# MATHEMATICS OF SQUARE CONSTRUCTION

CHRIS LONG
Bridgewater, New Jersey

## Theoretical Results

In the May 1992 issue of **Word Ways**, Ross Eckler raised the question of the minimum number of words needed before you'd expect to be able to form a word square, and he termed this number the **support**. A simple estimate may be obtained by assuming that the frequencies for each letter in each word are probabilistically independent (e.g., the fraction of words that have an A in the first position times the fraction that have B in the second position equals the fraction that have both an A in the first position and a B in the second position), and also that letter frequencies are positionally independent (e.g., the fraction of words that have an M in the first position equals the fraction that have M in the second position). For English words this is clearly not true (e.g., a Q is almost always followed by a U), but this estimate should still be reasonably accurate.

Under this assumption, we may apply Bayes' theorem from probability and compute that the expected number of fills for a form F which is a double n-square is approximately

$$E(F) = W^{2n}p^{n^2} \quad \text{where } p = f(a)^2 + \ldots + f(z)^2$$

W is the number of different n-letter words in our word list, and f(*) is the frequency of the letter * in our word list. Computations for various word-lists reveal that p is about 1/15.8 with very little variation. We may now obtain an estimate of the support by defining the support of F, denoted by S(F), to equal that value of W which yields an E(F) of one. Setting E(F) = 1 in the above formula, we solve for W, finding that

$$W = p^{-n/2} \quad \text{or } S(F) = 15.8^{n/2}$$

when F is a double n-square.

To compute S(F) for a regular n-square we must be careful not to include any redundancies. It turns out that

$$E(F) = W^{n}p^{n(n-1)/2} \text{ and therefore } S(F) = p^{-(n-1)/2}$$

where p = 1/15.8 as before. This is interesting, since, assuming that difficulty of square construction is directly proportional to the size of the word lists required, this implies that it is about as difficult to construct a regular square of order n as it is to construct a double square of order n-1. This was the opinion of Palmer Peterson of the National Puzzlers' League, one of the greatest formists of all time. Letting $F_1$ be an n-square and $F_2$ be a double n-square, quick work with a calculator gives

| $n$ | $S(F_1)$ | $S(F_2)$ |
|---|---|---|
| 4 | 63 | 250 |
| 5 | 250 | 992 |
| 6 | 992 | 3,944 |
| 7 | 3,944 | 15,678 |
| 8 | 15,678 | 62,320 |
| 9 | 62,320 | 247,718 |
| 10 | 247,718 | 984,658 |
| 11 | 984,658 | 3,913,938 |
| 12 | 3,913,938 | 15,557,597 |

Comparing these figures to the number of words of the appropriate size in the typical unabridged dictionary goes a long way toward explaining why Johnny hasn't constructed a 10-square or a double 9-square using only dictionary words.

More accurate comparisons result if we drop the condition that letter frequencies are independent of position. A simple computer program may be written to implement this; it states that a more accurate estimate for a 9-square F is S(F) = 75642 and for a 10-square F is S(F) = 256945, so the above figures tend to be on the low side.

## Experimental Results

Several sample runs were done to compute the number of single and double squares for the 4-square and 5-square cases. Random subsets of a master wordlist were generated by choosing each word with a selection probability q, and then two computer runs were done using each list, once for the regular squares, and once for the double squares. In the tables on the next page, the selection probability is given first, followed by the size of the wordlist which resulted. Then the number of squares produced from this list is given for both forms (the number of double squares found was actually twice the given number, since each square may be reflected around the diagonal). Two support estimates are then given. $E_1$ was generated by a program which looks at the wordlist and computes the frequencies of letters that occur in various positions and obtains an estimate based on Bayes' theorem; $E_2$ was generated by using the heuristic formulas derived in the first section, i.e., $E(F) = (W/S)^n$ for the single square and $E(F) = (W/S)^{2n}$ for the double square, by setting E(F) equal to the number of fills actually found and then solving for S. For this estimate the number of double squares found needed to be doubled (no pun intended), as the heuristic formula was derived under the assumption that reflective pairs are different fills.

Note that at the q = .30 level the number of double 4-squares found exceeded that of the single 4-squares, and that at the q = .40 level the number of double 5-squares found exceeded that of the single 5-squares, which verifies a somewhat unexpected prediction of the model. The relative stability of the $E_2$ estimates implies that the heuristic formula appears to be a good model of reality. The number of n-squares appears to go as the nth power of the number of words and the number of double n-squares appears to go as the 2nth power of the number of words.

How many 4-squares and 5-squares are out there? The total number of 4-letter words in my list is 7364, so extrapolating based on the above data yields the surprisingly large estimates of 45 million 4-squares and 11 billion double 4-squares. Furthermore, doing the same computations for the 5-letter case, there are 16536 words in my list, yielding the estimates of 240 million 5-squares and 37 billion double 5-squares.

$n = 4$

| $q$ | Words | Singles | $E_1$ | $E_2$ | Doubles | $E_1$ | $E_2$ |
|---|---|---|---|---|---|---|---|
| .10 | 726 | 3,587 | 96.36 | 93.81 | 95 | 358.23 | 376.78 |
| .15 | 1,041 | 16,671 | 93.85 | 91.61 | 1,780 | 355.67 | 374.56 |
| .20 | 1,445 | 59,308 | 93.61 | 92.60 | 20,736 | 355.39 | 382.52 |
| .25 | 1,833 | 196,875 | 89.48 | 87.02 | 192,866 | 334.01 | 367.18 |
| .30 | 2,260 | 398,897 | 90.91 | 89.93 | 1,014,462 | 340.93 | 367.87 |
| .35 | 2,587 | 741,522 | 91.04 | 88.16 | 2,981,655 | 343.54 | 368.01 |
| .40 | 2,973 | 1,236,104 | 91.35 | 89.16 | 8,460,854 | 346.15 | 371.23 |
| .45 | 3,310 | 1,777,351 | 92.25 | 90.65 | 17,910,777 | 349.75 | 376.32 |
| .50 | 3,718 | 2,937,478 | 91.63 | 89.81 | 47,781,879 | 347.37 | 373.91 |

$n = 5$

| $q$ | Words | Singles | $E_1$ | $E_2$ | Doubles | $E_1$ | $E_2$ |
|---|---|---|---|---|---|---|---|
| .10 | 1,654 | 2,651 | 345.18 | 341.86 | 0 | 1263.59 | |
| .15 | 2,493 | 19,788 | 352.80 | 344.70 | 199 | 1273.77 | 1370.04 |
| .20 | 3,281 | 67,255 | 354.88 | 355.19 | 2,525 | 1284.72 | 1398.55 |
| .25 | 4,250 | 285,411 | 351.16 | 344.58 | 52,080 | 1283.26 | 1338.50 |
| .40 | 6,515 | 2,277,345 | 352.30 | 348.68 | 3,347,557 | 1283.71 | 1353.13 |

**Further Research**

Similar computations may be done for other forms, such as rectangles, pyramids, and stars, but things get more complicated as these forms include words of different lengths. For such forms the support estimate is no longer unique, but it does have a parametric representation. However, it would be relatively straightforward to write a program that would, given the form shape, the various lists of words, and possibly other information such as fixed letters, compute the expected number of fills. As one such example, the support for an 8-square with the phrase WORD WAYS on the top line is about 114000.

Other logological problems may possibly be approached in a similar manner as that given in this article. Two examples are palindromic pairs and tautonymic words. The support for an n-letter palindromic pair (e.g., LIVE and EVIL) is about $S = 15.8^{n/2}$, which is the same as that of the double n-square. For tautonyms, the estimate $S = 15.8^{n/2}$ if n is even (e.g., BOBO), and $15.8^{(n-1)/2}$ if n is odd (e.g., SHUSH). However, it should be pointed out that these support estimates are likely to be substantially less accurate than the support estimates for squares for various reasons.

*A* ***NEW PATTERN WORDLIST***

*A recent catalogue of Aegean Press books includes two books by Sheila A. Carlisle, Pattern Words Three Letters to Eight Letters in Length (134 pages), containing 60,000 pattern words (like EXCESS and BAMBOO), and Pattern Words: Nine Letters in Length (164 pages). Each book sells for $16.80 softcover or $25.80 hardcover; telephone 800-736-3587 for more information on ordering.*