

2 Informationsquellen

2.1 Modellierung und Klassifizierung von Quellen

Im Abschn. 1 wurde schon darauf hingewiesen, dass die Berechnung von Informationsmengen eine Abstraktion von realen Vorgängen, d. h. eine Modellbildung erfordert. Das trifft im Besonderen auch für Informationsquellen zu. Man muss sich deshalb im Folgenden immer bewusst sein, dass die Berechnung der Quelleninformation auf **Modelle** bezogen ist, die von realen Quellen abgebildet sind. In diesem Zusammenhang sind zwei wesentliche Aspekte hervorzuheben:

- Im Modell werden nur spezifische Eigenschaften des realen Objekts abgebildet, so dass immer eine Überprüfung notwendig ist, ob die Abbildungsgenauigkeit für den vorgesehenen Zweck ausreicht.
- Aufgrund der Abstraktion können mit gleichen Modellen physisch sehr verschiedenartige Quellen beschrieben werden, wie wir in vielen Beispielen zeigen werden.

Gemäß der SHANNONschen Theorie erfolgt die Modellbildung auf wahrscheinlichkeitstheoretischer Grundlage: Die reale Informationsquelle wird durch eine vorgegebene Wahrscheinlichkeitsverteilung auf die Menge der möglichen Ereignisse der Quelle abgebildet.

Unter einem „Ereignis“ verstehen wir die Auswahl eines Symbols oder Zeichens aus der Quelle; das kann im konkreten Fall ein Buchstabe, eine Ziffer, ein Messwert, o. Ä. sein.

Beispiele für die Abbildung realer Informationsquellen sind:

- Anzahl der Zeichen einer alphanumerischen Tastatur mit der Wahrscheinlichkeitsverteilung der Tastenanschläge,
- Anzahl unterscheidbarer Helligkeitswerte jedes Bildpunktes (z. B. eines Fernsehbildes) und ihre Auftrittswahrscheinlichkeiten,
- Anzahl und Wahrscheinlichkeitsverteilung der Amplitudenstufen eines quantisierten analogen Signals.

Die Menge der möglichen Ereignisse einer Quelle wird als Zeichenvorrat oder **Alphabet** der Quelle bezeichnet.

Ein bestimmtes Ereignis, d. h. ein konkreter Auswahlvorgang aus einem Alphabet, wird dem Informationsbegriff entsprechend als **zufälliges Ereignis** betrachtet.

Damit kann das Modell einer Informationsquelle auch auf eine zufällige Variable mit einem vorgegebenen Wertevorrat und einer zugehörigen Wahrscheinlichkeitsverteilung zurückgeführt werden.

Der Wertevorrat bzw. die Anzahl der Elemente des Alphabets kann *endlich* sein, wie beim lateinischen Alphabet der Buchstaben, oder *unendlich*, wie z. B. die Amplitudenwerte eines analogen Signals. Werden jedoch nur ganz bestimmte Amplitudenwerte der analogen Signalquelle erfasst, dann hat das Alphabet nur endlich viele oder abzählbar unendlich viele Elemente.

Unter diesem Aspekt wird eine Einteilung in **diskrete** und **kontinuierliche** (bzw. **analoge**) Quellen vorgenommen.

In realen Informationsquellen sind die Ereignisse meistens voneinander *abhängig*, d. h., ein Ereignis ist durch ein anderes oder durch mehrere andere Ereignisse bedingt, wie z. B. die Anordnung der Buchstaben in einem sinnvollen Wort. Diese Abhängigkeit wird modellmäßig durch bedingte Wahrscheinlichkeiten berücksichtigt. In vielen Anwendungsfällen kann die Abhängigkeit jedoch vernachlässigt werden, und häufig reicht sogar eine Näherungslösung durch die zusätzliche Annahme gleichwahrscheinlicher Ereignisse aus.

Diese unterschiedlichen Aspekte der Modellbildung werden auch zur Klassifizierung der Informationsquellen (Bild 2.1.1) und für die weitere Gliederung im vorliegenden Abschnitt verwendet.

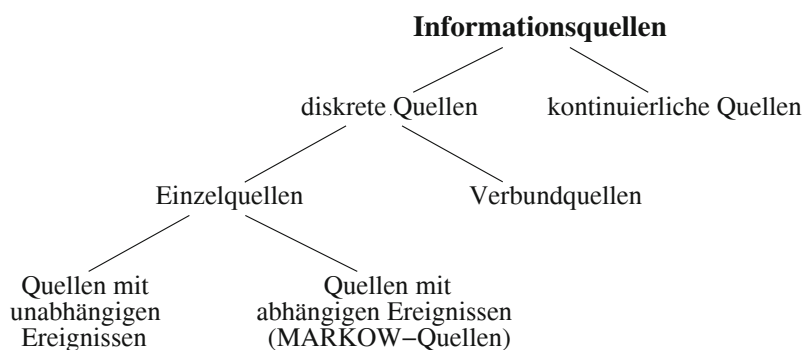


Bild 2.1.1 Klassifizierung von Informationsquellen

2.2 Diskrete Quellen

2.2.1 Diskrete Quellen mit unabhängigen Ereignissen

Im Abschn. 1.2 wurde ein Informationsmaß eingeführt (Gl. (1.1)), das nun zur Berechnung verschiedener Quellenmodelle genutzt und erweitert werden soll. Dabei werden die zunächst für ein Einzelereignis dargestellten Überlegungen auf eine endliche Menge unterschiedlicher Ereignisse übertragen.

Definition 2.2.1 *Eine Quelle mit dem Alphabet $X = \{x_1, x_2, \dots, x_N\}$ und der Verteilung der zugehörigen Auftretenswahrscheinlichkeiten $(p(x_i)) = (p(x_1), p(x_2), \dots, p(x_N))$, wobei*

$$\sum_{i=1}^N p(x_i) = 1, \quad (2.1)$$

*wird als **diskrete Quelle X mit unabhängigen Ereignissen** bezeichnet.*

Da Gl. (1.1) offensichtlich für alle $p(x_i)$ ($i = 1, 2, \dots, N$) gilt, ergeben sich

$$\begin{aligned} H_1 &= -\text{ld } p(x_1), \\ H_2 &= -\text{ld } p(x_2), \\ &\vdots \\ H_N &= -\text{ld } p(x_N). \end{aligned}$$

Die einzelnen Ereignisse liefern i. Allg. unterschiedliche Beiträge zur Unbestimmtheit bzw. zum Informationsgehalt der Quelle. Da die Ereignisse zufälligen Charakter haben, wie oben festgestellt wurde, ist auch H_i ($i = 1, 2, \dots, N$) eine Zufallsgröße, für die folgender *gewichteter* Mittelwert H_m berechnet werden kann:

$$H_m = \sum_{i=1}^N p(x_i) H_i$$

bzw.

$$H_m = \sum_{i=1}^N p(x_i) \text{ld } \frac{1}{p(x_i)} = - \sum_{i=1}^N p(x_i) \text{ld } p(x_i). \quad (2.2)$$

Anmerkungen zur Notation:

1. Im Weiteren wird zur einfacheren Schreibweise meistens $p(x_i) = p_i$ gesetzt.
2. Wo es zweckmäßig erscheint, wird $H_m = H(p_1, p_2, \dots, p_N)$ geschrieben. \square

Mit Gl. (2.2) wird die mittlere Unbestimmtheit der Quelle berechnet, die als **Entropie** [entropy] bzw. **Quellenentropie** bezeichnet wird. Aufgrund des Zusammenhangs zwischen Unbestimmtheit und Information stellt die Quellenentropie H_m damit gleichzeitig den **mittleren Informationsgehalt** [average information content] der Quelle dar.

H_m erhält die Maßeinheit *bit/Zeichen*, *bit/Messwert* u. Ä. Obwohl in der Literatur häufig nur die Einheit *bit* verwendet wird, werden wir im vorliegenden Buch aus methodischen Gründen auch immer die Bezugsgröße für *bit* angeben.

In den folgenden Sätzen werden einige wichtige Eigenschaften der Entropiefunktion gemäß Gl. (2.2) aufgeführt:

1. Die Quellenentropie H_m ist eine stetige Funktion von p_i für $0 \leq p_i \leq 1$.
2. H_m wird **maximal**, wenn alle Ereignisse der Quelle **gleichwahrscheinlich** sind.

Beweis:

Zur Extremwertbestimmung von H_m verwenden wir den LAGRANGESchen Multiplikator λ und erhalten die Hilfsfunktion

$$\begin{aligned} F &= - \sum_{i=1}^N p_i \lg p_i - \lambda \left(\sum_{i=1}^N p_i - 1 \right) \quad \text{mit} \quad \sum_{i=1}^N p_i = 1 \\ &= - \sum_{i=1}^N (p_i \lg p_i + \lambda p_i) + \lambda. \end{aligned}$$

Die partiellen Ableitungen nach p_i ($i = 1, 2, \dots, N$) ergeben

$$\frac{\partial F_i}{\partial p_i} = -\lg p_i - \lg e - \lambda.$$

Aus der Bedingung $\frac{\partial F_i}{\partial p_i} = 0$ erhalten wir schließlich

$$\lg p_i = -\lg e - \lambda \quad (i = 1, 2, \dots, N).$$

Da $\lg p_i$ für alle i gleich ist, muss

$$p_i = \frac{1}{N} \quad (i = 1, 2, \dots, N)$$

für den Extremwert gelten. Alle zweiten Ableitungen nach p_i ergeben negative Werte, d. h., es liegt tatsächlich ein Maximalwert vor.

Nach dem Einsetzen von $p_i = 1/N$ in Gl. (2.2) ergibt sich damit der **Maximalwert** der Quellenentropie

$$H_{max} = \lg N. \quad (2.3)$$

3. Eine Quelle, deren Alphabet ein *sicheres Ereignis* enthält, hat keine Unbestimmtheit:

$$H(0, \dots, 0, 1, 0, \dots, 0) = 0.$$

Der Beweis folgt unmittelbar aus Gl. (2.2).

4. Die Hinzufügung von *unmöglichen Ereignissen* zum Alphabet einer Quelle ändert nicht ihre Entropie:

$$H(p_1, p_2, \dots, p_N, 0, 0, \dots, 0) = H(p_1, p_2, \dots, p_N).$$

Der Beweis folgt unmittelbar aus Gl. (2.2).

5. Die Auflösung eines Ereignisses in Teilereignisse, für die $p_i = q_1 + q_2$ gilt, führt zu einer *Zunahme der Entropie*:

$$H_1(p_1, p_2, \dots, p_i, \dots, p_N) < H_2(p_1, \dots, q_1, q_2, \dots, p_N).$$

Beweis (für $i = N$, also $p_N = q_1 + q_2$):

$$H_1 = - \sum_{i=1}^{N-1} p_i \lg p_i - p_N \lg p_N = - \sum_{i=1}^{N-1} p_i \lg p_i - q_1 \lg p_N - q_2 \lg p_N,$$

$$H_2 = - \sum_{i=1}^{N-1} p_i \lg p_i - q_1 \lg q_1 - q_2 \lg q_2,$$

$$\begin{aligned} H_2 - H_1 &= -q_1 \lg q_1 - q_2 \lg q_2 + q_1 \lg p_N + q_2 \lg p_N, \\ &= q_1 \lg \frac{p_N}{q_1} + q_2 \lg \frac{p_N}{q_2} > 0. \end{aligned}$$

Der letzte Satz lässt folgende allgemeine Schlussfolgerung zu:

Je größer die Auflösung eines diskreten Systems ist, d. h. je feiner es strukturiert ist, um so größer ist seine Entropie bzw. sein mittlerer Informationsgehalt.

Abschließend kehren wir nochmal zum 2. Satz mit der Bemerkung zurück, dass der Maximalwert der Entropie auch als **Entscheidungsgehalt** H_0 [decision content] der Quelle bezeichnet wird. Darunter ist Folgendes zu verstehen:

In einem System zufälliger Ereignisse kann jedes Ereignis durch aufeinanderfolgende Binärentscheidungen (ja/nein, kleiner/größer, u. Ä.) bestimmt werden. Soll z. B. auf diese Weise aus einer Menge von N Zahlen eine bestimmte Zahl „erraten“ werden, so sind dazu im Mittel H_0 Fragen bzw. Binärentscheidungen erforderlich.

Von besonderem Interesse in diesem Zusammenhang ist der binäre Fall ($N = 2$) mit gleichwahrscheinlichen Ereignissen, z. B. der Wurf einer Münze mit den möglichen Ergebnissen „Kopf“ oder „Zahl“. Für diesen Fall wird die Einheit der Informationsmenge definiert.

Definition 2.2.2 *Der Entscheidungsgehalt von zwei unabhängigen und gleichwahrscheinlichen Ereignissen einer Quelle*

$$H_0 = \lg 2 = 1 \frac{\text{bit}}{\text{Ereignis}} \quad (2.4)$$

wird als **Einheit der Informationsmenge** bezeichnet.

Beispiel 2.2.1

Berechnung der Quellenentropie eines Würfels ($N = 6$) für folgende Fälle:

- idealer Würfel mit $p_i = \frac{1}{6}$ für $i = 1, 2, \dots, 6$.

Entsprechend Gl. (2.3) ergibt sich

$$H_0 = \lg 6 = \underline{2,58 \text{ bit/Ereignis}},$$

- vom Idealfall abweichender Würfel mit dem Wahrscheinlichkeitsvektor

$$(p_i) = (p_1, p_2, \dots, p_6) = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{3}{8}\right).$$

Entsprechend Gl. (2.2) erhält man

$$H_m = 5 \cdot \frac{1}{8} \lg 8 + \frac{3}{8} \lg \frac{8}{3} = \underline{2,41 \text{ bit/Ereignis}}. \quad \square$$

Diese Ergebnisse werden durch die praktische Erfahrung bestätigt, derzufolge die Unbestimmtheit (Entropie) geringer ist, wenn die Wahrscheinlichkeit für das Auftreten einer bestimmten Augenzahl beim Würfeln größer als die der anderen Augenzahlen ist.

Beispiel 2.2.2

Berechnung der Entropie einer **Binärquelle** ($N = 2$) mit $p_1 = p$ und $p_2 = 1 - p$. Durch Einsetzen der Wahrscheinlichkeiten in Gl. (2.2) ergibt sich

$$H_m = -p \lg p - (1 - p) \lg (1 - p). \quad (2.5)$$

Die logarithmische Funktion $H(p_1, p_2)$ hat die Extremwerte

$$H(0, 1) = H(1, 0) = 0 \quad \text{und}$$

$$H\left(\frac{1}{2}, \frac{1}{2}\right) = H_{\max} = 1 \text{ bit/BZ}. \quad \square$$

Anmerkung:

In der Rechentechnik und Datenverarbeitung wird grundsätzlich jedem *Bit* (binäres Element als Träger der Information) der Wert von 1 *bit* zugeordnet. Der entsprechende Informationsgehalt gemäß der SHANNONschen Theorie beträgt jedoch nur im Grenzfall 1 *bit*, d. h., wenn beide möglichen Binärzustände *gleichwahrscheinlich* sind. □

Wir betrachten abschließend noch ein **spezielles Quellenmodell**:

Ein besonderer Modellfall liegt vor, wenn sich das Wahrscheinlichkeitsfeld aus gleichwahrscheinlichen *und* nichtgleichwahrscheinlichen Ereignissen zusammensetzt. Als Beispiel kann man sich eine Skalanzeige (mit Grob- und Feinanzeige) vorstellen, bei der die ganzzahligen Werte mit sehr unterschiedlichen Wahrscheinlichkeiten auftreten können, während für die Zwischenwerte meistens Gleichwahrscheinlichkeit angenommen werden kann.

Das mathematische Modell dieses Falls kann wie folgt beschrieben werden: Gegeben sei eine diskrete Quelle mit N unabhängigen Objekten mit unterschiedlichen Auftrittswahrscheinlichkeiten p_i und M_i gleichwahrscheinlichen Elementen des i -ten Objektes ($i = 1, 2, \dots, N$).

Die Entropie des i -ten Objektes ist

$$H_i = \lg \frac{1}{p_i} + \lg M_i.$$

Als Mittelwert über alle N Objekte erhalten wir die Quellenentropie H_m in *bit/Element*:

$$H_m = \sum_{i=1}^N p_i \left(\lg \frac{1}{p_i} + \lg M_i \right). \quad (2.6)$$

Für den Fall $M_i = M$ für alle i gilt

$$H_m = \sum_{i=1}^N p_i \lg \frac{1}{p_i} + \lg M. \quad (2.7)$$

Die Gln. (2.6) und (2.7) zeigen nochmal deutlich, dass es sich bei diesem Quellenmodell um einen zweistufigen Entscheidungsprozess handelt:

1. Auswahl eines Objektes,
2. Auswahl eines Elementes aus dem entsprechenden Objekt.

Beispiel 2.2.3

Eine diskrete Quelle enthält 24 Zeichen, die in drei gleich große Gruppen mit den Auftrittswahrscheinlichkeiten $(p_i) = (0,80 \ 0,15 \ 0,05)$ unterteilt werden können. Innerhalb jeder Gruppe treten die Zeichen gleichwahrscheinlich auf. Es ist die Entropie dieser Informationsquelle zu bestimmen!

Lösung:

Mit $M = 8$, $N = 3$ und (p_i) ergibt sich nach Gl. (2.7)

$$H_m = \lg 8 - 0,8 \lg 0,8 - 0,15 \lg 0,15 - 0,05 \lg 0,05 = \underline{\underline{3,9 \text{ bit/Zeichen}}}.$$

Ohne Berücksichtigung der unterschiedlichen Gruppenwahrscheinlichkeiten wäre die Quellenentropie

$$H_0 = \text{ld } 24 = \underline{4,6 \text{ bit/Zichen}}.$$

□

Hinweis: **Aufgaben** s. Abschn. 2.4

2.2.2 Diskrete Quellen mit abhängigen Ereignissen (MARKOW-Quellen)

2.2.2.1 Beschreibung diskreter MARKOW-Quellen

Im vorangegangenen Abschnitt wurde angenommen, dass die aufeinanderfolgenden Ereignisse einer Quelle voneinander unabhängig sind. Man spricht in diesem Fall auch von einer „Quelle ohne Gedächtnis“. Diese Modellannahme ist zwar oft bei praktisch hinreichender Genauigkeit gerechtfertigt, weicht jedoch meistens erheblich von der Realität ab, wie z. B. bei der Verwendung des Alphabets der lateinischen Buchstaben. Schon aus Erfahrung wissen wir, dass in jedem sinnvollen Text bestimmte Buchstabenverbindungen (z. B. en, de, ch) sehr viel häufiger als andere vorkommen, d. h., zwischen diesen existieren stärkere Abhängigkeiten.

Die **Abhängigkeit** zeigt sich darin, dass das Eintreten eines Ereignisses (die Auswahl eines Quellenzeichens) von den *vorangegangenen* Ereignissen bestimmt wird. Mit anderen Worten: *Das Ereignis $x^{(m+1)}$ tritt unter der Bedingung ein, dass ganz bestimmte Ereignisse $x^{(1)}, x^{(2)}, \dots, x^{(m)}$ bereits eingetreten sind.*

Diese m Ereignisse stellen den **Zustand** der Quelle vor dem Eintreten des Ereignisses $x^{(m+1)}$ dar. Bei jedem Ereignis (Auswahl eines Quellenzeichens) geht die Quelle in einen Folgezustand über, der durch die m zuletzt ausgewählten Quellenzeichen bestimmt wird und von dem die Auswahl des jeweils nächsten Quellenzeichens abhängt.

Die Auswahl des Quellenzeichens $x^{(m+1)}$ erfolgt demnach mit der bedingten Wahrscheinlichkeit

$$p(x^{(m+1)} | x^{(m)} \dots x^{(2)} x^{(1)}). \quad (2.8)$$

Ein Quellenmodell, bei dem die Quellenzeichen entsprechend Gl. (2.8) ausgewählt werden, wird als „MARKOW-Quelle m -ter Ordnung“ oder auch als „Quelle mit Gedächtnis“ bezeichnet.

Wir werden uns hier auf den wichtigen Modellfall „MARKOW-Quellen **erster Ordnung**“ beschränken, bei dem die Auftrittswahrscheinlichkeiten der Ereignisse immer nur von dem *zuletzt* eingetretenen Ereignis $x^{(m)}$ abhängen.

Da der jeweilige Zustand der Quelle in diesem Fall nur vom Ereignis $x^{(m)}$ bestimmt wird, spricht man bei MARKOW-Quellen erster Ordnung auch meistens von **Zustandswahrscheinlichkeiten** anstelle von Ereignis- oder Zeichenwahrscheinlichkeiten.

Die bedingte Wahrscheinlichkeit ist in diesem Modellfall gemäß Gl. (2.8)

$$p(x^{(m+1)}|x^{(m)}),$$

wofür wir im Folgenden die Schreibweise

$$p(x_j|x_i) \quad (i, j = 1, 2, \dots, N) \quad (2.9)$$

verwenden werden. Sie ist die Wahrscheinlichkeit dafür, dass das Ereignis bzw. der Zustand x_j eintreten wird, wenn der Zustand x_i vorliegt. Weil damit der Übergang vom Zustand x_i in den Zustand x_j ausgedrückt wird, bezeichnet man $p(x_j|x_i)$ auch als **Übergangswahrscheinlichkeit**.

Definition 2.2.3 *Eine MARKOW-Quelle ist das mathematische Modell einer Informationsquelle, bei dem die aufeinanderfolgende Auswahl von Quellenzeichen, d. h. die Folge der Zustände, sowohl von der momentanen Verteilung der Auftritts- bzw. Zustandswahrscheinlichkeiten als auch von der Verteilung der Übergangswahrscheinlichkeiten abhängt.*

Der Begriff „momentane Verteilung“ weist darauf hin, dass die Auftritts- bzw. Zustandswahrscheinlichkeiten (im Gegensatz zu Quellen mit unabhängigen Ereignissen) bei MARKOW-Quellen i. Allg. zeitlich veränderlich sind. Die Folge dieser Wahrscheinlichkeitsverteilungen der Zustände bzw. Quellenzeichen nennen wir MARKOW-**Kette** (s. Bild 2.2.1).

Die MARKOW-Kette kann in jedem diskreten Zeitpunkt nach dem Satz von der vollständigen Wahrscheinlichkeit wie folgt berechnet werden:

$$p(x_j) = \sum_{i=1}^N p(x_i) p(x_j|x_i) \quad (j = 1, 2, \dots, N). \quad (2.10)$$

Anmerkung:

Im Gegensatz zu den Zustandswahrscheinlichkeiten soll die Verteilung der Übergangswahrscheinlichkeiten zeitlich invariant sein. \square

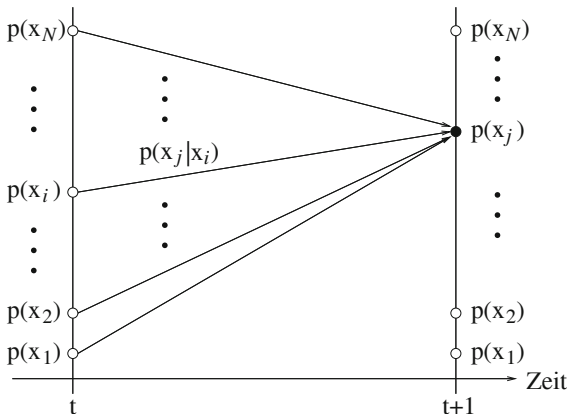


Bild 2.2.1 Zustandswahrscheinlichkeiten $p(x_i)$, $p(x_j)$ und Übergangswahrscheinlichkeiten $p(x_j|x_i)$ ($i, j = 1, 2, \dots, N$)

Beispiel 2.2.4

Eine diskrete Quelle sei durch die Anfangswahrscheinlichkeiten (Vektor der Zustandswahrscheinlichkeiten zum Zeitpunkt $t = 0$) $(p_i^{(0)}) = (1 \ 0 \ 0)$ und die folgende Matrix der Übergangswahrscheinlichkeiten bestimmt:

$$(p(x_j|x_i)) = \begin{pmatrix} p(x_1|x_1) & p(x_2|x_1) & p(x_3|x_1) \\ p(x_1|x_2) & p(x_2|x_2) & p(x_3|x_2) \\ p(x_1|x_3) & p(x_2|x_3) & p(x_3|x_3) \end{pmatrix} = \begin{pmatrix} 0 & 0,2 & 0,8 \\ 0,1 & 0,9 & 0 \\ 0,2 & 0,4 & 0,4 \end{pmatrix}.$$

Zu berechnen sind die Zustandswahrscheinlichkeiten der MARKOW-Kette für eine hinreichend große Zahl von Übergängen!

Lösung:

Entsprechend Gl. (2.10) wird die Verteilung der Zustandswahrscheinlichkeiten zum Zeitpunkt $(t + 1)$ wie folgt berechnet:

$$(p_j^{(t+1)}) = (p_i^{(t)}) (p(x_j|x_i)).$$

Zustandswahrscheinlichkeiten zum Zeitpunkt $t = 1$:

$$p_1^{(1)} = p_1^{(0)} p(x_1|x_1) + p_2^{(0)} p(x_1|x_2) + p_3^{(0)} p(x_1|x_3) = 1 \cdot 0 + 0 \cdot 0,1 + 0 \cdot 0,2 = 0,$$

$$p_2^{(1)} = p_1^{(0)} p(x_2|x_1) + p_2^{(0)} p(x_2|x_2) + p_3^{(0)} p(x_2|x_3) = 1 \cdot 0,2 + 0 \cdot 0,9 + 0 \cdot 0,4 = 0,2,$$

$$p_3^{(1)} = p_1^{(0)} p(x_3|x_1) + p_2^{(0)} p(x_3|x_2) + p_3^{(0)} p(x_3|x_3) = 1 \cdot 0,8 + 0 \cdot 0 + 0 \cdot 0,4 = 0,8.$$

Zustandswahrscheinlichkeiten zum Zeitpunkt $t = 2$:

$$p_1^{(2)} = 0 \cdot 0 + 0,2 \cdot 0,1 + 0,8 \cdot 0,2 = 0,18,$$

$$p_2^{(2)} = 0 \cdot 0,2 + 0,2 \cdot 0,9 + 0,8 \cdot 0,4 = 0,50,$$

$$p_3^{(2)} = 0 \cdot 0,8 + 0,2 \cdot 0 + 0,8 \cdot 0,4 = 0,32.$$

Für die weiteren Übergänge wollen wir uns nur die Berechnungsergebnisse ansehen:

Hervorzuheben ist hierbei der Fall **symmetrischer Übergangswahrscheinlichkeiten**

$$p(x_2|x_1) = p(x_1|x_2) = p,$$

$$p(x_1|x_1) = p(x_2|x_2) = 1 - p,$$

mit den stationären Zustandswahrscheinlichkeiten

$$\overline{p_1} = \overline{p_2} = \frac{1}{2}.$$

Eine Binärquelle mit gleichen Zustandswahrscheinlichkeiten und symmetrischen Übergangswahrscheinlichkeiten (was häufig, wenigstens näherungsweise, zutrifft) befindet sich demzufolge von Anfang an im stationären Zustand.

Eine anschauliche Beschreibung von MARKOW-Quellen ist auch durch **Zustandsgraphen** möglich (Bild 2.2.2), in denen die Ereignisse (Zustände) durch Knoten und die Übergänge (mit den zugehörigen Übergangswahrscheinlichkeiten) durch Kanten dargestellt werden (kantenbewerteter gerichteter Graph).

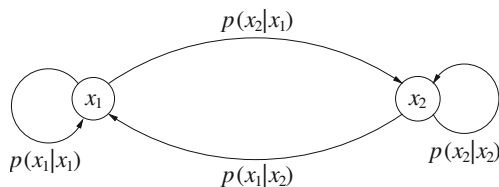


Bild 2.2.2 Zustandsgraph einer binären MARKOW-Quelle 1. Ordnung

2.2.2.2 Entropie diskreter MARKOW-Quellen

Wie wir wissen, ist die Entropie ein Maß für die Unbestimmtheit eines Systems zufälliger Ereignisse. Bei Quellen mit N nicht gleichwahrscheinlichen und voneinander abhängigen Zuständen liegt Unbestimmtheit in der Hinsicht vor, dass man für einen bestimmten Zeitpunkt nicht genau voraussagen kann,

- welcher von N möglichen Zuständen gerade vorliegt und
- welcher von N möglichen Übergängen als nächster eintreten wird.

Zunächst wollen wir feststellen, dass bei dem angenommenen Fall immer beide Teile der Unbestimmtheit vorliegen, und zwar unabhängig davon, welches Quellenmodell verwendet wird. Sobald man aber eine Unbestimmtheit durch eine Wahrscheinlichkeitsverteilung beschreibt und diese bei der Entropieberechnung berücksichtigt, wird ein Teil der ursprünglichen Unbestimmtheit beseitigt, d. h., die wirkliche Entropie wird kleiner.

Bei MARKOW-Quellen (Definition 2.2.3) werden sowohl die Zustands- als auch die Übergangswahrscheinlichkeiten berücksichtigt. Man kann demzufolge

erwarten, dass die Entropie kleiner als beim Quellenmodell mit unabhängigen Zuständen sein wird.

Zuerst wollen wir die Unbestimmtheit, die in den Übergangsmöglichkeiten von einem beliebigen x_i zu allen x_j ($j = 1, 2, \dots, N$) liegt, berechnen. Analog zur Gleichung für die Entropie unabhängiger Ereignisse (Gl. (2.2)) erhält man

$$H_i = \sum_{j=1}^N p(x_j|x_i) \lg \frac{1}{p(x_j|x_i)}.$$

Den anderen Teil der Unbestimmtheit erfasst man durch Mittelwertbildung über alle x_i ($i = 1, 2, \dots, N$), d. h. durch Wichtung der einzelnen Beträge H_i mit den entsprechenden Auftrittswahrscheinlichkeiten $p(x_i)$:

$$H_m = \sum_{i=1}^N p(x_i) H_i.$$

Der Mittelwert H_m , der die Entropie bzw. den mittleren Informationsgehalt der MARKOW-Quelle erster Ordnung darstellt, wird für den stationären Fall $p(x_i) = \overline{p(x_i)}$ als **MARKOW-Entropie** H_M bezeichnet:

$$H_M = \sum_{i=1}^N \sum_{j=1}^N \overline{p(x_i)} p(x_j|x_i) \lg \frac{1}{p(x_j|x_i)} \quad \text{in} \quad \frac{\text{bit}}{\text{Zustand}}. \quad (2.12)$$

Anmerkung:

Da wir uns auf MARKOW-Quellen *erster* Ordnung beschränken, könnte die Maßeinheit auch *bit/Ereignis* oder *bit/Quellenzeichen* heißen. \square

Beispiel 2.2.5

Berechnung der MARKOW-Entropie einer diskreten Quelle mit $N = 3$ voneinander abhängigen Zuständen (mit den im Beispiel 2.2.4 gegebenen Zustands- und Übergangswahrscheinlichkeiten) entsprechend Gl. (2.12):

$$\begin{aligned} H_M &= 0,10 \left(0,2 \lg \frac{1}{0,2} + 0,8 \lg \frac{1}{0,8} \right) + 0,76 \left(0,1 \lg \frac{1}{0,1} + 0,9 \lg \frac{1}{0,9} \right) \\ &\quad + 0,14 \left(0,2 \lg \frac{1}{0,2} + 2 \cdot 0,4 \lg \frac{1}{0,4} \right) \\ &= \underline{0,64 \text{ bit/Zustand}}. \end{aligned}$$

Zum Vergleich soll die Entropie dieser Quelle im stationären Zustand bestimmt werden, wenn die Abhängigkeiten dabei unberücksichtigt bleiben:

$$H_m = 0,10 \lg \frac{1}{0,10} + 0,76 \lg \frac{1}{0,76} + 0,14 \lg \frac{1}{0,14} = \underline{1,03 \text{ bit/Zustand}}.$$

Wird darüber hinaus noch Gleichwahrscheinlichkeit der Zustände angenommen, ergibt sich eine maximale Entropie

$$H_0 = \lg 3 = \underline{1,58 \text{ bit/Zustand}}. \quad \square$$

Die Ergebnisse des Beispiels 2.2.5 bestätigen den objektiven Zusammenhang zwischen dem Maß der Unbestimmtheit und der Entropie bzw. dem mittleren Informationsgehalt der Quelle: Je mehr „Vorinformation“ in Form von Wahrscheinlichkeitsverteilungen berücksichtigt wird, um so geringer ist die verbleibende Unbestimmtheit. Wir sollten aber nicht vergessen, dass es sich in diesem Beispiel nur um unterschiedliche Modelle *einer* Quelle handelt.

Die Anwendung des MARKOW-Modells (unter der Voraussetzung der Realisierbarkeit) würde zu einer beträchtlichen Reduzierung der zu verarbeitenden oder zu speichernden Informationsmengen führen. Auf diesen Aspekt, der mit der Kodierung zusammenhängt, werden wir im Abschn. 3.4.2.4 näher eingehen.

Hinweis: **Aufgaben** s. Abschn. 2.4

2.2.2.3 Spezielle MARKOW-Modelle

Bei dem oben betrachteten MARKOW-Modell werden zu jedem diskreten Zeitpunkt *alle möglichen* Zustandsübergänge berücksichtigt. Dadurch ist das Modell universell anwendbar. In speziellen Anwendungsfällen kann es aber sinnvoll sein, das Modell so an die konkrete Quelle anzupassen, dass viele Übergangsmöglichkeiten „verdeckt“ bleiben. Auf diese Weise kann die Modellkomplexität wesentlich verringert werden. Für diese angepassten Modelle wurde der Begriff **Hidden-MARKOW-Modell** (HMM) eingeführt.

Wir wollen uns hier auf die kurze Beschreibung eines praktischen Anwendungsfalls beschränken, und zwar auf ein „phonetisches Strukturmodell als Hidden-Markow-Modell“ [SCH 92].

Die automatische Sprachverarbeitung setzt eine Modellierung der natürlichen Sprache voraus. Eine Möglichkeit dazu ist, die in der Sprache enthaltenen **Phoneme** (kleinste bedeutungstragende Lauteinheiten) durch spezielle MARKOW-Modelle nachzubilden.

Man hat erkannt, dass jedes Phonem durch 6 aufeinanderfolgende Zustände (z_1, z_2, \dots, z_6) hinreichend genau beschrieben werden kann. Diese Zustände sind zeitlich so angeordnet, dass man in jedem Zustand nur drei Übergangsmöglichkeiten berücksichtigen muss: man verweilt im jeweiligen Zustand *oder* geht zum nächstfolgenden über *oder* überspringt einen Zustand (Bild 2.2.3).

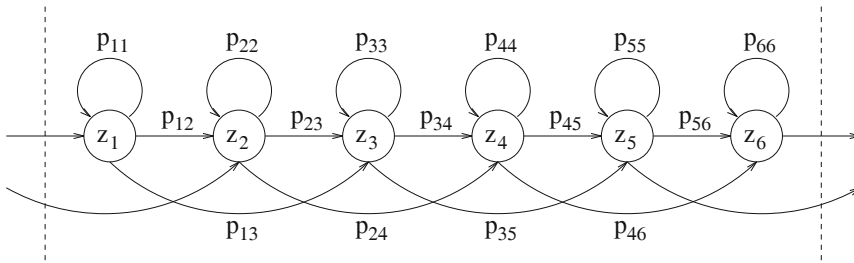


Bild 2.2.3 *Hidden-MARKOW-Modell eines Phonems* ($p_{ij} = p(z_j|z_i)$)

Hidden-MARKOW-Modelle vereinfachen damit auch das Darstellen von Zusammenhängen. Betreffs weiterer Einzelheiten sowie spezieller Modellberechnungen wird auf die oben angegebene Literatur verwiesen.

2.2.3 Verbundquellen

Wir betrachten in diesem Abschnitt gleichzeitig zwei diskrete Quellen X und Y mit den zugehörigen Verteilungen der Auftretswahrscheinlichkeiten $(p(x_i)) = (p(x_1), p(x_2), \dots, p(x_N))$ der Zeichen $x_i \in X$ und $(p(y_j)) = (p(y_1), p(y_2), \dots, p(y_M))$ der Zeichen $y_j \in Y$. Dabei wollen wir annehmen, dass die *Ereignisse innerhalb jeder Einzelquelle voneinander unabhängig* sind.

Für die Modellbeschreibung ist es dabei unwesentlich, ob es sich wirklich um zwei verschiedene Quellen handelt oder nur um eine Quelle mit zwei verschiedenen Ereignismengen. Wesentlich ist dagegen, dass die Ereignisse beider Quellen voneinander abhängig sind, d. h., dass ein Ereignis der einen Quelle von einem vorausgegangenen Ereignis der anderen Quelle bestimmt wird.

Wir wollen im Weiteren den Fall zugrunde legen, dass immer zuerst in der Quelle X ein Ereignis stattfindet, das unmittelbar danach ein *bedingtes Ereignis* in der Quelle Y mit der bedingten Wahrscheinlichkeit $p(y_j|x_i)$ auslöst.

Dieses Auftreten von zwei Ereignissen x_i ($i = 1, 2, \dots, N$) und y_j ($j = 1, 2, \dots, M$) bezeichnet man als **Verbundereignis** (x_i, y_j) , das durch eine **Verbundwahrscheinlichkeit** $p(x_i, y_j)$ bewertet wird.

Da bei der konjunktiven Verknüpfung von Ereignissen bekanntlich der Multiplikationssatz der Wahrscheinlichkeitslehre gilt, ergibt sich für die Verbundwahrscheinlichkeit

$$p(x_i, y_j) = p(x_i) \cdot p(y_j|x_i). \quad (2.13)$$

Definition 2.2.4 Die diskreten Quellen X und Y mit den Verbundwahrscheinlichkeiten $p(x_i, y_j)$ für $i = 1, 2, \dots, N$ und $j = 1, 2, \dots, M$ bilden eine **Verbundquelle** (X, Y) .

Da die Verbundquelle allein durch eine Menge diskreter Wahrscheinlichkeiten eindeutig beschrieben wird, kann für die Entropieberechnung prinzipiell der gleiche Ansatz wie für diskrete Einzelquellen verwendet werden. Indem in Gl. (2.2) $p(x_i)$ durch $p(x_i, y_j)$ ersetzt wird, erhält man die Entropie der Verbundquelle bzw. die **Verbundentropie** [joint entropy]

$$H(X, Y) = - \sum_{i=1}^N \sum_{j=1}^M p(x_i, y_j) \operatorname{ld} p(x_i, y_j). \quad (2.14)$$

Um weitere Aussagen zur Verbundentropie zu gewinnen, setzen wir Gl. (2.13) in Gl. (2.14) ein,

$$H(X, Y) = - \sum_{i=1}^N \sum_{j=1}^M p(x_i) p(y_j|x_i) \operatorname{ld} (p(x_i) p(y_j|x_i)),$$

und erhalten nach einigen Umformungen

$$H(X, Y) = - \sum_i p(x_i) \operatorname{ld} p(x_i) - \sum_i \sum_j p(x_i) p(y_j|x_i) \operatorname{ld} p(y_j|x_i).$$

Im ersten Term erkennen wir die Quellenentropie $H(X)$, der zweite Term stellt die **bedingte Entropie** [conditional entropy] $H(Y|X)$ dar:

$$H(Y|X) = - \sum_i \sum_j p(x_i) p(y_j|x_i) \operatorname{ld} p(y_j|x_i). \quad (2.15)$$

Damit erhalten wir schließlich folgende Formel für die Verbundentropie:

$$H(X, Y) = H(X) + H(Y|X). \quad (2.16)$$

Für den Fall, dass zuerst in der Quelle Y ein Ereignis auftritt, d. h., dass

$$p(x_i, y_j) = p(y_j) \cdot p(x_i|y_j)$$

ist, erhält man im Ergebnis

$$H(X, Y) = H(Y) + H(X|Y) \quad (2.17)$$

mit der bedingten Entropie

$$H(X|Y) = - \sum_j \sum_i p(y_j) p(x_i|y_j) \operatorname{ld} p(x_i|y_j). \quad (2.18)$$

Die gewonnenen Ergebnisse können in einem VENN-Diagramm (Flächendiagramm) (Bild 2.2.4) anschaulich dargestellt werden. Das Bild zeigt, dass der Grad der Abhängigkeit beider Quellen formal im Grad der Überdeckung beider Kreisflächen und damit in der Größe der bedingten Entropien zum Ausdruck kommt.

Unter Nutzung dieser Interpretation können wir bei der Angabe der folgenden Schranken für die bedingten Entropien auf einen mathematischen Beweis verzichten:

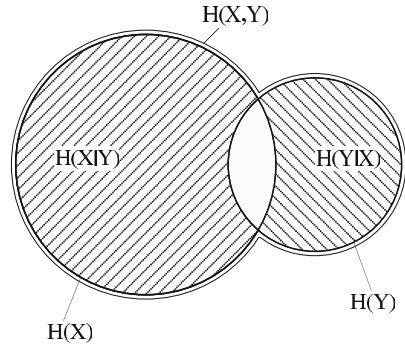


Bild 2.2.4 VENN-Diagramm einer Verbundquelle (X, Y)

$$H(X|Y) \leq H(X) \quad \text{und} \quad H(Y|X) \leq H(Y). \quad (2.19)$$

Es sollen jetzt noch zwei interessante *Grenzfälle* der Abhängigkeiten beider Quellen betrachtet werden (Bild 2.2.5):

a) Vollständige Unabhängigkeit:

Bei unabhängigen Ereignissen gilt $p(y_j|x_i) = p(y_j)$, d. h.

$H(Y|X) = H(Y)$ entsprechend Gl. (2.19) und damit

$$H(X, Y) = H(X) + H(Y).$$

b) Vollständige Abhängigkeit:

Bei vollständig abhängigen Ereignissen ist $H(Y|X) = 0$ und damit wird

$$H(X, Y) = H(X).$$

(Für den Fall, dass zuerst in der Quelle Y ein Ereignis stattfindet, wäre $H(X, Y) = H(Y)$.)

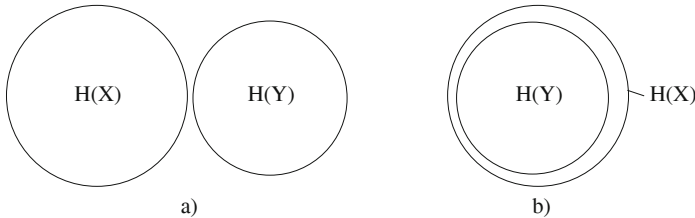


Bild 2.2.5 Grenzfälle für Verbundquellen

a) vollständig unabhängig, b) vollständig abhängig

Beispiel 2.2.6

Gegeben seien zwei diskrete Quellen X und Y mit folgenden Verbundwahrscheinlichkeiten:

$$(p(x_i, y_j)) = \begin{pmatrix} p(x_1, y_1) & p(x_1, y_2) & \dots & p(x_1, y_M) \\ p(x_2, y_1) & p(x_2, y_2) & \dots & p(x_2, y_M) \\ \dots & \dots & \dots & \dots \\ p(x_N, y_1) & p(x_N, y_2) & \dots & p(x_N, y_M) \end{pmatrix} = \begin{pmatrix} \frac{1}{8} & 0 & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{32} & \frac{1}{32} \\ 0 & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \end{pmatrix}.$$

Zu berechnen sind:

- a) Einzelwahrscheinlichkeiten $(p(x_i)), (p(y_j))$,
- b) bedingte Wahrscheinlichkeiten $(p(y_j|x_i))$,
- c) Entropien $H(X), H(Y), H(Y|X), H(X, Y)$.

Lösung:

zu a)

$$p(x_i) = \sum_{j=1}^M p(x_i, y_j) \text{ ergibt: } p(x_1) = \frac{1}{4}, p(x_2) = p(x_3) = \frac{1}{8}, p(x_4) = \frac{1}{2}.$$

$$p(y_j) = \sum_{i=1}^N p(x_i, y_j) \text{ ergibt: } p(y_1) = \frac{5}{16}, p(y_2) = \frac{9}{32}, p(y_3) = \frac{13}{32}.$$

zu b)

Nach Gl. (2.13) ist $p(y_j|x_i) = \frac{p(x_i, y_j)}{p(x_i)}$ und damit wird

$$(p(y_j|x_i)) = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

zu c)

$$\begin{aligned} H(X) &= \sum_{i=1}^N p(x_i) \lg \frac{1}{p(x_i)} = \frac{1}{4} \lg 4 + 2 \cdot \frac{1}{8} \lg 8 + \frac{1}{2} \lg 2 \\ &= \underline{1,75 \text{ bit/Ereignis}}, \end{aligned}$$

$$\begin{aligned} H(Y) &= \sum_{j=1}^M p(y_j) \lg \frac{1}{p(y_j)} = \frac{5}{16} \lg \frac{16}{5} + \frac{9}{32} \lg \frac{32}{9} + \frac{13}{32} \lg \frac{32}{13} \\ &= \underline{1,57 \text{ bit/Ereignis}}, \end{aligned}$$

$$\begin{aligned}
H(Y|X) &= \sum_i \sum_j p(x_i) p(y_j|x_i) \lg \frac{1}{p(y_j|x_i)} \\
&= \frac{1}{4} \left(2 \cdot \frac{1}{2} \lg 2 \right) + \frac{1}{8} \left(\frac{1}{2} \lg 2 + 2 \cdot \frac{1}{4} \lg 4 \right) + \frac{1}{8} \lg 1 + \frac{1}{2} \left(2 \cdot \frac{1}{4} \lg 4 + \frac{1}{2} \lg 2 \right) \\
&= \underline{1,19 \text{ bit/Ereignis}},
\end{aligned}$$

$$\begin{aligned}
H(X, Y) &= H(X) + H(Y|X) \\
&= 1,75 \text{ bit/Ereignis} + 1,19 \text{ bit/Ereignis} \\
&= \underline{2,94 \text{ bit/Verbundereignis}}.
\end{aligned}$$

Kontrolle: Berechnung der Verbundentropie nach Gl. (2.14)

$$\begin{aligned}
H(X, Y) &= \frac{1}{4} \lg 4 + 5 \cdot \frac{1}{8} \lg 8 + \frac{1}{16} \lg 16 + 2 \cdot \frac{1}{32} \lg 32 \\
&= \underline{2,94 \text{ bit/Verbundereignis}}.
\end{aligned}$$

Grenzfälle:

a) Vollständige Unabhängigkeit

$$H(X, Y) = H(X) + H(Y) = \underline{3,32 \text{ bit/Verbundereignis}}.$$

b) Vollständige Abhängigkeit

$$H(X, Y) = H(X) = \underline{1,75 \text{ bit/Verbundereignis}}. \quad \square$$

Wir wollen nun noch die **spezielle Verbundquelle** (X, X) betrachten, bei der beide Quellen identisch sind.

Stellt man sich diesen Modellfall als *eine Quelle mit zwei identischen Ereignismengen* vor, dann wird mit der Verbundwahrscheinlichkeit $p(x_i, x_j)$ die Abhängigkeit von zwei aufeinanderfolgenden Ereignissen einer Quelle X ausgedrückt, und die bedingte Entropie $H(X|X)$ ist nichts anderes als die bekannte MARKOW-Entropie H_M erster Ordnung (Bild 2.2.6).

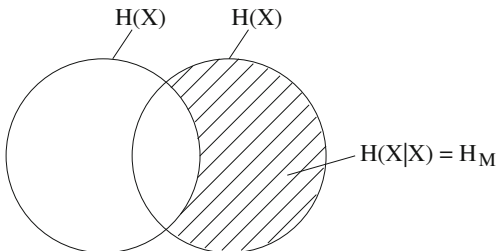


Bild 2.2.6 VENN-Diagramm der Verbundquelle (X, X)

In diesem Fall gilt für die Verbundentropie

$$H(X) \leq H(X, X) \leq 2 \cdot H(X),$$

und somit wird die MARKOW-Entropie

$$H_M = H(X, X) - H(X) = \begin{cases} H(X) & \text{bei vollständiger Unabhängigkeit,} \\ 0 & \text{bei vollständiger Abhängigkeit.} \end{cases}$$

Zur Demonstration dieser speziellen Verbundquelle dient das folgende Beispiel.

Beispiel 2.2.7

Eine diskrete Quelle $X = \{x_1, x_2, x_3\}$ sei durch folgende Matrix der Verbundwahrscheinlichkeiten beschrieben:

$$(p(x_i, x_j)) = \begin{pmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & \frac{7}{32} \\ \frac{1}{4} & \frac{3}{32} & \frac{1}{8} \end{pmatrix}.$$

Zu berechnen sind:

- Quellenentropie $H(X)$,
- bedingte Entropie $H(X|X)$ als MARKOW-Entropie.

Lösung:

zu a)

$$p(x_i) = \sum_j p(x_i, x_j) \text{ ergibt: } p(x_1) = \frac{5}{16}, p(x_2) = \frac{7}{32}, p(x_3) = \frac{15}{32}.$$

$$p(x_j) = \sum_i p(x_i, x_j) \text{ ergibt: } p(x_1) = \frac{5}{16}, p(x_2) = \frac{7}{32}, p(x_3) = \frac{15}{32}.$$

Die Gleichheit der Ergebnisse bedeutet, dass sich die Quelle im stationären Zustand befindet (Bedingung für die Berechnung der MARKOW-Entropie!).

$$H(X) = \frac{5}{16} \lg \frac{16}{5} + \frac{7}{32} \lg \frac{32}{7} + \frac{15}{32} \lg \frac{32}{15} = \underline{1,52 \text{ bit/Zeichen}}.$$

zu b)

$$H(X|X) = H_M = \sum_i \sum_j p(x_i) p(x_j|x_i) \lg \frac{1}{p(x_j|x_i)},$$

$$(p(x_j|x_i)) = \left(\frac{p(x_i, x_j)}{p(x_i)} \right) = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \\ 0 & 0 & 1 \\ \frac{8}{15} & \frac{3}{15} & \frac{4}{15} \end{pmatrix},$$

$$\begin{aligned} H_M &= \frac{5}{16} \left(\frac{1}{5} \lg 5 + 2 \cdot \frac{2}{5} \lg \frac{5}{2} \right) + \frac{15}{32} \left(\frac{8}{15} \lg \frac{15}{8} + \frac{3}{15} \lg \frac{15}{3} + \frac{4}{15} \lg \frac{15}{4} \right) \\ &= \underline{1,16 \text{ bit/Zeichen}}. \end{aligned}$$

□

Abschließend soll noch erwähnt werden, dass die Verbundquelle als Modellklasse nicht nur die Verbindung zu MARKOW-Quellen herstellt, sondern auch die Grundlage zur wahrscheinlichkeitstheoretischen Beschreibung der Informationsübertragung bildet (s. Abschn. 4.4.2, Kanalmodell).

Hinweis: **Aufgaben** s. Abschn. 2.4

2.3 Kontinuierliche Quellen

Das von einer kontinuierlichen Quelle ausgehende Signal kann in einem vorgegebenen Bereich jeden beliebigen Wert annehmen, d. h., die Menge der möglichen Ereignisse dieser Quelle ist unbegrenzt. Unter dem Informationsaspekt sind auch in diesem Fall nur *zufällige* Ereignisse von Bedeutung. Wir interessieren uns deshalb auch nur für zufällige Signale, deren Amplitudenwerte meistens eine charakteristische Verteilung (z. B. Gleich- oder Normalverteilung) aufweisen.

Aus der Mathematik wissen wir, dass für stetige Zufallsgrößen die **Wahrscheinlichkeitsdichte** eine charakteristische Kennfunktion darstellt. In Analogie zur Auftrittswahrscheinlichkeit bei diskreten Ereignissen kann die Wahrscheinlichkeitsdichte interpretiert werden als die Wahrscheinlichkeit, mit der ein zu einem bestimmten Zeitpunkt auftretender Funktionswert des zufälligen Signals $x(t)$ in ein bestimmtes Intervall Δx fällt (wobei $\Delta x \rightarrow 0$).

Zur Berechnung der **Entropie einer kontinuierlichen Quelle** bzw. eines kontinuierlichen Signals wollen wir von einer diskreten Betrachtung der stetigen Dichtefunktion ausgehen, damit bereits bekannte Beziehungen von den diskreten Quellen übernommen werden können.

Dazu denkt man sich die Fläche unter der Dichtefunktion $f(x)$ in Teile gleicher Breite Δx zerlegt (Bild 2.3.1).

Das Integral einer Teilfläche der Breite Δx gibt dann die Wahrscheinlichkeit $p(x_i)$ dafür an, dass die zufällige Größe x_i im Bereich Δx liegt:

$$p(x_i) = \int_{\Delta x} f(x) dx \approx f(x_i) \Delta x.$$

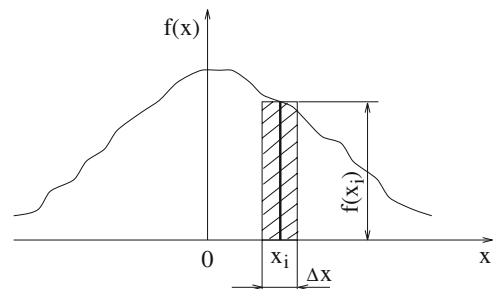


Bild 2.3.1 Wahrscheinlichkeitsdichtefunktion

Durch Einsetzen dieser Beziehung in Gl. (2.2) für diskrete Ereignisse erhält man

$$\begin{aligned} H(X) &= - \sum_i f(x_i) \Delta x \operatorname{ld} (f(x_i) \Delta x) \\ &= - \sum_i f(x_i) \Delta x \operatorname{ld} f(x_i) - \sum_i f(x_i) \Delta x \operatorname{ld} \Delta x. \end{aligned}$$

Um nun zur Entropie einer **kontinuierlichen** Quelle zu gelangen, muss der Grenzübergang $\Delta x \rightarrow 0$ durchgeführt werden. Das gelingt nicht vollständig, denn

$$\lim_{\Delta x \rightarrow 0} \operatorname{ld} \Delta x$$

würde eine unendlich große Entropie $H(X)$ ergeben, was offensichtlich der Realität widerspricht. Betrachtet man die Stufenbreite Δx als Maß für die Auflösung der stetigen Funktion in praktisch unterscheidbare Amplitudenwerte (was der praktisch möglichen Genauigkeit bei der Informationserfassung entspricht), dann hat Δx immer einen Wert, der größer als Null ist und, im Gegensatz zur Funktion $x(t)$, nicht zufällig ist.

Nach dem Grenzübergang für die übrigen Ausdrücke in der obigen Gleichung erhalten wir

$$H(X) = - \int_{-\infty}^{\infty} f(x) \operatorname{ld} f(x) dx - \operatorname{ld} \Delta x. \quad (2.20)$$

Da Δx unter gleichen Bedingungen als konstant angesehen werden kann, lässt man das Glied $\operatorname{ld} \Delta x$ in Gl. (2.20) meistens weg und spricht von der **relativen Entropie** einer kontinuierlichen Quelle:

$$H_{rel} = - \int_{-\infty}^{\infty} f(x) \operatorname{ld} f(x) dx. \quad (2.21)$$

Beispiel 2.3.1

Wir bestimmen die Entropie für zwei praktisch wichtige Fälle von kontinuierlichen (analogen) Zufallssignalen.

1. **Amplitudenbegrenztes** Signal mit einer **Gleichverteilung** der Funktionswerte im Bereich $-a \leq x \leq a$:

Einsetzen von $f(x) = \frac{1}{2a}$ in Gl. (2.21) ergibt

$$\begin{aligned} H_{rel} &= \int_{-a}^a \frac{1}{2a} \operatorname{ld} (2a) dx, \\ H_{rel} &= \operatorname{ld} (2a). \end{aligned} \quad (2.22)$$

2. **Leistungsbegrenztes** Signal mit einer **Normalverteilung** der Funktionswerte x bei gegebener mittlerer Leistung P :

Wir setzen die Dichtefunktion der Normalverteilung

$$f(x) = \frac{1}{\sqrt{2\pi P}} e^{-\frac{x^2}{2P}}$$

in Gl. (2.21) ein und erhalten:

$$\begin{aligned} H_{rel} &= \int_{-\infty}^{\infty} f(x) \operatorname{ld} \left(\sqrt{2\pi P} e^{\frac{x^2}{2P}} \right) dx \\ &= \operatorname{ld} \sqrt{2\pi P} \int_{-\infty}^{\infty} f(x) dx + \frac{1}{2P} \operatorname{ld} e \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \operatorname{ld} \sqrt{2\pi P} + \frac{1}{2} \operatorname{ld} e, \\ H_{rel} &= \frac{1}{2} \operatorname{ld} (2\pi e P). \end{aligned} \tag{2.23}$$

□

Aus dem Beispiel geht hervor, dass die Entropie neben dem Integrationsbereich vor allem von der Art der Dichtefunktion bestimmt wird. Wir haben in diesem Beispiel bereits die Dichtefunktionen gewählt, die im jeweiligen Fall eine **maximale Entropie** ergeben.

Die entsprechende Beweisführung, die hier nur grob skizziert werden kann, erfolgt nach der Methode der LAGRANGESchen Multiplikatoren λ_1 und λ_2 :

1. Für das **amplitudenbegrenzte Signal** mit der Nebenbedingung

$$\int_{-a}^a f(x) dx = 1$$

erhalten wir aus Gl. (2.21)

$$F = - \int_{-a}^a f(x) \operatorname{ld} f(x) dx - \lambda_1 \left(\int_{-a}^a f(x) dx - 1 \right).$$

Die partielle Ableitung

$$\frac{\partial F}{\partial f(x)} = 0$$

beinhaltet die Lösung eines Variationsproblems und liefert die Extremwertbedingung (Extremwert gleich Maximum, da zweite Ableitung negativ!):

$$-\ln f(x) - 1 - \lambda_1 \ln 2 = 0.$$

Daraus folgt

$$f(x) = e^{-(1+\lambda_1 \ln 2)} = \text{const}.$$

Dieses Ergebnis bedeutet, dass die **Gleichverteilung** den Maximalwert der Entropie ergibt.

2. Für das **leistungsbegrenzte Signal** erhält man durch die zusätzliche Nebenbedingung

$$\int_{-\infty}^{\infty} x^2 f(x) dx = P$$

in analoger Weise wie oben:

$$-\ln f(x) - 1 - \lambda_1 \ln 2 - (\lambda_2 \ln 2) x^2 = 0.$$

Daraus folgt

$$f(x) = e^{-(1+\lambda_1 \ln 2)} e^{-(\lambda_2 \ln 2) x^2}.$$

Indem dieser Ausdruck in die beiden Nebenbedingungen eingesetzt wird, können mit Hilfe bekannter Integralformeln für $e^{-(\lambda_2 \ln 2) x^2}$ die Unbekannten λ_1 und λ_2 bestimmt werden und man erhält schließlich

$$f(x) = \frac{1}{\sqrt{2\pi P}} e^{-\frac{x^2}{2P}}.$$

Dieses Ergebnis zeigt, dass bei leistungsbegrenzten Signalen die **Normalverteilung** zum Entropiemaximum führt.

Zum Vergleich wollen wir jetzt noch die Entropie bei **Gleichverteilung leistungsbegrenzter Signale** (H_{GV}) berechnen:

Für $f(x) = \frac{1}{2a}$ mit $-a \leq x \leq a$ wird die mittlere Leistung

$$P = \int_{-a}^a x^2 f(x) dx = \frac{1}{2a} \int_{-a}^a x^2 dx = \frac{a^2}{3}.$$

Setzt man $a = \sqrt{3P}$ in Gl. (2.22) ein, so ergibt sich eine Entropie

$$H_{GV} = \text{ld}(2a) = \frac{1}{2} \text{ld}(12P).$$

Die entsprechende Entropie bei Normalverteilung gemäß Gl. (2.23) ist

$$H_{NV} = \frac{1}{2} \lg(17 P).$$

Sollen die Entropien bei Normal- und Gleichverteilung gleich sein ($H_{NV} = H_{GV}$), dann ergibt sich für die mittleren Leistungen $17 P_{NV} = 12 P_{GV}$, d. h., es müsste $P_{GV} = 1,42 P_{NV}$ sein.

2.4 Aufgaben

Abschn. 2.2.1: Quellen mit unabhängigen Ereignissen

1. Berechnen Sie den mittleren Informationsgehalt (Entropie) einer diskreten Quelle mit sechs voneinander unabhängigen Zeichen, wenn
 - a) $p_1 = 0,5$ $p_2 = 0,2$ $p_3 = p_4 = 0,1$ $p_5 = p_6 = 0,05$
 - b) alle Zeichen gleichwahrscheinlich sind.
2. Bestimmen Sie den mittleren Informationsgehalt einer Buchseite!
Für die Berechnung sind anzunehmen:
45 unabhängige und gleichwahrscheinliche Zeichen,
40 Zeilen und 65 Zeichen/Zeile.
3. Eine automatische Waage mit binärer Messwerterfassung habe einen Messbereich von 0 ... 100 g bei einer Schrittweite von 1 g.
 - a) Bestimmen Sie den mittleren Informationsgehalt je Messwert!
 - b) Wie groß wird der mittlere Informationsgehalt bei einer Schrittweite von 0,1 g?
4. Ein Rasterbild bestehe aus 10^5 Bildpunkten mit folgender Wahrscheinlichkeitsverteilung der Helligkeitswerte H_i :

H_1 : 50,00 %,
 H_2 : 25,00 %,
 H_3 : 12,50 %,
 H_4 : 6,25 %,
 H_5 : 6,25 %.

 - a) Wie groß ist der mittlere Informationsgehalt eines Bildes?
 - b) Wie groß wäre der mittlere Informationsgehalt, wenn die Auftrittswahrscheinlichkeiten der Helligkeitsstufen unbekannt sind?
5. Der Amplitudenbereich eines zufälligen Signals soll in sieben Intervalle eingeteilt sein, für die folgende Auftrittswahrscheinlichkeiten der Amplitudenwerte ermittelt wurden:
 $(p(x_i)) = (0,47 \quad 0,25 \quad 0,13 \quad 0,07 \quad 0,04 \quad 0,02 \quad 0,02)$.
 Innerhalb jedes Intervalls, bestehend aus 16 Teilintervallen, wird gleichwahrscheinliches Auftreten der Amplitudenwerte angenommen.

- a) Bestimmen Sie die Quellenentropie, d. h. den mittleren Informationsgehalt je Amplitudenwert!
- b) Wie groß ist die maximale Entropie?
6. Das Alphabet einer Informationsquelle bestehe aus den Zahlen 1 bis 100, die mit folgenden Wahrscheinlichkeiten auftreten:
- 1 bis 25: $p(x_1) = \frac{1}{6}$,
- 26 bis 70: $p(x_2) = \frac{1}{3}$,
- 71 bis 100: $p(x_3) = \frac{1}{2}$.

Innerhalb der Teilbereiche treten die Zahlen mit gleichen Wahrscheinlichkeiten auf.

- a) Berechnen Sie die Entropie dieser Quelle!
- b) Wie groß wäre die Entropie bei gleichwahrscheinlichem Auftreten aller Zahlen?

Abschn. 2.2.2: MARKOW-Quellen

1. Eine ergodische Informationsquelle habe ein Alphabet mit drei Zeichen, wobei folgende Abhängigkeiten zwischen den Zeichen bestehen:

$$(p(x_j|x_i)) = \begin{pmatrix} 0,5 & 0,2 & 0,3 \\ 0,1 & 0,6 & 0,3 \\ 0,2 & 0,1 & 0,7 \end{pmatrix}.$$

- a) Bestimmen Sie die stationären (ergodischen) Wahrscheinlichkeiten $\overline{p(x_i)}$!
- b) Bestimmen Sie die MARKOW-Entropie dieser Quelle!
2. Die Steuerung eines automatischen Teilefertigungsprozesses erfordert die laufende Qualitätsprüfung der produzierten Teile. Dabei sollen drei Güteklassen (Zustände z_1, z_2, z_3) unterschieden werden, für die folgende Verteilung der Auftrittswahrscheinlichkeiten zum Zeitpunkt $t = 0$ anzunehmen ist:

$$p(z_1)^{(0)} = 0,9 \quad p(z_2)^{(0)} = 0,1 \quad p(z_3)^{(0)} = 0.$$

Für den Fertigungsprozess wurde folgendes Übergangsverhalten der Zustände statistisch ermittelt:

$$(p(z_j|z_i)) = \begin{pmatrix} 0,60 & 0,38 & 0,02 \\ 0,15 & 0,80 & 0,05 \\ 0,40 & 0,60 & 0 \end{pmatrix}.$$

Berechnen Sie

- a) die Zustandswahrscheinlichkeiten für die Zeitpunkte $t = 1, 2, \dots, 5$,
- b) den mittleren Informationsgehalt (MARKOW-Entropie) je Prüfergebnis, wobei die Zustandswahrscheinlichkeiten zum Zeitpunkt $t = 5$ als stationär anzunehmen sind.
3. Ein sogenanntes „System mit Erneuerung“ mit den Zuständen z_1 (Funktionstüchtigkeit) und z_2 (Ausfall) habe eine Ausfallrate λ und eine Reparaturrate μ . Das Übergangsverhalten des Systems soll durch folgende Matrix beschrieben sein:

$$(p(z_j|z_i)) = \begin{pmatrix} 1-\lambda & \lambda \\ \mu & 1-\mu \end{pmatrix}.$$

Bestimmen Sie (allgemein)

- die stationären Zustandswahrscheinlichkeiten,
- die MARKOW-Entropie des Systems.

Abschn. 2.2.3: Verbundquellen

- Gegeben seien zwei Signalquellen A und B mit jeweils drei Farbsignalen (rot, gelb, grün) und folgenden Verbundwahrscheinlichkeiten:

A	a_1	a_2	a_3
B	(rot)	(gelb)	(grün)
b_1 (rot)	$\frac{2}{9}$	$\frac{1}{18}$	$\frac{1}{9}$
b_2 (gelb)	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{9}$
b_3 (grün)	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{1}{6}$

Berechnen Sie

- Einzelentropien $H(A)$, $H(B)$,
 - bedingte Entropien $H(A|B)$, $H(B|A)$,
 - Verbundentropie $H(A, B)$.
- Von zwei diskreten Quellen X und Y sei folgende Matrix der Verbundwahrscheinlichkeiten gegeben:

$$(p(x_i, y_j)) = \begin{pmatrix} 0,12 & 0,10 & 0,08 & 0,05 & 0,03 \\ 0,02 & 0,04 & 0,12 & 0,04 & 0,02 \\ 0,03 & 0,05 & 0,08 & 0,10 & 0,12 \end{pmatrix}.$$

Gesucht sind:

- Einzelentropien $H(X)$, $H(Y)$,
- bedingte Entropie $H(X|Y)$,
- Verbundentropie $H(X, Y)$
 - für den gegebenen Fall,
 - für den Fall, dass y_j ($j = 1, 2, \dots, 5$) auf ein sicheres Ereignis $x_i \in X$ führt.

Abschn. 2.3: Kontinuierliche Quellen

- Berechnen Sie die Entropie eines kontinuierlichen Signals mit symmetrischer Exponentialverteilung

$$f(x) = \frac{a}{2} e^{-a|x|}, \quad -\infty < x < \infty.$$

Informations- und Kodierungstheorie

Schönfeld, D.; Klimant, H.; Piotraschke, R.

2012, VIII, 296 S. 40 Abb., Softcover

ISBN: 978-3-8348-0647-5