MATHEMATICAL MODELING IN WORD GAMES: APPORTIONMENT OF LETTER TILES

Benjamin Dickman

bmd2118@columbia.edu

In mathematical modeling, the topic of apportionment provides powerful tools for the design of games. In particular, letter-frequency based modeling can be used to compute appropriate letter distributions for a variety of word games. This paper will provide possible letter distributions for two different games: Scrabble and Boggle. Throughout the paper, a familiarity with the five standard methods of apportionment – Hamilton, Jefferson, Webster, Adams, and Huntington-Hill – will be assumed. For a review of these methods, please consult [1] or the following link:

http://www.ctl.ua.edu/math103/apportionment/appmeth.htm

Computations were done using various online applets, the most important of which can be found at:

http://www.cut-the-knot.org/Curriculum/SocialScience/ApportionmentApplet.shtml

Scrabble

For the purposes of this paper, letter distributions were calculated as follows: (1) regard the 26 English letters as states; (2) regard the 98 letter tiles (without blanks) as the total number of members in the house; (3) regard letters' relative frequency percentages (calculated to the thousandths) as their populations (after multiplication by 1000). A table of letter frequencies cited in [2] can be found below, followed by a full chart of the suggested letter distributions in Scrabble:

Frequency of letters in the English language

Letter	Frequency in English Language	Letter	Frequency in English Language
E	12.702%	M	2.406
T	9.056	W	2.360
A	8.167	F	2.228
O	7.507	G	2.015
I	6.966	Y	1.974
N	6.749	P	1.929
S	6.327	B	1.492
H	6.094	V	0.978
R	5.987	K	0.772
D	4.253	J	0.153
L	4.025	X	0.150
C	2.782	Q	0.095
U	2.758	Z	0.074

Letter distributions suggested by the five apportionment methods

Letter	Scrabble	Hamilton	Jefferson	Webster	Adams	Huntington-Hill
A	9	8	8	8	7	8
B	2	2	1	1	2	1
C	2	3	3	3	3	3
D	4	4	4	4	4	4
E	12	12	13	13	11	12
F	2	2	2	2	2	2
G	3	2	2	2	2	2
H	2	6	6	6	6	6
I	9	7	7	7	6	6
J	1	0	0	0	1	1
K	1	1	0	1	1	1
L	4	4	4	4	4	4
M	2	2	2	2	3	2
N	6	7	7	7	6	6
O	8	7	8	7	7	7
P	2	2	2	2	2	2
Q	1	0	0	0	1	1
R	6	6	6	6	6	6
S	4	6	6	6	6	6
T	6	9	9	9	8	8
U	4	3	3	3	3	3
V	2	1	1	1	1	1
W	2	2	2	2	2	2
X	1	0	0	0	1	1
Y	2	2	2	2	2	2
Z	1	0	0	0	1	1

Note that only Adams' method and the Huntington-Hill method give at least one tile to each letter. The latter is of particular interest, since it is currently used by the United States House of Representatives. Recall that Huntington-Hill will *never* assign a state (i.e., letter) 0 representatives (i.e., tiles), since the geometric mean of 0 and 1 is 0. Thus, computations in the following section will be performed only for Huntington-Hill.

Boggle

For a brief introduction to the rules of Boggle and some of the mathematics it entails, see [3]. For a discussion of the role of problem solving strategies in Boggle and similar games, see [4]. Using the frequency table from the previous section, distributions were computed as follows: (1) regard the 26 English letters as states; (2) regard the 96 cube faces as the total number of members in the house; (3) regard letters' relative frequency percentages (calculated to the thousandths) as their populations (after multiplication by 1000). Next, the cube faces were assigned to the 16 cubes one by one, beginning with the most frequent letter, and cycling through the cubes $c_1, \ldots, c_{16}$. For example, a total of 12 cube faces were assigned to the most frequent letter, E, so each of c_1 through c_{12} received an E-face. The letter of the next highest frequency was T, to which a total of 8 cube faces were assigned, so the next eight cubes – i.e., $c_{13}, c_{14}, c_{15}, c_{16}, c_1, c_2, c_3, c_4$ – each received a T-face. And so forth. The table below has been alphabetized to improve readability.

Letter distribution suggested by Huntington-Hill

Boggle	Huntington-Hill
AAEEGN	ABCEHN
ABBJOO	ACEHIP
ACHOPS	ACEHNP
AFFKPS	AEGHIL
AOOTTW	AEHILY
CIMOTV	AEHILY
DEILRX	AENRUV
DELRVY	DEFIST
DISTTY	DEFOST
EEGHNW	DEOSTW
EEINSU	DOSTWZ
EHRTVW	EGILST
EIOSST	EKONRU
ELRTTY	JNORTU
HIMNQU	MNORTX
HLNNRZ	MOQRST

Further information on how the creators of Boggle determined the letter distribution (which, incidentally, has changed among various editions) was unavailable. For the reader interested in programming, a possible follow-up would be to generate a large number of boards using the two distributions above, and answer questions such as: Which words appeared most often for each of the distributions? What was the average (mean, median) total score for each of the distributions? What was the average (mean, median) length of the longest word for each of the distributions? Answers to these questions or similar ones would be of great interest to the author, who welcomes any communication related to this paper.

Acknowledgment: The author wishes to thank Andrew Sanfratello for early discussion of apportionment methods in the context of Scrabble tile distribution.

References

[1] Consortium for Mathematics, & Its Applications (US). (2003). For all practical purposes: mathematical literacy in today's world. WH Freeman & Company.

[2] Beker, Henry; Piper, Fred (1982). Cipher Systems: The Protection of Communications. Wiley-Interscience. p. 397.

[3] Ash, C. Boggle. Mathematics in School, Vol. 16, No. 1 (Jan., 1987), pp. 41-43. Retrieved from http://www.jstor.org/stable/30214170.

[4] Dickman, B. (2013). "Problem Solving Strategies in Boggle-like Games." Word Ways, 46(1), 66–72.

" Willy boy, Willy boy, where are you going?
I will go with you, if I may."
" I am going to the meadows to see them mowing;
I am going to see them make the hay."

Find one of the mowers.