

Worksheet: Arithmetic Coding

The frequency of usage and sample probabilities of the alphabet in H.G. Well's "*War of the Worlds*" are:

a	b	c	d	e	f	g	h	i	j	k	l	m
23433	4283	6760	13373	35215	6436	6802	16959	19121	274	2154	10687	7210
0.0833	0.0152	0.024	0.0476	0.1252	0.0229	0.0242	0.0603	0.068000	0.001	0.0077	0.038	0.0256
n	o	p	q	r	s	t	u	v	w	x	y	z
20258	19962	5118	190	16847	16924	27193	7535	2470	6462	390	5076	107
0.072	0.071	0.0182	0.0007	0.0599	0.0602	0.0967	0.0268	0.0088	0.023	0.0014	0.018	0.0004

1. It will take too much time to manually process all the above symbols. So we will work with the six most probable symbols (with probabilities that have been normalized to sum to 1 for just these 6):

e	t	a	n	o	i
0.2426	0.1873	0.1614	0.1395	0.1375	0.1317

Determine the range of values needed to transmit the following symbol sequence:

'i', 't', 'a', 'o', 'e', 't', 'e'

[illegible]

2. Determine the sequence of integer numbers that can be used to transmit this sequence.

[illegible]

