EMBEDDING HAMLET IN PI

MIKE KEITH
Richmond, Virginia
A. ROSS ECKLER
Morristown, New Jersey

Take the text of Shakespeare's *Hamlet* beginning with its title "The Tragedy of Hamlet, Prince of Denmark...". Write down the digits of pi (after the decimal point) in successive slices (1, 14, 141, 1415...), reduce these mod 26 (1, 14, 11, 11...), and convert the resulting numbers to letters using the correspondence A=0, B=1...Z=25 (B, O, L, L...). Continue until the letter T is first achieved at slice 27: 1415926535 8979323846 2643383 mod 26 = 19 = T. Continue with the next pi-digits until the letter H is first encountered at slice 10: 179502841 mod 26 = 7 = H. The slice sizes corresponding to the first few words are

The (27,10,22) Tragedy (2,33,19, 70,9,18,13) Of (5,28) Hamlet (48,18,38,3,10,6)
Prince (12,4,11,7,40,1) Of (59,26) Denmark (16,56,13,53,97,13,21)

Act (27,37,4) I (50), Scene (38,1,3,12,75) I (30)
Elsinore (31,12,19,27,11,81,52,2)
A (69) Platform (46,36,37,5,7,25,3,61) Before (36,41,1,53,22,7) The (28,1,7)
Castle (74,1,16,14,17,66)

and so on, to

Exeunt (10,21,15,3,12,4) Marching (26,22,105,7,12,14,3,16) After (38,15,22,37,26)
The (8,31,17) Which (11,101,201,8,54) A (38) Peal (25,36,2,29) Of (25,36)
Ordnance (7,38,20,11,8,39,36,7) Are (1,71,61) Shot (5,1,50,2) Off (66,4,28)

In all, 3359924 digits of pi have been used to encode the 129529 letters of *Hamlet.* The average slice size per letter is 25.94, statistically indistinguishable from the expected value of 26 slices per letter (the digits of pi are believed to occur at random with equal frequencies). If one assumes that the slices are distributed according to the geometric distribution which has mean and variance both equal to 26, then the standard deviation of the average slice size has a standard deviation of $26/\sqrt{129529} = 26/360 = 0.07$.

There is another way one can transform the digits of pi to letters: sum the digits in each slice and reduce them mod 26, converting 1, 14, 141, 1415... to 1, 5, 6, 11... and B, G, G, L... This method of encoding *Hamlet* is a little less efficient; the average slice size per letter increases to 26.14. This is hardly surprising if one examines the probabilities of very small slice sizes for different letters. A slice size of one is achievable only for letters A through J, and in fact the letters T through Z (with values 19 through 25) cannot have slices of length one or two. The probabilities for slices of one through three for each letter are given in the table on the next page, along with a simulation of the average slice size required if one has texts consisting solely of the letters A, B, C, etc. (Finding the theoretical values of the individual letter averages appears mathematically difficult.)

	Probability That First Slice Is of Size One	Two	Three	Average Slice Size
A	0.1		0.003	25.9
B	0.1	0.01	0.002	25.9
C	0.1	0.02	0.003	25.8
D	0.1	0.03	0.006	25.6
E	0.1	0.04	0.010	25.4
F	0.1	0.05	0.015	25.1
G	0.1	0.06	0.021	24.8
H	0.1	0.07	0.028	24.4
I	0.1	0.08	0.036	23.9
J	0.1	0.09	0.045	23.4
K		0.09	0.053	26.0
L		0.08	0.059	26.0
M		0.07	0.063	26.1
N		0.06	0.065	26.1
O		0.05	0.066	26.2
P		0.04	0.065	26.3
Q		0.03	0.063	26.6
R		0.02	0.059	26.8
S		0.01	0.053	27.2
T			0.045	27.6
U			0.036	27.6
V			0.028	27.8
W			0.021	27.9
X			0.015	28.3
Y			0.010	28.5
Z			0.006	28.5

Notice that the last two probabilities of the third slice (0.003, 0.001) that would normally appear after Z are added back at the beginning (to A and B) because of the mod 26 requirement.

These inequalities suggest an interesting question: can one rearrange the alphabet so that the commonest letters in running text are associated with the smallest average slice sizes, thus creating a more efficient packing? We suggest DHSRI NOATE LUCMF PGWBY VKXJQ Z, which leads to an average slice size per letter of 25.31, an improvement of 0.8.