

### Caesar shift frequency analysis

To analyze a Caesar cipher, we can look at letter frequencies to determine the number of letters by which each letter has been shifted. This handout shows the outcome of our scoring calculation for plaintext with "perfect" letter distribution. The kth entry of table on the left contains the sum from n=0 to 25 of:

$$(\text{frequency of the } n\text{th letter}) * \log(\text{frequency of } (n-k)\text{th letter}).$$

Note that the result for k=0 is significantly larger than for all other k. The table on the right shows the calculation that produced this data for for k=8.

k	sum	n	freq nth letter		
0	-1.2573	0	A	0.0800	
1	-1.6666	1	B	0.0154	
2	-1.6971	2	C	0.0258	
3	-1.8034	3	D	0.0432	
4	-1.6445	4	E	0.1257	
5	-1.8246	5	F	0.0235	log(freq)
6	-1.6221	6	G	0.0198	(n-k)th lett. freq*log(shifted)
7	-1.6670	7	H	0.0624	
8	-1.7444	8	I	0.0692	-1.0969 -0.075906
9	-1.7719	9	J	0.0014	-1.8125 -0.002537
10	-1.7528	10	K	0.0074	-1.5884 -0.011754
11	-1.5576	11	L	0.0406	-1.3645 -0.055399
12	-1.5883	12	M	0.0256	-0.9007 -0.023057
13	-1.5871	13	N	0.0690	-1.6289 -0.112396
14	-1.7623	14	O	0.0759	-1.7033 -0.129283
15	-1.6141	15	P	0.0180	-1.2048 -0.021687
16	-1.6345	16	Q	0.0012	-1.1599 -0.001392
17	-1.7556	17	R	0.0596	-2.8539 -0.170091
18	-1.7540	18	S	0.0634	-2.1308 -0.135091
19	-1.6164	19	T	0.0909	-1.3915 -0.126485
20	-1.7811	20	U	0.0284	-1.5918 -0.045206
21	-1.8076	21	V	0.0098	-1.1612 -0.011379
22	-1.6089	22	W	0.0222	-1.1198 -0.024859
23	-1.6925	23	X	0.0018	-1.7447 -0.003141
24	-1.7704	24	Y	0.0190	-2.9208 -0.055496
25	-1.6285	25	Z	0.0008	-1.2248 -0.00098
		0	A	0.0800	-1.1979 -0.095833
		1	B	0.0154	-1.0414 -0.016038
		2	C	0.0258	-1.5467 -0.039904
		3	D	0.0432	-2.0088 -0.086779
		4	E	0.1257	-1.6536 -0.207863
		5	F	0.0235	-2.7447 -0.064501
		6	G	0.0198	-1.7212 -0.034081
		7	H	0.0624	-3.0969 -0.193247

-1.744385