		1		
0	00	000	0000	
			0001	get
		001	0010	
			0011	nd
	01	010	0100	he total symbol code budget
			0101	
		011	0110	
			0111	
1	10	100	1000	
			1001	
		101	1010	
			1011	tol
	11	110	1100	he
			1101	
		111	1110	
		111	1111	

Figure 5.1. The symbol coding budget. The 'cost'  $2^{-l}$  of each codeword (with length l) is indicated by the size of the box it is written in. The total budget available when making a uniquely decodeable code is 1. You can think of this diagram as showing a codeword supermarket, with the codewords arranged in aisles by their length, and the cost of each codeword indicated by the size of its box on the shelf. If the cost of the codewords that you take exceeds the budget then your code will not be uniquely decodeable.

## Book by David MacKay

$a_i$	$p_i$	$\log_2 \frac{1}{p_i}$	$l_i$	$c(a_i)$
a	0.0575	4.1	4	0000
b	0.0128	6.3	6	001000
С	0.0263	5.2	5	00101
d	0.0285	5.1	5	10000
е	0.0913	3.5	4	1100
f	0.0173	5.9	6	111000
g	0.0133	6.2	6	001001
h	0.0313	5.0	5	10001
i	0.0599	4.1	4	1001
j	0.0006	10.7	10	1101000000
k	0.0084	6.9	7	1010000
1	0.0335	4.9	5	11101
m	0.0235	5.4	6	110101
n	0.0596	4.1	4	0001
0	0.0689	3.9	4	1011
р	0.0192	5.7	6	111001
q	0.0008	10.3	9	110100001
r	0.0508	4.3	5	11011
S	0.0567	4.1	4	0011
t	0.0706	3.8	4	1111
u	0.0334	4.9	5	10101
v	0.0069	7.2	8	11010001
W	0.0119	6.4	7	1101001
x	0.0073	7.1	7	1010001
у	0.0164	5.9	6	101001
Z	0.0007	10.4	10	1101000001
—	0.1928	2.4	2	01

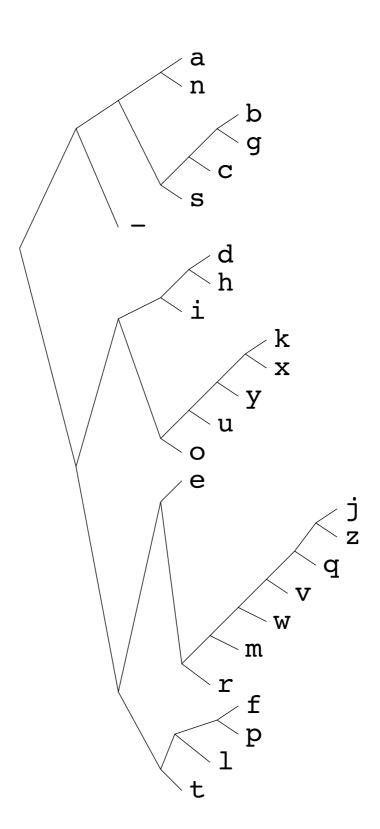


Figure 5.6. Huffman code for the English language ensemble (monogram statistics).

## Book by David MacKay

ure 5.6. This code has an expected length of 4.15 bits; the entropy of the ensemble is 4.11 bits. Observe the disparities between the assigned codelengths and the ideal codelengths  $\log_2 1/p_i$ .

		$= 00000 \\ = 00001 0000$	Context (sequence thus far)	Probability of next symbol			
		000		P(a) = 0.425	P(b) = 0.425	$P(\Box) = 0.15$	
		$= 00010 \\ = 00011 \\ 0001 \\ = 00011 \\ 00001 \\ 000001 \\ 000001 \\ 000001 \\ 0000000 \\ 0000000 \\ 00000000$	b	$P({\tt a} {\tt b}){=}0.28$	$P({f b} {f b}){=}0.57$	$P(\Box \mathbf{b}){=}0.15$	
a		$\frac{-00011}{-00100} 0010$	bb	$P({\tt a} {\tt bb}){=}0.21$	$P(\mathbf{b}   \mathbf{b} \mathbf{b}) {=} 0.64$	$P(\Box \texttt{bb}){=}0.15$	
		$\frac{1}{10000000000000000000000000000000000$	bbb	$P(\texttt{a} \texttt{bbb}){=}0.17$	$P(\mathbf{b} \mathbf{bbb}){=}0.68$	$P(\Box \texttt{bbb}){=}0.15$	
		$\frac{-00110}{-00111} 0011 = 0$	bbba	$P(\mathbf{a}   \mathbf{bbba}) {=} 0.28$	$P(\mathbf{b}   \mathbf{bbba}) {=} 0.57$	$P(\Box   \texttt{bbba}) {=} 0.15$	
		$ \begin{array}{r} = 01000 \\ = 01001 \\ 0100 \\ = 01010 \\ = 01011 \\ 0101 \\ = 01100 \\ = 01101 \\ 0110 \end{array} $ 01	Figure 6.4 shows the corresponding intervals. The interval <b>b</b> is the middle 0.425 of $[0, 1)$ . The interval <b>bb</b> is the middle 0.567 of <b>b</b> , and so forth.				
	ba	$\begin{array}{c} 0 \\ -$	bbbaa	$   \begin{array}{r} - 10010111 \\       - 10011000 \\       - 10011001 \\       \hline       10011001   \end{array} $			
_	bba <u>bbba</u> bb bbb bbbb	$= \frac{10010}{1001} 1001 / 1001 / 1001 / 1001 / 1001 / 1010 / 1010 / 1010 / 1010 / 10111 / 1011 / 1011 / 1011 / 1011 / 1011 / 1011 / 1011 / 1011 / 101$	bbba bbbab	$ \frac{-10011010}{-10011011} 1001 \\ -10011100 -10011101 \\ -10011101 \\ -10011110 \\ -10011111 $	1		
	<u>bbb□</u> bb□	$\frac{10111}{1000} = 11000 = 1100$		<u>-\1010000</u> 00 100111101			
		$\begin{array}{r} - 11010 \\ - 11011 \\ - 11011 \\ - 11100 \\ - 11101 \\ - 11110 \\ - 11111 \\ - 11111 \\ - 11111 \\ \end{array} $		arithme	5.4. Illustration etic coding prose bbba $\square$ is tra	cess as the	

Book by David MacKay