Prof. Eduard A. Jorswieck, Dr. Bile Peng 14. Dezember 2023

1. Übung
Grundlagen der Informationstechnik, Teil Nachrichtentechnik

Aufgabe 1.

Gegeben sei eine gedächtnislose Quelle, deren mögliche Ausgangssymbole $x_i, i = 1, \ldots, 8$ die Buchstaben A bis H sind. X sei dabei eine Zufallsvariable, die diese Quelle beschreibt. Die einzelnen Symbole treten dabei mit den folgenden Wahrscheinlichkeiten auf:

i	1	2	3	4	5	6	7	8
x_i	A	B	C	D	E	F	G	H
$P(X = x_i)$	0,2	0,05	0,01	0,1	0,4	0,1	0,06	0,08

Tabelle 1: Aufgabe Codebaum

a) Berechnen Sie die Unsicherheit $H(X)$ der Quelle bzw. der Zufallsvariablen X.

b) Konstruieren Sie einen präfixfreien Shannon-Fano Code um diese Quelle zu codieren. Geben Sie dabei für jedes Symbol $x_i (i = 1, \ldots, 8)$ das zugehörige Codewort an und zeichnen Sie den Codebaum.

c) Bestimmen Sie die mittlere Codewlrtlänge $\mathbb{E}(\omega_i)$, wobei ω_i die Länge des Codewortes ist, das zum Symbol x_i gehört.

Lösung 1.

a)

$$\begin{aligned}
H(X) = &- \sum_i P(X = x_i) \log_2(P(X = x_i)) \\
= &- 0,2 \log_2(0.2) - 0,05 \log_2(0,05) - 0,01 \log_2(0,01) - 0,1 \log_2(0,1) \\
&- 0,4 \log_2(0,4) - 0,1 \log_2(0,1) - 0,06 \log_2(0,06) - 0,08 \log_2(0,08) \\
= &2,4751
\end{aligned} \tag{1}$$

b) Die Codelänge von jedem Zeichen ist in Tabelle 2 aufgelistet.

Der Codebaum ist in Abbildung 1 dargestellt.

i	1	2	3	4	5	6	7	8
x_i	A	B	C	D	E	F	G	H
$P(X = x_i)$	0,2	0,05	0,01	0,1	0,4	0,1	0,06	0,08
$\left\lceil \log_2 \left(\frac{1}{P(X=x_i)} \right) \right\rceil$	3	5	7	4	2	4	5	4

Tabelle 2: Codelänge des Shannon-Fano-Codes

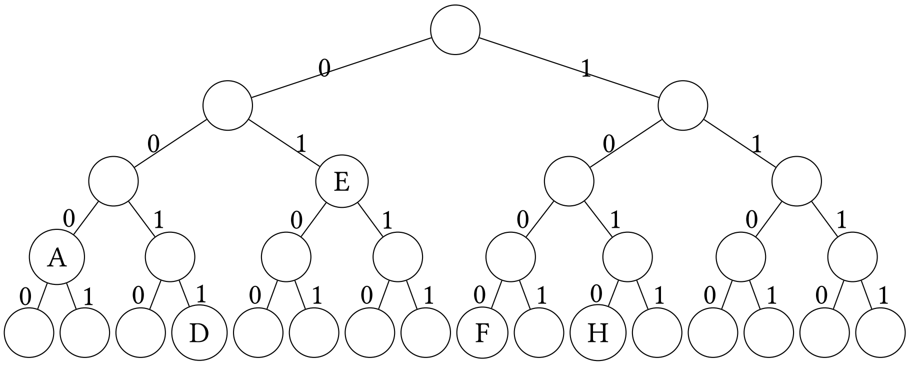


Abbildung 1: Der Codebaum des Shannon-Fano-Codes. Wegen der Einschränkung der Seitenlänge werden nur die ersten 5 Zeilen gezeigt.

c)

$$\mathbb{E}(\omega_i) = \sum_i p_i \left\lceil \log_2 \left(\frac{1}{P(X = x_i)} \right) \right\rceil = 3,1400 \tag{2}$$

Aufgabe 2.

Gegeben sei ein normaler Würfel mit den Augenzahlen 1 bis 6, die alle gleich wahrscheinlich sind. X_i seien Zufallsvariablen, die die geworfene Augenzahl beim i-ten Wurf angeben.

a) Der Würfel wird einmal geworfen. Berechnen Sie die Unsicherheit $H(X_1)$ der Zufallsvariable X_1 und die Anzahl der im Mittel nötigen Bits, um die Zufallsvariable X_1 mit einem Shannon-Fano Codierer zu codieren.

b) Nun wird der Würfel zwei Mal geworfen. Berechnen Sie die Unsicherheit $H(X_1, X_2)$ des Vektors (X_1, X_2).

c) Wie groß ist im Unterschied zu Teilaufgabe b) die Unsicherheit der Zufallsvariable $Y = X_1 + X_2$, welche die Summe der gworfenen Augenzahlen angibt. Erklären Sie diesen Unterschied kurz. Berechnen Sie außerdem die Anzahl der im Mittel zur Codierung nötigen Bits, wenn zur Codierung von Y ein Shannon-Fano Codierer verwendet wird.

Lösung 2.

a)

$$H(X_1) = -6 \times \frac{1}{6} \log_2 \left(\frac{1}{6}\right) = 2,585 \tag{3}$$

$$\left\lceil \log_2 \left(\frac{1}{P(X = x_i)}\right) \right\rceil = 3 \tag{4}$$

b) Es gibt insgesamt $6 \times 6 = 36$ mögliche Werte von $H(X_1, X_2)$, die gleich wahrscheinlich sind und eine Wahrscheinlichkeit von $1/36$ haben. Die Unsicherheit ist deswegen

$$H(X_1, X_2) = -36 \times \frac{1}{36} \log_2 \left(\frac{1}{36}\right) = 5,170 \tag{5}$$

c) Die Anzahl der Möglichkeiten bei jedem Wert der Summe ist in Tabelle 3 aufgelistet. Der Unterschied liegt daran, dass eine Summe von unterschiedlicher Kombinationen von X_1 und X_2 realisiert werden kann.

$X_1 + X_2$	2	3	4	5	6	7	8	9	10	11	12
Möglichkeiten	1	2	3	4	5	6	5	4	3	2	1

Tabelle 3: Anzahl der Möglichkeiten bei jedem Wert von Y

Da die Wahrscheinlichkeit von jedem Wert von (X_1, X_2) $1/36$ ist, ist die Wahrscheinlichkeit von jedem Wert der Summe $X_1 + X_2$ proportional zur Anzahl der Möglichkeiten.

$$H(Y) = -\sum_i P(Y = y_i) \log_2 \left(\frac{1}{P(Y = y_i)}\right) = 3,274. \tag{6}$$

Es ist wie erwartet, dass $H(X_1, X_2) > H(Y)$.

$$\sum_i P(Y = y_i) \left\lceil \log_2 \left(\frac{1}{P(Y = y_i)}\right) \right\rceil = 3,778. \tag{7}$$

Aufgabe 3.

In der folgenden Aufgabe sollen Methoden der Quellencodierung betrachtet werden.

a) In der nachfolgenden Tabelle sind die Codewörter von drei unterschiedlichen Quellencodes A, B und C aufgeführt. Welcher bzw. welche der Codes ist eindeutig decodierbar?

Code A	{0, 10, 11}
Code B	{01, 10, 11}
Code C	{0, 1, 11}

Tabelle 4: Aufgabe Codebaum

b) Code A aus Teilaufgabe a) kann zur Codierung einer gedächtnislosen, ergodischen Quelle mit Ausgangsalphabet $\{a, b, c\}$ und den zugehörigen Wahrscheinlichkeiten $P(a) = 0.7, P(b) = 0.25$ und $P(c) = 0.05$ verwendet werden, die durch eine Zufallsvariablen U beschrieben werden kann. Bestimmen Sie die Unsicherheit $H(U)$ der beschriebenen Quelle.

Lösung 3.

a) A und B

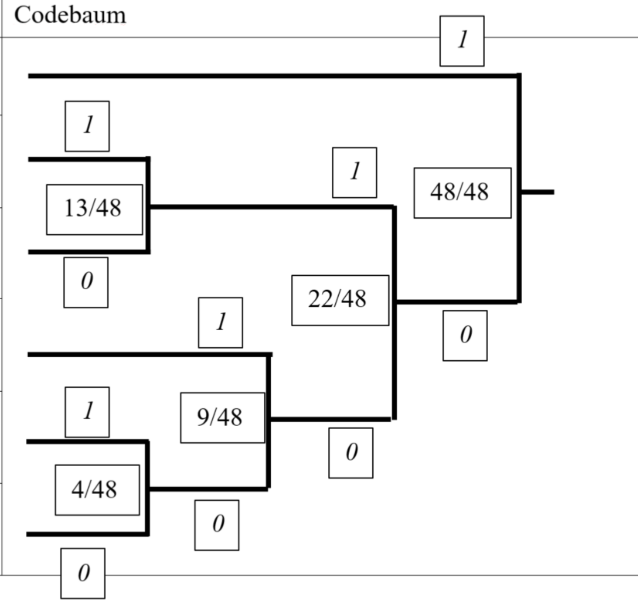
b)

$$\begin{aligned} H(U) &= -\sum_i p_i \log_2(p_i) \\ &= -0,7 \log_2(0,7) - 0,25 \log_2(0,25) - 0,05 \log_2(0,05) \\ &= 1,076 \end{aligned} \tag{8}$$

Aufgabe 4.

Es liegt das Alphabet $X = \{T, I, E, N, S, R\}$ mit den zugeordneten Wahrscheinlichkeiten $P = \{1/48, 5/48, 26/48, 7/48, 3/48, 6/48\}$ vor. Erstellen Sie nachvollziehbar den Codebaum entsprechend der Huffman-Codierung und nennen Sie die sich ergebenden Codewörter zur Darstellung des Alphabets X.

Lösung 4.

x_i	$p(x_i)$	Codebaum	Code
E	26/48		1
N	7/48		011
R	6/48		010
I	5/48		001
S	3/48		0001
T	1/48		0000

Aufgabe 5.

(Optional) wenn eine von 64 Drogen vergiftet ist. Wie viele Mäuse braucht man mindestens, um die vergiftete Droge herauszufinden und wie macht man das?

Lösung 5.

Die Informationstheorie beantwortet die Frage *wie viel* und die Codierungstheorie beantwortet die Frage *wie*.

Eine Maus hat zwei Zustände: lebendig oder tot. Daher benötigen wir $\log_2(64) = 6$ Mäuse, um 64 Zustände darzustellen.

Wir können die 64 Drogen in binärer Form darstellen. Zum Beispiel: Droge 0 ist 000000, Droge 1 ist 000001, Droge 2 ist 000010… Droge 63 ist 111111.

Dann lassen wir Maus 1 Drogen nehmen, deren erste Ziffer der Binärform 1 ist, lassen Maus 2 Drogen nehmen, deren zweite Ziffer der Binärform 1 ist usw. Auf diese Weise nimmt Maus 1 die Drogen 32, 33, 34, …, 63, Maus 2 die Drogen 16, 17, …, 31, Maus 6 die Drogen 1, 3, 5, …, 63.

Nehmen wir an, dass die Mäuse 1, 2 und 6 nach der Einnahme der Drogen sterben und die Mäuse 3, 4 und 5 leben. Dann ist die Droge 110001 (49) die vergiftete Droge, denn die Mäuse 1, 2 und 6 nehmen diese Droge und sterben, die Mäuse 3, 4 und 5 nehmen die Droge nicht und leben. Das gleiche Prinzip gilt für andere Binärzahlen.

Diese Methode wurde z. B. von Google angewandt, um kleinere Verbesserungen zu bewerten. Normalerweise wählt Google nach dem Zufallsprinzip 1% der Nutzer aus, um eine neue Verbesserung zu testen, und bittet sie um Feedback. Bei einem großen Unternehmen wie Google gibt es jedoch normalerweise mehr als 100 experimentelle Verbesserungen. Google nutzt diese Methode, um mehr als 100 Verbesserungen mit weniger als 100% der Nutzer zu testen.

Aufgabe 5.

(Optional) wenn eine von 64 Drogen vergiftet ist. Wie viele Mäuse braucht man mindestens, um die vergiftete Droge herauszufinden und wie macht man das?

Lösung 5.
Die Informationstheorie beantwortet die Frage *wie viel* und die Codierungstheorie beantwortet die Frage *wie*.

Eine Maus hat zwei Zustände: lebendig oder tot. Daher benötigen wir $\log_2(64) = 6$ Mäuse, um 64 Zustände darzustellen.

Wir können die 64 Drogen in binärer Form darstellen. Zum Beispiel: Droge 0 ist 000000, Droge 1 ist 000001, Droge 2 ist 000010… Droge 63 ist 111111.

Dann lassen wir Maus 1 Drogen nehmen, deren erste Ziffer der Binärform 1 ist, lassen Maus 2 Drogen nehmen, deren zweite Ziffer der Binärform 1 ist usw. Auf diese Weise nimmt Maus 1 die Drogen 32, 33, 34, …, 63, Maus 2 die Drogen 16, 17, …, 31, Maus 6 die Drogen 1, 3, 5, …, 63.

Nehmen wir an, dass die Mäuse 1, 2 und 6 nach der Einnahme der Drogen sterben und die Mäuse 3, 4 und 5 leben. Dann ist die Droge 110001 (49) die vergiftete Droge, denn die Mäuse 1, 2 und 6 nehmen diese Droge und sterben, die Mäuse 3, 4 und 5 nehmen die Droge nicht und leben. Das gleiche Prinzip gilt für andere Binärzahlen.

Diese Methode wurde z. B. von Google angewandt, um kleinere Verbesserungen zu bewerten. Normalerweise wählt Google nach dem Zufallsprinzip 1% der Nutzer aus, um eine neue Verbesserung zu testen, und bittet sie um Feedback. Bei einem großen Unternehmen wie Google gibt es jedoch normalerweise mehr als 100 experimentelle Verbesserungen. Google nutzt diese Methode, um mehr als 100 Verbesserungen mit weniger als 100% der Nutzer zu testen.