TK 6.1) 1

Informationsgehalt eines Ereignisses

3 Eigenschaften (Axiome):

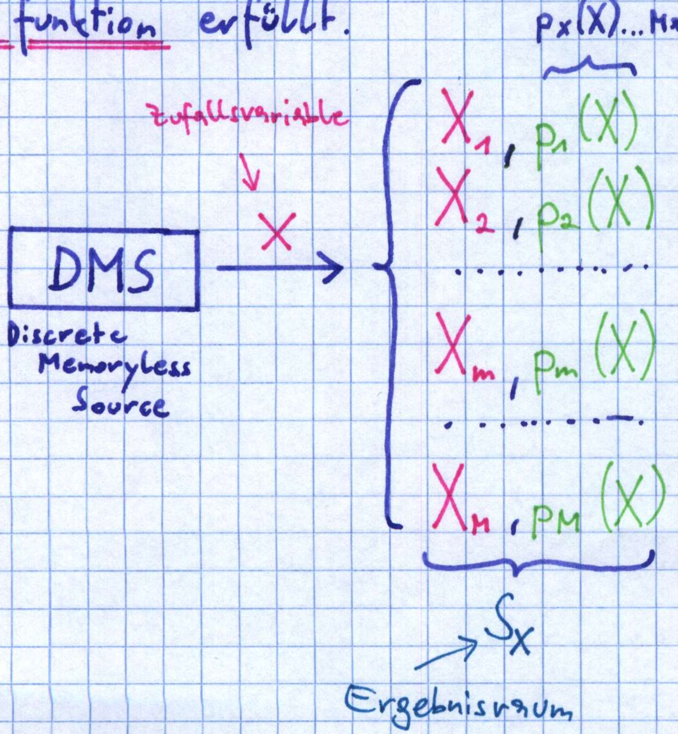
(1) Informationsgehalt kann nur positive Größe sein

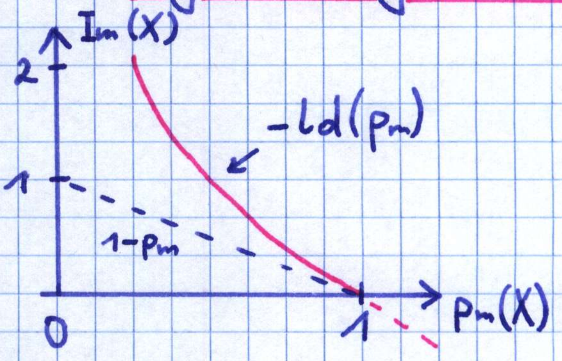
(2) Je vorhersagbarer ein Ereignis ist, desto weniger Information enthält es. Der Informationsgehalt eines Ereignisses ist monoton abnehmend, wenn dessen Häufigkeit zunimmt → Grenzfall: $p(X) = 1 \Rightarrow I(X) = 0$

(3) Informationsgehalt zweier unabhängiger Ereignisse ist die Summe der einzelnen Informationsgehälter → Additivitätseigenschaft der Information

⇒ Häufigkeiten werden multipliziert, Informationsgehalt wird addiert

Die geforderten Eigenschaften werden von der negativen Logarithmusfunktion erfüllt.





$$0 \leq p_m(X) \leq 1$$

$$\boxed{I_m(X) = -Ld(p_m) = Ld\left(\frac{1}{p_m}\right)}$$

$p_m(X)$.... Primärwahrscheinlichkeit für das Auftreten des m-ten Symbols

TK 6.1)[2]

Da eine übliche Datenquelle mehrere Symbole mit, in der Regel, unterschiedlichen Häufigkeiten ausgibt, ist die Definition des mittleren Informationsgehaltes der Zufallsvariablen X sinnvoll.

Man bezeichnet diesen Erwartungswert der Information als die Entropie der Information.

$$\boxed{H(X) = E\left[I_m(X) \right] = -\sum_{m=1}^{M} p_m \cdot ld\left[p_m \right]}$$

$H(X)$ ...... Entropie [Bit/Symbol]

Die Entropie einer Quelle mit M Symbolen ist immer kleiner oder gleich dem Zweierlogarithmus von M. Das Gleichheitszeichen gilt für gleichwahrscheinliche Symbole $p_m = \frac{1}{M}$.

$$H(X) = -\sum_{m=1}^{M} p_m \cdot ld(p_m) \implies H_{max} = -M \cdot \frac{1}{M} \, ld\left(\frac{1}{M}\right) = ld(M)$$

$$\Rightarrow \quad 0 \leq H(X) \leq ld(M)$$

Fundamentale Theoreme der Informationstheorie:

[1] Die minimale mittlere Codewortlänge kann nicht kleiner sein als die Entropie.

[2] Besteht die Massenfunktion aus Potenzen von ½ dann existiert ein Codebaum der der Entropie entspricht.

TK 6.1)₃

Informationsrate einer diskreten Quelle:

$$R_i = R_s \cdot H(X) \quad \ldots\ldots \text{ Informationsrate } [\text{Bit}/s]$$

R_s ...... Symbolrate [Symbole/s] [Baud]

Bedingte Entropie und Redundanz einer diskreten Quelle:

Eine Quelle mit Gedächtnis wählt das aktuelle Symbol in Abhängigkeit des vorangegangenen Symbols aus [Markov-Quelle 1. Ordnung].

Mathematisch ist dies mit der gemeinsamen Entropie zweier Symbole beschreibbar.

⇒ Abhängigkeit der Symbole reduziert die maximale Entropie!
Die Differenz wird Redundanz genannt.

$$R = H_{max} - H = ld(M) - H \quad \ldots \text{ [Bit/Symbol]}$$

TK 6.2) 1

Quellkodierung

Die Entropie einer Quelle ist eine Fundamentale Größe der Quelle, und wird durch die Quellkodierung nicht verändert.

Quellkodierung liefert eine höhere Entropie in den Codewörtern als die Entropie der Quellsymbole ist, es wird Fluktuation im Informationsgehalt vermieden und so der Kanal nicht mit Symbolen belastet welche einen geringen Informationsgehalt besitzen.

Man bezeichnet Quellkodierung auch als „Noiseless Coding" da eine volle Informationsrückgewinnung der komprimierten Symbole möglich ist (im gegensatz zu z.B. Quantisierung wo Information verloren geht)

Quellkodierung reduziert die Redundanz der Quelle vor der Übertragung!

Beispiele: Morse-Alphabet (e = .), FAX (Huffmann), String Coding (Lempel-Ziv)

$$\eta_{code} = \frac{H}{H_{max}} \quad \ldots\ldots \text{ Code-Effizienz}$$

Code-Effizienz ist 1 wenn Redundanz Null wird: $\max\{\eta_{code}\} = \frac{H_{max} - \overset{0}{R}}{H_{max}} = 1$

Das Ziel der Quellkodierung ist es die Auftrittswahrscheinlichkeit aller Codewörter gleich zu machen!

$$L = \sum_{m=1}^{M} p_m \cdot l_m \quad \ldots\ldots \text{ Mittlere Codewortlänge } \left[\frac{\text{Binärstellen}}{\text{Symbol}} \right]$$

$$L = H_{max} \quad \Rightarrow \quad \eta_{code} = \frac{H}{L}$$

F33 Warum?

TK 6.2)₂

Ein effizienter Code kommt, für die gleiche Information, mit weniger Binärstellen im Mittel aus.

⇒ Codewörter mit unterschiedlicher Länge !

Jene Symbole die häufiger vorkommen, bekommen die kürzeren Codewörter!

Folgende Eigenschaften müssen beachtet werden:

[A] Eindeutige Decodierung

d.h eindeutige Umkehrung der Codierung

$M=4$ $A=[0]$
$B=[01]$
$C=[11]$
$D=[00]$

Wenn man „0011" empfängt, könnte dies „D,C" oder A,A,C sein ! ☹

[B] Fortlaufende Decodierung

$M=4$ $A=[0]$
$B=[10]$
$C=[110]$
$D=[111]$

Ist dann anwendbar, wenn kein kurzes Codewort als Vorspann (Präfix) eines Längeren Codewortes vorkommt.

⇒ Komma - Code

TK 6.3)₁

Huffman - Codierung

Huffman-Codierung ist die optimale Quellcodierung für variable Wortlänge (diskrete, gedächtnislose Quellen [DMS])!

Die Huffman-Codierung erfolgt in 2 Schritten:

(1) Schreibe alle Symbole untereinander, sortiert nach abnehmenden Wahrscheinlichkeiten. Fasse die letzten beiden zusammen und sortiere sie in der nächsten Spalte ein. → Wiederholen bis nur noch 2 Symbole vorhanden sind. Ordne den letzten beiden Symbolen eine „0" und eine „1" zu.

(2) Beschrifte die Klammern derart, das eine „0" oben und eine „1" unten steht. Der Huffman-Code ergibt sie wie folgt:

(1) Gehe vom Symbol aus waagrecht nach hinten, bis man auf Pfeile trifft. Folge diesen Pfeilen und folge der Klammer bis zum Ende.

(2) Die Anzahl der Klammern ergibt die Anzahl der Stellen des Codes.

(3) Lies den Code von hinten beginnend.

TK(6.3)₂

Eine diskrete Quelle erzeugt 8 Symbole mit folgenden Auftrittswahrscheinlichk.

Symbol	A	B	C	D	E	F	G	H
p_m	0,10	0,18	0,40	0,05	0,06	0,10	0,07	0,04

$M = 8$

C	0,40 [1]	0,40 [1]	0,40 [1]	0,40 [1]	0,40 [1]	0,40 [1]	0,60 [0]	1
B	0,18 [001]	0,18 [001]	0,18 [001]	0,19 [000]	0,23 [01]	0,37 [00]	0,40 [1]	
A	0,10 [011]	0,10 [011]	0,13 [010]	0,18 [001]	0,19 [000]	0,23 [01]		
F	0,10 [0000]	0,10 [0000]	0,10 [011]	0,13 [010]	0,18 [001]			
G	0,07 [0100]	0,09 [0001]	0,10 [0000]	0,10 [011]				
E	0,06 [0101]	0,07 [0100]	0,09 [0001]					
D	0,05 [00010]	0,06 [0101]						
H	0,04 [00011]							