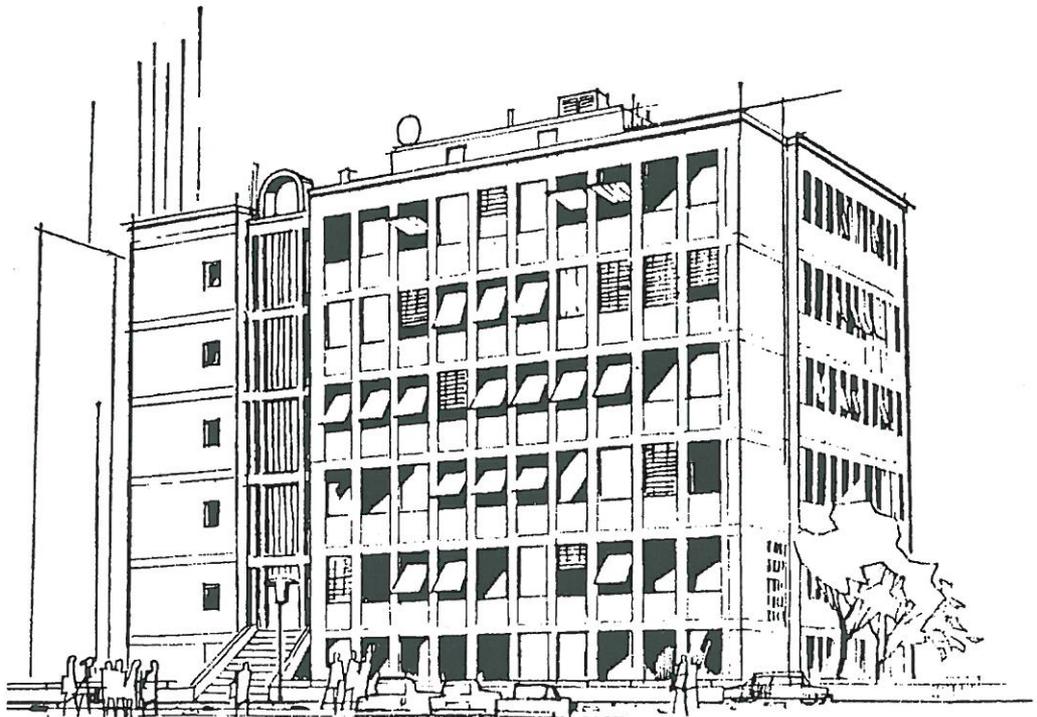

G. Söder: Simulation digitaler Übertragungssysteme

Wertdiskrete Informationstheorie

Lehrstuhl für Nachrichtentechnik
Technische Universität München



Vorwort

Den Abschluß des Praktikums *Simulation digitaler Übertragungssysteme* bildet ein Versuch zur Informationstheorie, wie sie beispielsweise an der Technischen Universität München in den Vorlesungen [7] und [12] gelehrt wird. Da erfahrungsgemäß doch einige Studierende mit dem Verständnis dieses teilweise abstrakten Lehrgebietes Probleme haben, ist die Bereitstellung eines anschaulichen Lehrprogramms für dieses Spezialfach besonders wichtig.

Die Informationstheorie ist die Wissenschaftsdisziplin, die sich mit den theoretischen Grundlagen der Kommunikations- und Informationstechnik befaßt. Dazu gehören beispielsweise die heute üblichen Kompressions- und Übertragungsverfahren, die Mehrbenutzerkommunikation und die Kryptographie. Es ist das überragende Verdienst einer einzelnen Person, nämlich von *Claude E. Shannon*, all diese Gebiete begründet zu haben. Auf ihn gehen die Definitionen der Entropie einer Nachricht und der Kapazität eines Übertragungskanals ebenso zurück wie grundlegende Aussagen zur Quellen- und Kanalcodierung.

Eine der berühmtesten Aussagen Shannons ist das sogenannte Kanalcodierungstheorem, das besagt, daß man bei Anwendung der Kanalcodierung ab einer bestimmten Grenze fehlerfrei übertragen kann, auch wenn der Kanal starke Störungen aufweist, wie es z.B. beim Mobilfunk der Fall ist (vgl. Versuch „*Mobilfunkkanal*“ in diesem Praktikum).

Die Shannonsche Informationstheorie läßt sich sowohl auf wertdiskrete als auch auf wertkontinuierliche Quellen anwenden. In diesem Versuch beschränken wir uns aus Zeitgründen nur auf den ersteren Fall, wie er beispielsweise bei einem geschriebenen Text vorliegt.

Die vorliegende Anleitung gliedert sich wieder in die drei Kapitel *Theoretische Grundlagen*, *Vorbereitungsfragen* und *Versuchsdurchführung*. Anhand dieser Anleitung und des zugrundeliegenden Windows-Lehrprogramms „*wdit*“ sollte ein Teilnehmer mit Grundkenntnissen der Nachrichtentechnik in der Lage sein, sich die wesentlichen Merkmale der Informationstheorie in ca. 8 Stunden (4 Stunden Vorbereitung sowie 4 Stunden Versuchsdurchführung) zu erarbeiten.

Die Musterlösungen der Vorbereitungsfragen und der Versuchsdurchführung sind am Ende des Heftes (ab Seite 47) beigelegt. Auf der nächsten Seite finden Sie eine Zusammenfassung einschlägiger Literaturstellen, die für eine erfolgreiche Versuchsdurchführung durchaus nützlich sein können, jedoch nicht unbedingt vorausgesetzt werden.

Das Programm „*Wertdiskrete Informationstheorie*“ inklusive der dazugehörigen Anleitung wurde 1994 von Herrn *Dipl.-Ing. Georg Delle* im Rahmen seiner von mir betreuten Diplomarbeit am Lehrstuhl für Nachrichtentechnik der Technischen Universität München mit großer Sorgfalt konzipiert und implementiert, wofür ich ihm an dieser Stelle herzlich danken möchte.

Dank gebührt auch Herrn *Dr.-Ing. Michael Fleischmann*, der diese Diplomarbeit mitbetreut hat und viele Ideen in das Programm eingebracht hat. Bei den theoretischen Grundlagen wurde oft auf das Vorlesungsmanuskript von Herrn *Prof. Dr.-Ing. Joachim Hagenauer* zurückgegriffen; auch ihm möchte ich für sein Interesse herzlich danken.

Ich wünsche Ihnen nun viel Erfolg und Spaß beim Durcharbeiten dieses Versuchs und bedanke mich gleichzeitig für Ihr Engagement während des gesamten Praktikums. Sollten Sie an weiteren Lehrprogrammen zur Informationstechnik interessiert sein, so verweise ich auf das Praktikum *Simulationenmethoden in der Nachrichtentechnik*, das im Sommersemester stattfindet.

Literaturverzeichnis

Zur Vorbereitung dieses Praktikumsversuchs genügt es durchaus, die theoretischen Grundlagen (Kapitel 1) durcharbeiten und die Vorbereitungsfragen (Kapitel 2) zu beantworten. Die nachfolgende Liste gibt eine Übersicht von zusätzlichen Fachbüchern und Aufsätzen, die die hier behandelte Thematik zum Inhalt haben.

- [1] Anderson, J. B.; Mohan, S.: *Source and Channel Coding*. Norwell (Mass.): Kluwer Academic Publisher, 1990.
- [2] Blahut, R. E.: *Principles and Practice of Information Theory*. Massachusetts: Addison-Wesley, 1987.
- [3] Cover, T.; Thomas J.A.: *Elements of Information Theory*. New York: John Wiley & Sons, 1991.
- [4] Fano, R. M.: *Transmission of Information: A Statistical Theory of Communication*. New York: John Wiley & Sons, 1961.
- [5] Forney, G. D.: *Information Theory*. Stanford University, 1972.
- [6] Gallager, R. G.: *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.
- [7] Hagenauer, J.: *Informationstheorie und Quellencodierung*. Vorlesungsmanuskript, Lehrstuhl für Nachrichtentechnik, Technische Universität München, 1997.
- [8] Johannesson, R.: *Informationstheorie*. Lund: Studentlitteratur, 1992.
- [9] Küpfmüller, K.: *Die Entropie der deutschen Sprache*. Fernmeldetechnische Zeitung 7 (1954), S. 265-272.
- [10] Lüke, H. D.: *Signalübertragung*. Berlin: Springer-Verlag, 1992.
- [11] McEliece, R. J.: *The Theory of Information Theory and Coding*. Massachusetts: Addison-Wesley, 1977.
- [12] Mecking, M.: *Information Theory and Source Coding*. Vorlesungsmanuskript, Lehrstuhl für Nachrichtentechnik, Technische Universität München, 1999.
- [13] Shannon, C. E.: *A Mathematical Theory of Communication*. Bell Syst. Techn. J., Band 27 (1948), S. 379-432 (Teil I) und S. 623-656 (Teil II).
- [14] Shannon, C. E.: *Prediction and Entropy of Printed English*. Bell Syst. Techn. J., Band 30 (1951), S. 50-64.
- [15] Söder, G.: *Modellierung, Simulation und Optimierung von Nachrichtensystemen*. Berlin: Springer-Verlag, 1993.

Hinweis: Das interaktive Graphikprogramm „*wdit*“ ist Bestandteil des Lehrsoftwarepakets *LNTwin*, das wir - ebenso wie die Programme der Reihe *LNTsim* - an Hochschuleinrichtungen zu günstigen Konditionen weitergeben. Bei diesbezüglichen Fragen - aber auch bei Kritik jeder Art - wenden Sie sich bitte an:

Priv.-Doz. Dr.-Ing. habil. Günter Söder
Lehrstuhl für Nachrichtentechnik, Technische Universität München
Arcisstr. 21, D-80290 München, Tel.: (089) 289-23486, Fax: (089) 289-23490
Email: guenter.soeder@ei.tum.de

Inhalt

1	Theoretische Grundlagen	1
1.1	Allgemeines.....	1
1.2	Definitionen einiger Grundbegriffe	2
1.2.1	Wahrscheinlichkeitsdichtefunktion.....	2
1.2.2	Erwartungswerte und Momente	3
1.2.3	Informationsgehalt und Entropie.....	3
1.2.4	Entscheidungsgehalt und Redundanz.....	5
1.2.5	Verbundentropie und Transinformationsgehalt	5
1.3	Gedächtnislose wertdiskrete Nachrichtenquellen	7
1.3.1	Transinformation zwischen zwei Folgen	7
1.3.2	Kanalkapazität eines Digitalkanals	11
1.3.3	Grundlagen der Quellencodierung	13
1.3.4	Mittlere Codewortlänge.....	14
1.3.5	Codierung nach Huffman	16
1.3.6	Codierung nach Lempel und Ziv.....	18
1.4	Nachrichtenquellen mit inneren statistischen Bindungen	19
1.4.1	Entropiebestimmung	19
1.4.2	Quellen mit Markoveigenschaften	21
1.4.3	Entropie von Texten.....	23
1.4.4	Entropieabschätzung nach Küpfmüller	24
1.4.5	Einige Simulationsergebnisse	27
2	Vorbereitungsfragen	29
V1:	Shannonfunktion	29
V2:	Entropie einer Ternärquelle (M=3).....	29
V3:	Transinformationsgehalt	30
V4:	Kanalkapazität eines unsymmetrischen Binärkanals	31
V5:	Huffman-Codierung	32
V6:	Lempel-Ziv-Codierung.....	33
V7:	Entropie von Markovprozessen.....	34
3	Versuchsdurchführung.....	35
D1:	Shannonfunktion	35
D2:	Statistisch unabhängige Symbole.....	36
D3:	Huffman-Codierung.....	37
D4:	Mittlere Codewortlänge.....	38
D5:	Entropie der Binärübertragung.....	39
D6:	Kanalkapazität eines Binärkanals.....	40
D7:	Kanalkapazität eines Ternärkanals.....	41
D8:	Entropie einer binären Markovquelle.....	43
D9:	Entropie einer ternären Markovquelle.....	44
D10:	Analyse einer Textdatei.....	45
D11:	Synthetische Erzeugung von Texten	46

4	Musterlösungen der Vorbereitungsfragen	47
	Lösung V1	47
	Lösung V2	47
	Lösung V3	48
	Lösung V4	48
	Lösung V5	49
	Lösung V6	50
	Lösung V7	50
5	Musterlösungen der Versuchsdurchführung	51
	Ergebnisse D1.....	51
	Ergebnisse D2.....	51
	Ergebnisse D3.....	52
	Ergebnisse D4.....	52
	Ergebnisse D5.....	53
	Ergebnisse D6.....	53
	Ergebnisse D7.....	53
	Ergebnisse D8.....	54
	Ergebnisse D9.....	55
	Ergebnisse D10.....	55
	Ergebnisse D11.....	56

1 Theoretische Grundlagen

Inhalt: Im Mittelpunkt dieses Versuches steht die Informationstheorie, die theoretische Grenzen für die Digitalsignalübertragung bereitstellt. Behandelt wird hier nur der wertdiskrete Fall, also Signale, die auch durch Symbolfolgen dargestellt werden können. Nach einigen allgemeinen Definitionen wie Informations- und Entscheidungsgehalt, Entropie und Redundanz werden die Transinformation zwischen zwei Folgen berechnet und die Kanalkapazität eines Digitalkanals angegeben. Anschließend folgen die Grundlagen der Quellencodierung, z.B. die praktisch relevanten Kompressionsverfahren nach Huffman und Lempel/Ziv. Der theoretische Teil endet mit einigen Anmerkungen zur Entropieabschätzung von Texten.

1.1 Allgemeines

Seit den ersten Anfängen der Nachrichtenübertragung als naturwissenschaftliche Disziplin war es das Bestreben vieler Ingenieure und Mathematiker, ein quantitatives Maß für die in einer *Nachricht* enthaltene *Information* zu finden. Hierbei soll unter „Information“ ganz allgemein die Kenntnis über irgend etwas verstanden werden, während wir im folgenden eine „Nachricht“ stets als eine Zusammenstellung von Symbolen und/oder Zuständen betrachten, die zur Übermittlung von Information dient. Die (abstrakte) Information wird durch die (konkrete) Nachricht mitgeteilt. Information kann in vielerlei Hinsicht als Interpretation einer Nachricht aufgefaßt werden.

Aber erst *Claude E. Shannon* gelang es im Jahre 1948, eine in sich konsistente Theorie über den Informationsgehalt von Nachrichten zu begründen, die zu ihrer Zeit revolutionär war und ein neues, bis heute hochaktuelles Wissenschaftsgebiet kreierte: die nach ihm benannte *Shannonsche Informationstheorie*.

Entsprechend dieser Theorie ist die in einem Nachrichtensymbol enthaltene *Information* um so größer, je unwahrscheinlicher das Symbol ist. Daß als mathematische Umsetzung dieser Regel der negative Logarithmus herangezogen wird, erscheint zunächst als willkürlich - und ist es in mancher Hinsicht tatsächlich auch. Es zeigt sich aber, daß die konsequente Anwendung dieses logarithmischen Informationsmaßes nicht zu Inkonsistenzen führt, d.h. die hierfür zugrundegelegte Mathematik liefert stets informationstheoretisch interpretierbare Ergebnisse.

Zur Verdeutlichung der informationstheoretischen Grundlagen wird stets eine stochastische *zeitdiskrete Nachrichtenquelle* betrachtet, die eine Folge von Quellensymbolen abgibt:

$$\langle q_n \rangle = q_1, q_2, \dots, q_N . \quad (1.1)$$

Je größer N ist, d.h. je länger die Folge beobachtet wird, um so fundierter sind die Aussagen einer auf diese Quelle angewandten statistischen Analyse. Desweiteren wird für das Folgende vorausgesetzt, daß die Nachrichtenquelle wertdiskret ist und die einzelnen Symbole einem endlichen Wertevorrat entstammen. Das bedeutet:

$$q_n = S_m \quad \text{mit} \quad n = 1, 2, \dots, N \quad \text{und} \quad m = 1, 2, \dots, M . \quad (1.2)$$

Hierbei wird M als *Symbolumfang* (aus informationstheoretischer Sicht) bzw. als *Stufenzahl* (entsprechend den Konventionen der Übertragungstechnik) bezeichnet. Es ist anzumerken, daß gesicherte statistische Aussagen hinsichtlich der Nachrichtenquelle nur dann möglich sind, wenn die Folgenlänge N um ein Vielfaches größer ist als der Symbolumfang M .

1.2 Definitionen einiger Grundbegriffe

1.2.1 Wahrscheinlichkeitsdichtefunktion

Die Wahrscheinlichkeit, daß das n -te Symbol der zeitlichen Folge gleich dem m -ten Element des Symbolvorrats ist, sei gleich

$$p_{nm} = p(q_n = S_m). \quad (1.3)$$

Die Definition gilt allgemein, also auch für nichtstationäre Nachrichtenquellen. Bei Stationarität sind alle Wahrscheinlichkeiten unabhängig vom Zeitpunkt n . Für diesen meist betrachteten Sonderfall sind die *Symbolwahrscheinlichkeiten* (mit nur einer Laufvariablen $m = 1, 2, \dots, M$)

$$p_m = p(q_n = S_m) \quad (1.4)$$

für alle Zeitpunkte n identisch. Für den Sonderfall $M = 2$ spricht man von einer Binärquelle, z.B. mit dem Symbolvorrat $S_1 = „A“$ und $S_2 = „B“$.

Ist die Folgenlänge N hinreichend groß und die Quelle stationär und ergodisch, so können die Wahrscheinlichkeiten nach (1.4) durch *relative Häufigkeiten* angenähert werden. Es gilt:

$$h_m = \frac{N_m}{N} \quad (m = 1, 2, \dots, M), \quad (1.5)$$

wobei N_m die empirisch ermittelte Anzahl des Symbols S_m bezeichnet. Aufgrund der Beziehung

$$\sum_{m=1}^M N_m = N \quad (1.6)$$

gilt für die relativen Häufigkeiten ebenso wie für die Wahrscheinlichkeiten der M Symbole:

$$\sum_{m=1}^M h_m = \sum_{m=1}^M p_m = 1. \quad (1.7)$$

Die durch (1.4) bzw. (1.5) definierten Beschreibungsgrößen lassen sich für jeden beliebigen Symbolvorrat angeben, z.B. auch für eine binäre Quelle mit den zwei möglichen Symbolen „A“ und „B“. Dagegen ist die Angabe der Wahrscheinlichkeitsdichtefunktion gemäß (1.8) und der Erwartungswerte (siehe nachfolgenden Abschnitt) nur bei reellwertigem Symbolvorrat möglich.

Bei beliebigen Alphabeten umgeht man diese Schwierigkeit durch die Einführung einer geeignet definierten *Zufallsgröße*. Darunter versteht man die Abbildung der möglichen Symbole auf reellwertige Zahlen, wie das nachfolgende einfache Beispiel zeigen soll.

Beispiel: Binärquelle ($M = 2$)

$$q_n \in \{S_1, S_2\} \Rightarrow x_n \in \{x_1, x_2\}$$

$$S_1 = „A“ \Rightarrow x_1 = 0,$$

$$S_2 = „B“ \Rightarrow x_2 = 1.$$

Die Abbildung der Symbole auf Zahlenwerte ist prinzipiell beliebig. Aus der Aufgabenstellung ergeben sich jedoch meist von selbst sinnvolle Zuordnungen: So ist die im obigen Beispiel gewählte Abbildung zur Beschreibung eines binären unipolaren Übertragungssystems angemessen, wenn die beiden möglichen Symbole durch die Spannungswerte 0V bzw. 1V dargestellt werden.

Die Erwartungswerte erster und zweiter Ordnung (siehe nächster Abschnitt) geben dann den Gleichanteil bzw. die Leistung des Nachrichtensignals wieder (jeweils normiert).

Dagegen würde sich bei bipolarer Übertragung mit den möglichen Amplitudenstufen $-1V$ und $+1V$ folgende Zufallsgröße anbieten:

$$x_n \in \{-1, +1\}.$$

Durch die Einführung der Zufallsgröße x läßt sich die *Wahrscheinlichkeitsdichtefunktion* einer wertdiskreten Quelle wie folgt angeben (vgl. [15]):

$$f_x(x) = \sum_{m=1}^M p_m \cdot \delta(x - x_m). \quad (1.8)$$

Diese besteht aus einer Summe von M Diracfunktionen, deren Lagen durch die möglichen Werte der Zufallsgröße festliegen. Die Gewichtung erfolgt nach den Auftrittswahrscheinlichkeiten.

Im folgenden ist stets Ergodizität vorausgesetzt. Diese besagt, daß die statistischen Eigenschaften der Nachrichtenquelle sowohl aus der zeitlichen Folge (Laufvariable $n = 1, 2, \dots, N$) als auch anhand des Symbolvorrats (Laufvariable $m = 1, 2, \dots, M$) abgeleitet werden können.

1.2.2 Erwartungswerte und Momente

Die WDF liefert weitreichende Informationen über eine betrachtete Zufallsgröße. Reduzierte Informationen erhält man durch die Erwartungswerte und Momente. Der *Erwartungswert* bezüglich einer beliebigen Funktion $g(x)$ berechnet sich mit der WDF in folgender Weise:

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) \cdot f_x(x) dx. \quad (1.9)$$

Handelt es sich um eine diskrete Zufallsgröße x mit dem Symbolvorrat M , so kann mit den Symbolwahrscheinlichkeiten gemäß (1.4) hierfür auch geschrieben werden:

$$E[g(x)] = \sum_{m=1}^M p_m \cdot g(x_m). \quad (1.10)$$

Sonderfälle von Erwartungswerten sind die *Momente*. Mit

$$g(x) = x^k \quad (1.11)$$

erhält man im Fall einer diskreten Zufallsgröße für das Moment k -ter Ordnung:

$$m_k = E[x^k] = \sum_{m=1}^M p_m \cdot x_m^k. \quad (1.12)$$

Beispielsweise liefert diese Gleichung für $k = 1$ den *linearen Mittelwert (Gleichanteil)* und für $k = 2$ den *quadratischen Mittelwert (Leistung)* einer diskreten Zufallsgröße.

Anzumerken ist, daß die Gln. (1.9) bis (1.12) nicht angewandt werden können, wenn der Symbolvorrat nicht aus (reellwertigen) Zahlen besteht, sondern aus Symbolen wie „A“, „B“, usw.

1.2.3 Informationsgehalt und Entropie

Mit der Wahrscheinlichkeit (1.3), daß das n -te Symbol der zeitlichen Folge gleich dem m -ten Element des Symbolvorrats ist, wird der *Informationsgehalt* dieses Symbols wie folgt berechnet:

$$I_n = \text{ld} \frac{1}{p_{nm}} = -\text{ld}(p_{nm}). \quad (1.13)$$

Hierbei ist $\text{ld}(\cdot)$ der Logarithmus zur Basis zwei (*logarithmus dualis*), der aus dem natürlichen Logarithmus $\ln(\cdot)$ nach folgender Gleichung berechnet werden kann:

$$\text{ld}(x) = \frac{\ln(x)}{\ln(2)}. \quad (1.14)$$

Bei der numerischen Auswertung von (1.13) wird meist die Hinweiseinheit „bit“ hinzugefügt. Nach neueren Definitionen (Stand 1994) war hierfür auch die Einheit „Shannon“ im Gespräch; dies hat sich jedoch nicht durchgesetzt.

Der Informationsgehalt eines Symbols ist stets positiv bzw. im Grenzfall (d.h. beim sicheren Symbol mit der Symbolwahrscheinlichkeit 1) gleich 0. Je unwahrscheinlicher ein Symbol ist, desto größer ist sein Informationsgehalt.

Durch Zeitmittelung über alle Symbole der unendlich langen Folge erhält man die *Entropie* der digitalen Quelle:

$$H = \overline{I_n} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N I_n. \quad (1.15)$$

Diese entspricht somit dem mittleren Informationsgehalt eines Symbols und besitzt dementsprechend ebenfalls die Einheit „bit“. H ist auch ein Maß für die „Zufälligkeit“ der Symbole.

Für die nachfolgenden Abschnitte wird vorausgesetzt, daß die Folgeelemente statistisch voneinander unabhängig sind und somit die zeitabhängigen Wahrscheinlichkeiten nach (1.3) durch die zeitunabhängigen Symbolwahrscheinlichkeiten entsprechend (1.4) ersetzt werden können:

$$p_{nm} = p(q_n = S_m) = p_m. \quad (1.16)$$

Unter dieser Einschränkung kann anstelle der Zeitmittelung in (1.15) auch eine Scharmittelung erfolgen, und es ergibt sich für die Entropie unter Ausnutzung der Ergodizität:

$$H = \sum_{m=1}^M p_m \cdot \text{ld} \frac{1}{p_m} = - \sum_{m=1}^M p_m \cdot \text{ld}(p_m). \quad (1.17)$$

Ein Vergleich mit (1.10) macht deutlich, daß H als der Erwartungswert hinsichtlich der Funktion von Gl. (1.13), also dem Informationsgehalt, interpretiert werden kann:

$$H = E[I_m] = E \left[\text{ld} \frac{1}{p_m} \right]. \quad (1.18)$$

Bei statistisch unabhängigen Quellensymbolen führen die Gleichungen (1.15) und (1.17) zum gleichen Ergebnis. Bei einer Nachrichtenquelle mit inneren statistischen Bindungen liefert dagegen (1.17) stets einen größeren Wert als die tatsächliche Entropie gemäß (1.15). Diese ist somit nur als eine (oft grobe) Näherung zu gebrauchen.

Beispiel: Eine binäre Quelle ($M = 2$) erzeugt das Symbol „A“ mit der Wahrscheinlichkeit p und das Symbol „B“ mit der hierzu komplementären Wahrscheinlichkeit $1 - p$. Sind die Symbole der Folge statistisch unabhängig, so ergibt sich für die Quellenentropie:

$$H = p \cdot \text{ld} \frac{1}{p} + (1 - p) \cdot \text{ld} \frac{1}{1 - p} = S(p). \quad (1.19)$$

Die durch (1.19) definierte Funktion $S(p)$ wird häufig als die *Shannonfunktion* bezeichnet. In [7] wird diese wichtige informationstheoretische Funktion $H_B(p)$ genannt. Die Eigenschaften dieser Funktion werden in Vorbereitungsfrage V1 ausführlich behandelt.

1.2.4 Entscheidungsgehalt und Redundanz

Unabhängig vom Symbolumfang gilt, daß der Maximalwert der Entropie dann erreicht wird, wenn alle Symbole gleichwahrscheinlich sind, d.h. wenn für alle $m = 1, 2, \dots, M$ gilt:

$$p_m = \frac{1}{M}. \quad (1.20)$$

Unter dieser Voraussetzung erhält man aus (1.17) für die Entropie:

$$H_0 = \sum_{m=1}^M \frac{1}{M} \text{ld}(M) = \text{ld}(M). \quad (1.21)$$

Dieser Maximalwert der Entropie wird in der Literatur auch als *Entscheidungsgehalt* bezeichnet. Der Index „0“ weist darauf hin, daß es sich dabei um die 0-te Näherung der Entropie handelt.

Sind die M möglichen Symbole nicht gleichwahrscheinlich oder bestehen statistische Bindungen zwischen den einzelnen Symbolen, so besitzt die Nachrichtenquelle stets eine kleinere Entropie als dieser Grenzwert $\text{ld}(M)$. Somit ergibt sich als Wertebereich der Entropie:

$$0 \leq H \leq \text{ld}(M). \quad (1.22)$$

Die Differenz zwischen der maximalen und der tatsächlichen Entropie bezeichnet man als die *Redundanz*:

$$\Delta H = H_0 - H. \quad (1.23)$$

Diese beschreibt den Teil der Nachricht, der keine Information enthält. Wird beispielsweise immer nur das gleiche Symbol ausgegeben, so ist die gesamte Nachricht Redundanz.

Dagegen bezeichnet man eine Quelle mit statistisch unabhängigen, gleichwahrscheinlichen Symbolen als redundanzfrei. Eine solche Folge ist anfällig für nichterkennbare und damit auch nichtkorrigierbare Übertragungsfehler, wenn nicht durch Kanalcodierung Redundanz hinzugefügt wird. Das bedeutet, daß Redundanz stets Voraussetzung für Fehlererkennung und -korrektur ist.

Die Redundanz hat ebenfalls die Einheit „bit“. Häufig wird jedoch die relative Redundanz

$$r = \frac{\Delta H}{H_0} = \frac{H_0 - H}{H_0} \quad (1.24)$$

angegeben, die dimensionslos ist und wegen (1.22) nur Werte zwischen $0 \leq r \leq 1$ annehmen kann.

1.2.5 Verbundentropie und Transinformationsgehalt

Weitere wichtige informationstheoretische Größen sind die *Verbundentropie* und der *mittlere Transinformationsgehalt*. Hier sollen diese beiden Beschreibungsgrößen anhand zweier Ereignisse „A“ und „B“ kurz allgemein erläutert werden. Im Abschnitt 1.3.1 werden sie für zwei Symbolfolgen berechnet, zwischen denen gewisse statistische Bindungen bestehen.

Die Ereignisse „A“ und „B“ seien wertdiskret, wobei sich der Symbolumfang von „A“ (hier mit I bezeichnet) durchaus von dem von „B“ (dieser sei J) unterscheiden kann. Betrachtet man die 2 Ereignisse gemeinsam, so kann man insgesamt $I \cdot J$ *Verbundwahrscheinlichkeiten* angeben:

$$p(A_i \cap B_j) \quad \text{mit} \quad i = 1, 2, \dots, I \quad \text{und} \quad j = 1, 2, \dots, J.$$

Analog zu (1.17) kann daraus die *Verbundentropie* zwischen „A“ und „B“ berechnet werden:

$$H_{AB} = \sum_{i=1}^I \sum_{j=1}^J p(A_i \cap B_j) \cdot \text{ld} \frac{1}{p(A_i \cap B_j)}. \quad (1.25)$$

Diese Größe ist eine untere Schranke für die Summe der beiden Einzelentropien:

$$H_A + H_B \geq H_{AB}, \quad (1.26)$$

die jeweils entsprechend (1.17) zu bestimmen sind. Das Gleichheitszeichen gilt für den Fall, daß die beiden Ereignisse statistisch unabhängig sind.

Bestehen dagegen statistische Bindungen, so lassen sich - zumindest in beschränktem Maße - durch Beobachtung des Ereignisses „B“ statistische Aussagen über das Ereignis „A“ machen. Die Information, die man vom Ereignis „A“ erhält, indem man „B“ beobachtet, bezeichnet man als *Transinformationsgehalt*. Dieser ist für das Wertepaar (i, j) folgendermaßen definiert:

$$I_T(A_i; B_j) = \text{ld} \frac{p(A_i | B_j)}{p(A_i)} = \text{ld} \frac{p(A_i \cap B_j)}{p(A_i) \cdot p(B_j)}. \quad (1.27)$$

Natürlich ist hierfür vorauszusetzen, daß die Wahrscheinlichkeiten im Nenner jeweils ungleich 0 sind. Das heißt, daß die beiden Ereignisse tatsächlich vorkommen können.

Aus (1.27) ist ersichtlich, daß $I_T(A_i; B_j)$ symmetrisch bezüglich „A“ und „B“ ist. Das heißt, es gilt stets die Beziehung:

$$I_T(A_i; B_j) = I_T(B_j; A_i). \quad (1.28)$$

Es ist also egal, ob man die wechselseitige Information durch Beobachtung des Ereignisses „B“ hinsichtlich „A“ gewinnt oder umgekehrt.

Durch Mittelung über alle möglichen Kombinationen erhält man aus (1.28) den *mittleren Transinformationsgehalt*, teilweise in der Literatur auch als *Synentropie* bezeichnet:

$$H_T = E[I_T(A_i; B_j)]. \quad (1.29)$$

Analog zu (1.25) läßt sich hierfür mit (1.27) auch schreiben:

$$H_T = \sum_{i=1}^I \sum_{j=1}^J p(A_i \cap B_j) \cdot \text{ld} \frac{p(A_i \cap B_j)}{p(A_i) \cdot p(B_j)}. \quad (1.30)$$

Bei statistisch unabhängigen Ereignissen ist der mittlere Transinformationsgehalt gleich Null, da hier die Verbundwahrscheinlichkeiten gleich dem Produkt der Einzelwahrscheinlichkeiten sind. Ansonsten ergibt sich stets ein positiver Wert.

Aus (1.27) und (1.29) kann ein Zusammenhang zwischen den Einzelentropien, der Verbundentropie und dem mittleren Informationsgehalt hergeleitet werden. Wegen

$$E \left[\text{ld} \frac{p(A_i \cap B_j)}{p(A_i) \cdot p(B_j)} \right] = E \left[\text{ld} \frac{1}{p(A_i)} \right] + E \left[\text{ld} \frac{1}{p(B_j)} \right] - E \left[\text{ld} \frac{1}{p(A_i \cap B_j)} \right] \quad (1.31)$$

gilt nämlich:

$$H_T = H_A + H_B - H_{AB}. \quad (1.32)$$

Im Abschnitt 1.3.1 werden diese Beschreibungsgrößen anhand zweier korrelierter Symbolfolgen eingehend diskutiert. Anzumerken ist weiterhin, daß in [7] der mittlere Transinformationsgehalt H_T gemäß der Definition (1.29) als die *wechselseitige Transinformation* $I(A; B)$ bezeichnet wird.

1.3 Gedächtnislose wertdiskrete Nachrichtenquellen

1.3.1 Transinformation zwischen zwei Folgen

Für den gesamten Abschnitt 1.3 wird vorausgesetzt, daß zwischen den einzelnen Symbolen der betrachteten Folgen keine statistischen Bindungen bestehen. Eine solche Quelle wird durch die Symbolwahrscheinlichkeiten gemäß (1.4) vollständig beschrieben, wobei diese unabhängig vom Zeitpunkt n (*Stationarität*) als auch von den vorher abgegebenen Symbolen sind.

Zunächst wird die in Kap. 1.2.5 eingeführte Transinformation zwischen zwei sich gegenseitig beeinflussende Folgen behandelt. Zur Verdeutlichung dieser Größe betrachten wir beispielhaft ein binäres Übertragungssystem mit der Quellensymbolfolge am Eingang und der Sinkensymbolfolge am Ausgang:

$$\begin{aligned} \langle q_n \rangle &= q_1, q_2, \dots, q_N, \\ \langle v_n \rangle &= v_1, v_2, \dots, v_N. \end{aligned} \tag{1.33}$$

Die einzelnen Quellensymbole q_n seien statistisch voneinander unabhängig, ebenso wie die Übertragungsfehler des Kanals. Daraus folgt natürlich auch die statistische Unabhängigkeit der Sinkensymbole v_n . Aufgrund dieser Voraussetzungen kann dieses Übertragungssystem gemäß Bild 1.1 modelliert werden.

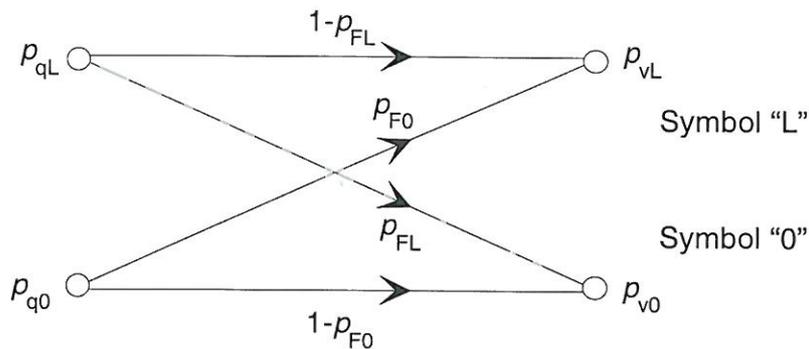


Bild 1.1: Betrachtetes Kanalmodell zur Interpretation der Transinformation.

Es wird zugelassen, daß sich sowohl die Auftretswahrscheinlichkeiten der beiden möglichen Symbole „0“ und „L“ als auch deren Verfälschungswahrscheinlichkeiten unterscheiden. Ist der Kanal ideal, d.h. beide Verfälschungswahrscheinlichkeiten sind identisch 0, so stimmt jedes Sinkensymbol mit dem entsprechenden Quellensymbol überein. In diesem Fall ist selbstverständlich die Entropie der Sinkensymbolfolge gleich der Quellenentropie, wobei beide Größen entsprechend (1.19) zu berechnen sind.

Wir betrachten hier jedoch folgenden Parametersatz:

$$\begin{aligned} p_{q0} &= p(q_n = "0") = 0.1, & p_{qL} &= p(q_n = "L") = 0.9, \\ p_{F0} &= p(v_n = "L" | q_n = "0") = 0.01, & p_{FL} &= p(v_n = "0" | q_n = "L") = 0.2. \end{aligned}$$

Die Auftretswahrscheinlichkeiten der beiden möglichen Sinkensymbole ergeben sich daraus zu:

$$\begin{aligned} p_{v0} &= p_{q0} \cdot (1 - p_{F0}) + p_{qL} \cdot p_{FL} = 0.279, \\ p_{vL} &= p_{q0} \cdot p_{F0} + p_{qL} \cdot (1 - p_{FL}) = 0.721. \end{aligned}$$

Für die Entropien, d.h. die jeweils mittleren Informationsgehalte, von Quellen- und Sinkensymbolfolge erhält man dann entsprechend (1.19):

$$H_q = 0,4690 \text{ bit und } H_v = 0,8541 \text{ bit.}$$

Es ist anzumerken, daß das Sinkensignal bei Vorhandensein von Übertragungsfehlern stets eine größere Entropie aufweisen wird als das Quellensignal. Das bedeutet: Durch die Störungen (Fehler) wird stets Information hinzugefügt, allerdings keine brauchbare Information.

Betrachten wir jeweils ein Quellen- und ein Sinkensymbol als Wertepaar, so kann man folgende vier *Verbundwahrscheinlichkeiten* berechnen:

$$p_{00} = p(q_n = "0" \cap v_n = "0") = p_{q0} \cdot (1 - p_{F0}) = 0.099,$$

$$p_{0L} = p(q_n = "0" \cap v_n = "L") = p_{q0} \cdot p_{F0} = 0.001,$$

$$p_{L0} = p(q_n = "L" \cap v_n = "0") = p_{qL} \cdot p_{FL} = 0.180,$$

$$p_{LL} = p(q_n = "L" \cap v_n = "L") = p_{qL} \cdot (1 - p_{FL}) = 0.720.$$

Mit diesen Verbundwahrscheinlichkeiten ist die *Verbundentropie* gemäß (1.25) berechenbar:

$$H_{qv} = \sum_q \sum_v p(q_n \cap v_n) \cdot \text{ld} \frac{1}{p(q_n \cap v_n)}. \quad (1.34)$$

Diese ist, wie bereits in Abschnitt 1.2.5 allgemein gezeigt, eine untere Schranke für die Summe der beiden Einzelentropien:

$$H_q + H_v \geq H_{qv}. \quad (1.35)$$

Das Gleichheitszeichen gilt, wenn die beiden betrachteten Folgen statistisch unabhängig sind. Bei dem hier betrachteten Beispiel tritt dieser Fall ein, wenn beide Kanalfehlerwahrscheinlichkeiten gleich 0.5 sind. Dann kann von einem Sinkensymbol nicht mehr auf das tatsächlich gesendete Symbol geschlossen werden. Dagegen würde bei fehlerfreier Übertragung gelten:

$$H_q = H_v = H_{qv}. \quad (1.36)$$

Mit den obigen Zahlenwerten ergibt sich für die Verbundentropie und dem nach (1.32) berechenbaren mittleren Transinformationsgehalt:

$$H_{qv} = 1.1268 \text{ bit,}$$

$$H_T = (0.4690 + 0.8541 - 1.1268) \text{ bit} = 0.1963 \text{ bit.}$$

Desweiteren werden in der Informationstheorie häufig noch zwei bedingte Entropien verwendet, nämlich die *Streuentropie (Irrelevanz)*

$$H_{v|q} = \sum_q \sum_v p(q_n \cap v_n) \cdot \text{ld} \frac{1}{p(v_n|q_n)} \quad (1.37)$$

und die *Rückschlußentropie (Äquivokation)*

$$H_{q|v} = \sum_q \sum_v p(q_n \cap v_n) \cdot \text{ld} \frac{1}{p(q_n|v_n)}. \quad (1.38)$$

In diesen beiden Gleichungen sind die bedingten Wahrscheinlichkeiten nur im Nenner des Logarithmus' einzusetzen. Die Gewichtung erfolgt dagegen, wie in (1.34) allgemein angegeben, entsprechend den Verbundwahrscheinlichkeiten.

Die Irrelevanz gibt den mittleren Informationsgehalt an, den ein Beobachter, der über die Quellensymbolfolge genau Bescheid weiß, durch die Beobachtung der Sinkensymbolfolge

gewinnt. Diese Information ist ausschließlich im Kanal entstanden, also von den Kanalfehlern verursacht worden. Im Beispiel ergibt sich für die Irrelevanz der Wert

$$H_{v|q} = 0.6578 \text{ bit.}$$

Zur Berechnung der Rückschlußentropie müssen zunächst die Rückschlußwahrscheinlichkeiten berechnet werden. Mit dem Satz von Bayes gilt hierfür:

$$p(q_n|v_n) = \frac{p(q_n \cap v_n)}{p(v_n)}. \tag{1.39}$$

Setzt man die errechneten Werte in (1.38) ein, so erhält man den Zahlenwert:

$$H_{q|v} = 0.2727 \text{ bit.}$$

Diese als Äquivokation bezeichnete Größe gibt den mittleren Informationsgehalt an, den ein Beobachter, der über die Ausgabe der Sinkensymbole genau Bescheid weiß, durch die Beobachtung der Quellensymbolfolge gewinnt. Dieser Anteil wird durch die Kanalstörungen vernichtet, geht also im Kanal verloren.

Ein Vergleich der oben ermittelten Zahlenwerte macht deutlich, daß die hier betrachtete gestörte Binärübertragung durch das informationstheoretische Modell von Bild 1.2 charakterisiert werden kann. Ein Teil der Quellenentropie, nämlich die Äquivokation „fließt ab“, so daß nur noch der mittlere Transinformationsgehalt verbleibt. Aufgrund der „zufließenden“ Irrelevanz, die nie kleiner ist als die Äquivokation, erhält man die Sinkenentropie $H_v \geq H_q$.

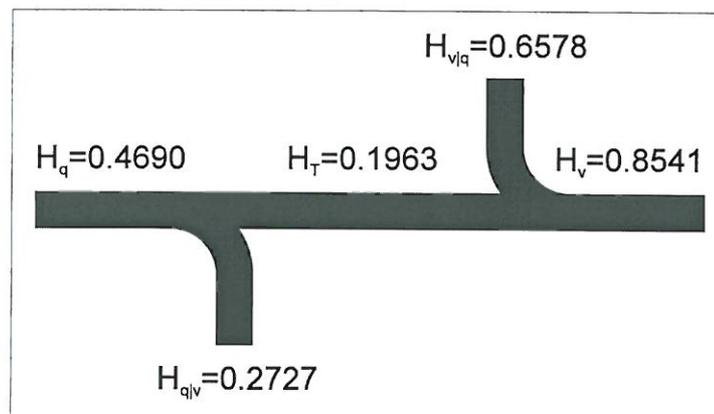


Bild 1.2: Informationstheoretisches Modell der gestörten Binärübertragung.

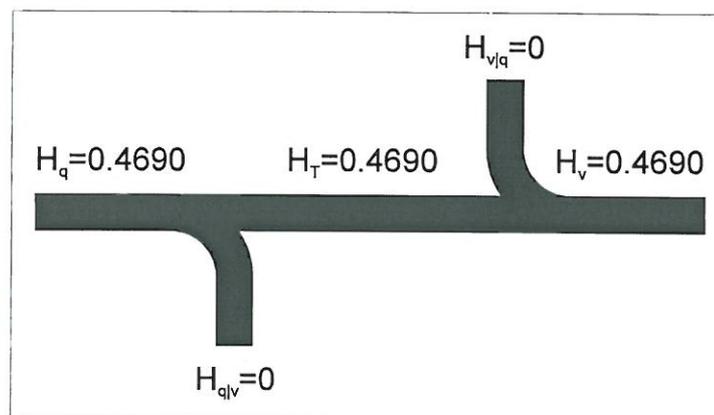


Bild 1.3: Informationstheoretisches Modell der ungestörten Binärübertragung (die Fehlerwahrscheinlichkeiten für „0“ und „L“ sind beide Null).

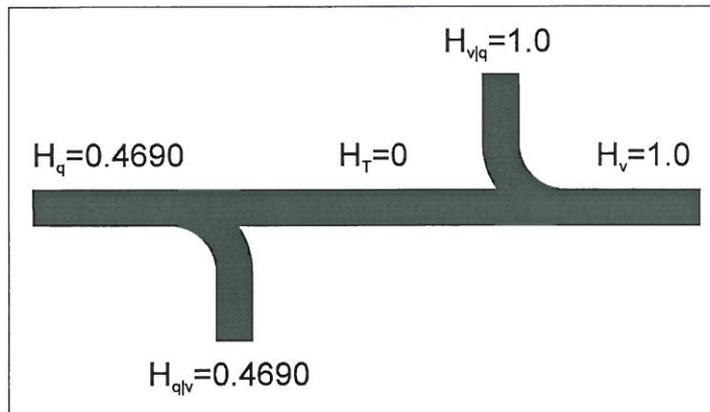


Bild 1.4: Informationstheoretisches Modell der vollkommen gestörten Binärübertragung (die Fehlerwahrscheinlichkeiten für „0“ und „L“ sind beide 0.5).

Zum Vergleich sind in Bild 1.3 und 1.4 die entsprechenden Modelle für den störungsfreien sowie den vollständig gestörten Binärkanal dargestellt, wobei für die Quellensymbole die gleichen Auftretswahrscheinlichkeiten gewählt sind. Die Quellenentropie ist deshalb stets 0.4690 bit.

Aus dem Vergleich der Zahlenwerte in den Bildern 1.2 bis 1.4 ist offensichtlich, daß der mittlere Transinformationsgehalt möglichst groß sein sollte. Dieser Wert gibt an, welcher Anteil der Quellenentropie die Übertragungsstörungen überlebt, und somit der Sinke zur weiteren Verarbeitung zur Verfügung steht. Er sagt weiterhin aus, wieviel Nutzinformation pro Zeiteinheit im Mittel über den Kanal transportiert wird.

Zur Berechnung dieser wichtigen informationstheoretischen Größe können neben der Definitionsgleichung (1.30) noch folgende Gleichungen benutzt werden:

$$H_T = H_q - H_{q|v}, \quad (1.40)$$

$$H_T = H_v - H_{v|q}, \quad (1.41)$$

$$H_T = H_q + H_v - H_{qv}. \quad (1.42)$$

Abschließend wird noch der Transinformationsgehalt der einzelnen Symbolpaare zahlenmäßig angegeben, der nach der Definitionsgleichung (1.27) ermittelt wurde:

$$\begin{aligned} I_T(q_n = "0"; v_n = "0") &= 1.827; & I_T(q_n = "L"; v_n = "0") &= -0.480; \\ I_T(q_n = "0"; v_n = "L") &= -6.172; & I_T(q_n = "L"; v_n = "L") &= 0.150. \end{aligned}$$

Durch Mittelung über diese vier Transinformationsgehalte (gewichtet mit den dazugehörigen Verbundwahrscheinlichkeiten) ergibt sich wiederum der Zahlenwert 0.1963.

Die beiden Symbolpaare „00“ und „LL“, die für eine fehlerfreie Übertragung stehen, leisten demzufolge einen positiven Beitrag zum mittleren Transinformationsgehalt, wobei in diesem Beispiel der Beitrag der weniger gestörten „0“ trotz des selteneren Auftretens größer ist. Dagegen besitzen die Symbolpaare „0L“ bzw. „L0“ jeweils einen negativen Transinformationsgehalt. Dies ist auch einsichtig, da diese Symbolpaare ja Übertragungsfehler kennzeichnen. Deshalb wäre dem Empfänger zu raten, diese Informationen, die ausschließlich von Übertragungsfehlern herühren, zu verwerfen. In diesem Fall würde sich der mittlere Transinformationsgehalt auf 0.2889 erhöhen. Leider kann der Empfänger diese informationstheoretischen Kenntnisse nicht nutzen, da ihm nicht bekannt ist, welches Empfangssymbol auf dem Kanal verfälscht wurde.

1.3.2 Kanalkapazität eines Digitalkanals

Wie im Abschnitt 1.3.1 betrachten wir einen Kanal mit einer endlichen Anzahl von Ein- und Ausgängen (digitaler Kanal). Der Binärkanal ist beispielsweise in Bild 1.1 dargestellt. Eine Grenze für die Ausnutzbarkeit eines solchen Digitalkanals stellt die *Kanalkapazität* dar.

Zur Herleitung dieser Grenze muß zunächst die *mittlere Informationsrate* definiert werden. Diese ist der Quotient aus dem mittleren Transinformationsgehalt (wie in Abschnitt 1.3.1 berechnet) und der mittleren Übertragungszeit:

$$F = \frac{H_T}{T}. \quad (1.43)$$

Der allgemeine Fall ist dabei, daß die Übertragung der unterschiedlichen Quellensymbole auch verschieden lange dauert. Dann gilt für die *mittlere Übertragungszeit*:

$$T = \sum_{m=1}^M p_m \cdot T_m. \quad (1.44)$$

Hierbei ist T_m die Zeit, die zur Übertragung des m -ten Quellensymbols benötigt wird. Sind die Übertragungszeiten alle gleich, so kann auf die Mittelung in (1.44) verzichtet werden. T ist dann die von den Quellensymbolen unabhängige Symboldauer.

Beispiel: Es wird die in Abschnitt 3.1 behandelte Binärübertragung betrachtet. Ist die Symboldauer für die beiden möglichen Symbole „0“ und „L“ jeweils gleich 2 Mikrosekunden, so errechnet sich die mittlere Informationsrate zu

$$F = \frac{0.1963 \text{ bit/Symbol}}{2 \cdot 10^{-6} \text{ s/Symbol}} = 98.15 \text{ kbit/s}.$$

Benötigt jedoch das Symbol „0“ die dreifache Übertragungszeit wie das Symbol „L“ (z.B., weil die Übertragung des Symbols „0“ besser gegen Fehler gesichert ist), so erhält man nach (1.44):

$$T = p_{q0} \cdot T_0 + p_{qL} \cdot T_L = 0.1 \cdot 6 \mu\text{s} + 0.9 \cdot 2 \mu\text{s} = 2.4 \mu\text{s}.$$

Die mittlere Informationsrate beträgt somit nur noch etwa 81.8 kbit/s.

Die mittlere Informationsrate hängt stark von den Symbolwahrscheinlichkeiten der Quelle ab. Dies erkennt man, wenn man z.B. die allgemeingültige Gleichung (1.30) für den mittleren Transinformationsgehalt auf einen digitalen Übertragungskanal anwendet und in (1.43) einsetzt:

$$F = \frac{\sum_q \sum_v p(q) \cdot p(v|q) \cdot [\text{ld}(p(v|q)) - \text{ld}(p(v))]}{\sum_q p(q) \cdot T(q)}. \quad (1.45)$$

Bei einem Digitalkanal mit M Eingängen und M Ausgängen besteht dabei die Summe im Nenner aus M , die Doppelsumme im Zähler aus M^2 Summanden.

Die Abhängigkeit von der Statistik der Quelle ist in (1.45) an den Wahrscheinlichkeiten $p(q)$ ersichtlich. Zu berücksichtigen ist, daß die Wahrscheinlichkeiten $p(v)$ am Ausgang ebenfalls von den Quellensymbolwahrscheinlichkeiten $p(q)$ abhängen. Dagegen werden die bedingten Wahrscheinlichkeiten $p(v|q)$ sowie die Übertragungszeiten $T(q)$ allein vom Kanal bestimmt.

Optimiert man die statistischen Kenngrößen der Quelle im Hinblick auf eine Maximierung der mittleren Informationsrate F , so kommt man zur Kanalkapazität:

$$C = F_{\max} = \operatorname{Max}_{p(q)} \left(\frac{H_T}{T} \right). \quad (1.46)$$

Diese wichtige, von Shannon eingeführte informationstheoretische Größe gilt somit für ein optimal an den Kanal angepaßtes Sendesignal. Sie ist somit nur noch von den Kanalparametern abhängig und stellt eine Grenze für die Kanalcodierung dar.

Das *Kanalcodierungstheorem* für diskrete, gedächtnislose Kanäle besagt, daß die Information der Nachrichtenquelle mit beliebig kleiner Verfälschungswahrscheinlichkeit zur Sinke übertragen werden kann, solange die Übertragungsrate R kleiner als die Kanalkapazität C ist. Über den dabei zu wählenden Code macht das Kanalcodierungstheorem allerdings keine Aussagen.

Die Einheiten der Kanalkapazität C gemäß (1.46) sowie der Informationsübertragungsrate R sind jeweils „bit/s“. Sind die Übertragungszeiten für alle Symbole gleich (nämlich T), so kann sich die Maximierung auf den Zähler von (1.46) beschränken. Häufig wird deshalb in der Literatur mit der auf $1/T$ normierten Kanalkapazität

$$C' = \operatorname{Max}_{p(q)} (H_T) = \frac{C}{1/T} = C \cdot T \quad (1.47)$$

gearbeitet. Dies ist, wie gesagt, nur dann möglich, wenn die Dauer T für alle Symbole gleich ist. Zur Unterscheidung wird hier - im Gegensatz zu mancher Literatur - ein Hochkomma angefügt.

Zur Formulierung des Kanalcodierungstheorems muß dann auch die normierte Rate

$$R' = \frac{R}{1/T} = 1 - r \quad (1.48)$$

herangezogen werden, die sich als der Komplementärwert der in (1.24) definierten relativen Redundanz r berechnen läßt.

Teilt man eine binäre redundanzfreie Quellensymbolfolge in k -Tupel auf und weist jedem k -Tupel ein Codewort der Länge $n > k$ zu, so hat dieser Code die (normierte) Coderate $R' = k/n$. Aus dem Kanalcodierungstheorem folgt dann die Aussage, daß zumindest ein Code (n, k) existiert, dessen Codewort-Fehlerwahrscheinlichkeit gegen 0 geht, vorausgesetzt, es gilt $R' < C'$. Bei einem solchen Code handelt es sich meist um einen sehr langen Code (unendlich große Werte für die Codeparameter n und k bei gegebenem R'). Ist dagegen $R' > C'$, so läßt sich eine beliebig kleine Fehlerwahrscheinlichkeit nicht erreichen.

In den Jahren nach der Veröffentlichung der Shannonschen Theorie zur Kanalcodierung (1948/49) wurden große Anstrengungen unternommen, um Codeklassen zu finden, die die von Shannon versprochene beliebig kleine Fehlerwahrscheinlichkeit erreichen; die Resultate waren meist mager. Erst die sogenannten BCH-Codes (1959 von Bose, Ray-Chaudhuri und Hocquenghem entwickelt) und eine verwandte Codeklasse für mehrstufige Digitalkanäle (von Reed und Solomon im Jahr 1960 vorgestellt) brachten den Durchbruch.

Beim *symmetrischen Binärkanal* (BSC) ist die Berechnung der Kanalkapazität sehr einfach. Das BSC-Modell ist ein Sonderfall des in Bild 1.1 dargestellten Modells mit gleichen Fehlerwahrscheinlichkeiten für die beiden möglichen Symbole „0“ und „1“: $p_{F0} = p_{F1} = p$.

Aufgrund der Symmetrie des Modells wird das Maximum des mittleren Transinformationsgehalts bei einer Binärquelle mit gleichwahrscheinlichen Symbolen erreicht. Die Quellenentropie beträgt somit 1 bit, ebenso wie die Sinkenentropie, da am Ausgang das Symbol „0“ bzw. „L“ ebenfalls jeweils mit der Wahrscheinlichkeit 0.5 auftritt. Die vier Verbundwahrscheinlichkeiten ergeben sich hier zu $p/2$ (zweimal) bzw. $(1-p)/2$ (ebenfalls zweimal), so daß man für die Verbundentropie H_{qv} mit der Shannonfunktion $S(p)$ entsprechend (1.19) den Wert $1 + S(p)$ erhält. Mit (1.32) und (1.47) folgt dann für die (normierte) Kanalkapazität:

$$C' = H_T = H_q + H_v - H_{qv} = 1 - S(p). \quad (1.49)$$

Die Optimierung der Quellenstatistik wurde in diesem Sonderfall durch die geeignete Wahl der Symbolwahrscheinlichkeiten bereits vorweggenommen.

In Bild 1.5 ist die Kanalkapazität des BSC-Kanals in Abhängigkeit der Fehlerwahrscheinlichkeit p dargestellt. Dieses Bild ist wie folgt zu interpretieren: Über einen BSC-Kanal mit $p = 0.5$ kann keine Information übertragen werden (vgl. auch Bild 1.4). Bei $p = 1$ ist ebenso wie bei $p = 0$ die Übertragung von einem Informationsbit/Binärzeichen möglich, wobei mit $p = 1$ implizit eine Umcodierung vorgenommen wird („0“ nach „L“ und umgekehrt). Typische Werte der Fehlerwahrscheinlichkeit in einem Telegraphie- oder Funkkanal liegen in der Größenordnung von $p = 0.001$. Hier liefert (1.49) etwa den Wert $C' = 0.989$. Bis zu dieser Coderate kann theoretisch ein Code gefunden werden, so daß die Nutzinformation fehlerfrei (genauer: mit beliebig kleiner Fehlerwahrscheinlichkeit) über diesen mit 0.1% verfälschenden Kanal übertragen werden kann.

In Vorbereitungsfrage V5 wird die Berechnung der Kanalkapazität eines unsymmetrischen Binärkanals behandelt.

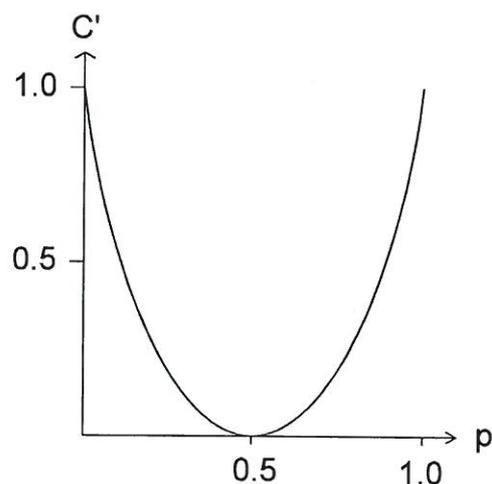


Bild 1.5: Kanalkapazität des BSC-Modells.

1.3.3 Grundlagen der Quellencodierung

Im letzten Abschnitt wurde das von Shannon formulierte Kanalcodierungstheorem für diskrete gedächtnislose Kanäle angegeben. Nun wird kurz auf das zweite wichtige Shannonsche Theorem eingegangen, das die *Quellencodierung* zum Inhalt hat.

Das *Quellencodierungstheorem* besagt, daß es bei einer redundanten Nachrichtenquelle Q , deren Entropie H beträgt, ausreichend ist, im Mittel H bit/Symbol zu übertragen. In diesem Fall geht die Wahrscheinlichkeit für eine perfekte Rekonstruktion beim Empfänger gegen 1, wenn die Länge der (codierten) Quellensymbolfolge gegen unendlich geht.

Betrachten wir einen deutschen Text in Großbuchstaben, der außer den 29 Buchstaben (incl. Umlaute) nur das Leerzeichen und zwei Satzzeichen (z.B. Punkt und Komma) beinhaltet, so können diese 32 Symbole auch ohne Kenntnis der Informationstheorie mit 5 Binärzeichen pro Symbol dargestellt werden. Berücksichtigt man den von Küpfmüller empirisch gewonnenen Entropiewert von 1.3 bit/Symbol bei der deutschen Sprache (genauer: bei einem deutschen Text), so besagt das Quellencodierungstheorem, daß es zumindest eine Codierung gibt, so daß der Text im Mittel mit nur 1.3 Binärzeichen pro Symbol dargestellt werden kann. Wird keines der Binärzeichen verfälscht, so kann dieser codierte (und ohne Kenntnis des Codes völlig unverständliche Text) vollständig rekonstruiert werden. Der binär dargestellte Text wird auf diese Weise um ca. den Faktor 3.8 kürzer.

Aufgabe der Quellencodierung ist es, der Quellensymbolfolge möglichst alle Redundanz zu entziehen, so daß für die codierte Folge eine normierte Übertragungsrate (vgl. Definition (1.48))

$$R' = H_q \quad (1.50)$$

ausreichend ist. Bei einer niedrigeren Rate ist keine richtige Rekonstruktion möglich, so daß (1.50) gleichzeitig die untere Grenze für die (normierte) Übertragungsrate angibt.

Die redundanzfreie Codesymbolfolge ist äußerst anfällig gegen Übertragungsfehler. Deshalb wird meist durch Kanalcodierung wieder Redundanz hinzugefügt. Bei einem Faltungscoder, der aus jedem Eingangsbit zwei Ausgangsbit entsprechend einer festen Codiervorschrift gewinnt (d.h.: Coderate 1/2), erhöht sich z. B. die benötigte Übertragungsrate um den Faktor 2.

Im ersten Moment mag es unlogisch erscheinen, daß die Quellencodierung zuerst Redundanz entzieht, die dann durch Kanalcodierung wieder (zumindest teilweise) hinzugefügt wird. Im Gegensatz zu der natürlichen Redundanz einer Nachrichtenfolge, die oft nicht zur Fehlererkennung und -korrektur genutzt werden kann, wird bei der Kanalcodierung jedoch solche Redundanz gezielt zugesetzt, die vom Empfänger genutzt werden kann.

Shannon hat gezeigt, daß die Probleme der Quellen- und Kanalcodierung (fast) immer ohne Qualitätsverlust separiert werden können. Im folgenden wird die Quellencodierung, d.h. die Repräsentation der Quellensymbolfolge durch (möglichst wenige) Binärsymbole ohne Informationsverlust, näher betrachtet. Dabei beschränken wir uns weiter auf gedächtnislose Quellen.

1.3.4 Mittlere Codewortlänge

Im folgenden wird von dem in Bild 1.6 dargestellten Modell ausgegangen: Die Quellensymbolfolge sei M -wertig, die einzelnen Folgeelemente statistisch voneinander unabhängig. Somit ist die Entropie mit (1.17) und der Entscheidungsgehalt mit (1.21) berechenbar:

$$H_0 = \text{ld}(M), \quad H_q = \sum_{m=1}^M p_m \cdot \text{ld} \frac{1}{p_m}.$$

Sind die M Symbole gleichwahrscheinlich, so stimmen beide Größen überein und es liegt eine redundanzfreie Quelle vor. Eine Quellencodierung ist in diesem Fall nicht erforderlich.

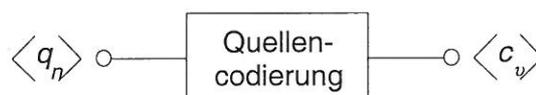


Bild 1.6: Zur Untersuchung der Quellencodierung.

Im folgenden werden deshalb stets Quellen mit unterschiedlichen Symbolwahrscheinlichkeiten betrachtet, deren M -wertige Ausgangsfolgen durch die Quellencodierung in Binärfolgen umgewandelt werden:

$$\begin{aligned} \langle q_n \rangle &= q_1, q_2, \dots, q_N \quad \text{mit} \quad q_n = \{S_1, S_2, \dots, S_M\}, \\ &\Downarrow \text{ Quellencodierung} \\ \langle c_v \rangle &= c_1, c_2, \dots, c_{N_c} \quad \text{mit} \quad c_v = \{ "0", "L" \}. \end{aligned} \quad (1.51)$$

Die einzelnen Quellensymbole können dabei durchaus durch unterschiedlich lange Codeworte dargestellt werden; dies ist in (1.51) durch die unterschiedliche Indizierung der beiden Folgen angedeutet. Die Umsetzung soll dabei in der Weise erfolgen, daß die *mittlere Codewortlänge* möglichst gering ist:

$$L_{\text{CW}} = \frac{N_c}{N} = \text{Minimum}. \quad (1.52)$$

Nach dem Quellencodierungstheorem ist eindeutig, daß die mittlere Codewortlänge nicht kleiner als die Quellenentropie sein wird:

$$L_{\text{CW}} \geq H_q. \quad (1.53)$$

Die mittlere Codewortlänge kann aber auch als Scharmittelwert über die M möglichen Codeworte ermittelt werden:

$$L_{\text{CW}} = \sum_{m=1}^M p_m \cdot L_m \quad \text{mit} \quad L_m : \text{Anzahl der Binärzeichen des } m\text{-ten Codewortes}. \quad (1.54)$$

Um diesen Wert zu minimieren, ist es zweckmäßig, häufige Symbole durch kurze Codeworte, seltenere dagegen durch entsprechend längere Worte darzustellen. Dieses Prinzip soll an einigen Beispielen verdeutlicht werden.

Beispiel: Betrachtet wird eine quaternäre Nachrichtenquelle ($M = 4$) mit den möglichen Symbolen „A“, „B“, „C“ und „D“, die mit folgenden Wahrscheinlichkeiten auftreten:

$$p_A = \frac{1}{2}; p_B = \frac{1}{4}; p_C = p_D = \frac{1}{8}.$$

Der Entscheidungsgehalt dieser Quelle ist $\text{Id}(4) = 2$ bit/Symbol, während die Entropie, berechnet nach (1.17), den Wert 1.75 bit/Symbol ergibt. Ohne Quellencodierung kann eine Folge der Länge N durch $2N$ Binärzeichen dargestellt werden.

Wählt man nun zwischen den Quellen- und den Codesymbolen die Zuordnung

$$\text{„A“} \Rightarrow 0, \text{„B“} \Rightarrow L0, \text{„C“} \Rightarrow LL0, \text{„D“} \Rightarrow LLL,$$

so ergibt sich für die mittlere Codewortlänge entsprechend (1.54):

$$L_{\text{CW}} = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 1.75 \text{ bit/Symbol}.$$

Zur Übermittlung der Information einer N Symbole langen Folge sind aufgrund der ungleichförmig verteilten Quellensymbole nun nur mehr $1.75 \cdot N$ Binärsymbole zu übertragen (dieser Wert gilt allerdings für kleine Werte von N nur näherungsweise). Ein effektiverer Code kann nicht gefunden werden, da die mittlere Codewortlänge bereits mit der Quellenentropie übereinstimmt.

Bei der Codekonstruktion ist darauf zu achten, daß die ursprüngliche M -stufige Folge aus der (codierten) Binärfolge wieder rekonstruiert werden kann. Man spricht dann von präfixfreien Codes. Hierauf wird im Abschnitt 1.3.5 noch näher eingegangen.

Bei einer Ternärquelle ($M = 3$) mit den Symbolwahrscheinlichkeiten

$$p_A = \frac{1}{2}; p_B = p_C = \frac{1}{4}$$

(d.h.: $H_0 = \text{ld}(3) = 1.585$ bit/Symbol, $H = 1.5$ bit/Symbol) wird mit der präfixfreien Zuweisung

$$\text{„A“} \Rightarrow 0, \text{„B“} \Rightarrow L0, \text{„C“} \Rightarrow LL$$

ebenfalls die mittlere Codewortlänge minimiert (auf 1.5 bit/Symbol). Obwohl in diesem Fall die Redundanz der Quelle mit $(1.585-1.5)/1.585 = 5.36\%$ kleiner ist als bei $M = 4$, ist hier die Quellencodierung noch effizienter als im ersten Beispiel, vorausgesetzt, die Übertragung soll binär erfolgen; ohne Quellencodierung müßten nämlich ebenfalls 2 bit/Symbol übertragen werden. Die Übertragungsrate kann damit um 25% gesenkt werden.

Bei diesen beiden Beispielen war die Codekonstruktion sehr einfach, da die Symbolwahrscheinlichkeiten als negative Potenzen zur Basis 2 geeignet vorgegeben waren. In den nächsten beiden Abschnitten werden zwei Algorithmen vorgestellt, die effektive Codes auch für beliebige Symbolwahrscheinlichkeiten erzeugen.

1.3.5 Codierung nach Huffman

Für das Folgende setzen wir weiter voraus, daß die Quellensymbole einem Alphabet mit dem Symbolumfang M entstammen und statistisch voneinander unabhängig seien. Huffman hat unter diesen Voraussetzungen im Jahre 1952 (also kurz nach Shannon's bahnbrechenden Veröffentlichungen) einen Algorithmus zur Konstruktion von optimalen präfixfreien Codes angegeben. Wir beschränken uns hier auf die Binärcodes.

Optimal bedeutet in diesem Zusammenhang, daß die mittlere Codewortlänge gemäß (1.52) für die gegebenen Symbolwahrscheinlichkeiten minimal ist. Die Eigenschaft „präfixfrei“ sagt aus, daß kein Codewort der Präfix (d.h. Beginn) eines längeren Codewortes sein darf. Ein präfixfreier Code ist sofort decodierbar.

Beispiel: Der Binärcode für das Alphabet „A“, „B“, „C“,

$$\text{„A“} \Rightarrow 0, \text{„B“} \Rightarrow L0, \text{„C“} \Rightarrow OL$$

ist nicht präfixfrei. Hier kann die Codefolge 0L00 nicht eindeutig decodiert werden, da sowohl „ABA“ als auch „CAA“ durch diese Binärfolge dargestellt werden. Dagegen beschreiben die im Abschnitt 1.3.4 angegebenen Zuordnungen jeweils präfixfreie Codes.

Der Huffman-Algorithmus soll hier ohne Ableitung und Beweis rezeptartig angegeben werden:

1. Man ordne die Symbole nach fallender Wahrscheinlichkeit.
2. Man fasse die zwei Symbole mit der kleinsten Wahrscheinlichkeit zu einem neuen Symbol zusammen.
3. Man wiederhole die Schritte 1 und 2 solange, bis nur mehr zwei Symbole übrig bleiben.
4. Mit dem letzten Symbolpaar beginnend, wird schrittweise binär codiert, indem man den Code jedes aufgespaltenen Teilcodes mit „0“ und „L“ ergänzt.

Beispiel: Ohne Einschränkung der Allgemeingültigkeit setzen wir voraus, daß die Symbole bereits entsprechend ihren Wahrscheinlichkeiten geordnet sind:

$$p_A = 0.27, \quad p_B = 0.23, \quad p_C = 0.20, \quad p_D = 0.15, \quad p_E = 0.10, \quad p_F = 0.05.$$

Durch paarweises Zusammenfassen und Sortieren erhält man in 4 Schritten die folgenden Symbolkombinationen (resultierende Wahrscheinlichkeiten in Klammern):

1. A(0.27), B(0.23), C(0.20), D(0.15), EF(0.15)
2. DEF(0.30), A(0.27), B(0.23), C(0.20)
3. BC(0.43), DEF(0.30), A(0.27)
4. ADEF(0.57), BC(0.43)

Rückwärts gemäß den Schritten 4 bis 1 erfolgt dann die Zuordnung der Symbole:

4. ADEF \rightarrow 0x ; BC \rightarrow Lx ;
3. DEF \rightarrow 00x ; A \rightarrow 0L ; B \rightarrow L0 ; C \rightarrow LL ;
2. D \rightarrow 000 ; EF \rightarrow 00Lx ;
1. E \rightarrow 00L0 ; F \rightarrow 00LL ;

In dieser Darstellung kennzeichnet ein x, daß im nächsten Schritt noch Bits hinzugefügt werden müssen. Dagegen sind die endgültigen Codezuordnungen unterstrichen.

Aus dem Vergleich der mittleren Codewortlänge gemäß (1.54),

$$L_{\text{CW}} = 2 \cdot (0.27 + 0.23 + 0.20) + 3 \cdot 0.15 + 4 \cdot (0.10 + 0.05) = 2.45 \text{ bit/Symbol},$$

mit der Entropie (hier 2.42 bit/Symbol) erkennt man die Effizienz der Huffman-Codierung bei dieser Wahrscheinlichkeitsverteilung.

Häufig wird die Konstruktion des Huffman-Codes durch eine Baumstruktur angegeben. Bild 1.7 zeigt eine solche für obiges Beispiel.

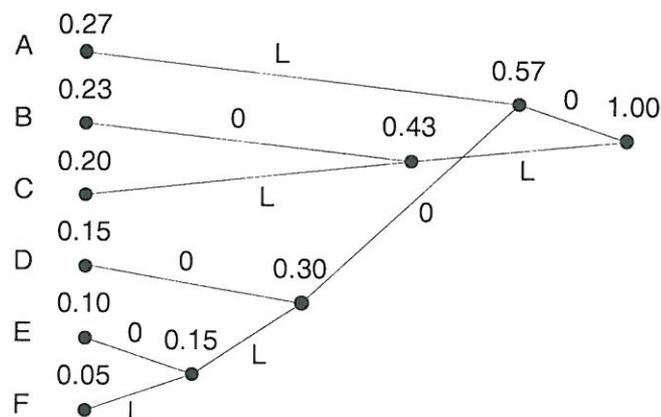


Bild 1.7: Darstellung der Huffman-Codierung durch eine Baumstruktur.

Anzumerken bleibt, daß der Huffman-Code auch zur Datenreduktion einer binären Nachrichtenquelle herangezogen werden kann, wenn mehrere Quellensymbole gemeinsam codiert werden.

1.3.6 Codierung nach Lempel und Ziv

Der Algorithmus nach Lempel und Ziv wird häufig zur Datenkompression auf Computern eingesetzt, da er sehr effektiv und relativ einfach zu realisieren ist. Wie beim Huffman-Code ist für sehr großes N die mittlere Codewortlänge gleich der Quellenentropie. Ein wesentlicher Vorteil des Lempel-Ziv-Algorithmus' ist, daß keine Information über die Quelle erforderlich ist.

Im nachfolgenden Beispiel gehen wir von einer Binärquelle aus. Der Algorithmus läßt sich jedoch auch auf andere Quellen analog anwenden. Bei der Codierung nach Lempel-Ziv sind hier - genauso wie bei der Huffman-Codierung - zwei Durchgänge vorgesehen. Im ersten Durchgang wird die Binärfolge in sogenannte *Phrasen* eingeteilt, die im vorangegangenen Teil der Folge noch nicht aufgetreten sind. Prinzipiell kann der Algorithmus nach Lempel-Ziv auch mit nur einem Durchgang realisiert werden.

Beispiel: Das Binärsignal „1011010100010“ wird in folgende Phrasen eingeteilt:

1, 0, 11, 01, 010, 00, 10.

Die aktuell ermittelte Phrase ist also der kürzeste String, der zuvor noch nicht als Phrase eingeteilt wurde. Im obigen Beispiel wird die Binärfolge der Länge $N = 13$ somit durch $C = 7$ Phrasen beschrieben. Im zweiten Durchgang werden nun die Phrasen jeweils durch $L+1$ Bits codiert, wobei $L = \text{ld}[C]$ ist. Ist C keine Zweierpotenz, so ist für L die nächstgrößere Integerzahl zu nehmen.

Die ersten L Bits bezeichnet man als den Präfix. Dieser ist der Teil der Phrase ohne das letzte Bit. Diesem Präfix wird das letzte Bit der Phrase angehängt. Im obigen Beispiel mit $C = 7$ Phrasen sind $L = 3$ Bit nötig, um den Präfix zu beschreiben. Das sieht dann folgendermaßen aus:

(000,1) (000,0) (001,1) (010,1) (100,0) (010,0) (001,0).

Da die Phrasen „1“ und „0“ noch nicht aufgetreten sind, erhalten beide den Präfix „000“. Die dritte Phrase „11“ ergibt sich aus der ersten Phrase (daher Präfix „001“) durch Anhängen einer „1“, die vierte aus der zweiten (Präfix „010“) ebenfalls durch Anhängen einer „1“ usw.

In diesem Beispiel wird durch diese Codierung die Anzahl der Binärzeichen von $N = 13$ auf $7 \cdot (L+1) = 28$ vergrößert. Das bedeutet, daß der Algorithmus nach Lempel und Ziv bei kurzen Symbolfolgen nicht zweckmäßig ist. Bei langen Folgen ist dagegen mit diesem Verfahren eine mit Huffman vergleichbare Datenreduktion möglich.

Anzumerken ist, daß es sehr viele Varianten des Lempel-Ziv-Algorithmus' gibt. In [3] wird z.B. gezeigt, daß man durchaus mit einem Durchgang auskommen kann. Eine weitere Verbesserung hinsichtlich der Datenreduktion ergibt sich, wenn man für den Präfix keine feste Länge vorsieht, sondern dieser entsprechend einer (im Coder und Decoder zu implementierenden) Vorschrift je nach Bedarf verlängert wird.

Ein Vorteil gegenüber der Huffman-Codierung ist nach Aussagen einiger Benutzer die höhere Geschwindigkeit. Nachteilig ist dagegen der hohe Speicherbedarf. Aus diesem Grund ist dieser Algorithmus im Programm nicht realisiert.

1.4 Nachrichtenquellen mit inneren statistischen Bindungen

1.4.1 Entropiebestimmung

Im Kapitel 1.3 wurde stets vorausgesetzt, daß die einzelnen Symbole der Folge statistisch voneinander unabhängig seien. Die Redundanz einer Folge war deshalb allein auf die unterschiedlichen Auftrittswahrscheinlichkeiten zurückzuführen.

Reale Nachrichtenquellen lassen sich so nur unzureichend modellieren. Bei diesen sind meist *statistische Bindungen* zwischen den einzelnen Folgeelementen gegeben. Betrachtet man z.B. einen deutschen Text, so ist hierin der Buchstabe „u“ eher unwahrscheinlich. Dessen Auftrittswahrscheinlichkeit beträgt etwa 2.7% (insgesamt für Klein- und Großbuchstaben). Dagegen beträgt die bedingte Wahrscheinlichkeit, daß nach einem „Q“ bzw. „q“ ein „u“ folgt, nahezu 100%. Dies ist eine Folge der Redundanz von Texten aufgrund statistischer Bindungen zwischen den einzelnen Zeichen.

Auch bei Analogsignalen (Sprache, Bilder usw) existieren statistische Bindungen, die dann z.B. durch die *Autokorrelationsfunktion* erfaßt werden. Diese Bindungen bleiben auch nach der Digitalisierung der Signale erhalten und können dann mit den Methoden der wertdiskreten Informationstheorie erfaßt werden.

Bestehen zwischen den einzelnen Symbolen statistische Bindungen, so liegt eine *Nachrichtenquelle mit Gedächtnis* vor. Auch in diesem Fall wird der mittlere Informationsgehalt pro Symbol durch die Entropie beschrieben, doch ist diese dann nicht mehr nach der für statistisch unabhängige Symbole gültigen Gleichung (1.17) zu berechnen. Diese Gleichung ist ebenso wie der Entscheidungsgehalt nach (1.21),

$$H_0 = \text{ld}(M),$$

nur eine (meist grobe) Näherung für die tatsächliche Entropie H . Um dies auch in der Nomenklatur zu verdeutlichen, wird die Entropie nach (1.17) im folgenden mit

$$H_1 = \sum_{m=1}^M p_m \cdot \text{ld} \frac{1}{p_m} \geq H \quad (1.55)$$

bezeichnet. Der Index „1“ soll darauf hinweisen, daß zur Berechnung dieser Näherung nur jeweils ein isoliertes Symbol herangezogen wird, während die statistischen Bindungen der Folge nicht berücksichtigt werden.

Betrachtet man zwei aufeinanderfolgende Folgeelemente, von denen jedes einzelne eines von M möglichen Symbolen sein kann, so gibt es M^2 Verbundwahrscheinlichkeiten

$$p(q_n \cap q_{n+1}) \leq p(q_n) \cdot p(q_{n+1}). \quad (1.56)$$

Daraus ist die Verbundentropie eines Zweier-Tupels berechenbar:

$$H_{q_n, q_{n+1}} = \sum_{q_n} \sum_{q_{n+1}} p(q_n \cap q_{n+1}) \cdot \text{ld} \frac{1}{p(q_n \cap q_{n+1})}. \quad (1.57)$$

Die hier gewählte Nomenklatur deutet auf einen Scharmittelwert hin. Die Verbundwahrscheinlichkeiten nach (1.56) sowie die Verbundentropie gemäß (1.57) lassen sich aber auch als Zeitmittelwerte aus einer sehr langen Folge berechnen (besser gesagt, durch die entsprechenden Häufigkeiten annähern), wenn der Index n alle Werte zwischen 1 und N durchläuft.

Berücksichtigt man, daß sich die oben definierte Verbundentropie auf zwei Symbole bezieht, so muß dieser Wert noch halbiert werden, um zum mittleren Informationsgehalt pro Symbol zu kommen. Somit erhält man als weitere obere Schranke der Entropie:

$$H_2 = \frac{1}{2} \cdot H_{q_n, q_{n+1}} < H_1 < H_0. \quad (1.58)$$

Um die Bindungen zwischen k aufeinanderfolgenden Symbolen mit vertretbarem Schreibaufwand angeben zu können, führen wir folgende abkürzende Symbolik ein:

$$H_2 = \frac{1}{2} \cdot \sum_{[2]} p_2 \cdot \text{ld} \frac{1}{p_2}. \quad (1.59)$$

Hierbei steht die Wahrscheinlichkeit p_2 stellvertretend für die (insgesamt M^2) Verbundwahrscheinlichkeiten; die [2] an der Summe soll darauf hinweisen, daß es sich hier eigentlich um eine Doppelsumme handelt.

Mit dieser Schreibweise ergibt sich für die Verbundentropie eines k -Tupels, d.h. von k aufeinanderfolgenden Symbolen $q_n, q_{n+1}, \dots, q_{n+k-1}$, mit den M^k Verbundwahrscheinlichkeiten p_k :

$$H_k = \frac{1}{k} \cdot \sum_{[k]} p_k \cdot \text{ld} \frac{1}{p_k}. \quad (1.60)$$

Summiert wird in dieser Gleichung über k -fach indizierte Größen. Je größer k gewählt wird, um so kleiner wird der erhaltene Entropiewert sein (falls die Folgeelemente überhaupt statistische Bindungen besitzen).

Als die Entropie einer Nachrichtenquelle mit Gedächtnis bezeichnet man den Grenzwert:

$$H = \lim_{k \rightarrow \infty} H_k. \quad (1.61)$$

Die nach den Gleichungen (1.55) bis (1.61) berechenbaren statistischen Kenngrößen zeigen folgende Größenrelationen:

$$H \leq \dots H_k \leq \dots \leq H_2 \leq H_1 \leq H_0. \quad (1.62)$$

Ein sehr einfaches Beispiel soll die Bedeutung dieser Entropien etwas durchsichtiger machen.

Beispiel: Wir betrachten eine binäre alternierende Folge:

...0L0L0L0L0L0L0...

Der Entscheidungsgehalt beträgt hier 1 bit/Symbol, ebenso die nach (1.55) berechnete Entropie H_1 , die nur die Auftrittshäufigkeiten berücksichtigt. Mögliche Wertepaare sind hier nur „0L“ und „L0“, so daß Gleichung (1.59) für die zweite Entropienäherung den Wert 0.5 bit/Symbol liefert. Da für beliebige k -Tupel ebenfalls nur jeweils zwei Kombinationen möglich sind (jede mit der Wahrscheinlichkeit 0.5), ergibt sich für deren Entropie nach (1.60) der Wert $1/k$. Die Entropie dieser alternierenden Binärfolge, berechenbar als Grenzwert für k gegen unendlich, ist dementsprechend 0. Das heißt, die unendliche alternierende Folge liefert keine Information.

1.4.2 Quellen mit Markoveigenschaften

Wertdiskrete Folgen mit statistischen Bindungen werden meist durch *Markovprozesse* modelliert. Bild 1.8 zeigt das Übergangsdiagramm eines binären Markovprozesses erster Ordnung mit den beiden möglichen Symbolen „A“ und „B“.

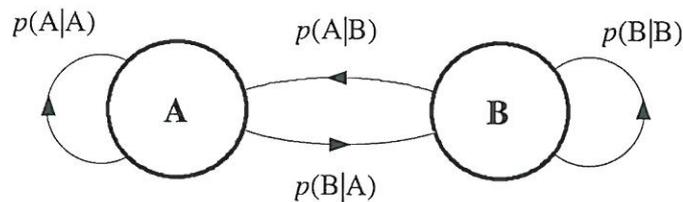


Bild 1.8: Markovprozeß mit $M = 2$ Zuständen.

Von den 4 Übergangswahrscheinlichkeiten sind nur zwei frei wählbar, da jeweils die Summe von zweien den Wert 1 ergibt:

$$p(A|A) + p(B|A) = 1, \quad (1.63)$$

$$p(A|B) + p(B|B) = 1. \quad (1.64)$$

Der Entscheidungsgehalt des Markovprozesses beträgt 1 bit/Symbol (wegen $M = 2$). Zur Berechnung der ersten Entropienäherung entsprechend (1.55) benötigt man die *stationären Auftretts-wahrscheinlichkeiten*. Diese berechnen sich entsprechend den Gleichungen (vgl. [15]):

$$p(A) = \frac{p(A|B)}{p(A|B) + p(B|A)}, \quad (1.65)$$

$$p(B) = \frac{p(B|A)}{p(A|B) + p(B|A)}. \quad (1.66)$$

Die beiden Symbole sind gleichwahrscheinlich, wenn $p(A|B) = p(B|A)$ gilt, d.h. wenn das Markovdiagramm symmetrisch aufgebaut ist (vgl. Bild 1.9).

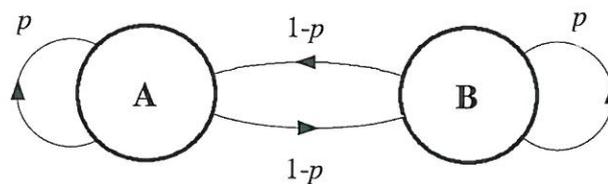


Bild 1.9: Betrachteter binärer symmetrischer Markovprozeß

Für alle Werte von p (zwischen 0 und 1 incl. der beiden Grenzwerte) gilt hier $p(A) = p(B) = 0.5$ und dementsprechend ist:

$$H_1 = H_0 = 1 \text{ bit/Symbol}. \quad (1.67)$$

Bei unsymmetrischem Markovdiagramm (und folglich ungleichen Auftretts-häufigkeiten) ergibt sich hierfür ein kleinerer Wert.

Betrachtet man ein Symbolpaar, so erhält man folgende Verbundwahrscheinlichkeiten:

$$p(A \cap A) = p(A) \cdot p(A|A), \quad (1.68)$$

$$p(A \cap B) = p(A) \cdot p(B|A), \quad (1.69)$$

$$p(B \cap A) = p(B) \cdot p(A|B), \quad (1.70)$$

$$p(B \cap B) = p(B) \cdot p(B|B). \quad (1.71)$$

Daraus folgt für die Verbundentropie:

$$\begin{aligned} H_2' &= p(A) \cdot p(A|A) \cdot \text{ld} \frac{1}{p(A) \cdot p(A|A)} + p(A) \cdot p(B|A) \cdot \text{ld} \frac{1}{p(A) \cdot p(B|A)} + \\ &+ p(B) \cdot p(A|B) \cdot \text{ld} \frac{1}{p(B) \cdot p(A|B)} + p(B) \cdot p(B|B) \cdot \text{ld} \frac{1}{p(B) \cdot p(B|B)}. \end{aligned} \quad (1.72)$$

Die Kennzeichnung mit einem Hochkomma soll darauf hinweisen, daß es sich hier um die Entropie eines Tupels handelt. Diese ist mit (1.57) vergleichbar.

Ersetzt man die Logarithmen der Produkte durch Summen von Logarithmen, so kann man für (1.72) schreiben:

$$H_2' = H_1 + H_M \quad (1.73)$$

mit

$$\begin{aligned} H_1 &= p(A) \cdot (p(A|A) + p(B|A)) \cdot \text{ld} \frac{1}{p(A)} + p(B) \cdot (p(A|B) + p(B|B)) \cdot \text{ld} \frac{1}{p(B)} = \\ &p(A) \cdot \text{ld} \frac{1}{p(A)} + p(B) \cdot \text{ld} \frac{1}{p(B)} = S(p(A)) = S(p(B)) \end{aligned} \quad (1.74)$$

und

$$\begin{aligned} H_M &= p(A) \cdot p(A|A) \cdot \text{ld} \frac{1}{p(A|A)} + p(A) \cdot p(B|A) \cdot \text{ld} \frac{1}{p(B|A)} + \\ &p(B) \cdot p(A|B) \cdot \text{ld} \frac{1}{p(A|B)} + p(B) \cdot p(B|B) \cdot \text{ld} \frac{1}{p(B|B)}. \end{aligned} \quad (1.75)$$

Der erste Term ist gleich der ersten Entropienäherung nach (1.55), bei binären Markovprozessen ist dieser gleich der Shannonfunktion $S(p)$. Dagegen muß der zweite Term tatsächlich entsprechend (1.75) berechnet werden.

Beim symmetrischen Markovprozeß nach Bild 1.9 erhält man:

$$H_1 = S(0.5) = 1, \quad (1.76)$$

$$H_M = S(p), \quad (1.77)$$

$$H_2' = 1 + S(p). \quad (1.78)$$

Die Entropie des Zweiertupels pro Symbol ergibt sich daraus zu

$$H_2 = \frac{1}{2} \cdot H_2' = 0.5 \cdot (1 + S(p)). \quad (1.79)$$

In ähnlicher Weise kann man die Entropie eines Tripels berechnen (vgl. Vorbereitungsfrage V7):

$$H_3' = H_1 + 2 \cdot H_M, \quad (1.80)$$

$$H_3 = \frac{1}{3} \cdot (H_1 + 2 \cdot H_M) = \frac{1}{3} + \frac{2}{3} \cdot S(p). \quad (1.81)$$

Setzt man diese Berechnungen für immer größere Werte von k fort, so erhält man schließlich:

$$H_k = \frac{1}{k} \cdot (H_1 + (k-1) \cdot H_M). \quad (1.82)$$

Die Entropie der Folge ist entsprechend (1.61) der Grenzwert dieser k -ten Näherung für k gegen unendlich:

$$H = \lim_{k \rightarrow \infty} H_k = H_M = S(p). \quad (1.83)$$

Die in (1.79) bis (1.83) jeweils zuerst angegebenen Gleichungen gelten für Markovquellen mit beliebigem M und beliebigen Übergangswahrscheinlichkeiten. Daraus folgt, daß zur Bestimmung der Entropie einer Markovquelle nur die ersten beiden Näherungen ermittelt werden müssen, da stets folgender Zusammenhang gilt:

$$H = H_M = 2 \cdot H_2 - H_1. \quad (1.84)$$

Dadurch wird der Rechenaufwand durch die Kenntnis, daß die Quelle Markoveigenschaften besitzt, gegenüber einer Quelle mit unbekanntem statistischen Bindungen drastisch reduziert.

Die in (1.79) bis (1.83) rechts angegebenen Gleichungen gelten nur für eine symmetrische binäre Markovquelle. Bei diesen ist $H = S(p)$. Für $p = 0.5$ sind die Symbole der Folge statistisch voneinander unabhängig und die Entropie mit $H = 1$ bit/Symbol maximal. Der Grenzwert $p = 1$ kennzeichnet die lange „A“- bzw. die lange „B“-Folge (je nach Startwert), während $p = 0$ für die alternierende Folge ...ABABABABABA... steht. In all diesen Fällen ist die Entropie $H = 0$.

1.4.3 Entropie von Texten

Bisher haben wir uns vorwiegend mit künstlich erzeugten Symbolfolgen beschäftigt. Eine natürliche wertdiskrete Nachrichtenquelle ist z.B. geschriebener Text, der natürlich auch informationstheoretisch analysiert werden kann.

Betrachtet man ein Buch mit 300 Seiten, 42 Zeilen pro Seite und 80 Zeichen pro Zeile, so besteht bei sequentiellem Lesen die hier vorliegende Symbolfolge aus ca. $N = 10^6$ Zeichen. Wird dieser Text entsprechend dem 8 Bit-ASCII-Code binär abgespeichert, so benötigt man hierfür etwa $8 \cdot 10^6$ Binärzeichen (ca. 8 MByte). Wird diese Datenmenge mit einer Baudrate von 9600 bit/s übertragen, so benötigt man hierfür etwa 14 Minuten.

Bei deutschsprachigem Text könnte man bei Unterscheidung von Klein- und Großbuchstaben, Berücksichtigung von Umlauten und Interpunktion vielleicht mit 64 unterschiedlichen Zeichen auskommen. Dadurch würde sich der benötigte Speicherplatz um $1 - \log_2(64)/\log_2(256) = 0.25$ vermindern und die Übertragungszeit sich um den gleichen Faktor verkürzen.

Der Entscheidungsgehalt dieser Nachrichtenquelle mit 64 Zeichen beträgt 6 bit/Zeichen. Will man die erste Entropienäherung entsprechend (1.55) ermitteln, so müßte man alle 64 Symbolwahrscheinlichkeiten empirisch (d.h. als Häufigkeiten) bestimmen. Für die zweite und dritte Näherung, die von Zeichentupeln bzw. -tripeln ausgehen und somit die statistischen Bindungen zum vorangegangenen Symbol bzw. zu den beiden vorangegangenen Symbolen mit berücksichtigen, sind bereits 4096 bzw. 262144 Häufigkeiten zu ermitteln. Ebenso viele Speicherplätze sind erforderlich.

Geht man weiterhin davon aus, daß für jedes mögliche k -Tupel im Mittel etwa 100 Entsprechungen für eine einigermaßen sichere Statistik erforderlich sind, so benötigt man bereits zur

Bestimmung der 3. Entropienäherung eine Nachrichtenquelle mit etwas mehr als 26 Millionen Zeichen, also ein Buch mit ca. 8000 Seiten.

Die Entropie von englischen bzw. deutschen Texten wurde schon früh von Shannon [14] bzw. Küpfmüller [9] zwischen 1.0 und 1.6 bit/Zeichen angegeben. Hierbei wurden jeweils 26 Buchstaben und das Leerzeichen berücksichtigt. Aus einer von Küpfmüller angegebenen Skizze, in der die einzelnen Entropienäherungen gemäß (1.60) in Abhängigkeit der Ordnung k dargestellt sind (vgl. Bild 1.10), läßt sich abschätzen, daß die statistischen Bindungen der deutschen Sprache mindestens bis $k = 100$ reichen.

Zur Ermittlung der 100. Entropienäherung benötigt man 2^{600} (etwa 10^{180}) Wahrscheinlichkeiten, ebenso viele Speicherplätze und - für eine gesicherte Statistik - nochmals um den Faktor 100 mehr Zeichen.

Eine berechnete Frage ist nun: Wie hat Küpfmüller im Jahre 1954 (und vor ihm bereits Shannon für die englische Sprache) die Entropie der deutschen Sprache ermittelt? Eine richtige Antwort darauf lautet: So nicht!

1.4.4 Entropieabschätzung nach Küpfmüller

Küpfmüller geht von einem Alphabet mit 26 Buchstaben aus (keine Unterscheidung zwischen Groß- und Kleinschreibung, keine Umlaute und Satzzeichen, Nichtberücksichtigung des Leerzeichens). Der Entscheidungsgehalt ergibt sich somit zu

$$H_0 = \text{ld}(26) = 4.7 \text{ bit/Buchstabe.} \quad (1.85)$$

Die Häufigkeiten der einzelnen Buchstaben in deutschen Texten wurden bereits 1939 veröffentlicht (häufigster Buchstabe: „e“ mit 16.7%, seltenster Buchstabe „x“ mit 0.02%). Aus diesen Häufigkeiten errechnete Küpfmüller die erste Entropienäherung zu:

$$H_1 \approx 4.1 \text{ bit/Buchstabe.} \quad (1.86)$$

Als nächstes wurden - ausgehend von dem 1898 von F.W. Kaeding herausgegebenen Werk „Häufigkeitswörterbuch der deutschen Sprache“ - die Silbenhäufigkeiten ausgewertet. Dabei wurde zwischen Stamm-, Vor- und Endsilben unterschieden.

Die 400 häufigsten Stammsilben (beginnend mit „de“) bilden etwa die Hälfte (47 %) eines deutschen Textes und tragen zur Entropie mit 4.15 bit/Silbe bei. Der Beitrag der 242 häufigsten Vorsilben (an der Spitze: „ge“, Gesamtanteil 9 %) ist laut Küpfmüller 0.82 bit/Silbe, die der 118 meistgebrauchten Endsilben („en“, 30 %) 1.62 bit/Silbe.

Der fehlende Rest von 14 % verteilt sich auf nicht gezählte Silben. Nimmt man an, daß es davon 4000 gibt und daß diese gleichverteilt sind, so liefert dieser Rest einen Beitrag von ca. 2 bit/Silbe. Die gesamte Silbenentropie ist somit

$$H_{\text{Silbe}} \approx 4.15 + 0.82 + 1.62 + 2.0 \approx 8.6 \text{ bit/Silbe.} \quad (1.87)$$

Die durchschnittliche Buchstabenzahl je Silbe wurde zu 3.03 ermittelt. Daher ergibt sich als Näherungswert der Entropie für $k = 3$:

$$H_3 \approx \frac{8.6}{3.03} \approx 2.8 \text{ bit/Buchstabe.} \quad (1.88)$$

In ähnlicher Weise kann man mit den Wörtern verfahren. Die Hälfte eines deutschen Textes wird durch nur 322 Wörter gebildet, wobei die fünf häufigsten „die“, „der“, „und“, „zu“, „in“ sind.

Diese 322 häufigsten Wörter liefern einen Beitrag von 4.5 bit/Wort. Die restlichen 50 % wurden geschätzt, wobei von 40000 Wörtern ausgegangen wurde. Weiterhin wurde angenommen, daß die Häufigkeit der selteneren Wörter umgekehrt proportional zu ihrer Ordnungszahl abnimmt. Mit diesen Voraussetzungen ist der Gesamtbeitrag 11.0 bit/Wort. Die durchschnittliche Anzahl der Buchstaben pro Wort ist nach dieser Auszählung 5.53. Daraus ergibt sich als weiterer Näherungswert für $k = 5.5$:

$$H_{5.5} \approx 2.0 \text{ bit/Buchstabe.} \quad (1.89)$$

Natürlich kann k entsprechend der Definition (1.60) nur ganzzahlige Werte annehmen. Die Gleichung (1.89) ist deshalb so zu interpretieren, daß sich für $k = 5$ ein etwas größerer, für $k = 6$ dagegen ein etwas kleinerer Wert ergibt.

In Bild 1.10 sind die oben berechneten Werte mit logarithmischer Abszisse aufgetragen. Denkt man sich durch die 3 berechneten Punkte eine Kurve gelegt, so kommt man für k gegen unendlich zur tatsächlichen Entropie. Die Extrapolation ist allerdings sehr vage.

Da Küpfmüller für die Statistik von Wortgruppen oder ganzen Sätzen keine Unterlagen vorfand, hat er den Grenzwert abgeschätzt. Die Abschätzung basiert auf folgender Überlegung: ein beliebiger zusammenhängender Text wird hinter einem bestimmten Wort abgedeckt. Der vorhergehende Text wird gelesen, und es wird versucht, das folgende Wort aus diesem Text und dem Zusammenhang zu ermitteln. Bei einer großen Zahl solcher Versuche ergibt die Zahl der Treffer, bezogen auf die Gesamtzahl der Versuche, ein Maß für die Bindungen zwischen den Wörtern und Sätzen. Es zeigt sich, daß bei ein und derselben Textart, z.B. Romane, wissenschaftliche Schriften usw. ein und desselben Autors, verhältnismäßig rasch ein konstanter Endwert dieses Trefferverhältnisses erreicht wird (von etwa 100 ... 200 Versuchen ab). Das Trefferverhältnis hängt aber ziemlich stark von der Art des Textes ab. Für verschiedenartige Texte ergeben sich Werte zwischen 15 und 33 %, mit dem Mittelwert bei 22 %. Im Durchschnitt können also 22 % der Wörter aus dem Zusammenhang ermittelt werden. Anders ausgedrückt, die Zahl der Wörter eines langen Textes kann mit dem Faktor 0.78 reduziert werden, ohne daß der Nachrichtengehalt des Textes eine Einbuße erfährt. Damit wird definitionsgemäß die Entropie zu

$$H \approx 0.78 \cdot 2.0 = 1.56 \text{ bit/Buchstabe.} \quad (1.90)$$

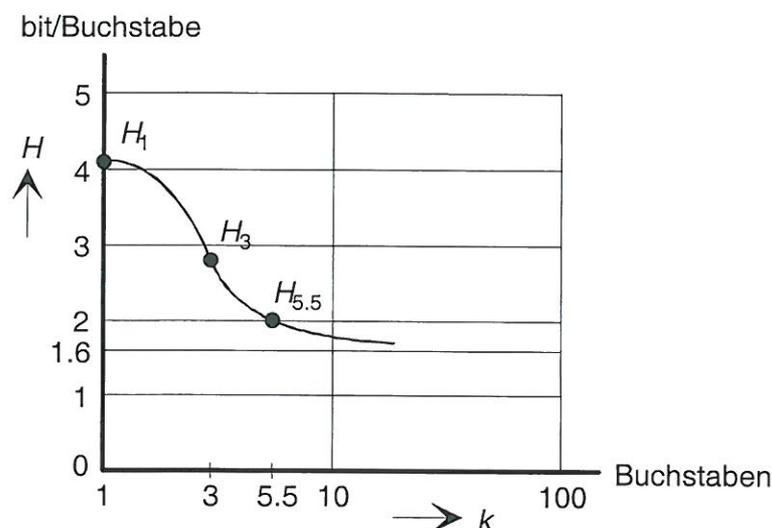


Bild 1.10: Näherungswerte der Entropie der deutschen Sprache (entnommen aus [9])

Küpfmüller hat diesen Wert mit einer vergleichbaren empirischen Untersuchung der Silben überprüft und kam mit (1.88) und dem ermittelten Reduktionsfaktor von 0.54 auf die Entropie 1.51 bit/Buchstabe. Um einen Vergleich mit Shannons informationstheoretischer Analyse von englischem Text anstellen zu können, hat Küpfmüller das Leerzeichen in sein Ergebnis mit eingerechnet, indem er seine empirisch gefundene Entropie noch mit dem Faktor 5.5/6.5 (mittlere Wortlängen ohne bzw. mit Berücksichtigung des Leerzeichens) multipliziert. So kommt er zum Wert $H = 1.3$ bit/Buchstabe, einem etwas höheren Wert, als Shannon für die Entropie der englischen Sprache gefunden hat ($H = 1$ bit/Buchstabe).

Dieser Richtwert von 1.3 bit/Buchstabe wird auch heute noch für informationstheoretische Abschätzungen herangezogen. Dieser besagt, daß bei einer idealen Codierung (die keiner kennt und die auch niemand angeben kann) ein deutscher Text entsprechend der relativen Redundanz von $(4.7 - 1.3)/4.7 = 0.72$ verkürzt werden könnte.

	deutsch	englisch
Buchstabennäherung 0. Ordnung	ITVWDGAKNAJTSQOSRM OIA QVFWTKHXD	ITVWDGAKNAJTSQOSR MOIA QVFWTKHXD
Buchstabennäherung 1. Ordnung	EME GKNEET ERS TITBL BTZENFNDBGD EAI E LASZ BETEATR IASMIRCH EGEOM	OCRO HLI RGWR NMIEL WIS EU LL NBNESEBYA TH EEI ALHENHHTTPA
Buchstabennäherung 2. Ordnung	AUSZ KEINU WONDINGLIN DUFNRN ISAR STEISBERER ITEHM ANORER	ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE
Buchstabennäherung 3. Ordnung	PLANZEUNDGES PHIN INE UNDEN ÜBBEICHT GES AUF ES SO UNG GAN DICH	IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF
Buchstabennäherung 4. Ordnung	ICH FOLGEMÄSZIG BIS STEHEN DISPONON SEELE NAMEN	
Wortnäherung 1. Ordnung	DENKEN ES ENTSAGEN ICH ZU WENN AUS DIESE VERANSTALTET ZEIT	REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT
Wortnäherung 2. Ordnung	WEIL JEDER ANLAGE HAT NACH DEM PFERDE NICHT ALLEIN DER HERR WILL ALS OB ICH FAST JEDES HAUS ZU SITZEN	THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITE THAT THE CHARACTER OF THIS

Bild 1.11: Künstlich erzeugte deutsche und englische Texte (entnommen aus [9]).

Es soll nochmals darauf hingewiesen werden, daß sich dieser Reduktionsfaktor auf ein Alphabet mit 27 Zeichen bezieht. Da heutzutage jedes Textverarbeitungssystem mit 256 unterschiedlichen darstellbaren Zeichen arbeitet, ergibt sich für die Entropie sicher ein größerer Wert als 1.3 bit/Symbol.

Abschließend sind in Bild 1.11 noch künstlich erzeugte deutsche und englische Texte angegeben. Die Buchstabennäherung 0. Ordnung geht von 27 gleichwahrscheinlichen Zeichen aus. Hier ist fast kein Unterschied zwischen deutschem und englischem Text feststellbar. Bei der ersten Buchstabennäherung werden bereits die unterschiedlichen Auftrittswahrscheinlichkeiten berücksichtigt, bei den Näherungen höherer Ordnung auch die vorangegangenen Zeichen. Bei einer Synthese 4. Ordnung (die Wahrscheinlichkeiten richten sich nach den drei zuletzt ausgewählten Zeichen) ergeben sich bereits erkennbare Worte. Entsprechend gibt die Wortnäherung 1. Ordnung einen Satz entsprechend den Wortwahrscheinlichkeiten wieder, die 2. Ordnung berücksichtigt auch das vorangegangene Wort.

1.4.5 Einige Simulationsergebnisse

Die Angaben von Küpfmüller hinsichtlich der Entropie der deutschen Sprache (vgl. Abschnitt 1.4.4) sollen nun mit einigen Simulationsergebnissen des Programms „*wdit*“ verglichen werden. Dieses berechnet neben dem Entscheidungsgehalt nach (1.21) die Entropienäherungen entsprechend (1.60) für $k = 1, 2$ und 3 . Als Symbolvorrat sind alle Buchstaben des deutschen Alphabets inklusive „ä“, „ö“, „ü“ und „ß“ zugelassen, außerdem Ziffern (ZI), Interpunktionszeichen (IP), Sonderzeichen (SZ) und Leerzeichen (LZ). Zwischen Klein- und Großschreibung wird nicht unterschieden. Der Symbolumfang beträgt dementsprechend $M = 34$.

Analysiert wurde z.B. die deutschsprachige Bibel mit $N = 4368593$ Symbolen. Die Ergebnisse sind in Bild 1.12 zusammengestellt. Betrachten wir zunächst die Spalte mit $M = 34$. Da Küpfmüller nur 26 Buchstaben zuließ, ist hier der Entscheidungsgehalt mit etwa 5.1 gegenüber 4.7 bit/Buchstabe geringfügig größer. Die Entropienäherung H_3 liefert hier ca. den Wert 3.3 bit/Buchstabe, während von Küpfmüller hierfür 2.8 bit/Buchstabe (also ebenfalls um ca. 0.5 weniger) angegeben wurde. Dieser größere Wert ist sicher u.a. auf den größeren Symbolumfang des Programms „*wdit*“ zurückzuführen.

Läßt man die Symbolklassen „ZI“, „IP“ und „SZ“ ($M = 31$, mittlere Spalte) außer Betracht, so ergibt sich bei allen Entropienäherungen ein geringfügig kleinerer Wert. Da aber auch Küpfmüller den Wert H_3 ohne Berücksichtigung des Leerzeichens ermittelt hat, muß dieser eigentlich mit dem in der letzten Spalte ($M = 30$) angegebenen Wert verglichen werden. Danach ist die Abweichung mit mehr als 0.66 bit/Buchstabe noch stärker.

Ein Vergleich der Spalten ($M = 30$) und ($M = 31$) macht deutlich, daß die Entropie bei Berücksichtigung des Leerzeichens trotz größerem Entscheidungsgehalt um mehr als 0.2 bit/Buchstabe abnimmt. Dieses Ergebnis wurde auch von Küpfmüller in etwa berücksichtigt (Subtraktion um 0.3 bit/Buchstabe).

In Bild 1.13 sind die vergleichbaren Ergebnisse der englischsprachigen Bibel zusammengestellt. Nach diesen Zahlenwerten ist aus informationstheoretischer Sicht kein entscheidender Unterschied zwischen deutschen und englischen Texten festzustellen.

	$M = 34$	$M = 31$ (ohne ZI, IP, SZ)	$M = 30$ (ohne ZI, IP, SZ, LZ)
N	4368593	4178307	3422356
H_0	5.087	4.954	4.907
H_1	4.185	4.062	4.126
H_2	3.682	3.622	3.786
H_3	3.281	3.245	3.466

*Bild 1.12: Einige Simulationsergebnisse der deutschsprachigen Bibel
(Entropie in bit/Buchstabe).*

	$M = 34$	$M = 31$ (ohne ZI, IP, SZ)	$M = 30$ (ohne ZI, IP, SZ, LZ)
N	42887381	4059343	3234995
H_0	5.044	4.907	4.858
H_1	4.174	4.121	4.012
H_2	3.699	3.600	3.795
H_3	3.288	3.222	3.506

*Bild 1.13: Einige Simulationsergebnisse der englischsprachigen Bibel
(Entropie in bit/Buchstabe).*

2 Vorbereitungsfragen

V1: Shannonfunktion

- a) Zeigen Sie, daß die durch Gl. (1.19) definierte Shannonfunktion symmetrisch um den Punkt $p = 0.5$ ist, d.h. daß gilt: $S(p) = S(1 - p)$.
- b) Für welchen p -Wert ist $S(p)$ maximal? Wie groß ist dieser Maximalwert?
- c) Wie groß ist $S(p)$ an den beiden Rändern $p = 0$ bzw. $p = 1$?
- d) Berechnen Sie $S(p)$ für $p = 0.1, 0.25, 0.5, 0.75, 0.9$.

p	0.1	0.25	0.5	0.75	0.9
$S(p)$					

V2: Entropie einer Ternärquelle ($M = 3$)

Betrachten Sie das Werfen einer Roulettekugel als ternäre Nachrichtenquelle mit den Symbolen R(Rot), S(Schwarz), W(Weiß). Werden Manipulationen am Gerät ausgeschlossen, so sind die einzelnen Würfe statistisch voneinander unabhängig, und es gilt für die Auftrittswahrscheinlichkeiten:

$$p_R = p_S = \frac{18}{37}; p_W = \frac{1}{37}.$$

- a) Berechnen Sie die Entropie dieser „Quelle“ unter Benutzung der Shannonfunktion (1.19).

- b) Wie groß ist der Informationsgehalt von Rot, Schwarz und Weiß?
- c) Zeigen Sie an diesem Beispiel, daß das unwahrscheinlichste Symbol trotz des größeren Informationsgehaltes einen kleineren Beitrag zur Entropie liefern kann als jedes der beiden anderen Symbole.
- d) Wie groß ist die maximale Entropie (Entscheidungsgehalt) sowie die relative Redundanz dieser Nachrichtenquelle?
- e) Geben Sie die Werte von H , H_0 und r an, wenn das „Roulette“ nicht hinsichtlich der Farben „R“, „S“ und „W“ als ternäre Nachrichtenquelle aufgefaßt wird, sondern bezüglich der Zahlen „0“, „1“, ..., „36“.

V3: Transinformationsgehalt

Durch Würfeln werden parallel zwei Symbolfolgen erzeugt, wobei der Symbolumfang jeweils $M = 2$ beträgt. Die Folge X hat die beiden Elemente „T“ (falls die Augenzahl drei oder kleiner) und „H“ (falls die Augenzahl vier oder größer). Dagegen ist $Y = „G“$, falls die Augenzahl gerade ist, andernfalls ist $Y = „U“$.

- a) Geben Sie die speziellen Folgen X und Y an, wenn folgende Zahlen gewürfelt werden:

A = 3 5 2 6 3 1 4 4 1 2 6 1 3 2 2 6 1 4 5 3

X =

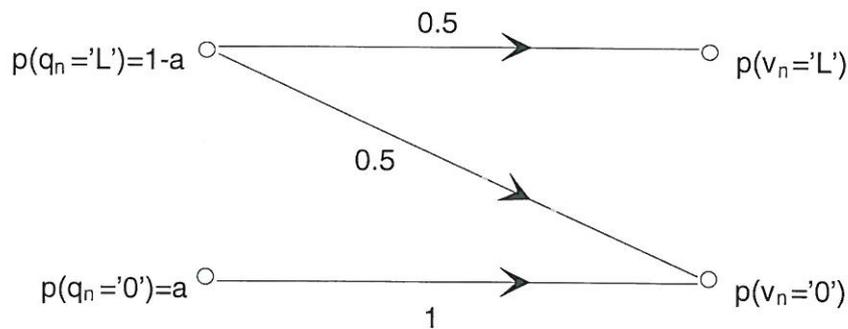
Y =

- b.) Wie groß sind die Entropien von X und Y ?
- c) Berechnen Sie die Verbundentropie.

- d) Wie groß ist der mittlere Transinformationsgehalt zwischen X und Y ?
- e) Berechnen Sie die Einzelbeiträge zum mittleren Transinformationsgehalt nach den Definitionsgleichungen (1.27) bzw. (1.30). Interpretieren Sie die Zahlenwerte.

V4: Kanalkapazität eines unsymmetrischen Binärkanals

Gegeben sei der folgende Binärkanal:



- a) Berechnen Sie die Quellen-, Sinken- und Verbundentropie in Abhängigkeit von a .
- b) Berechnen Sie den mittleren Transinformationsgehalt in Abhängigkeit von a .

- c) Optimieren Sie die Quellensymbolwahrscheinlichkeiten (Parameter a), so daß der mittlere Transinformationsgehalt H_T maximal wird. Geben Sie die (normierte) Kanalkapazität C' des obigen Binärkanals an.

V5: Huffman-Codierung

Gegeben sei eine Nachrichtenquelle mit statistisch unabhängigen Symbolen. Der Symbolumfang sei $M = 8$, die Auftretswahrscheinlichkeiten sind wie folgt gegeben:

$$p(„A“) = 0.01, \quad p(„B“) = 0.15, \quad p(„C“) = 0.05, \quad p(„D“) = 0.2, \\ p(„E“) = 0.02, \quad p(„F“) = 0.02, \quad p(„G“) = 0.05, \quad p(„LZ“) = 0.5.$$

Somit beträgt die Quellenentropie $H = 2.0993$ bit/Symbol.

- a) Wie groß ist die mittlere Codewortlänge eines Binärcodes, wenn die unterschiedlichen Auftretswahrscheinlichkeiten der Quellensymbole bei der Codierung nicht berücksichtigt werden?

- b) Wie groß ist in diesem Fall die Entropie pro Binärzeichen (Näherung 1. Ordnung), wenn obige Symbole fortlaufend binär codiert werden:

$$„A“ \rightarrow 000, „B“ \rightarrow 001, \dots, „LZ“ \rightarrow 111?$$

- c) Erstellen Sie einen Huffman-Code für obige Auftretswahrscheinlichkeiten. Zeichnen Sie die dazugehörige Baumstruktur.

- d) Berechnen Sie die mittlere Wortlänge des Huffman-Codes.

- e) Wie groß sind die Auftretswahrscheinlichkeiten der Binärsymbole „0“ und „L“ bei der Huffman-codierten Folge?
- f) Berechnen Sie die Entropie pro Binärsymbol (1. Näherung).

V6: Lempel-Ziv-Codierung

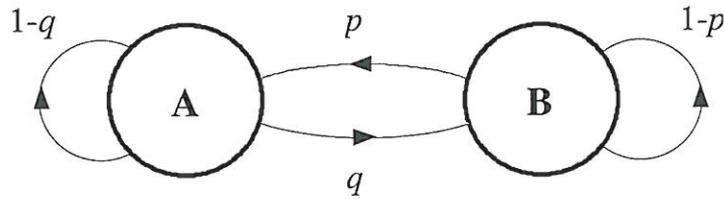
Erstellen Sie den Code nach Lempel-Ziv für die folgende Binärfolge mit $N = 45$:

110100011101101011110010101001011010110011011.

- a) Teilen Sie die Folge in Phrasen ein.
- b) Wieviele Phrasen gibt es? Mit wievielen Bit muß somit jede Phrase codiert werden?
- c) Geben Sie die Ausgangsfolge an.
- d.) Wieviele Binärzeichen (Bit) müssen entsprechend der Lempel-Ziv-Codierung übertragen werden?
- e.) Begründen Sie, daß bei längeren Folgen weniger als N Binärzeichen übertragen werden müssen.

V7: Entropie von Markovprozessen

Ein Markovprozeß erster Ordnung sei durch die beiden Parameter $p = p(A|B)$ und $q = p(B|A)$ vollständig beschrieben:



- Berechnen Sie die stationären Auftrittswahrscheinlichkeiten.
- Berechnen Sie die Entropienäherung 1. Ordnung entsprechend Gl. (1.74).
- Welcher Wert ergibt sich für die Entropienäherung 2. Ordnung?
- Berechnen Sie die beiden Entropienäherungen 1. und 2. Ordnung für eine symmetrische Markovquelle ($p = q$). Welcher Wert ergibt sich für die tatsächliche Entropie H ?
- Welche Entropie erhält man für $p = q = 0.25$?

3. Versuchsdurchführung

D1: Shannonfunktion

Einstellungen:

Parametersatz: Statistisch unabhängige Symbole.

Geöffnete Fenster: Alle.

Erzeugen Sie eine Binärfolge mit statistisch unabhängigen Symbolen („A“ und „B“), wobei die Auftretswahrscheinlichkeiten $p(A) = 0.10$ und $p(B) = 0.90$ betragen.

- a) Wie groß ist der Informationsgehalt der Symbole „A“ bzw. „B“ ?
- b) Wie groß sind die Entropie der Symbolfolge sowie der Entscheidungsgehalt? Welcher Wert ergibt sich damit für die relative Redundanz?
- c) Bestimmen Sie die Auftretswahrscheinlichkeiten sowie die Quellenentropie empirisch durch Simulation einer Folge der Länge $N = 1000$, $N = 5000$ bzw. $N = 20000$. Interpretieren Sie die Simulationsergebnisse. Hinweis: Da die Symbole als statistisch unabhängig vorausgesetzt werden, gilt: $H = H_1 = H_2 = H_3 = \dots = H_k = \dots$

N	1000	5000	20000	theor.
$p('A')$				
$p('B')$				
H				

- d) Wiederholen Sie den Versuch c.) für gleichwahrscheinliche Symbole.

N	1000	5000	20000	theor.
$p('A')$				
$p('B')$				
H				

- e) Bei welcher Auftrittswahrscheinlichkeit, $p(A) = 0.1$ oder $p(A) = 0.5$, liegt bei gleicher Folgenlänge N der empirisch ermittelte Entropiewert im Mittel näher am theoretisch erwarteten Wert? Begründen Sie dieses Ergebnis.

D2: Statistisch unabhängige Symbole

Einstellungen:

Parametersatz: Aus Datei \Rightarrow statistisch unabhängige Symbole.

Geöffnete Fenster: Alle.

Im Verzeichnis „DATA“ unterhalb des Hauptverzeichnisses sind vier Dateien (D2a.sun, D2b.sun, D2c.sun und D2d.sun) abgelegt, die jeweils eine M -stufige Symbolfolge der Länge $N = 20000$ enthalten. Analysieren Sie diese Folgen im Hinblick auf Symbolumfang, Auftrittswahrscheinlichkeiten und statistische Unabhängigkeit.

Anmerkung:

Obwohl alle Dateien die Kennung .sun aufweisen, stammen nicht alle Folgen von Quellen mit statistisch unabhängigen Symbolen. Die Wahrscheinlichkeiten im Parameterfenster sind - um nicht die Antworten vorwegzunehmen - alle mit Null angegeben.

D4 Mittlere Codewortlänge

Einstellungen:

Parametersatz : Statistisch unabhängige Symbole.

Geöffnete Fenster: Alle.

Erzeugen Sie nun eine ternäre Symbolfolge der Länge $N = 20000$ mit annähernd gleichen Wahrscheinlichkeiten: $p(A) = 0.333$, $p(B) = 0.333$, $p(C) = 0.334$.

a) Wie groß ist die Entropie?

b) Codieren Sie diese Folge nach Huffman. Wie lautet die Zuordnung?

c) Berechnen Sie die mittlere Codewortlänge. Vergleichen Sie diesen Wert mit der vom Programm numerisch ermittelten Codewortlänge.

d) Vergewissern Sie sich, daß folgender Zusammenhang gilt:

$$H_{\text{CW}} = \frac{H_q}{L_{\text{CW}}}.$$

Hierbei bezeichnet H_{CW} die Entropie der binären Codefolge.

e) Überprüfen Sie diese Aussage auch für die (nicht optimalen) Zuordnungen „C“ \Rightarrow „00“ bzw. „C“ \Rightarrow „01“.

D5 Entropie der Binärübertragung

Einstellungen:

Parametersatz : 2 statistisch abhängige Folgen.

Geöffnete Fenster: Alle.

Betrachten Sie den Binärkanal von Bild 1.1 mit den beiden Fehlerwahrscheinlichkeiten

$$p_{F0} = 0.01 \text{ und } p_{FL} = 0.2.$$

Simulieren Sie zunächst diesen Binärkanal mit den in Abschnitt 3.1 angegebenen Quellsymbolwahrscheinlichkeiten $p_{q0} = 0.1$ und $p_{qL} = 0.9$.

Anmerkung:

Im Programm ist dies mit dem Menüpunkt „Parameterwahl \Rightarrow 2 statistisch abhängige Folgen“ möglich. Ersetzen Sie z.B. „0“ durch „A“ und „L“ durch „B“. Die Eingangsfolge wird hier durch Großbuchstaben dargestellt, die Ausgangsfolge durch entsprechende Kleinbuchstaben.

- a) Welche Wahrscheinlichkeiten müssen Sie eingeben?
- b) Welche Werte liefert die Simulation für die einzelnen Entropien (siehe nachfolgende Tabelle) für $N = 1000$, $N = 10000$ und $N = 60000$.

N	H_q	H_v	H_T	$H_{q v}$	$H_{v q}$
1000					
10000					
60000					
theor.					

- c) Vergleichen Sie für $N = 60000$ die vom Programm ausgegebenen Auftrittshäufigkeiten (Verbund-, bedingte-, Rückschluß-) sowie die Entropiewerte von b) mit den in Abschnitt 1.3.1 ermittelten theoretischen Werten.

- d) Interpretieren Sie die ausgegebenen Auftrittswahrscheinlichkeiten für den vollständig gestörten Kanal. Welche Übergangswahrscheinlichkeiten müssen hier eingestellt werden?
- e) Interpretieren Sie die ausgegebenen Auftrittswahrscheinlichkeiten für den idealen Kanal. Welche Übergangswahrscheinlichkeiten müssen nun eingestellt werden?

D6 Kanalkapazität eines Binärkanals

Einstellungen:

Parametersatz : 2 statistisch abhängige Folgen.

Geöffnete Fenster: Alle.

Stellen Sie die gleichen Parameterwerte ein wie zu Beginn der Versuchsdurchführung D5 (siehe Teilaufgabe D5a, Länge 1000). Bestimmen Sie die Transinformation in Abhängigkeit der Symbolwahrscheinlichkeiten auf zwei Dezimalstellen genau. Verwenden Sie dabei den theoretischen Wert. Wie groß ist dementsprechend die Kanalkapazität?

$p(A)$						
H_T						

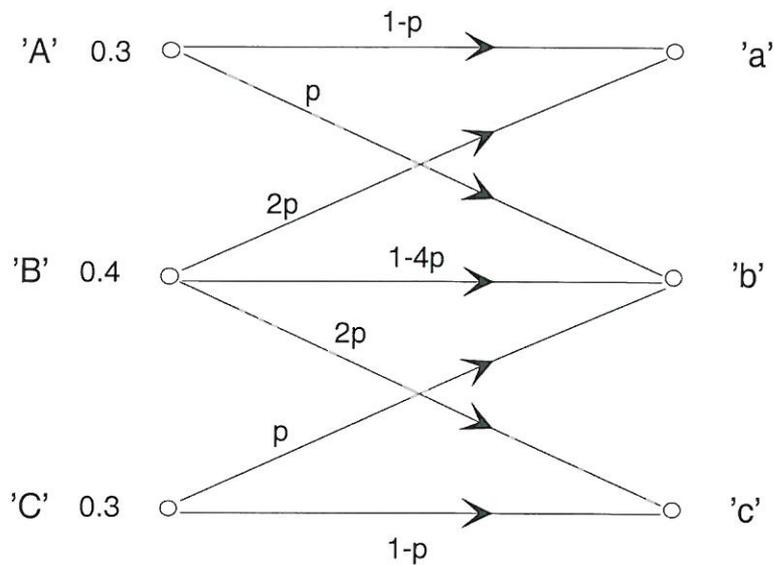
D7 Kanalkapazität eines Ternärkanals

Einstellungen:

Parametersatz: 2 statistisch abhängige Folgen.

Geöffnete Fenster: Alle.

Gegeben ist ein Digitalkanal mit den Symbolen „A“, „B“ und „C“ und nachfolgendem Übergangsdiagramm. Es sei $N = 10000$ und $p = 0.1$. Dieser Parametersatz ist im Programm bei dem Menüpunkt „2 statistisch abhängige Symbolquellen“ voreingestellt.



a) Welche Übergangswahrscheinlichkeiten müssen Sie einstellen?

b) Was ist aus den Verbundwahrscheinlichkeiten ersichtlich?

c) Was ist aus den bedingten Wahrscheinlichkeiten ersichtlich?

- d) Was ist aus den Rückschlußwahrscheinlichkeiten ersichtlich?
- e) Welcher Wert ergibt sich für den mittleren Transinformationsgehalt?
- f) Bestimmen Sie die Transinformation in Abhängigkeit der Symbolwahrscheinlichkeiten auf zwei Dezimalstellen genau. Verwenden Sie dabei wieder den theoretischen Wert. Wie groß ist dementsprechend die Kanalkapazität? Berücksichtigen Sie bei der Optimierung die Symmetrie des vorliegenden Kanals.

$p(A)$					
H_T					

$p(A)$				
H_T				

- g) Berechnen Sie den Transinformationsgehalt für $p(A) = p(C) = 0.5$ und $p(B) = 0$.

D8 Entropie einer binären Markovquelle

Einstellungen:

Parametersatz : Markovfolge (2 Symbole).

Göffnete Fenster: Alle.

Betrachtet wird wie in Vorbereitungsfrage V7 ein binärer Markovprozeß mit den beiden Symbolen „A“ und „B“ und den Übergangswahrscheinlichkeiten $p = p(A|B)$ und $q = p(B|A)$.

- a) Berechnen Sie mit den Ergebnissen von V7 die Entropienäherungen erster und zweiter Ordnung für $p = 0.5$ und $q = 0.25$.

- b) Überprüfen Sie diese Ergebnisse durch Simulation einer Folge der Länge $N = 60000$.

- c) Berechnen Sie die tatsächliche Entropie H der Folge.

- d) Welcher Wert wird sich für die Entropienäherung dritter Ordnung ergeben?

- e) Es sei nach wie vor $p = 0.5$. Für welchen Wert von q erhält man die maximale Entropie? Begründung.

- f) Für welchen Wert von p erhält man die maximale Entropie H , wenn $q = 0.25$ ist?

- g) Interpretieren Sie die Entropiewerte für $q = 0.25$ und $p = 0.25$ bzw. $p = 0.75$.

D9 Entropie einer ternären Markovquelle

Einstellungen:

Parametersatz : Markovfolge (3 Symbole).

Geöffnete Fenster: Alle.

Gegeben sei eine Markovquelle mit den drei Symbolen „A“, „B“ und „C“ sowie den nachfolgenden Übergangswahrscheinlichkeiten

$$p(A | A) = k \cdot p, \quad p(B | A) = k \cdot p, \quad p(C | A) = 1 - 2 \cdot k \cdot p,$$

$$p(A | B) = 1 - 2 \cdot p, \quad p(B | B) = p, \quad p(C | B) = p,$$

$$p(A | C) = p, \quad p(B | C) = 1 - 2 \cdot p, \quad p(C | C) = p.$$

Der Parameter k kann Werte zwischen 0 und 1 annehmen, der Parameter p zwischen 0 und 0.5 (Grenzwerte sind jeweils möglich). Für die Teilaufgaben a) und b) gelte $p = 0.25$.

a) Diskutieren Sie die Ergebnisse für $k = 1$.

b) Diskutieren Sie den Kurvenverlauf $H(k)$.

c) Für welchen Parametersatz ergibt sich die maximal mögliche Entropie?

4 Musterlösungen der Vorbereitungsfragen

Lösung V1:

- a) Ersetzt man in (1.19) p durch $1 - p$ und umgekehrt, so ergibt sich der gleiche Wert.
 b) Die Ableitung von $S(p)$ ergibt:

$$\frac{dS(p)}{dp} = \text{ld}\left(\frac{1}{p}\right) - p \cdot \frac{\text{ld}(e)}{p} - \text{ld}\left(\frac{1}{1-p}\right) + (1-p) \cdot \frac{\text{ld}(e)}{1-p} = \text{ld}\left(\frac{1}{p}\right) - \text{ld}\left(\frac{1}{1-p}\right) = \text{ld}\left(\frac{1-p}{p}\right)$$

$$\frac{dS(p)}{dp} \stackrel{!}{=} 0: \frac{1-p}{p} = 1 \Rightarrow \underline{\underline{p = 0,5}}$$

Für $p < 0,5$ ist die Ableitung stets positiv, d.h. das Maximum liegt bei $p = 0,5$:

$$S(p = 0,5) = 0,5 \cdot \text{ld}(2) + 0,5 \cdot \text{ld}(2) = \underline{\underline{1}}$$

- c) Für $p = 1$ gilt: $p \cdot \text{ld}\left(\frac{1}{p}\right) = 0$

Für $p = 0$ gilt nach de l'Hospital: $\lim_{p \rightarrow 0} p \cdot \ln(p) = \lim_{p \rightarrow 0} \frac{\ln(p)}{1/p} = \lim_{p \rightarrow 0} \frac{1/p}{-1/p^2} = \lim_{p \rightarrow 0} (-p) = 0$

$$\Rightarrow p \cdot \text{ld}\frac{1}{p} = 0 \Rightarrow \underline{\underline{S(p = 0) = S(p = 1) = 0}}$$

d)

p	0,1	0,25	0,5	0,75	0,9
S(p)	0,469	0,811	1,000	0,811	0,469

Lösung V2:

- a) $H = p_R \cdot \text{ld}\left(\frac{1}{p_R}\right) + p_S \cdot \text{ld}\left(\frac{1}{p_S}\right) + p_W \cdot \text{ld}\left(\frac{1}{p_W}\right) = 2 \cdot p_R \cdot \text{ld}\left(\frac{1}{p_R}\right) + p_W \cdot \text{ld}\left(\frac{1}{p_W}\right)$.

Mit $\text{ld}\left(\frac{1}{p_R}\right) = \text{ld}\left(\frac{1}{2 \cdot p_R}\right) + 1$ und $2 \cdot p_R = 1 - p_W$ folgt daraus:

$$H = 1 - p_W + S(p_W) = \frac{36}{37} + S\left(\frac{1}{37}\right) = \underline{\underline{1,1522 \text{ bit}}}$$

- b) $I_R = I_S = \underline{\underline{1,0395 \text{ bit}}}$, $I_W = \underline{\underline{5,2095 \text{ bit}}}$.
 c) $H = H_W + H_R + H_S = (0,1407 + 0,5057 + 0,5057) \text{ bit} = \underline{\underline{1,1522 \text{ bit}}}$.

Obwohl I_W etwa fünfmal größer ist als I_R bzw. I_S , ist H_W deutlich kleiner als $H_R = H_S$.

- d) $H_0 = \text{ld}(3) = \underline{\underline{1,585 \text{ bit}}} \Rightarrow r = \frac{H_0 - H}{H_0} = \frac{1,585 - 1,152}{1,585} = \underline{\underline{27,3\%}}$.

- e) $H = H_0 = \text{ld}(37) = \underline{\underline{5,2095 \text{ bit}}} \Rightarrow r = 0$.

Lösung V3:

a) $A = 3 \ 5 \ 2 \ 6 \ 3 \ 1 \ 4 \ 4 \ 1 \ 2 \ 6 \ 1 \ 3 \ 2 \ 2 \ 6 \ 1 \ 4 \ 5 \ 3$

$X = T \ H \ T \ H \ T \ T \ H \ H \ T \ T \ H \ T \ T \ T \ T \ H \ T \ H \ H \ T$
$Y = U \ U \ G \ G \ U \ U \ G \ G \ U \ G \ G \ U \ U \ G \ G \ G \ U \ G \ U \ U$

b) $H_X = H_Y = S(0.5) = \underline{1}$

c) $p(T \cap G) = \frac{1}{6}, p(T \cap U) = \frac{2}{6}, p(H \cap G) = \frac{2}{6}, p(H \cap U) = \frac{1}{6}$

$$H_{XY} = \frac{1}{6} \cdot \text{ld} \frac{1}{1/6} + \frac{2}{6} \cdot \text{ld} \frac{1}{2/6} + \frac{2}{6} \cdot \text{ld} \frac{1}{2/6} + \frac{1}{6} \cdot \text{ld} \frac{1}{1/6} = \underline{1.9183 \text{ bit}}$$

d) $H_T = H_X + H_Y - H_{XY} = 1 + 1 - 1.9183 = \underline{0.0817 \text{ bit}}$

e) $I_T(T, G) = \text{ld} \frac{1/6}{0.5 \cdot 0.5} = \underline{-0.5850} \quad I_T(T, U) = \text{ld} \frac{2/6}{0.5 \cdot 0.5} = \underline{0.4150}$

$$I_T(H, G) = \text{ld} \frac{2/6}{0.5 \cdot 0.5} = \underline{0.4150} \quad I_T(H, U) = \text{ld} \frac{1/6}{0.5 \cdot 0.5} = \underline{-0.5850}$$

Die Kombinationen (T; G) bzw. (H; U) sind eher unwahrscheinlich. Die entsprechenden Transinformationen sind negativ. Der entsprechend den Verbundwahrscheinlichkeiten gewichtete Mittelwert ist gleich dem mittleren Transinformationsgehalt:

$$H_T = 2 \cdot \frac{1}{6} \cdot (-0.585) + 2 \cdot \frac{2}{6} \cdot 0.415 = \underline{0.0817}$$

Lösung V4:

a) $H_q = S(a) = \underline{a \cdot \text{ld} \frac{1}{a} + (1-a) \cdot \text{ld} \frac{1}{1-a}}$

Mit $p(v_n = "0") = (1-a) \cdot 0.5 + a = 0.5 \cdot (1+a)$ und $p(v_n = "L") = 0.5 \cdot (1-a)$:

$$H_v = S\left(\frac{1+a}{2}\right) = \underline{0.5 \cdot (1+a) \cdot \text{ld} \frac{2}{1+a} + 0.5 \cdot (1-a) \cdot \text{ld} \frac{2}{1-a}}$$

$$p(q_n = "0" \cap v_n = "0") = a \quad p(q_n = "0" \cap v_n = "L") = 0$$

$$p(q_n = "L" \cap v_n = "0") = 0.5 \cdot (1-a) \quad p(q_n = "L" \cap v_n = "L") = 0.5 \cdot (1-a)$$

$$H_{qv} = \underline{a \cdot \text{ld} \frac{1}{a} + (1-a) \cdot \text{ld} \frac{2}{(1-a)} = S(a) + 1 - a}$$

b) $H_T = H_q + H_v - H_{qv} = S(a) + S\left(\frac{1+a}{2}\right) - S(a) - 1 + a = \underline{S\left(\frac{1+a}{2}\right) - 1 + a}$

c) Mit dem Ergebnis von V1b) gilt: $\frac{dH_T}{da} = \frac{1}{2} \cdot \text{ld} \frac{1-(1+a)/2}{(1+a)/2} + 1 = 0$
 $\Rightarrow \frac{1}{2} \cdot \text{ld} \frac{1-a}{1+a} + 1 = 0 \Rightarrow \text{ld} \frac{1+a}{1-a} = 2 \Rightarrow \frac{1+a}{1-a} = 4 \Rightarrow \underline{\underline{a = 0.6}}$.

a	0,40	0,50	0,59	0,60	0,61	0,70
H_T	0.2813	0.3113	0.3218	0.3219	0.3218	0.3098

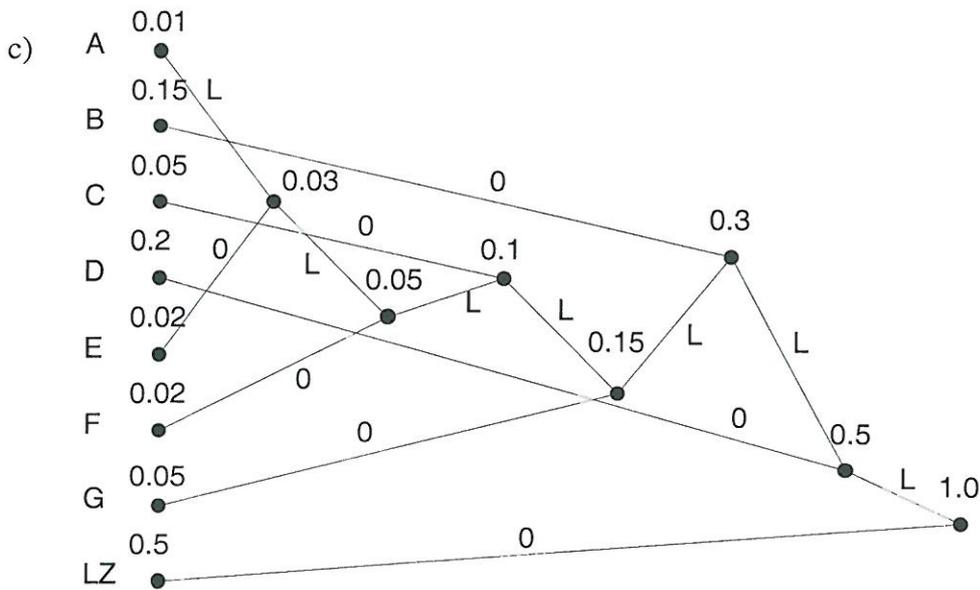
$C' = \text{Max}_a(H_T(a)) = S(0.8) - 0.4.$

Lösung V5:

a) $L_{CW} = 3 \text{ bit/Symbol}.$

b) $p("L") = \frac{0.15 + 0.05 + 2 \cdot 0.2 + 0.02 + 2 \cdot 0.02 + 2 \cdot 0.05 + 3 \cdot 0.5}{3} = 0.7533.$

$H_1 = S(p("L")) = \underline{\underline{0.8060 \text{ bit/Binärzeichen} .}}$



Somit lautet die Zuordnung:

A: 1111111, B: 110, C: 11110, D: 10,
 E: 1111110, F: 111110, G: 1110, LZ: 0.

d) $L_{CW} = 0.01 \cdot 7 + 0.15 \cdot 3 + 0.05 \cdot 5 + 0.2 \cdot 2 + 0.02 \cdot 7 + 0.02 \cdot 6 + 0.05 \cdot 4 + 0.5 \cdot 1 = \underline{\underline{2.13}}.$

e) $p("0") = 0.02 + 0.02 + 0.05 + 0.05 + 0.15 + 0.2 + 0.5 = \frac{0.99}{2.13} = \underline{\underline{0.4648}},$

$p("L") = 1 - p("0") = \frac{1.14}{2.13} = \underline{\underline{0.5352}}.$

f) $H_1 = S(p("L")) = \underline{\underline{0.9964 \text{ bit/Binärsymbol} .}}$

Lösung V6:

- a) 1,10,100,0,11,101,1010,111,1001,01,010,0101,10101,10011,011
- b) Es gibt $C = 15$ Phrasen. Jede Phrase muß mit 5 Bit codiert werden.
- c) (0000,1)(0001,0)(0010,0)(0000,0)(0001,1)(0010,1)(0110,0)(0101,1)(0011,1)(0100,1)
(1010,0) (1011,1)(0111,1)(1001,1)(1010,1)
- d) Insgesamt müssen $15 \cdot 5 = 75$ Bit übertragen werden.
- e) Bei langen Binärfolgen werden die Phrasen länger. Da die Phrasen durch die Position des Präfix beschrieben werden, welche eine kürzere Bitfolge aufweist als die Phrase selbst, wird eine Datenreduktion erreicht.

Lösung V7:

- a) Nach (1.64) und (1.65) gilt mit der hier gewählten Nomenklatur:

$$p(A) = \frac{p}{p+q}, \quad p(B) = \frac{q}{p+q}.$$

- b) Es gilt:

$$H_1 = S(p(A)) = S\left(\frac{p}{p+q}\right).$$

- c) Für die 4 Verbundwahrscheinlichkeiten eines Tupels ergeben sich folgende Werte:

$$p(A \cap A) = p(A) \cdot p(A|A) = \frac{p \cdot (1-q)}{p+q} \quad p(A \cap B) = p(A) \cdot p(B|A) = \frac{p \cdot q}{p+q}$$

$$p(B \cap A) = p(B) \cdot p(A|B) = \frac{q \cdot p}{p+q} \quad p(B \cap B) = p(B) \cdot p(B|B) = \frac{q \cdot (1-p)}{p+q}$$

Daraus folgt für die Entropie eines Zweiertupels:

$$H_2' = \frac{p \cdot (1-q)}{p+q} \cdot \text{ld} \frac{p+q}{p \cdot (1-q)} + \frac{q \cdot (1-p)}{p+q} \cdot \text{ld} \frac{p+q}{q \cdot (1-p)} + 2 \cdot \frac{p \cdot q}{p+q} \cdot \text{ld} \frac{p+q}{p \cdot q}.$$

Nach einigen Umformungen erhält man hierfür

$$H_2' = \frac{p \cdot (1-q)}{p+q} \left[S(p) + S(q) + p \cdot q \cdot \text{ld} \frac{(1-p) \cdot (1-q)}{p \cdot q} + (1-p-q) \cdot \text{ld}(1-p) \cdot (1-q) \right] + \text{ld}(p+q).$$

Für die Entropie pro Symbol gilt nach (1.79): $H_2 = \frac{1}{2} \cdot H_2'$.

- d) $H_1 = S(0.5) = 1$ bzw. $H_2' = 1 + \frac{1+p}{p} \cdot S(p) + \text{ld}[p \cdot (1-p)^{(1/p-1)}]$.

Bei Berücksichtigung aller Bindungen erhält man für die Entropie:

$$H = H_M = H_2' - H_1 = \frac{1+p}{p} \cdot S(p) + \text{ld}[p \cdot (1-p)^{(1/p-1)}].$$

- e) $H = 5 \cdot S(0.25) + \text{ld}\left(\frac{1}{4} \cdot \left(\frac{3}{4}\right)^3\right) = 5 \cdot S(0.25) + \text{ld}\left(\frac{27}{256}\right) \approx \underline{\underline{0.81}}.$

5 Musterlösungen der Versuchsdurchführung

In diesem Kapitel sind die Ergebnisse der Versuchsdurchführung zusammengestellt. Einige Simulationsergebnisse haben hier allerdings nur beispielhaften Charakter.

Ergebnisse D1:

a) $I_1 = I("A") = \text{ld} \frac{1}{0.1} = \underline{\underline{3.322 \text{ bit}}}$; $I_2 = I("B") = \text{ld} \frac{1}{0.9} = \underline{\underline{0.152 \text{ bit}}}$.

b) Entropie entsprechend (1.17): $H = \underline{\underline{0.469 \text{ bit}}}$ (siehe V1d);

Entscheidungsgehalt nach (1.21): $H_0 = \underline{\underline{1 \text{ bit}}}$;

relative Redundanz (1.24): $r = \frac{1 - 0.469}{1} = \underline{\underline{53.1\%}}$.

- c) Bei größerem N ist die Abweichung der durch Simulation ermittelten Werte von $p(„A“)$, $p(„B“)$ und H von den theoretisch berechneten Werten meist (d.h. im statistischen Mittel) geringer. Für N gegen unendlich wären die empirisch ermittelten Werte exakt gleich den berechneten Werten.

N	1000	5000	20000	theor.
$p(„A“)$	0.8930	0.9018	0.8997	0.9
$p(„B“)$	0.1070	0.0982	0.1003	0.1
H	0.490	0.463	0.470	0.469

d)

N	1000	5000	20000	theor.
$p(„A“)$	0.5120	0.5070	0.5005	0.5
$p(„B“)$	0.4880	0.4930	0.4995	0.5
H	0.997	1.000	1.000	1.000

- e) Bei $p(„A“) = 0.5$ liegt der ermittelte Entropiewert näher am theoretisch ermittelten Wert, da die Shannonfunktion in diesem Bereich eine geringere Steigung besitzt. Das bedeutet, daß die Abweichungen um diesen Wert geringer sind.

Ergebnisse D2:

- a) D2a.sun, es ist $M = 3$. Da hier die Entropien H_1, H_2, H_3 gleich dem Entscheidungsgehalt H_0 sind (zumindest innerhalb der Simulationsgenauigkeit), stammt diese Folge von einer redundanzfreien Quelle, das heißt, die Symbole „A“, „B“, „C“ sind statistisch unabhängig und gleichwahrscheinlich.

- b) D2b.sun, es ist $M = 3$. Die Symbole sind nicht gleichwahrscheinlich, vielmehr überwiegt das Symbol „C“. Aus der Symbolfolge erkennt man, daß auf jedes der drei Symbole wieder jedes folgen kann. Da die Entropiewerte H_1 , H_2 , H_3 gleich groß sind, folgt, daß die Symbole „A“, „B“ und „C“ statistisch unabhängig sind (d.h. keine statistischen Bindungen aufweisen).
- c) D2c.sun, es ist $M = 3$. Die Symbole sind gleichwahrscheinlich und statistisch abhängig. Wären sie statistisch unabhängig, so müßten die Entropien H_1 , H_2 , H_3 gleich groß und identisch mit dem Entscheidungsgehalt H_0 sein. Da sie jedoch unterschiedliche Werte besitzen, folgt daraus, daß die einzelnen Symbole statistische Bindungen aufweisen.
Dies erkennt man auch an der Symbolfolge: nach dem Symbol „A“ folgt nie das Symbol „C“, nach „B“ nie „A“ und nach „C“ nie „B“.
- d) D2d.sun, es ist $M = 8$. Die Auftrittswahrscheinlichkeiten sind unterschiedlich. Am häufigsten tritt das Leerzeichen auf (ca. 30 %), die Wahrscheinlichkeiten für „G“, „E“ und „F“ betragen dagegen nur jeweils ca. 5 %. Da die Entropien H_1 , H_2 und H_3 (im Rahmen der statistischen Genauigkeit) gleich sind, folgt daraus die statistische Unabhängigkeit der Symbole. Die relative Redundanz beträgt ca. 9.6 %.

Ergebnisse D3:

- a) Mit der Zuordnung nach V5c ergibt sich $L_{CW} = 2.13$.
- b) Vom Programm wird vorgeschlagen:
A: 1101010, B: 111, C: 11011, D: 10,
E: 1101011, F: 110100, G: 1100, LZ: 0.
Hierfür ergibt sich die gleiche mittlere Codewortlänge.
- c) Wird das letzte Symbol eines langen Codewortes verfälscht, so gibt es nur einen Symbolfehler, z.B. von „A“ nach „E“. Dagegen hat die Verfälschung eines kurzen Symbols (z.B. des Leerzeichens) schlimmere Auswirkungen. Der nachfolgende Text wird zwar eventuell wieder richtig sein, aber versetzt. Bei symbolweisem Vergleich sind somit alle nachfolgenden Symbole falsch.

Ergebnisse D4:

- a) $H = \underline{1.585 \text{ bit/Symbol}}$.
- b) „A“ -> 10; „B“ -> 11; „C“ -> 0.
- c) *Theoretischer Wert* : $L_{CW} \approx \frac{1}{3}(2 + 2 + 1) = \underline{1.667}$;
Empirischer Wert : $L_{CW} = \underline{1.6623}$.
- d) $H_{CW} = \frac{1.585}{1.667} = \underline{0.951}$. Das Programm liefert (als 10. Entropienäherung) den Wert 0.955.

Dieser Zusammenhang ist verständlich: Besitzt die Quelle nur die Information 1.585 bit/Symbol, so ist eine Binärfolge mit der mittleren Codewortlänge 1.667 natürlich redundant. Die Bindungen reichen allerdings sehr weit.

- e) Die mittlere Codewortlänge beträgt nun 2. Somit müßte die Entropie der Codefolge ca. 0.793 betragen. Mit der Näherung wird dieser Endwert noch nicht erreicht, doch ist aus dem Abfall der einzelnen Näherungen ersichtlich, daß dieser Wert erreicht werden kann.

Ergebnisse D5:

- a) $p(A) = 0.1, p(B) = 0.9, p(a|A) = 0.99, p(b|A) = 0.01, p(a|B) = 0.2, p(b|B) = 0.8$.

b)

N	H_q	H_v	H_T	$H_{q v}$	$H_{v q}$
1000	0.4464	0.8485	0.1926	0.2538	0.6560
10000	0.4652	0.8530	0.1923	0.2729	0.6607
60000	0.4671	0.8538	0.1935	0.2736	0.6603
theor.	0.4690	0.8541	0.1963	0.2727	0.6578

- c) Bei $N = 60000$ kommen die simulierten Werte ziemlich nahe an die theoretisch ermittelten Werte, die im Abschnitt 1.3.1 der Theorie angegeben sind.
- d) Alle Übergangswahrscheinlichkeiten müssen den Wert 0.5 haben. Von den empfangenen Symbolen kann hier in keiner Weise auf die tatsächlich gesendeten Symbole zurückgeschlossen werden. Der mittlere Transinformationsgehalt H_T ist 0. Es ergeben sich die Entropiewerte nach Bild 1.4.
- e) Nun sind die Übergangswahrscheinlichkeiten $p(a|A) = p(b|B) = 1$ zu setzen, alle anderen müssen den Wert 0 haben. Die beiden Folgen sind identisch. Die Transinformation ist gleich der Quellenentropie und auch gleich der Sinkenentropie. Es gilt Bild 1.3.

Ergebnisse D6:

Durch Variation der Symbolwahrscheinlichkeiten erhält man folgende Transinformationswerte:

$p(A)$	0.1	0.5	0.6	0.55	0.56	0.54
H_T	0.1963	0.5724	0.5735	0.5779	0.5778	0.5774

Die Kanalkapazität ist gleich dem maximalen Transinformationsgehalt bei optimierten Quellen-symbolwahrscheinlichkeiten. Es ergibt sich bei dem gegebenen Kanal $C' = 0.5779$ bit/Symbol für $p(„A“) = 0.55$. Da „A“ weniger gestört wird als „B“, ist der optimale Wert für $p(„A“)$ größer 0.5.

Ergebnisse D7:

- a)
- | | | |
|----------------|----------------|----------------|
| $p(a A) = 0.9$ | $p(b A) = 0.1$ | $p(c A) = 0$ |
| $p(a B) = 0.2$ | $p(b B) = 0.6$ | $p(c B) = 0.2$ |
| $p(a C) = 0$ | $p(b C) = 0.1$ | $p(c C) = 0.9$ |

- b) Das Symbol „A“ kann nie in „c“, das Symbol „C“ nie in „a“ verfälscht werden.

- c) Die durch Simulation empirisch gewonnenen bedingten Wahrscheinlichkeiten stimmen im Rahmen der Simulationsgenauigkeit gut mit den eingegebenen Werten (siehe Punkt a) überein.
- d) Bei den Rückschlußwahrscheinlichkeiten ist jeweils die Summe einer Spalte gleich 1 (im Gegensatz dazu ist bei den bedingten Wahrscheinlichkeiten jede Zeilensumme 1). Es gilt z.B. $p(b|B) = 0.61$ und $p(B|b) = 0.80$. Das bedeutet, die Wahrscheinlichkeit, daß das Quellensymbol „B“ in das Sinkensymbol „b“ übergeht, ist kleiner als die Wahrscheinlichkeit, daß bei empfangenem „b“ auf das Quellensymbol „B“ rückgeschlossen werden kann.
- e) $H_T = 0.7515$ bit/Symbol.
- f) Aufgrund der Symmetrie kann davon ausgegangen werden, daß die Transinformation für $p(A) = p(C)$ maximal ist. Es genügt also, $p(B)$ zu optimieren und $p(A) = p(C) = (1-p(B))/2$ zu setzen. Man erhält:

p(B)	0.4	0.3	0.2	0.1	0.0
H_T	0.7515	0.8217	0.8725	0.9006	0.9000

p(B)	0.11	0.05	0.04	0.06
H_T	0.8916	0.9045	0.9043	0.9043

Das Optimum von 0.9045 bit/Symbol ist bei $p(B) = 5\%$ und $p(A) = p(C) = 47.5\%$. Verzichtet man ganz auf das Symbol „B“, so ist der mittlere Transinformationsgehalt nur unwesentlich kleiner (0.9000 anstelle von 0.9045).

- g) $H_q = S(0.5) = 1$
 $p(a) = 0.5 \cdot 0.9 = 0.45$; $p(b) = 0.5 \cdot 0.1 + 0.5 \cdot 0.1 = 0.1$; $p(c) = 0.5 \cdot 0.9 = 0.45 \Rightarrow H_v = 1.3690$
 $p(a, A) = 0.9 \cdot 0.5 = 0.45$; $p(b, A) = 0.1 \cdot 0.5 = 0.05$;
 $p(b, C) = 0.1 \cdot 0.5 = 0.05$; $p(c, C) = 0.9 \cdot 0.5 = 0.45$
 $\Rightarrow H_{qv} = 1.4690$
 $H_T = H_q + H_v - H_{qv} = 2.3690 - 1.4690 = \underline{\underline{0.9000}}$

Ergebnisse D8:

- a) Nach V7b) gilt: $H_1 = S\left(\frac{2}{3}\right) = 0.9183$.

Entsprechend folgt aus V7c): $H_2' = 0.5 \cdot (1 + \lg(6)) = 1.7925$, $H_2 = 0.5 \cdot H_2' = 0.8962$.

- b) Die Simulation liefert $H_1 = 0.9170$ und $H_2 = 0.8947$.
- c) Aus (1.84) erhält man: $H = H_M = H_2' - H_1 = 1.7925 - 0.9183 = 0.8742$.

Die Simulation liefert $H = 0.8724$. Dieser Wert ergibt sich exakt, wenn man in obige Gleichung die durch Simulation gewonnenen Entropienäherungen erster und zweiter Ordnung einsetzt.

d) Nach (1.81) erhält man hierfür:

$$H_3 = \frac{1}{3} \cdot (H_1 + 2 \cdot H_M) = \frac{1}{3} \cdot 0.9183 + \frac{2}{3} \cdot 0.8742 = 0.8889.$$

- e) Die Simulation liefert für $q = 0.5$ den maximalen Wert 1.0. Dies zeigt, daß die so entstandene Symbolfolge redundanzfrei ist, das heißt, daß bei diesen Parameterwerten „A“ und „B“ jeweils die Auftretswahrscheinlichkeit 0.5 besitzen und die Symbole der Folge statistisch voneinander unabhängig sind. Diese beiden Aussagen belegen auch die angegebenen absoluten bzw. bedingten Häufigkeiten.
- f) Das Maximum ist nun bei $p = 0.45$. Die Entropiezunahme ist gegenüber $p = 0.5$ jedoch sehr gering.
- g) Für $p = q = 0.25$ sind beide Symbolwahrscheinlichkeiten $p(A) = p(B) = 0.5$. Dies ergibt sich auch aus den Gleichungen (1.64) und (1.65). Deshalb ist die erste Entropienäherung $H_1 = 1$. Aus den bedingten Wahrscheinlichkeiten erkennt man, daß statistische Bindungen vorhanden sind; deshalb werden die Entropienäherungen mit wachsendem k geringer. Der Grenzwert ist $H = 0.8$.

Für $q = 0.25$ und $p = 0.75$ laufen alle Entropiekurven durch den gleichen Punkt. Dies deutet darauf hin, daß in der Folge keine statistischen Bindungen vorhanden sind. Dies erkennt man auch an den Häufigkeiten, da nun $p(A) = p(A|A) = p(A|B) = 0.75$ ist. Die Entropie ist jedoch nicht 1, sondern es gilt mit den Auftretswahrscheinlichkeiten $p(A)=2/3$ und $p(B)=1/3$: $H = S(0.75) = 0.8113$.

Ergebnisse D9:

- a) Die maximale Entropie ergibt sich bei $p = 0.25$ für $k = 1$. In diesem Fall ist das Markovdiagramm symmetrisch, die Auftretswahrscheinlichkeiten sind gleich

$$p(A) = p(B) = p(C) = \frac{1}{3}.$$

Daraus folgt $H_1 = \text{ld}(3) = 1.585$ bit/Symbol. Nach jedem dieser drei Symbole folgt jeweils ein Symbol mit der Wahrscheinlichkeit $1/2$ und die beiden anderen mit $1/4$. Aufgrund dieser statistischen Bindungen ist die tatsächliche Entropie etwas kleiner:

$$H = H_M = 3 \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \text{ld}(2) + 6 \cdot \frac{1}{3} \cdot \frac{1}{4} \cdot \text{ld}(4) = \underline{\underline{1.5 \text{ bit/Symbol}}}$$

- b) Mit abnehmendem k (bei festem $p = 0.25$) folgt nach einem „A“ immer häufiger das Symbol „C“, für $k = 0$ sogar determiniert. Deshalb sinkt $H(k)$ mit fallendem k monoton bis zum Endwert $H(k = 0) = 1.111$ bit/Symbol.
- c) Für $p = 1/3$ und $k = 1$ erhält man $H = 1.585$ bit/Symbol. Ein größerer Wert ist mit $M = 3$ Symbolen nicht möglich. In diesem Sonderfall sind alle Übergangswahrscheinlichkeiten zwischen den Symbolen gleich $1/3$ ebenso wie die 3 Symbolwahrscheinlichkeiten.

Ergebnisse D10:

- a) Die drei häufigsten Symbole sind das Leerzeichen (LZ) mit 17.79 %, der Buchstabe „e“ mit 12.83 % und der Buchstabe „n“ mit 8.32 %.
- b) Am häufigsten vor einem Leerzeichen sind „IP“ mit 16.7 %, „n“ mit 14.8 % und „e“ mit 12.7 %. Nach einem Leerzeichen sind „d“ mit 13.6 %, „s“ mit 11.6 % und „u“ mit 9.8 % am wahrscheinlichsten.
- c) Die Auftretswahrscheinlichkeit von „q“ beträgt 0.01 % (seltenstes Zeichen). Unter der Bedingung, daß das nachfolgende Zeichen ein „u“ ist, erhöht sich die Wahrscheinlichkeit jedoch auf 0.14 %. Dagegen ist die bedingte Wahrscheinlichkeit, daß nach „q“ ein „u“ folgt, gleich 100 %.
- d) Ohne „e“ und „n“ ist der Text fast nicht lesbar.
- e) Für $M = 34$ gilt: $H_0 = \text{ld}(34) = 5.087$, $H_1 = 4.177$, $H_2 = 3.658$, $H_3 = 3.217$.
Für $M = 30$ gilt: $H_0 = \text{ld}(30) = 4.907$, $H_1 = 4.132$, $H_2 = 3.755$, $H_3 = 3.414$.

Der Anstieg ist vor allem darauf zurückzuführen, daß das Leerzeichen (LZ) nicht mit berücksichtigt wird. Der Einfluß der anderen Ausnahmezeichen („ZI“, „IP“, „SZ“) ist deutlich geringer.

Ergebnisse D11:

Bei den synthetisch erzeugten Texten kann man durchaus erkennen, ob von „deutschen“ oder „englischen“ Tripelwahrscheinlichkeiten ausgegangen wurde.