alphabet, i.e., a letter is produced with a probability dependent upon the underlying state. The states themselves are not assumed to be observable, and indeed we can only determine probabilistically which state we are in at a given time. Since this is the case we call these models Hidden Markov Models.[1]

Now for each $S = 2, 3, \ldots, 12$ we have found (we believe) that $S \times S$ matrix, and that set of S probability distributions which maximize the probability of observing some particular long stretch of English text.[7]

III. The Models Themselves

In the tables following, we present the raw results of the calculations. There is given for each $S = 2, 3, \ldots, 12$, a model, that is, an $S \times S$ transition matrix, and an $S \times 27$ output matrix giving the output distributions in each state. The stationary probabilities which are the probabilities of being in each state after a long period of time are also shown.

IV. Discussion of the Models

What is displayed in Tables I-1 — I-11 is merely a collection of parameters for a sequence of statistical models of English. What is important to keep in mind in

[1] They have been called probabilistic functions of Markov chains, but we find this a bit unwieldy.

TABLE I-1. HIDDEN MARKOV MODEL*

S = 2

Transition Probabilities

1	.275	.725
2	.780	.220

Output Probabilities

	1	2
A	—	.133
B	.022	—
C	.063	—
D	.056	—
E	—	.218
F	.037	—
G	.015	.010
H	.074	—
I	—	.150
J	—	—
K	.009	—
L	.060	—
M	.041	—
N	.140	—
O	—	.136
P	.030	.001
Q	.001	—
R	.087	—
S	.105	—
T	.157	.019
U	—	.045
V	.016	—
W	.020	—
X	.002	—
Y	.004	.018
Z	.001	—
#	.060	.269

Stationary Probabilities

.52	.48

* # denotes "Word Space".