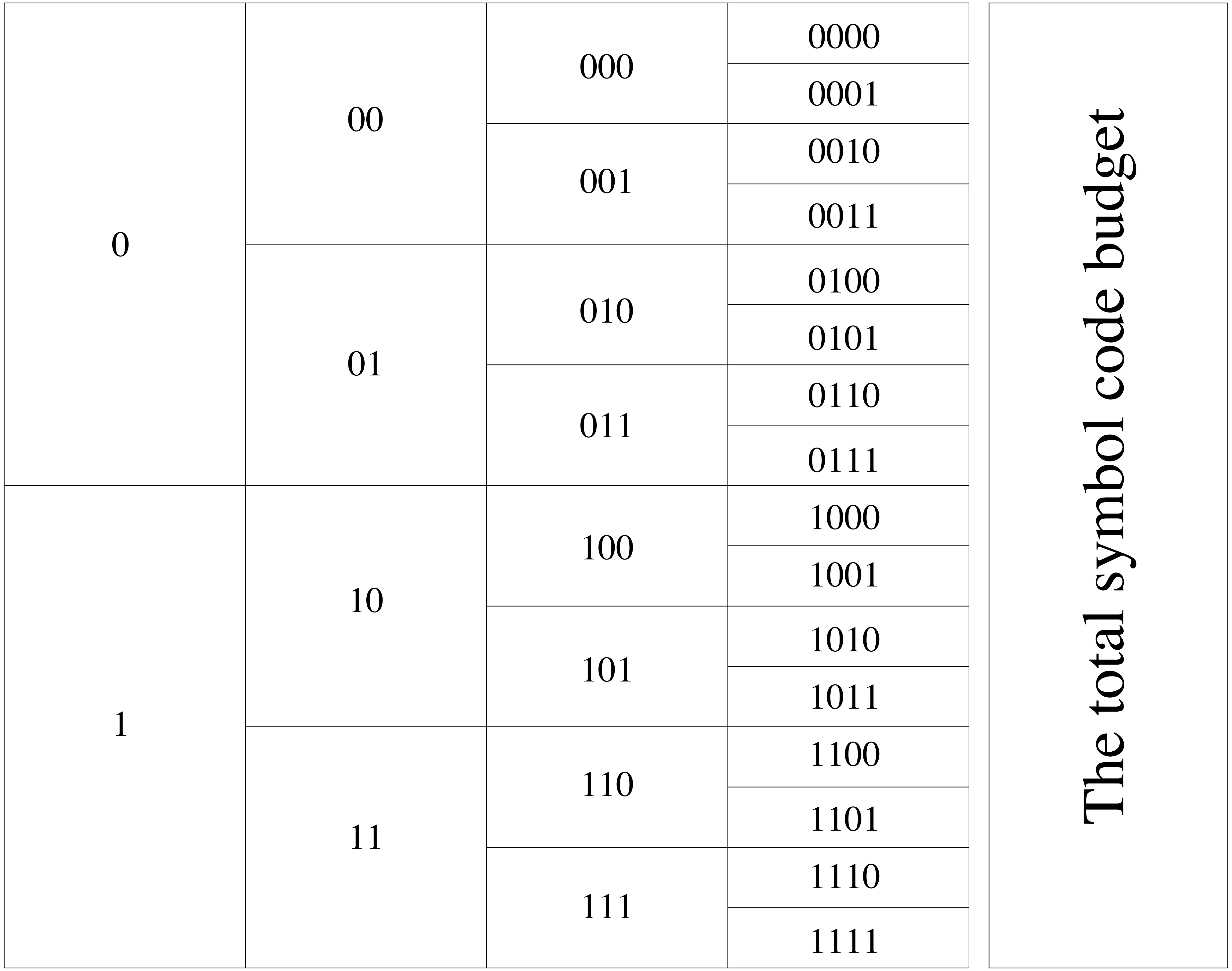


Figure 5.1. The symbol coding budget. The 'cost' 2^{-l} of each codeword (with length l) is indicated by the size of the box it is written in. The total budget available when making a uniquely decodeable code is 1.
You can think of this diagram as showing a *codeword supermarket*, with the codewords 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

C_0:

a_i	$c(a_i)$	l_i
a	1000	4
b	0100	4
c	0010	4
d	0001	4

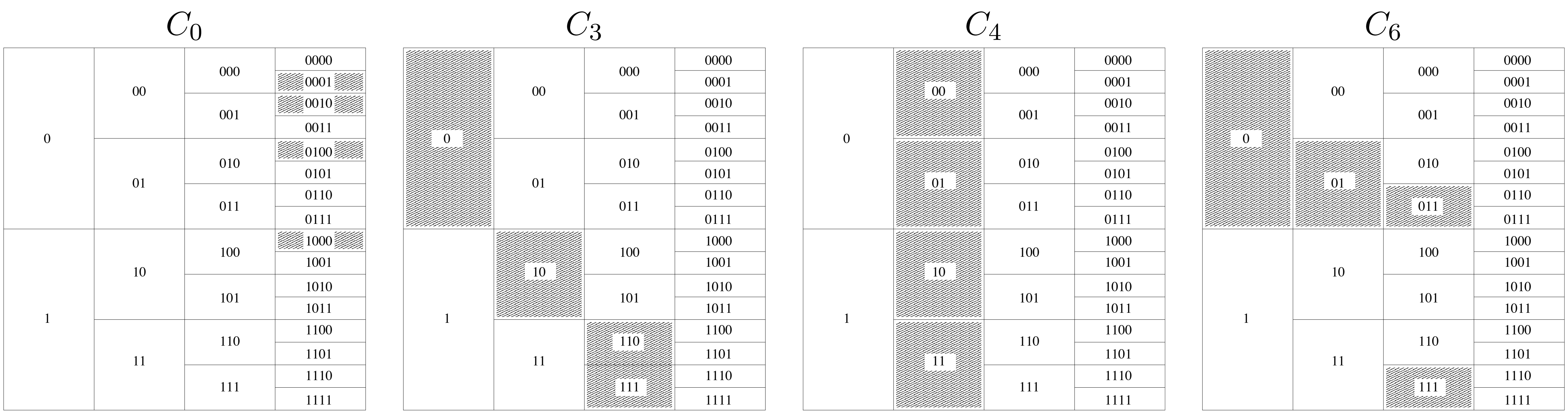
C_3:

a_i	$c(a_i)$	p_i	$h(p_i)$	l_i
a	0	$1/2$	1.0	1
b	10	$1/4$	2.0	2
c	110	$1/8$	3.0	3
d	111	$1/8$	3.0	3

	C_4	C_5
a	00	0
b	01	1
c	10	00
d	11	11

C_6:

a_i	$c(a_i)$	p_i	$h(p_i)$	l_i
a	0	$1/2$	1.0	1
b	01	$1/4$	2.0	2
c	011	$1/8$	3.0	3
d	111	$1/8$	3.0	3



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
c	0.0263	5.2	5	00101
d	0.0285	5.1	5	10000
e	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
l	0.0335	4.9	5	11101
m	0.0235	5.4	6	110101
n	0.0596	4.1	4	0001
o	0.0689	3.9	4	1011
p	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
y	0.0164	5.9	6	101001
z	0.0007	10.4	10	1101000001
–	0.1928	2.4	2	01

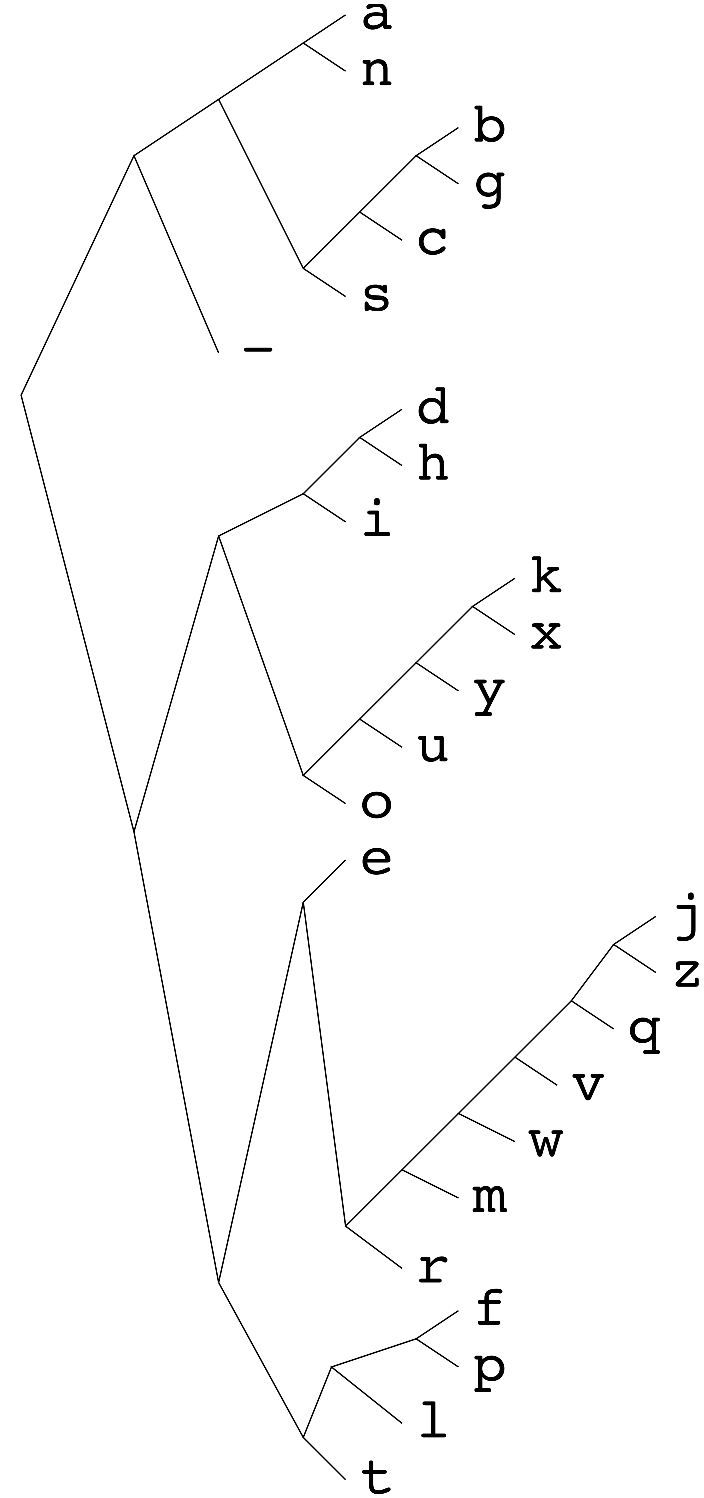


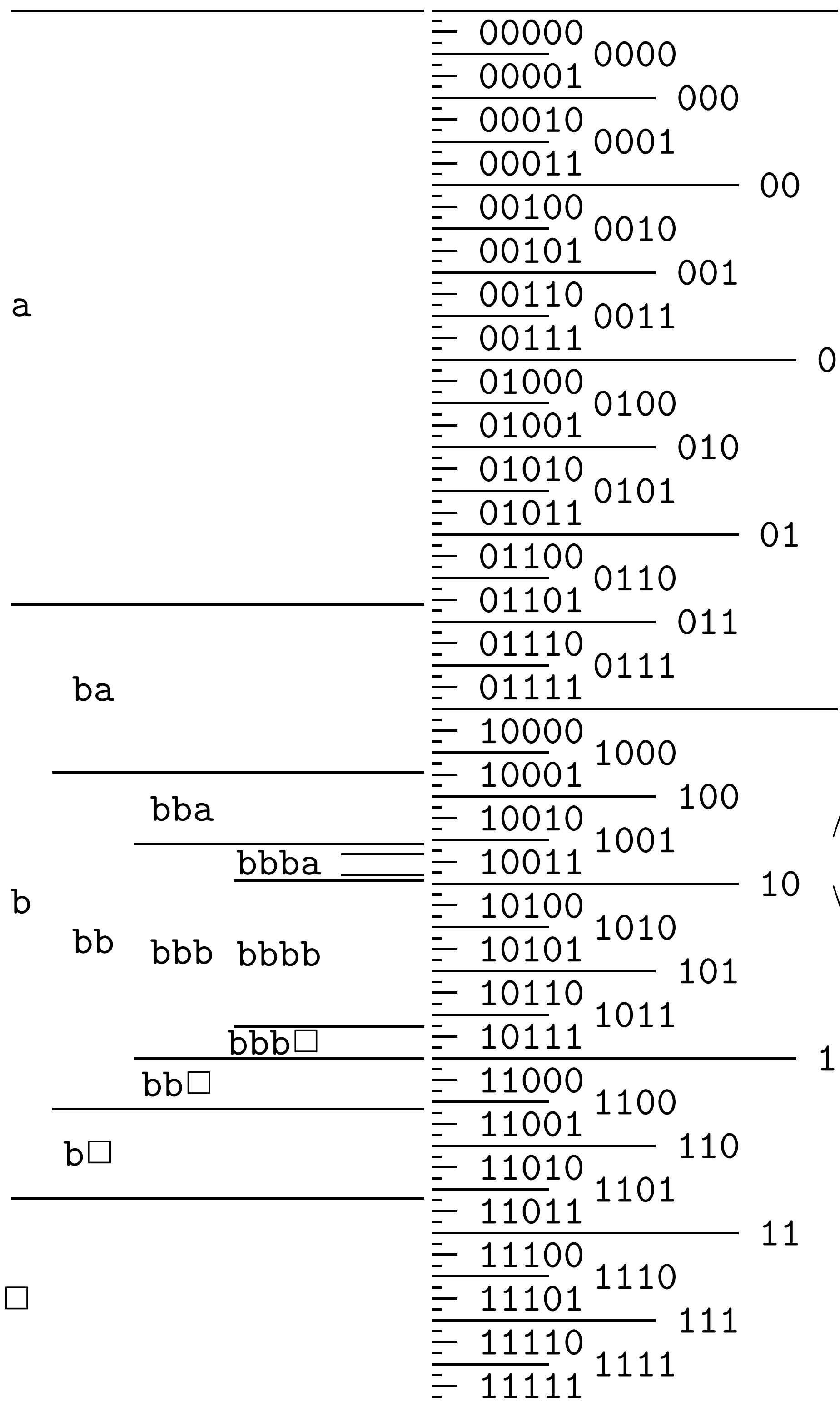
Figure 5.6. Huffman code for the English language ensemble (monogram statistics).

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 $\log_2 1/p_i$.

Book by David MacKay

Context (sequence thus far)	Probability of next symbol		
	$P(\texttt{a}) = 0.425$	$P(\texttt{b}) = 0.425$	$P(\square) = 0.15$
`b`	$P(\texttt{a} \mid \texttt{b}) = 0.28$	$P(\texttt{b} \mid \texttt{b}) = 0.57$	$P(\square \mid \texttt{b}) = 0.15$
`bb`	$P(\texttt{a} \mid \texttt{bb}) = 0.21$	$P(\texttt{b} \mid \texttt{bb}) = 0.64$	$P(\square \mid \texttt{bb}) = 0.15$
`bbb`	$P(\texttt{a} \mid \texttt{bbb}) = 0.17$	$P(\texttt{b} \mid \texttt{bbb}) = 0.68$	$P(\square \mid \texttt{bbb}) = 0.15$
`bbba`	$P(\texttt{a} \mid \texttt{bbba}) = 0.28$	$P(\texttt{b} \mid \texttt{bbba}) = 0.57$	$P(\square \mid \texttt{bbba}) = 0.15$

Figure 6.4 shows the corresponding intervals. The interval `b` is the middle 0.425 of $[0, 1)$. The interval `bb` is the middle 0.567 of `b`, and so forth.



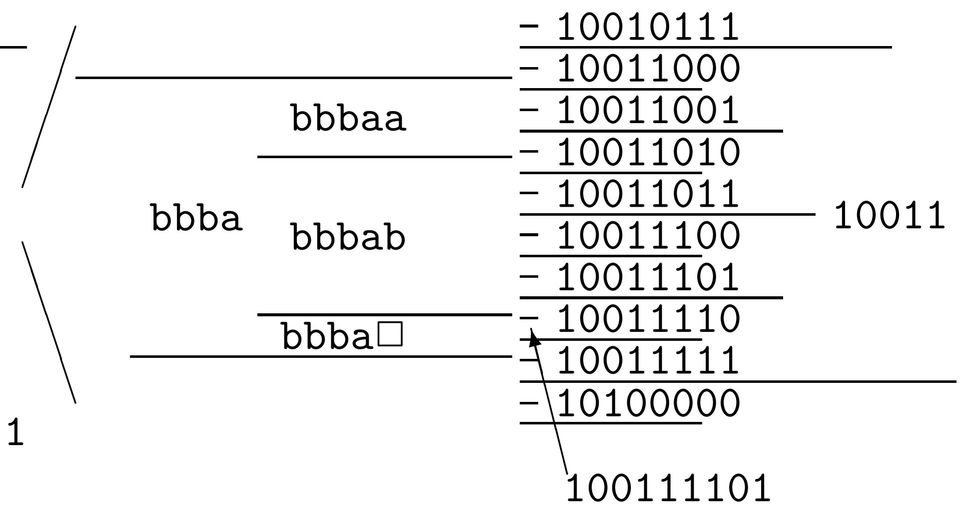


Figure 6.4. Illustration of the arithmetic coding process as the sequence `bbba□` is transmitted.