code.

Run your program several times for each choice of parameters n and d, finding the best (i.e., largest) r that you can.

Question 3 Take the output from the two previous questions and plot the corresponding points on a graph with information rate and error-control rate as the two axes. Comment on your results.

We call a code *linear* if it forms a subspace of the Hamming space, regarded as a vector space over the field F of 2 elements. The *weight* $w(\mathbf{x})$ of a vector $\mathbf{x}$ is the number of non-zero components, that is, the Hamming distance $d(\mathbf{x},\mathbf{0})$. The minimum distance of a linear code is just the minimum non-zero weight. The *rank* k of a linear code is the dimension of the code as a subspace, and the size of a linear code is $r = 2^k$.

Question 4 Write a procedure to find the minimum non-zero weight of a code generated over F by a set $g_1, \ldots, g_k$ of k generators. Use your procedure to find linear codes of given length n and either given rank k or given minimum distance d by considering random sets of generators. As before, run your programs several times to plot the information and error-control rates and comment on the results.

Question 5 Comment on your results in relation to known constraints on the design of codes and the Shannon coding theorems.

Comment on these methods as a way of designing effective error-control codes.

References

[1] C.M. Goldie and R.G.E. Pinch, *Communication theory*, CUP, 1991.

