Überblick zu Kapitel 1 von „Einführung in die Informationstheorie"

Dieses erste Kapitel beschreibt die Berechnung und die Bedeutung der **Entropie**, die entsprechend Shannons Definition von Information ein Maß ist für die mittlere Unsicherheit über den Ausgang eines statistischen Ereignisses oder die Unsicherheit bei der Messung einer stochastischen Größe. Etwas salopp ausgedrückt quantifiziert die Entropie einer Zufallsgröße deren „Zufälligkeit".

Im Einzelnen werden behandelt:

- der *Entscheidungsgehalt* und die *Entropie* einer gedächtnislosen Nachrichtenquelle,
- die *binäre Entropiefunktion* und deren Anwendung auf *nichtbinäre Quellen*,
- die Entropieberechnung bei *gedächtnisbehafteten Quellen* und geeignete Näherungen,
- die Besonderheiten von *Markovquellen* hinsichtlich der Entropieberechnung,
- die Vorgehensweise bei Quellen mit großem Symbolumfang, zum Beispiel *natürliche Texte*,
- die *Entropieabschätzungen* nach Shannon und Küpfmüller.

Geeignete Literatur: [AM90] – [Bla87] – [CT06] – [Fan61] – [For72] – [Gal68] – [Har28] – [Joh92b] – [Kra13] – [Küp54] – [McE77] – [Meck09] – [PS02] – [Sha48] – [Sha51] – [WZ73]

Die grundlegende Theorie wird auf 28 Seiten dargelegt. Außerdem beinhaltet dieses erste Kapitel noch 32 Grafiken, acht Aufgaben und vier Zusatzaufgaben mit insgesamt 66 Teilaufgaben, sowie drei Lernvideos und drei Interaktionsmodule.

Zusammenstellung der **Lernvideos** (LV) zu den Grundlagen:

- **Klassische Definition der Wahrscheinlichkeit** (Dauer 5:19)
- **Bedeutung und Berechnung der Momente bei diskreten Zufallsgrößen** (Dauer 6:32)
- **Statistische Abhängigkeit und Unabhängigkeit** (3 Teile – Dauer 4:17, 3:48, 3:48)

Zusammenstellung der **Interaktionsmodule** (IM) zu den Grundlagen und Kapitel 1:

- **Entropien von Nachrichtenquellen** (zu Kapitel 1.1)
- **Ereigniswahrscheinlichkeiten einer Markovkette** (zu Kapitel 1.2)
- **Signale, AKF und LDS der Pseudoternärcodes** (zu Kapitel 1.2)

Weitere Informationen zum Thema sowie grafikbasierte Simulationsprogramme und Aufgaben mit ausführlichen Musterlösungen finden Sie im Versuch „Wertdiskrete Informationstheorie" des Praktikums *Simulation digitaler Übertragungssysteme*, das von Prof. Günter Söder (Lehrstuhl für Nachrichtentechnik) für Studierende der Elektro– und Informationstechnik an der TU München angeboten wird:

Herunterladen des Windows–Programms „WDIT" (Zip–Version)

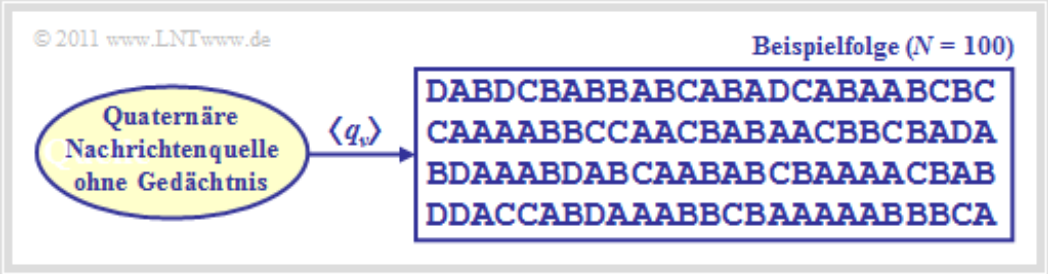
Herunterladen der dazugehörigen Texte (PDF–Datei)

Modell und Voraussetzungen für Kapitel 1.1 (1)

Wir betrachten eine wertdiskrete Nachrichtenquelle Q, die eine Folge $\langle q_\nu \rangle$ von Symbolen abgibt. Für die Laufvariable gilt $\nu = 1, \ldots, N$, wobei N „hinreichend groß" sein sollte. Jedes einzelne Quellensymbol q_ν entstammt einem Symbolvorrat $\{q_\mu\}$ mit $\mu = 1, \ldots, M$, wobei M den *Symbolumfang* bezeichnet:

$$q_\nu \in \{q_\mu\}\,, \text{ mit } \nu = 1, \ldots, N \text{ und } \mu = 1, \ldots, M\,.$$

Die Grafik zeigt eine quaternäre Nachrichtenquelle ($M = 4$) mit dem Alphabet {**A**, **B**, **C**, **D**}. Rechts ist eine beispielhafte Folge der Länge $N = 100$ angegeben.



Es gelten folgende Voraussetzungen:

- Die quaternäre Nachrichtenquelle wird durch $M = 4$ *Symbolwahrscheinlichkeiten* p_μ vollständig beschrieben. Allgemein gilt:

$$\sum_{\mu=1}^{M} p_\mu = 1\,.$$

- Die Nachrichtenquelle sei gedächtnislos, das heißt, die einzelnen Folgenelemente seien statistisch voneinander unabhängig:

$$\Pr\left(q_\nu = q_\mu\right) = \Pr\left(q_\nu = q_\mu \,|\, q_{\nu-1}, q_{\nu-2}, \ldots\right)\,.$$

- Da das Alphabet aus Symbolen (und nicht aus Zufallsgrößen) besteht, ist hier die Angabe von Erwartungswerten (linearer Mittelwert, quadratischer Mittelwert, Streuung, usw.) nicht möglich, aber auch aus informationstheoretischer Sicht nicht nötig.

Diese Eigenschaften werden auf der nächsten Seite mit einem Beispiel verdeutlicht.

Modell und Voraussetzungen für Kapitel 1.1 (2)

Beispiel: Für die Symbolwahrscheinlichkeiten einer Quaternärquelle gelte:

$$p_{\rm A} = 0.4,\ \ p_{\rm B} = 0.3,\ \ p_{\rm C} = 0.2,\ \ p_{\rm D} = 0.1\,.$$

Bei einer unendlich langen Folge ($N \to \infty$) wären die **relativen Häufigkeiten** $h_{\rm A}$, $h_{\rm B}$, $h_{\rm C}$ und $h_{\rm D}$ – also die a–posteriori–Kenngrößen – identisch mit den a–priori–Wahrscheinlichkeiten $p_{\rm A}$, $p_{\rm B}$, $p_{\rm C}$ und $p_{\rm D}$. Bei kleinerem N kann es aber durchaus zu Abweichungen kommen, wie die folgende Tabelle (Ergebnis einer Simulation) zeigt. Die Folge für $N = 100$ ist auf der letzten Seite angegeben.

	$N = 10^2$	$N = 10^3$	$N = 10^4$	$N = 10^5$	$N = 10^6$	$N \to \infty$
$h_{\rm A}$	42%	38.8%	40.77%	39.923%	39.967%	40%
$h_{\rm B}$	31%	29.1%	28.91%	29.887%	30.043%	30%
$h_{\rm C}$	18%	21.5%	19.93%	20.062%	19.948%	20%
$h_{\rm D}$	9%	10.6%	10.39%	10.128%	10.042%	10%

Aufgrund der Mengenelemente **A**, **B**, **C** und **D** können keine Mittelwerte angegeben werden. Ersetzt man die Symbole durch Zahlenwerte, zum Beispiel **A** ⇒ 1, **B** ⇒ 2, **C** ⇒ 3, **D** ⇒ 4, so ergeben sich

- für den **linearen Mittelwert:**

$$m_1 = {\rm E}\left[q_\nu\right] = {\rm E}\left[q_\mu\right] = 0.4 \cdot 1 + 0.3 \cdot 2 + 0.2 \cdot 3 + 0.1 \cdot 4 = 2\,,$$

- für den **quadratischen Mittelwert:**

$$m_2 = {\rm E}\left[q_\nu^2\right] = {\rm E}\left[q_\mu^2\right] = 0.4 \cdot 1^2 + 0.3 \cdot 2^2 + 0.2 \cdot 3^2 + 0.1 \cdot 4^2 = 5\,,$$

- für die **Standardabweichung** (Streuung) nach dem „Satz von Steiner”:

$$\sigma = \sqrt{m_2 - m_1^2} = \sqrt{5 - 2^2} = 1\,.$$

Entscheidungsgehalt – Nachrichtengehalt

Claude E. Shannon definierte 1948 im Standardwerk der Informationstheorie **[Sha48]** den Informationsbegriff als „Abnahme der Ungewissheit über das Eintreten eines statistischen Ereignisses". Machen wir hierzu ein gedankliches Experiment mit M möglichen Ergebnissen, die alle gleichwahrscheinlich seien:

$$p_1 = p_2 = \ldots = p_M = 1/M\,.$$

Unter dieser Annahme gilt:

- Ist $M = 1$, so wird jeder einzelne Versuch das gleiche Ergebnis liefern und demzufolge besteht keine Unsicherheit hinsichtlich des Ausgangs. Wird uns das Versuchsergebnis mitgeteilt, so haben wir dadurch natürlich auch keinen Informationsgewinn.
- Dagegen erfährt ein Beobachter bei einem Experiment mit $M = 2$, zum Beispiel dem „Münzwurf" mit der Ereignismenge {**Z**(ahl), **W**(app)} und den Wahrscheinlichkeiten $p_Z = p_W = 0.5$, durchaus einen Informationsgewinn. Die Unsicherheit, ob **Z** oder **W** geworfen wurde, wird aufgelöst.
- Beim Experiment „Würfeln" ($M = 6$) und noch mehr beim Roulette ($M = 37$) ist die gewonnene Information für den Beobachter noch deutlich größer als beim „Münzwurf", wenn er erfährt, welche Zahl gewürfelt bzw. welche Kugel gefallen ist.
- Schließlich sollte noch berücksichtigt werden, dass das Experiment „Dreifacher Münzwurf" mit den $M = 8$ möglichen Ergebnissen **ZZZ**, **ZZW**, **ZWZ**, **ZWW**, **WZZ**, **WZW**, **WWZ**, **WWW** die dreifache Information liefert wie der einfache Münzwurf ($M = 2$).

Die nachfolgende Festlegung erfüllt alle hier verbal aufgeführten Anforderungen an ein quantitatives Informationsmaß bei gleichwahrscheinlichen Ereignissen, gekennzeichnet durch den Symbolumfang M.

Definition: Der **Entscheidungsgehalt** einer Nachrichtenquelle hängt nur vom Symbolumfang M ab und ergibt sich zu

$$H_0 = \log M = \log_2 M \ \text{(in "bit")} = \ln M \ \text{(in "nat")} = \lg M \ \text{(in "Hartley")}\,.$$

Gebräuchlich ist hierfür auch die Bezeichnung *Nachrichtengehalt*. Da H_0 gleichzeitig den Maximalwert der **Entropie** H angibt, wird in *LNTwww* teilweise auch H_{max} als Kurzzeichen verwendet.

Anzumerken ist:

- Der Logarithmus wird in unserem Tutorial unabhängig von der Basis mit „log" bezeichnet. Die vier oben aufgestellten Kriterien werden aufgrund folgender Eigenschaften erfüllt:

$$\log 1 = 0\,,\ \ \log 37 > \log 6 > \log 2\,,\ \ \log M^k = k \cdot \log M\,.$$

- Meist verwenden wir den Logarithmus zur Basis 2 ⇒ *Logarithmus dualis* (ld), wobei dann die Pseudoeinheit „bit" – genauer: „bit/Symbol" – hinzugefügt wird:

$$\mathrm{ld}\ M = \log_2 M = \frac{\lg M}{\lg 2} = \frac{\ln M}{\ln 2}\,.$$

- Weiter findet man in der Literatur auch Definitionen, basierend auf dem natürlichen Logarithmus

(„ln”) oder dem Zehnerlogarithmus („lg”) entsprechend obigen Definitionen.

Informationsgehalt und Entropie

Wir verzichten nun auf die bisherige Voraussetzung, dass alle M möglichen Ergebnisse eines Versuchs gleichwahrscheinlich seien. Im Hinblick auf eine möglichst kompakte Schreibweise legen wir für diese Seite lediglich fest:

$$p_1 > p_2 > \ldots > p_\mu > \ldots > p_{M-1} > p_M\,, \quad \sum_{\mu=1}^{M} p_\mu = 1\,.$$

Unter dieser Voraussetzung betrachten wir nun den **Informationsgehalt** der einzelnen Symbole, wobei wir den *Logarithmus dualis* mit „ld"(manchmal auch mit „$\log_2$") bezeichnen :

$$I_\mu = \text{ld}\,\frac{1}{p_\mu} = -\,\text{ld}\,p_\mu \quad \text{(Einheit: bit oder bit/Symbol)}.$$

Man erkennt:

- Wegen $p_\mu \le 1$ ist der Informationsgehalt nie negativ. Im Grenzfall $p_\mu \to 1$ geht $I_\mu \to 0$. Allerdings ist für $I_\mu = 0 \to p_\mu = 1 \to M = 1$ auch der Entscheidungsgehalt $H_0 = 0$.
- Bei abfallenden Wahrscheinlichkeiten p_μ nimmt der Informationsgehalt kontinuierlich zu:

$$I_1 < I_2 < \ldots < I_\mu < \ldots < I_{M-1} < I_M\,.$$

Das heißt: Je weniger wahrscheinlich ein Ereignis ist, desto größer ist sein Informationsgehalt. Dieser Sachverhalt ist auch im täglichen Leben festzustellen:

- „6 Richtige" im Lotto nimmt man sicher eher war als „3 Richtige" oder gar keinen Gewinn.
- Ein Tsunami in Asien dominiert auch die Nachrichten in Deutschland über Wochen im Gegensatz zu den fast standardmäßigen Verspätungen der Deutschen Bahn.
- Eine Niederlagenserie von Bayern München führt zu Riesen–Schlagzeilen im Gegensatz zu einer Siegesserie. Bei 1860 München ist genau das Gegenteil der Fall.

Der Informationsgehalt eines einzelnen Symbols (oder Ereignisses) ist allerdings nicht sehr interessant. Durch Scharmittelung über alle möglichen Symbole q_μ bzw. durch Zeitmittelung über alle Folgenelemente q_ν erhält man dagegen eine der zentralen Größen der Informationstheorie.

Definition: Die **Entropie** einer Quelle gibt den mittleren Informationsgehalt aller Symbole an:

$$H = \overline{I_\nu} = \text{E}[I_\mu] = \sum_{\mu=1}^{M} p_\mu \cdot \text{ld}\,\frac{1}{p_\mu} = -\sum_{\mu=1}^{M} p_\mu \cdot \text{ld}\,p_\mu \quad \text{(Einheit: bit[/Symbol])}.$$

Die überstreichende Linie kennzeichnet eine Zeitmittelung und E[...] eine Scharmittelung.

Die Entropie ist ein Maß für

- die mittlere Unsicherheit über den Ausgang eines statistischen Ereignisses,
- die „Zufälligkeit" dieses Ereignisses,
- den mittleren Informationsgehalt einer Zufallsgröße.

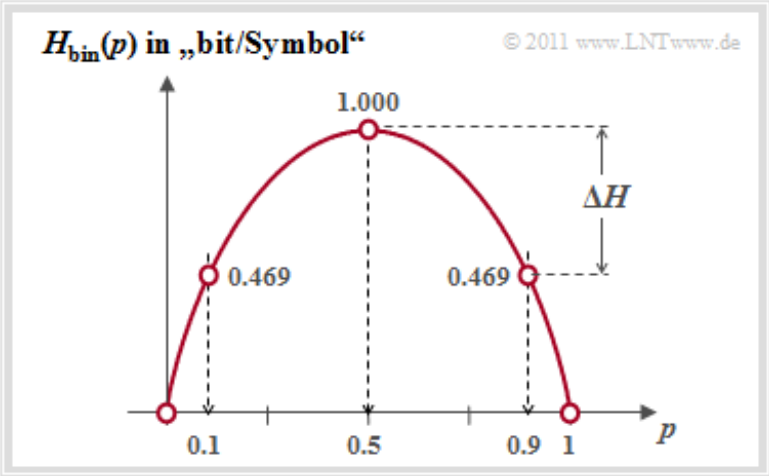
Binäre Entropiefunktion

Wir beschränken uns zunächst auf den Sonderfall $M = 2$ und betrachten eine binäre Quelle, die die beiden Symbole **A** und **B** abgibt. Die Auftrittwahrscheinlichkeiten seien $p_A = p$ und $p_B = 1 - p$.

Für die Entropie dieser Quelle gilt:

$$H_{\rm bin}(p) = p \cdot {\rm ld}\,\frac{1}{p} + (1-p) \cdot {\rm ld}\,\frac{1}{1-p} \quad \text{(Einheit: bit oder bit/Symbol)}.$$

Man nennt diese Funktion $H_{\rm bin}(p)$ die **binäre Entropiefunktion**. Die Entropie einer Quelle mit größerem Symbolumfang M lässt sich häufig unter Verwendung von $H_{\rm bin}(p)$ ausdrücken.



Die Grafik zeigt die Funktion $H_{\rm bin}(p)$ für die Werte $0 \le p \le 1$ der Symbolwahrscheinlichkeit von **A** (oder **B**). Man erkennt:

- Der Maximalwert $H_{\rm max} = 1$ bit ergibt sich für $p = 0.5$, also für gleichwahrscheinliche Binärsymbole. Dann liefern **A** und **B** jeweils den gleichen Beitrag zur Entropie.
- $H_{\rm bin}(p)$ ist symmetrisch um $p = 0.5$. Eine Quelle mit $p_A = 0.1$ und $p_B = 0.9$ hat die gleiche Entropie (Zufälligkeit) $H = 0.469$ bit wie eine Quelle mit $p_A = 0.9$ und $p_B = 0.1$.
- Die Differenz $\Delta H = H_{\rm max} - H$ gibt die **Redundanz** der Quelle an und $r = \Delta H/H_{\rm max}$ die relative Redundanz. Im genannten Beispiel ergeben sich $\Delta H = 0.531$ bit bzw. $r = 53.1\%$.
- Für $p = 0$ ergibt sich $H = 0$, da hier die Ausgangsfolge „**B B B** ...” sicher vorhersagbar ist. Eigentlich beträgt nun der Symbolumfang nur noch $M = 1$. Gleiches gilt für $p = 1$.

Es sollte noch erwähnt werden, dass die binäre Entropiefunktion *konkav* ist, da deren zweite Ableitung nach dem Parameter p für alle Werte von p negativ ist:

$$\frac{{\rm d}^2 H_{\rm bin}(p)}{{\rm d}\,p^2} = \frac{-1}{\ln(2) \cdot p \cdot (1-p)} < 0.$$

Nachrichtenquellen mit größerem Symbolumfang (1)

Auf der **ersten Seite** dieses Kapitels haben wir eine quaternäre Nachrichtenquelle ($M = 4$) mit den Symbolwahrscheinlichkeiten $p_A = 0.4$, $p_B = 0.3$, $p_C = 0.2$ und $p_D = 0.1$ betrachtet. Diese besitzt die folgende Entropie:

$$H_{\rm quat} = 0.4 \cdot \log_2 \frac{1}{0.4} + 0.3 \cdot \log_2 \frac{1}{0.3} + 0.2 \cdot \log_2 \frac{1}{0.2} + 0.1 \cdot \log_2 \frac{1}{0.1} =$$
$$= \frac{1}{\lg 2} \cdot \left[0.4 \cdot \lg \frac{1}{0.4} + 0.3 \cdot \lg \frac{1}{0.3} + 0.2 \cdot \lg \frac{1}{0.2} + 0.1 \cdot \lg \frac{1}{0.1}\right] = 1.845\ \text{bit}.$$

Oft ist der Umweg über den Zehnerlogarithmus $\lg x = \log_{10} x$ sinnvoll, da meist der *Logarithmus dualis* $\log_2 x$ (oder auch ld x) auf Taschenrechnern nicht zu finden ist.

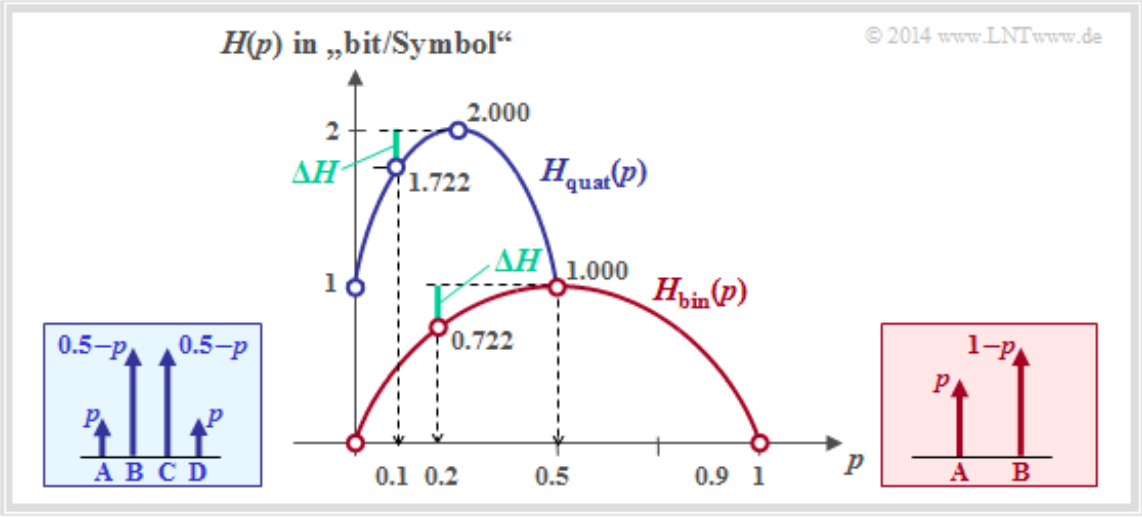
Bestehen zwischen den einzelnen Symbolwahrscheinlichkeiten Symmetrien wie im Beispiel

$$p_A = p_D = p, \quad p_B = p_C = 0.5 - p, \quad \text{mit } 0 \le p \le 0.5,$$

so kann zur Entropieberechnung auf die binäre Entropiefunktion zurückgegriffen werden:

$$H_{\rm quat} = 2 \cdot p \cdot \log_2 \frac{1}{p} + 2 \cdot (0.5 - p) \cdot \log_2 \frac{1}{0.5 - p} = 1 + H_{\rm bin}(2p).$$

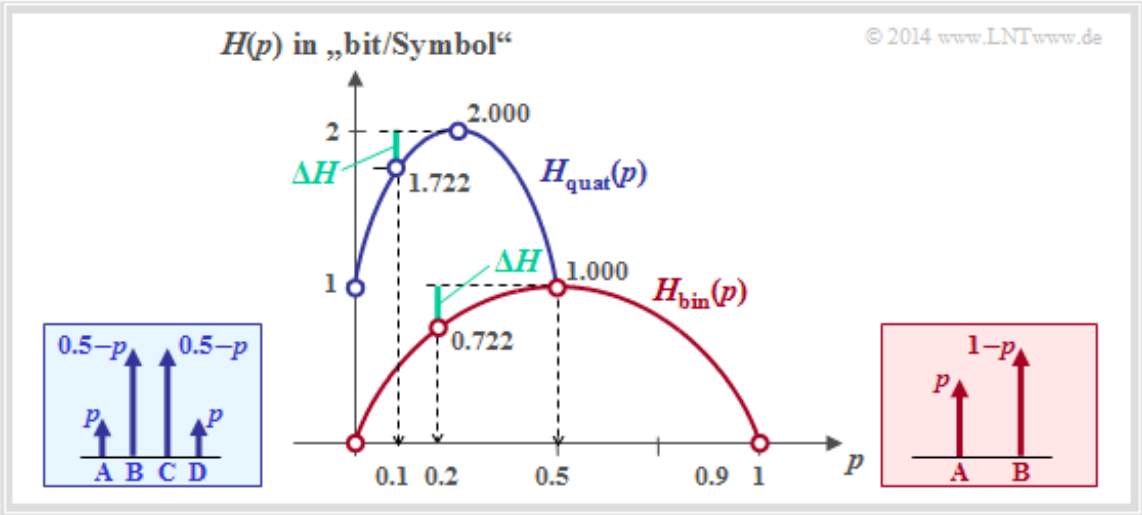
Die Grafik zeigt den Entropieverlauf der Quaternärquelle (blau) im Vergleich zur Binärquelle (rot) abhängig von p. Für die Quaternärquelle ist nur der Abszissenbereich $0 \le p \le 0.5$ zulässig.



Die Bildbeschreibung folgt auf der nächsten Seite.

Nachrichtenquellen mit größerem Symbolumfang (2)

Diskussion der Kurvenverläufe:



Man erkennt aus der blauen Kurve für die Quaternärquelle:

- Die maximale Entropie $H_{\rm max} = 2$ bit ergibt sich für $p = 0.25 \Rightarrow p_{\rm A} = p_{\rm B} = p_{\rm C} = p_{\rm D} = 0.25$, also wieder für gleichwahrscheinliche Symbole.
- Mit $p = 0$ bzw. $p = 0.5$ entartet die Quaternärquelle zu einer Binärquelle mit $p_{\rm B} = p_{\rm C} = 0.5$ bzw. $p_{\rm A} = p_{\rm D} = 0.5$. In diesem Fall ergibt sich die Entropie zu $H = 1$ bit.
- Die Quelle mit $p_{\rm A} = p_{\rm D} = p = 0.1$ und $p_{\rm B} = p_{\rm C} = 0.4$ weist folgende Entropie und (relative) Redundanz auf:

$$\begin{aligned} H &= 1 + H_{\rm bin}(2p) = 1 + H_{\rm bin}(0.2) = 1.722\,{\rm bit}\,,\\ \Delta H &= {\rm ld}\,M - H = 2\,{\rm bit} - 1.722\,{\rm bit} = 0.278\,{\rm bit}\,,\\ r &= \Delta H/({\rm ld}\,M) = 0.139\,. \end{aligned}$$

- Die Redundanz ΔH der Quaternärquelle mit $p = 0.1$ ist gleich 0.278 bit und damit genau so groß wie die Redundanz der Binärquelle mit $p = 0.2$.

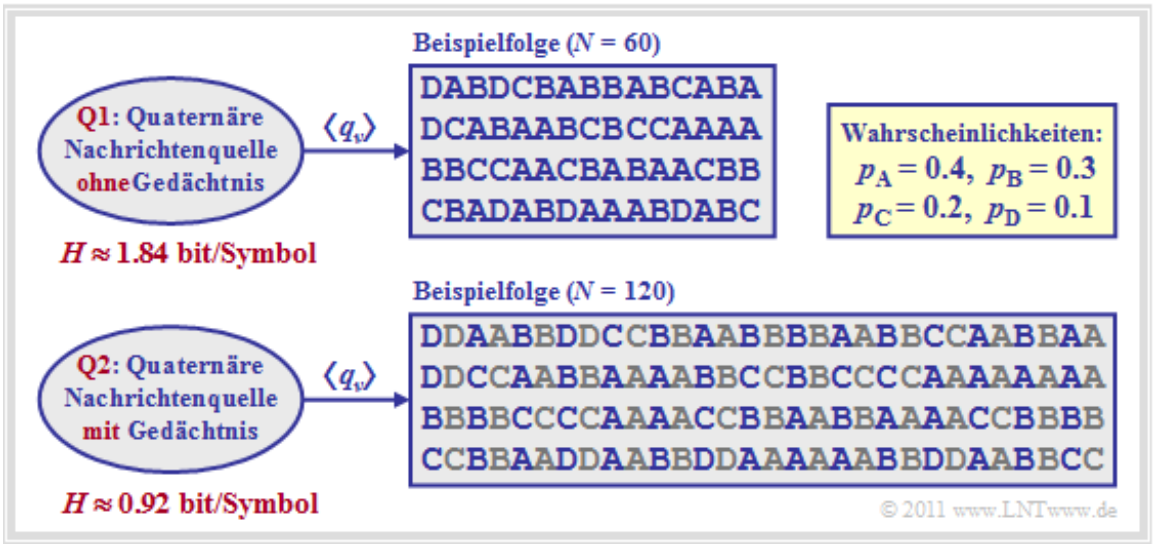
Anmerkung: Als Pseudoeinheit ist hier stets „bit” angegeben. Genauer wäre „bit/Symbol”.

Ein einfaches einführendes Beispiel

Zu Beginn des ersten Kapitels haben wir eine gedächtnislose Nachrichtenquelle mit dem Symbolvorrat {**A**, **B**, **C**, **D**} ⇒ $M = 4$ betrachtet. Eine beispielhafte Symbolfolge ist in der nachfolgenden Grafik als Quelle **Q1** nochmals dargestellt. Mit den Symbolwahrscheinlichkeiten $p_A = 0.4$, $p_B = 0.3$, $p_C = 0.2$ und $p_D = 0.1$ ergibt sich die Entropie zu

$$H = 0.4 \cdot \log_2 \frac{1}{0.4} + 0.3 \cdot \log_2 \frac{1}{0.3} + 0.2 \cdot \log_2 \frac{1}{0.2} + 0.1 \cdot \log_2 \frac{1}{0.1} \approx 1.84\,\text{bit/Symbol}.$$

Aufgrund der ungleichen Auftrittswahrscheinlichkeiten der Symbole ist die Entropie kleiner als der Entscheidungsgehalt $H_0 = \log_2 M = 2$ bit/Symbol.



Die Quelle **Q2** ist weitgehend identisch mit der Quelle Q1, außer, dass jedes einzelne Symbol nicht nur einmal, sondern zweimal nacheinander ausgegeben wird: **A** ⇒ **AA**, **B** ⇒ **BB**, usw.. Es ist offensichtlich, dass Q2 eine kleinere Entropie (Unsicherheit) aufweist als Q1. Aufgrund des einfachen Wiederholungscodes ist nun $H = 1.84/2 = 0.92$ bit/Symbol nur halb so groß, obwohl sich an den Auftrittswahrscheinlichkeiten nichts geändert hat.

Dieses Beispiel zeigt:

- Die Entropie einer gedächtnisbehafteten Quelle ist kleiner als die Entropie einer gedächtnislosen Quelle mit gleichen Symbolwahrscheinlichkeiten.
- Es müssen nun auch die statistischen Bindungen innerhalb der Folge $\langle q_\nu \rangle$ berücksichtigt werden, nämlich die Abhängigkeit des Symbols q_ν von den Vorgängersymbolen $q_{\nu-1}$, $q_{\nu-2}$, ...

Entropie hinsichtlich Zweiertupel (1)

Wir betrachten weiterhin die Quellensymbolfolge $\langle q_1, q_2, \ldots, q_{\nu-1}, q_\nu, q_{\nu+1}, \ldots\rangle$, interessieren uns aber nun für die Entropie zweier aufeinanderfolgender Quellensymbole. Alle Quellensymbole q_ν entstammen einem Alphabet mit dem Symbolunfang M, so dass es für die Kombination $(q_\nu, q_{\nu+1})$ genau M^2 mögliche Symbolpaare mit folgenden **Verbundwahrscheinlichkeiten** gibt:

$$\Pr(q_\nu \cap q_{\nu+1}) \le \Pr(q_\nu) \cdot \Pr(q_{\nu+1}).$$

Daraus ist die *Verbundentropie eines Zweier–Tupels* berechenbar:

$$H_2' = \sum_{q_\nu \in \{q_\mu\}} \sum_{q_{\nu+1} \in \{q_\mu\}} \Pr(q_\nu \cap q_{\nu+1}) \cdot \log_2 \frac{1}{\Pr(q_\nu \cap q_{\nu+1})} \quad \text{(Einheit: bit/Zweiertupel)}.$$

Der Index 2 symbolisiert, dass sich die so berechnete Entropie auf Zweiertupel bezieht. Um den mittleren Informationsgehalt pro Symbol zu erhalten, muss H_2' noch halbiert werden:

$$H_2 = \frac{H_2'}{2} \quad \text{(Einheit: bit/Symbol)}.$$

Um eine konsistente Nomenklatur zu erreichen, benennen wir nun die in **Kapitel 1.1** definierte Entropie mit H_1:

$$H_1 = \sum_{q_\nu \in \{q_\mu\}} \Pr(q_\nu) \cdot \text{ld} \frac{1}{\Pr(q_\nu)} \quad \text{(Einheit: bit/Symbol)}.$$

Der Index 1 soll darauf hinweisen, dass H_1 ausschließlich die Symbolwahrscheinlichkeiten berücksichtigt und nicht statistischen Bindungen zwischen Symbolen innerhalb der Folge. Mit dem Entscheidungsgehalt $H_0 = \log_2 M$ ergibt sich dann folgende Größenbeziehung:

$$H_0 \ge H_1 \ge H_2.$$

Bei statistischer Unabhängigkeit der Folgenelemente ist H_2 gleich H_1.

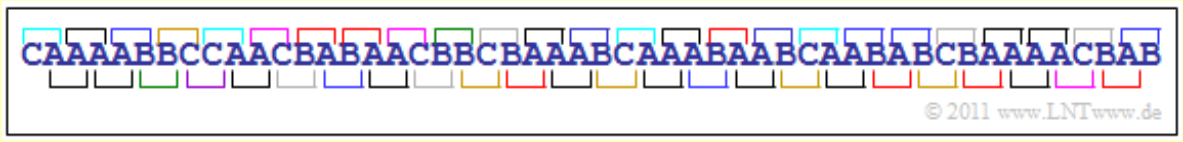
Die bisherigen Gleichungen geben jeweils einen Scharmittelwert an. Die für die Berechnung von H_1 und H_2 benötigten Wahrscheinlichkeiten lassen sich aber auch als Zeitmittelwerte aus einer sehr langen Folge berechnen oder – etwas genauer ausgedrückt – durch die entsprechenden **relativen Häufigkeiten** annähern.

Auf den nächsten Seiten werden die Aussagen dieser Seite anhand von Beispielen verdeutlicht.

Entropie hinsichtlich Zweiertupel (2)

Beispiel 1: Wir betrachten die Folge $\langle q_1, \text{ ... } , q_{50}\rangle$ entsprechend der folgernden Grafik:

- Die Folgenlänge ist $N = 50$.
- Die Folgenelemente q_ν entstammen dem Alphabet {**A**, **B**, **C**} ⇒ Symbolumfang $M = 3$.



Durch Zeitmittelung über die 50 Symbole erhält man die Symbolwahrscheinlichkeiten $p_A \approx 0.5$, $p_B \approx 0.3$ und $p_C \approx 0.2$, womit man die Entropienäherung erster Ordnung berechnen kann:

$$H_1 = 0.5 \cdot \log_2 \frac{1}{0.5} + 0.3 \cdot \log_2 \frac{1}{0.3} + 0.2 \cdot \log_2 \frac{1}{0.2} \approx 1.486\ \text{bit/Symbol}\,.$$

Aufgrund der nicht gleichwahrscheinlichen Symbole ist $H_1 < H_0 = 1.585$ bit/Symbol. Als Näherung für die Wahrscheinlichkeiten von Zweiertupeln erhält man aus der obigen Folge:

$$\begin{aligned} p_{AA} &= 14/49\,, \quad p_{AB} = 8/49\,, \quad p_{AC} = 3/49\,,\\ p_{BA} &= 7/49\,, \quad p_{BB} = 2/49\,, \quad p_{BC} = 5/49\,,\\ p_{CA} &= 4/49\,, \quad p_{CB} = 5/49\,, \quad p_{CC} = 1/49\,. \end{aligned}$$

Beachten Sie, dass aus den 50 Folgenelementen nur 49 Zweiertupel (**AA**, ... , **CC**) gebildet werden können, die in der obigen Grafik farblich unterschiedlich markiert sind.

Die daraus berechenbare Entropienäherung H_2 sollte eigentlich gleich H_1 sein, da die gegebene Symbolfolge von einer gedächtnislosen Quelle stammt. Aufgrund der kurzen Folgenlänge $N = 50$ und der daraus resultierenden statistischen Ungenauigkeit ergibt sich ein etwas kleinerer Wert: $H_2 \approx 1.39$ bit/Symbol.

Auf der nächsten Seite folgen noch zwei weitere Beispiele.

Entropie hinsichtlich Zweiertupel (3)

Verdeutlichen wir uns die Berechnung der Entropienäherungen H_1 und H_2 an weiteren Beispielen.

Beispiel 2: Wir betrachten eine **gedächtnislose Binärquelle** mit gleichwahrscheinlichen Symbolen, das heißt es gelte $p_A = p_B = 1/2$.

- Die ersten zwanzig Folgeelemente lauten:

  $\langle q_\nu \rangle$ = **BBAAABAABBBBBAAAABAB** ...

- Aufgrund der gleichwahrscheinlichen Symbole und $M = 2$ gilt:

  $H_1 = H_0 = 1$ bit/Symbol.

- Die Verbundwahrscheinlichkeit p_{AB} der Kombination **AB** ist gleich $p_A \cdot p_B = 1/4$. Ebenso gilt $p_{AA} = p_{BB} = p_{BA} = 1/4$. Damit erhält man für die zweite Entropienäherung

$$H_2 = \frac{1}{2} \cdot \left[\frac{1}{4} \cdot \log_2 4 + \frac{1}{4} \cdot \log_2 4 + \frac{1}{4} \cdot \log_2 4 + \frac{1}{4} \cdot \log_2 4\right] = 1\,\text{bit/Symbol}\,.$$

Hinweis: Aus der oben angegebenen Folge ergeben sich aufgrund der kurzen Länge etwas andere Verbundwahrscheinlichkeiten, nämlich $p_{AA} = 6/19$, $p_{BB} = 5/19$ und $p_{AB} = p_{BA} = 4/19$.

Das nächste Beispiel liefert dagegen das Ergebnis $H_2 < H_1$.

Beispiel 3: Die zweite hier betrachtete Folge ergibt sich aus der oberen Folge durch Anwendung eines einfachen **Wiederholungscodes** (wiederholte Symbole in Grau):

$\langle q_\nu \rangle$ = **BBBBAAAAAABBAAAABBBB** ...

- Aufgrund der gleichwahrscheinlichen Symbole und $M = 2$ ergibt sich auch hier:

  $H_1 = H_0 = 1$ bit/Symbol.

- Wie in **Aufgabe A1.3** gezeigt wird, gilt aber nun für die Verbundwahrscheinlichkeiten $p_{AA} = p_{BB} = 3/8$ und $p_{AB} = p_{BA} = 1/8$. Daraus folgt:

$$H_2 = \frac{1}{2} \cdot \left[2 \cdot \frac{3}{8} \cdot \log_2 \frac{8}{3} + 2 \cdot \frac{1}{8} \cdot \log_2 8\right] = \frac{3}{8} \cdot \log_2 8 - \frac{3}{8} \cdot \log_2 3 + \frac{1}{8} \cdot \log_2 8 =$$
$$= 1.5 - 0.375 \cdot 1.585 = 0.906\,\text{bit/Symbol} < H_1\,.$$

Wenn man sich die Aufgabenstellung genauer betrachtet, kommt man zu dem Schluss, dass hier die Entropie $H = 0.5$ bit/Symbol sein müsste (jedes zweite Symbol liefert keine neue Information). Die zweite Entropienäherung $H_2 = 0.906$ bit/Symbol ist aber deutlich größer als die Entropie H.

Dieses Beispiel legt den Schluss nahe, dass zur Entropiebestimmung die Näherung zweiter Ordnung nicht ausreicht. Vielmehr muss man größere zusammenhängende Blöcke mit $k > 2$ Symbolen betrachten. Einen solchen Block bezeichnen wir im Folgenden als k–Tupel.

Verallgemeinerung auf k–Tupel und Grenzübergang (1)

Zur Abkürzung schreiben wir mit der Verbundwahrscheinlichkeit $p_i^{(k)}$ eines k–Tupels allgemein:

$$H_k = \frac{1}{k} \cdot \sum_{i=1}^{M^k} p_i^{(k)} \cdot \log_2 \frac{1}{p_i^{(k)}} \quad \text{(Einheit: bit/Symbol)}.$$

Die Laufvariable i steht jeweils für eines der M^k Tupel. Die Näherung H_2 ergibt sich mit $k = 2$.

Definition: Die **Entropie** einer Nachrichtenquelle **mit Gedächtnis** ist der folgende Grenzwert:

$$H = \lim_{k \to \infty} H_k .$$

Für die Entropienäherungen H_k gelten folgende Größenrelationen (H_0: Entscheidungsgehalt):

$$H \le ... \le H_k \le ... \le H_2 \le H_1 \le H_0 .$$

Der Rechenaufwand wird bis auf wenige Sonderfälle (siehe nachfolgendes Beispiel) mit zunehmendem k immer größer und hängt natürlich auch vom Symbolumfang M ab:

- Zur Berechnung von H_{10} einer Binärquelle ($M = 2$) ist über $2^{10} = 1024$ Terme zu mitteln. Mit jeder weiteren Erhöhung von k um 1 verdoppelt sich die Anzahl der Summenterme.
- Bei einer Quaternärquelle ($M = 4$) muss zur H_{10}–Bestimmung bereits über $4^{10} = 1.048.576$ Summenterme gemittelt werden.
- Berücksichtigt man, dass jedes dieser $4^{10} = 2^{20} > 10^6$ k–Tupel bei Simulation und Zeitmittelung etwa 100 mal (statistischer Richtwert) vorkommen sollte, um ausreichende Simulationsgenauigkeit zu gewährleisten, so folgt daraus, dass die Folgenlänge größer als $N = 10^8$ sein sollte.

Beispiel: Wir betrachten eine alternierende Binärfolge $\Rightarrow$ $\langle q_\nu \rangle$ = **ABABABAB** ... entsprechend $H_0 = H_1 = 1$ bit/Symbol. In diesem Sonderfall muss zur Bestimmung der H_k–Näherung unabhängig von k stets nur über zwei Verbundwahrscheinlichkeiten gemittelt werden:

- $k = 2$: $p_{\rm AB} = p_{\rm BA} = 1/2$ $\Rightarrow$ $H_2 = 1/2$ bit/Symbol,
- $k = 3$: $p_{\rm ABA} = p_{\rm BAB} = 1/2$ $\Rightarrow$ $H_3 = 1/3$ bit/Symbol,
- $k = 4$: $p_{\rm ABAB} = p_{\rm BABA} = 1/2$ $\Rightarrow$ $H_4 = 1/4$ bit/Symbol.

Die Entropie dieser alternierenden Binärfolge ist demzufolge

$$H = \lim_{k \to \infty} 1/k = 0 .$$

Dieses Ergebnis war zu erwarten, da die betrachtete Folge nur minimale Information besitzt, die sich allerdings im Entropie–Endwert H nicht auswirkt, nämlich: „Tritt **A** zu den geraden oder ungeraden Zeitpunkten auf?"

Man erkennt aber auch, dass H_k diesem Endwert $H = 0$ nur sehr langsam näher kommt: Die Näherung H_{20} liefert immer noch 0.05 bit/Symbol.

Verallgemeinerung auf *k*–Tupel und Grenzübergang (2)

Die Ergebnisse der letzten Seiten sollen hier kurz zusammengefasst werden:

- Allgemein gilt für die Entropie einer Nachrichtenquelle:

$$H \le \ldots \le H_3 \le H_2 \le H_1 \le H_0 \,.$$

- Eine **redundanzfreie Quelle** liegt vor, falls alle M Symbole gleichwahrscheinlich sind und es keine statistischen Bindungen innerhalb der Folge gibt. Für diese gilt (r nennt man *relative Redundanz*):

$$H = H_0 = H_1 = H_2 = H_3 = \ldots$$

$$\Rightarrow \quad r = \frac{H - H_0}{H_0} = 0 \,.$$

- Eine **gedächtnislose Quelle** kann durchaus redundant sein ($r > 0$). Diese Redundanz geht dann allerdings allein auf die Abweichung der Symbolwahrscheinlichkeiten von der Gleichverteilung zurück. Hier gelten folgende Relationen:

$$H = H_1 = H_2 = H_3 = \ldots \le H_0$$

$$\Rightarrow \quad 0 \le r = \frac{H_1 - H_0}{H_0} < 1 \,.$$

- Die entsprechende Bedingung für eine **gedächtnisbehaftete Quelle** lautet:

$$H < \ldots < H_3 < H_2 < H_1 \le H_0$$

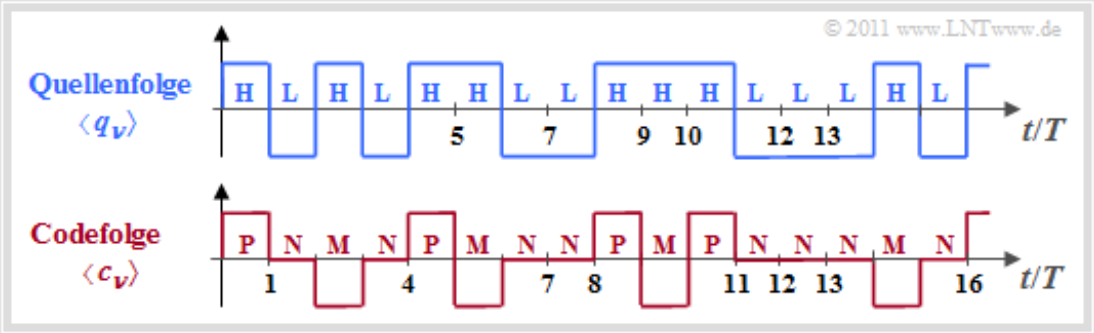
$$\Rightarrow \quad 0 < r = \frac{H_1 - H_0}{H_0} \le 1 \,.$$

- Ist $H_2 < H_1$, dann gilt (nach Meinung des Autors) auch $H_3 < H_2$, $H_4 < H_3$, ... , also es ist das „≤"–Zeichen in der allgemeinen Gleichung durch das „<"–Zeichen zu ersetzen. Sind die Symbole gleichwahrscheinlich, so gilt wieder $H_1 = H_0$, bei nicht gleichwahrscheinlichen Symbolen $H_1 < H_0$.

Die Entropie des AMI–Codes

Im Buch „Digitalsignalübertragung" – **Kapitel 2.4** wurde der AMI–Pseudoternärcode behandelt. Dieser wandelt die Binärfolge $\langle q_\nu \rangle$ mit $q_\nu \in \{\mathbf{L}, \mathbf{H}\}$ in die Ternärfolge $\langle c_\nu \rangle$ mit $c_\nu \in \{\mathbf{M}, \mathbf{N}, \mathbf{P}\}$. Die Bezeichnungen der Quellensymbole stehen für „Low" und „High" und die der Codesymbole für „Minus", „Null" und „Plus". Die Codierregel des AMI–Codes (diese Kurzform steht für „Alternate Mark Inversion") lautet:

- Jedes Binärsymbol $q_\nu = \mathbf{L}$ wird durch das Codesymbol $c_\nu = \mathbf{N}$ dargestellt.
- Dagegen wird $q_\nu = \mathbf{H}$ abwechselnd mit $c_\nu = \mathbf{P}$ und $c_\nu = \mathbf{M}$ codiert ⇒ Name „AMI".



Durch die Codierung wird Redundanz hinzugefügt mit dem Ziel, dass die Codefolge keinen Gleichanteil beinhaltet. Wir betrachten hier jedoch nicht die spektralen Eigenschaften des AMI–Codes, sondern interpretieren diesen Code informationstheoretisch:

- Aufgrund der Stufenzahl $M = 3$ ist der Entscheidungsgehalt der (ternären) Codefolge gleich $H_0 = \log_2 3 \approx 1.585$ bit/Symbol. Die erste Entropienäherung liefert $H_1 = 1.5$ bit/Symbol, wie nachfolgende Rechnung zeigt:

$$p_{\rm H} = p_{\rm L} = 1/2 \quad \Rightarrow \quad p_{\rm N} = p_{\rm L} = 1/2\,,\ p_{\rm M} = p_{\rm P} = p_{\rm H}/2 = 1/4\,,$$
$$\Rightarrow \quad H_1 = 1/2 \cdot \log_2 2 + 2 \cdot 1/4 \cdot \log_2 4 = 1.5\,{\rm bit/Symbol}\,.$$

- Betrachten wir nun Zweiertupel. Beim AMI–Code kann „P" nicht auf „P" und „M" nicht auf „M" folgen. Die Wahrscheinlichkeit für „NN" ist gleich $p_{\rm L} \cdot p_{\rm L} = 1/4$. Alle anderen (sechs) Zweiertupel treten mit der Wahrscheinlichkeit 1/8 auf. Daraus folgt für die zweite Entropienäherung:

$$H_2 = 1/2 \cdot [1/4 \cdot {\rm ld}\, 4 + 6 \cdot 1/8 \cdot {\rm ld}\, 8] = 1.375\,{\rm bit/Symbol}\,.$$

- Für die weiteren Entropienäherungen und die tatsächliche Entropie H wird gelten:

$$H < \ldots < H_5 < H_4 < H_3 < H_2 = 1.375\,{\rm bit/Symbol}\,.$$

- Bei diesem Beispiel kennt man die tatsächliche Entropie H der Codesymbolfolge $\langle c_\nu \rangle$. Da durch den Coder keine neue Information hinzukommt, aber auch keine verloren geht, ergibt sich die gleiche Entropie wie für die redundanzfreie Binärfolge $\langle q_\nu \rangle$:

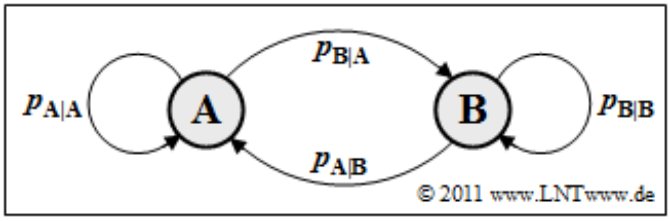
$$H = 1\,{\rm bit/Symbol}\,.$$

Aufgabe A1.4 zeigt den bereits beträchtlichen Aufwand zur Berechnung der Entropienäherung H_3; zudem weicht H_3 noch deutlich vom Endwert $H = 1$ bit/Symbol ab. Schneller kommt man zum Ergebnis,

wenn man den AMI–Code durch eine Markovkette beschreibt. Hierzu mehr auf der nächsten Seite.

Binärquellen mit Markoveigenschaften (1)

Folgen mit statistischen Bindungen zwischen den Folgenelementen (Symbolen) werden oft durch **Markovprozesse** modelliert, wobei wir uns hier auf Markovprozesse erster Ordnung beschränken. Zunächst betrachten wir einen binären Markovprozess ($M = 2$) mit den Zuständen (Symbolen) **A** und **B**.



Oben sehen Sie das Übergangsdiagramm für einen binären Markovprozess erster Ordnung. Von den vier angegebenen Übertragungswahrscheinlichkeiten sind allerdings nur zwei frei wählbar, zum Beispiel

- $p_{A|B} = \Pr(A|B)$ ⇒ bedingte Wahrscheinlichkeit, dass **A** auf **B** folgt.
- $p_{B|A} = \Pr(B|A)$ ⇒ bedingte Wahrscheinlichkeit, dass **B** auf **A** folgt.

Für die beiden weiteren Übergangswahrscheinlichkeiten gilt dann

$$p_{A|A} = 1 - p_{B|A}\,, \quad p_{B|B} = 1 - p_{A|B}\,.$$

Aufgrund der vorausgesetzten Eigenschaften **Stationarität** und **Ergodizität** gilt für die Zustands– bzw. Symbolwahrscheinlichkeiten:

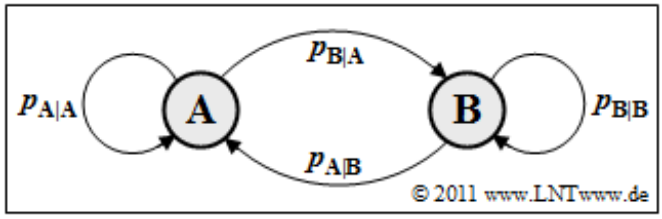
$$p_A = \Pr(A) = \frac{p_{A|B}}{p_{A|B} + p_{B|A}}\,, \quad p_B = \Pr(B) = \frac{p_{B|A}}{p_{A|B} + p_{B|A}}\,.$$

Diese Gleichungen erlauben erste informationstheoretische Aussagen über Markovprozesse:

- Für $p_{A|B} = p_{B|A}$ ergeben sich gleichwahrscheinliche Symbole ⇒ $p_A = p_B = 0.5$. Damit liefert die erste Entropienäherung $H_1 = H_0 = 1$ bit/Symbol, und zwar unabhängig von den tatsächlichen Werten der (bedingten) Übergangswahrscheinlichkeiten $p_{A|B}$ bzw. $p_{B|A}$.
- Die Quellenentropie H als der Grenzwert der **Entropienäherung *k*–ter Ordnung** H_k für $k \to \infty$ hängt aber sehr wohl von den tatsächlichen Werten von $p_{A|B}$ und $p_{B|A}$ ab und nicht nur von ihrem Quotienten. Dies zeigt das Beispiel auf der folgenden Seite.

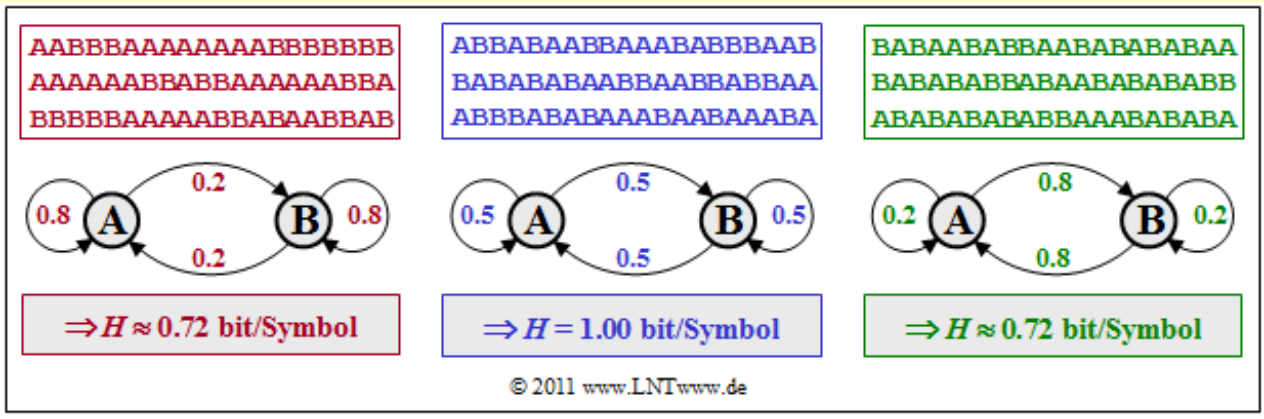
Binärquellen mit Markoveigenschaften (2)

Wir gehen von einer binären Markovquelle erster Ordnung aus und setzen nun voraus:



- Die 4 bedingten Wahrscheinlichkeiten seien symmetrisch, das heißt, es gelte $p_{A|B} = p_{B|A}$, $p_{A|A} = p_{B|B}$.
- Für die beiden Symbolwahrscheinlichkeiten gilt somit: $p_A = p_B = 0.5$.

Beispiel: Wir betrachten hier drei solche binäre Markovquellen, die sich durch die Zahlenwerte der symmetrischen Übergangswahrscheinlichkeiten $p_{A|B} = p_{B|A}$ unterscheiden. Die beiden anderen Übergangswahrscheinlichkeiten haben dann folgende Werte: $p_{A|A} = 1 - p_{B|A} = p_{B|B}$.



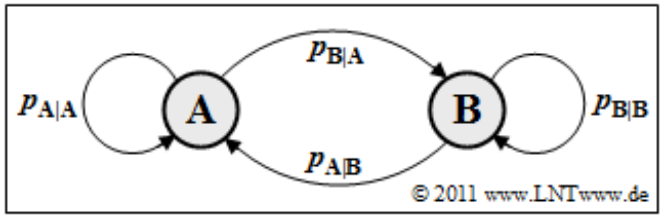
- Die mittlere Symbolfolge (mit $p_{A|B} = p_{B|A} = 0.5$) besitzt die Entropie $H = 1$ bit/Symbol. Das heißt: In diesem Sonderfall gibt es keine statistischen Bindungen innerhalb der Folge.
- Die linke (rote) Folge mit $p_{A|B} = p_{B|A} = 0.2$ weist weniger Wechsel zwischen **A** und **B** auf. Aufgrund von statistischen Abhängigkeiten zwischen benachbarten Symbolen ist nun $H \approx 0.72$ bit/Symbol kleiner.
- Die rechte (grüne) Symbolfolge mit $p_{A|B} = p_{B|A} = 0.8$ hat die genau gleiche Entropie wie die rote Folge. Hier erkennt man viele Bereiche mit sich stets abwechselnden Symbolen (... **ABABAB** ...).

Zu diesem Beispiel ist noch anzumerken:

- Hätte man nicht die Markoveigenschaften der roten und der grünen Folge ausgenutzt, so hätte man das Ergebnis $H \approx 0.72$ bit/Symbol erst nach langwierigen Berechnungen erhalten.
- Auf den nächsten Seiten wird gezeigt, dass bei einer Quelle mit Markoveigenschaften dieser Endwert H allein aus den Entropienäherungen H_1 und H_2 ermittelt werden kann.
- Ebenso lassen sich aus H_1 und H_2 alle Entropienäherungen H_k für k–Tupel in einfacher Weise berechnen $\Rightarrow$ $H_3, H_4, H_5, \text{...}, H_{100}, \text{...}$

Binärquellen mit Markoveigenschaften (3)

Wir gehen weiterhin von der symmetrischen binären Markovquelle erster Ordnung aus. Wie auf der vorherigen Seite verwenden wir folgende Nomenklatur:



- Übergangswahrscheinlichkeiten $p_{B|A}$, ...
- ergodische Wahrscheinlichkeiten p_A und p_B,
- Verbundwahrscheinlichkeiten, zum Beispiel $p_{AB} = p_A \cdot p_{B|A}$.

Wir berechnen nun die **Entropie eines Zweiertupels** (mit der Einheit „bit/Zweiertupel"):

$$H_2' = p_A \cdot p_{A|A} \cdot \log_2 \frac{1}{p_A \cdot p_{A|A}} + p_A \cdot p_{B|A} \cdot \log_2 \frac{1}{p_A \cdot p_{B|A}} + $$
$$+ \; p_B \cdot p_{A|B} \cdot \log_2 \frac{1}{p_B \cdot p_{A|B}} + p_B \cdot p_{B|B} \cdot \log_2 \frac{1}{p_B \cdot p_{B|B}}.$$

Ersetzt man nun die Logarithmen der Produkte durch entsprechende Summen von Logarithmen, so erhält man das Ergebnis $H_2' = H_1 + H_M$ mit

$$H_1 = p_A \cdot (p_{A|A} + p_{B|A}) \cdot \log_2 \frac{1}{p_A} + p_B \cdot (p_{A|B} + p_{B|B}) \cdot \log_2 \frac{1}{p_B} =$$
$$= p_A \cdot \log_2 \frac{1}{p_A} + p_B \cdot \log_2 \frac{1}{p_B} = H_{bin}(p_A) = H_{bin}(p_B),$$
$$H_M = p_A \cdot p_{A|A} \cdot \log_2 \frac{1}{p_{A|A}} + p_A \cdot p_{B|A} \cdot \log_2 \frac{1}{p_{B|A}} +$$
$$+ \; p_B \cdot p_{A|B} \cdot \log_2 \frac{1}{p_{A|B}} + p_B \cdot p_{B|B} \cdot \log_2 \frac{1}{p_{B|B}}.$$

Damit lautet die zweite Entropienäherung (mit der Einheit „bit/Symbol"):

$$H_2 = \frac{H_2'}{2} = \frac{1}{2} \cdot [H_1 + H_M].$$

Anzumerken ist:

- Der erste Summand wurde nicht zufällig mit H_1 abgekürzt, sondern ist tatsächlich gleich der ersten Entropienäherung, allein abhängig von den Symbolwahrscheinlichkeiten.
- Bei einem symmetrischen Markovprozess ($p_{A|B} = p_{B|A} \;\Rightarrow\; p_A = p_B = 1/2$) ergibt sich für diesen ersten Summanden $H_1 = 1$ bit/Symbol.
- Der zweite Summand (H_M) muss gemäß der zweiten der oberen Gleichungen berechnet werden. Bei einem symmetrischen Markovprozess erhält man $H_M = H_{bin}(p_{A|B})$.

Auf der nächsten Seite wird dieses Ergebnis auf die k–te Entropienäherung erweitert.

Binärquellen mit Markoveigenschaften (4)

Der Vorteil von Markovquellen gegenüber anderen Quellen ist, dass sich die Entropieberechnung für k–Tupel sehr einfach gestaltet. Für jede Markovquelle gilt:

$$H_k = \frac{1}{k} \cdot [H_1 + (k-1) \cdot H_{\rm M}] \quad \Rightarrow \quad H_2 = \frac{1}{2} \cdot [H_1 + H_{\rm M}],$$
$$H_3 = \frac{1}{3} \cdot [H_1 + 2 \cdot H_{\rm M}], \quad H_4 = \frac{1}{4} \cdot [H_1 + 3 \cdot H_{\rm M}], \ \text{usw.}$$

Bildet man den Grenzübergang für $k \to \infty$, so erhält man für die tatsächliche Quellenentropie:

$$H = \lim_{k \to \infty} H_k = H_{\rm M}.$$

Aus diesem einfachen Ergebnis folgen wichtige Erkenntnisse für die Entropieberechnung:

- Bei Markovquellen genügt die Bestimmung der Entropienäherungen H_1 und H_2. Damit lautet die Entropie einer Markovquelle:

  $$H = 2 \cdot H_2 - H_1.$$

- Durch H_1 und H_2 liegen auch alle weiteren Entropienäherungen H_k fest:

  $$H_k = \frac{2-k}{k} \cdot H_1 + \frac{2 \cdot (k-1)}{k} \cdot H_2.$$

- Diese Näherungen haben allerdings keine große Bedeutung. Wichtig ist meist nur der Grenzwert H. Bei Quellen ohne Markoveigenschaften berechnet man die Näherungen H_k nur deshalb, um den Grenzwert, also die tatsächliche Entropie, abschätzen zu können.
- Alle auf dieser Seite angegebenen Gleichungen gelten auch für nichtbinäre Markovquellen ($M > 2$), wie auf der nächsten Seite gezeigt wird.

Hinweis: In der **Aufgabe A1.5** werden die obigen Gleichungen auf den allgemeineren Fall einer unsymmetrischen Binärquelle angewendet.

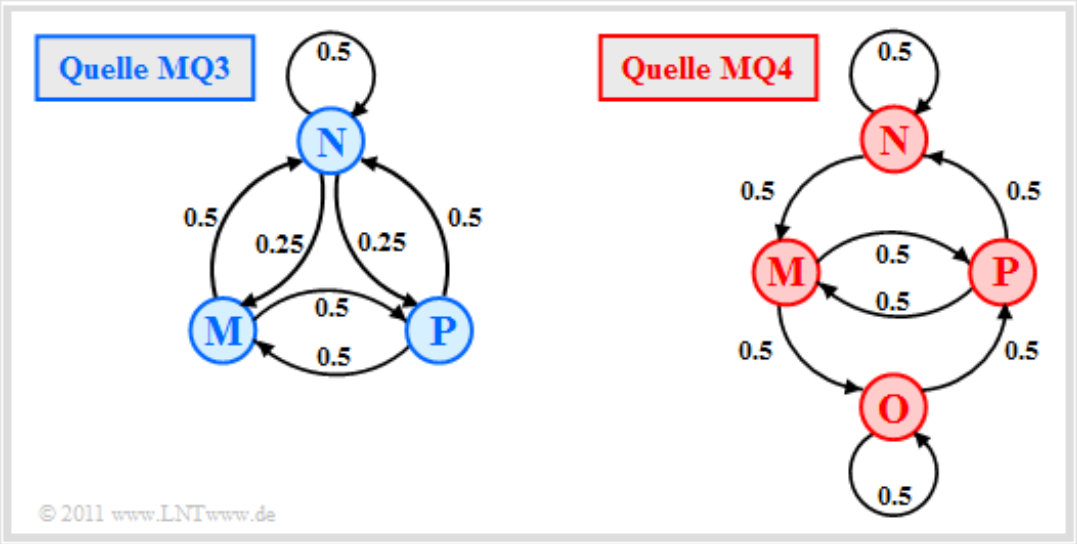
Nichtbinäre Markovquellen (1)

Für jede Markovquelle gelten unabhängig vom Symbolumfang die folgenden Gleichungen:

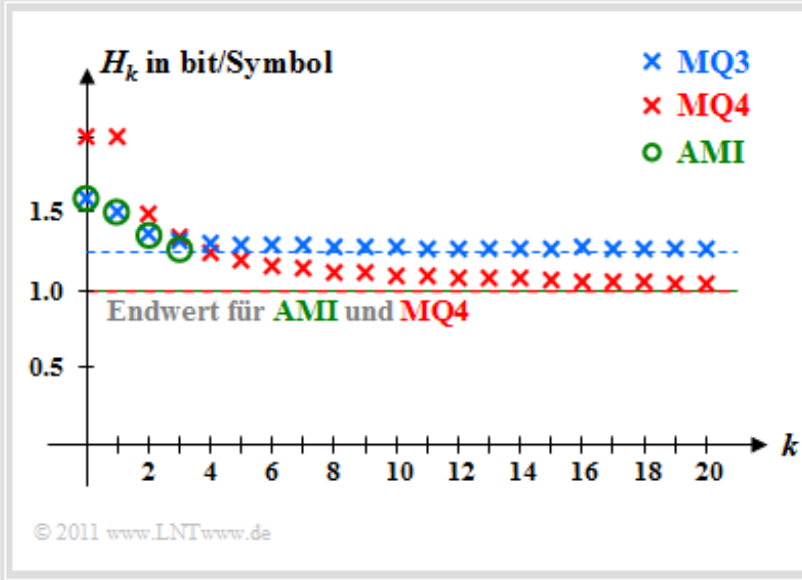
$$H = 2 \cdot H_2 - H_1\,, \quad H_k = \frac{1}{k} \cdot \left[H_1 + (k-1) \cdot H_{\rm M}\right], \quad \lim_{k \to \infty} H_k = H\,.$$

Diese ermöglichen die einfache Berechnung der Entropie H aus den Näherungen H_1 und H_2.

Wir betrachten nun eine ternäre Markovquelle MQ3 (Stufenzahl $M = 3$, blaue Farbgebung) und eine quaternäre Markovquelle MQ4 ($M = 4$, rot) mit folgenden Übergangsdiagrammen:



In der **Aufgabe A1.6** werden die Entropienäherungen H_k und die jeweiligen Quellenentropien H als der Grenzwert von H_k für $k \to \infty$ berechnet. Die Ergebnisse sind in der folgenden Grafik zusammengestellt. Alle Entropien haben die Einheit „bit/Symbol".



H_0	1.585	2.000	1.585
H_1	1.500	2.000	1.500
H_2	1.375	1.500	1.375
H_3	1.333	1.333	1.292
H_4	1.313	1.250	???
...			
H_{10}	1.275	1.100	???
...			
H_{50}	1.255	1.020	???
...			
H	1.250	1.000	1.000

Die Bildbeschreibung folgt auf der nächsten Seite.

Nichtbinäre Markovquellen (2)

Die Ergebnisse der letzten Seite – siehe **unteren Grafik** – lassen sich wie folgt interpretieren:

- Bei der ternären Markovquelle MQ3 nehmen die Entropienäherungen von $H_1 = 1.500$ über $H_2 = 1.375$ bis zum Grenzwert $H = 1.25$ kontinuierlich ab. Wegen $M = 3$ beträgt der Entscheidungsgehalt $H_0 = 1.585$ (alle Entropien in „bit/Symbol") .
- Für die quaternäre Markovquelle MQ4 (rote Markierungen) erhält man $H_0 = H_1 = 2$ (wegen den vier gleichwahrscheinlichen Zuständen) und $H_2 = 1.5$. Aus dem H_1– und H_2–Wert lassen sich auch hier alle Entropienäherungen H_k und auch der Endwert $H = 1$ berechnen.
- Die beiden Quellenmodelle MQ3 und MQ4 entstanden bei dem Versuch, den **AMI–Code** informationstheoretisch durch Markovquellen zu beschreiben. Die Symbole **M**, **N** und **P** stehen hierbei für „Minus", „Null" und „Plus".
- Die Entropienäherungen H_1, H_2 und H_3 des AMI–Codes (grüne Markierungen) wurden in **Aufgabe A1.4** berechnet. Auf die Berechnung von H_4, H_5, ... musste aus Aufwandsgründen verzichtet werden. Bekannt ist aber der Endwert von H_k für $k \to \infty \;\Rightarrow\; H = 1$.
- Man erkennt, dass das Markovmodell MQ3 für $H_0 = 1.585$, $H_1 = 1.500$ und $H_2 = 1.375$ genau die gleichen Werte liefert wie der AMI–Code. Dagegen unterscheiden sich H_3 (1.333 gegenüber 1.292) und insbesondere der Endwert H (1.25 gegenüber 1).
- Das Modell MQ4 ($M = 4$) unterscheidet sich vom AMI–Code ($M = 3$) hinsichtlich des Entscheidungsgehaltes H_0 und auch bezüglich aller Entropienäherungen H_k. Trotzdem ist MQ4 das geeignete Modell für den AMI–Code, da der Endwert $H = 1$ übereinstimmt.
- Das **Modell MQ3** liefert deshalb zu große Entropiewerte, da hier die Folgen **PNP** und **MNM** möglich sind, die beim AMI–Code nicht auftreten können. Bereits bei H_3 macht sich der Unterschied geringfügig bemerkbar, im Endwert H deutlich (1.25 gegenüber 1).

Beim **Modell MQ4** wurde der Zustand „Null" aufgespalten in zwei Zustände **N** und **O**:

- Hierbei gilt für den Zustand **N**: Das aktuelle Binärsymbol **L** wird mit dem Amplitudenwert „0" dargestellt, wie es der AMI–Regel entspricht. Das nächste auftretende **H**–Symbol wird als **M** (Minus) dargestellt, weil das letzte **H**–Symbol als **P** (Plus) codiert wurde.
- Auch beim Zustand **O** wird das aktuelle Binärsymbol **L** mit dem Ternärwert „0" dargestellt. Im Unterschied zum Zustand **N** wird aber nun das nächste auftretende **H**–Symbol als **P** (Plus) dargestellt werden, da das letzte **H**–Symbol als **M** (Minus) codiert wurde.

Die von MQ4 ausgegebene Symbolfolge entspricht tatsächlich den Regeln des AMI–Codes und weist die Entropie $H = 1$ bit/Symbol auf. Aufgrund des neuen Zustandes **O** ist nun allerdings $H_0 = 2$ bit/Symbol (gegenüber 1.585 bit/Symbol) deutlich zu groß und auch alle H_k–Näherungen sind größer als beim AMI–Code. Erst für $k \to \infty$ stimmen beide überein: $H = 1$ bit/Symbol.

Schwierigkeiten bei der Entropiebestimmung

Bisher haben wir uns ausschließlich mit künstlich erzeugten Symbolfolgen beschäftigt. Nun betrachten wir geschriebene Texte. Ein solcher Text kann als eine natürliche wertdiskrete Nachrichtenquelle aufgefasst werden, die natürlich auch informationstheoretisch analysiert werden kann, indem man ihre Entropie ermittelt.

Natürliche Texte werden auch in heutiger Zeit (2011) noch oft mit dem 8 Bit–Zeichensatz nach ANSI (*American National Standard Institute*) dargestellt, obwohl es etliche „modernere" Codierungen gibt. Die $M = 2^8 = 256$ ANSI–Zeichen sind dabei wie folgt belegt:

- **Nr. 0 bis 31**: nicht druck– und darstellbare Steuerbefehle,
- **Nr. 32 bis 127**: identisch mit den Zeichen des 7 Bit–ASCII–Codes,
- **Nr. 128 bis 159**: weitere Steuerzeichen bzw. Alphanumerikzeichen für Windows,
- **Nr. 160 bis 255**: identisch mit Unicode–Charts.

Theoretisch könnte man auch hier die Entropie entsprechend der Vorgehensweise in **Kapitel 1.2** als den Grenzübergang der Entropienäherung H_k für $k \to \infty$ ermitteln. Praktisch ergeben sich aber nach dieser Rezeptur unüberwindbare numerische Grenzen:

- Bereits für die Entropienäherung H_2 gibt es $M^2 = 256^2 = 65536$ mögliche Zweiertupel. Für die Berechnung sind somit ebenso viele Speicherplätze (in Byte) erforderlich. Geht man davon aus, dass man für eine ausreichend sichere Statistik im Mittel 100 Entsprechungen pro Tupel benötigt, so sollte die Länge der Quellensymbolfolge bereits $N > 6.5 \cdot 10^6$ sein.
- Die Anzahl der möglichen Dreiertupel ergibt sich zu $M^3 > 16 \cdot 10^7$ und damit ist die erforderliche Quellensymbollänge N schon größer als $1.6 \cdot 10^9$. Dies entspricht bei 42 Zeilen pro Seite und 80 Zeichen pro Zeile einem Buch mit etwa 500.000 Seiten.
- Bei einem natürlichen Text reichen die statistischen Bindungen aber sehr viel weiter als zwei oder drei Zeichen. Küpfmüller gibt für die deutsche Sprache einen Wert von 100 an **[Küp54].** Zur Ermittlung der 100. Entropienäherung benötigt man aber $2^{800} \approx 10^{240}$ Häufigkeiten und für die gesicherte Statistik nochmals um den Faktor 100 mehr Zeichen.

Eine berechtigte Frage ist deshalb: Wie hat **Karl Küpfmüller** im Jahre 1954 die Entropie der deutschen Sprache ermittelt, und vor ihm schon **Claude E. Shannon** die Entropie der englischen Sprache? Eines sei vorweg verraten: Nicht mit dem oben beschriebenen Ansatz.

Entropieabschätzung nach Küpfmüller (1)

Karl Küpfmüller hat die Entropie von deutschen Texten untersucht. Er geht bei seiner in **[Küp54]** veröffentlichten Abschätzung von folgenden Voraussetzungen aus:

- ein Alphabet mit 26 Buchstaben (keine Umlaute und Satzzeichen),
- Nichtberücksichtigung des Leerzeichens,
- keine Unterscheidung zwischen Groß– und Kleinschreibung.

Der Entscheidungsgehalt ergibt sich somit zu $H_0 = \log_2(26) \approx 4.7$ bit/Buchstabe.

Seine Abschätzung basiert auf den folgenden Überlegungen:

1.) Die **erste Entropienäherung** ergibt sich aus den Buchstabenhäufigkeiten in deutschen Texten. Nach einer Studie von 1939 ist „e" mit 16.7% am häufigsten, am seltensten ist „x" mit 0.02%.

$$H_1 \approx 4.1 \text{ bit/Buchstabe}.$$

2.) Hinsichtlich der **Silbenhäufigkeit** wertet Küpfmüller das von F.W. Kaeding herausgegebene „Häufigkeitswörterbuch der deutschen Sprache" aus. Er unterscheidet zwischen Stammsilben, Vorsilben und Endsilben. Er kommt so auf den mittleren Informationsgehalt aller Silben:

$$\begin{aligned} H_{\text{Silbe}} &= H_{\text{Stamm}} + H_{\text{Vor}} + H_{\text{End}} + H_{\text{Rest}} \approx \\ &\approx 4.15 + 0.82 + 1.62 + 2.0 \approx 8.6 \text{ bit/Silbe}. \end{aligned}$$

- Nach der Kaeding–Studie von 1898 bilden die 400 häufigsten Stammsilben (beginnend mit „de") 47% eines deutschen Textes und tragen zur Entropie mit $H_{\text{Stamm}} \approx 4.15$ bit/Silbe bei.
- Der Beitrag der 242 häufigsten Vorsilben – an erster Stelle „ge" mit 9% – wird von Küpfmüller mit $H_{\text{Vor}} \approx 0.82$ bit/Silbe beziffert.
- Der Beitrag der 118 meistgebrauchten Endsilben ist $H_{\text{End}} \approx 1.62$ bit/Silbe. Am häufigsten tritt „en" am Ende eines Wortes mit 30% auf.
- Der Rest von 14% verteilt sich auf bisher nicht erfasste Silben. Küpfmüller nimmt dazu an, dass es davon 4.000 gibt und diese gleichverteilt sind. Er setzt dafür $H_{\text{Rest}} \approx 2$ bit/Silbe an.

3.) Für die durchschnittliche Buchstabenzahl je Silbe ermittelte Küpfmüller den Wert 3.03. Daraus schloss er auf die **dritte Entropienäherung** hinsichtlich der Buchstaben:

$$H_3 \approx \frac{8.6}{3.03} \approx 2.8 \text{ bit/Buchstabe}.$$

Die Beschreibung wird auf der nächsten Seite fortgesetzt.

Entropieabschätzung nach Küpfmüller (2)

Küpfmüllers Abschätzung der Entropienäherung H_3 basierte vor allem auf den Silbenhäufigkeiten und dem Mittelwert von 3.03 Buchstaben pro Silbe. Um eine weitere Entropienäherung H_k mit größerem k zu erhalten, analysierte Küpfmüller zusätzlich die Wörter in deutschen Texten. Er kam zu folgenden Ergebnissen:

4.) Die 322 häufigsten Wörter liefern einen Entropiebeitrag von 4.5 bit/Wort. Die Beiträge der restlichen 40.000 Wörter wurden geschätzt, wobei angenommen wurde, dass die Häufigkeiten von seltenen Wörtern reziprok zu ihrer Ordnungszahl sind. Mit diesen Voraussetzungen ergibt sich der mittlere Informationsgehalt eines Wortes zu ca. 11 bit.

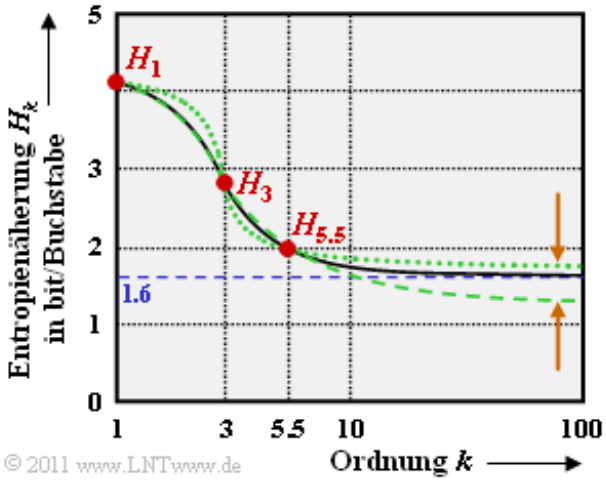
5.) Die Auszählung ergab im Mittel 5.5 Buchstaben pro Wort. Analog zu Punkt (3) wurde so die Entropienäherung für $k = 5.5$ angenähert:

$$H_{5.5} \approx \frac{11}{5.5} \approx 2 \text{ bit/Buchstabe}.$$

Natürlich kann k gemäß **Definition** nur ganzzahlige Werte annehmen. Diese Gleichung ist deshalb so zu interpretieren, dass sich für H_5 ein etwas größerer und für H_6 ein etwas kleinerer Wert ergeben wird.

6.) Man kann nun versuchen, aus diesen drei Punkten durch Extrapolation den Endwert der Entropie für $k \to \infty$ zu ermitteln. In folgender Grafik wird dies bei logarithmisch aufgetragener Abszisse versucht:

Die durchgehende Linie ist der Originalarbeit von Küpfmüller **[Küp54]** entnommen und führt zum Endwert $H = 1.6$ bit/Buchstabe. Die grünen Kurven (vom *LNTwww*–Autor hinzugefügt) und die braunen Pfeile zeigen aber, dass eine solche Extrapolation nur sehr vage ist.



7.) Küpfmüller versuchte anschließend, den von ihm gefundenen Endwert $H = 1.6$ bit/Buchstabe mit völlig anderer Methodik – siehe **nächste Seite** – zu verifizieren. Nach dieser Abschätzung revidierte er sein Ergebnis geringfügig auf $H = 1.51$ bit/Buchstabe.

8.) Shannon hatte drei Jahre vorher nach völlig anderer Vorgehensweise für die englische Sprache den Entropiewert $H \approx 1$ bit/Buchstabe angegeben, allerdings unter Berücksichtigung des Leerzeichens. Um seine Ergebnisse mit Shannom vergleichen zu können, hat Küpfmüller das Leerzeichen nachträglich in sein Ergebnis eingerechnet:

$$H = 1.51 \cdot \frac{5.5}{6.5} \approx 1.3 \text{ bit/Buchstabe}.$$

Der Korrekturfaktor ist der Quotient aus der mittleren Wortlänge ohne Berücksichtigung des

Leerzeichens (5.5) und der mittleren Wortlänge mit Berücksichtigung des Leerzeichens (6.5).

Entropieabschätzung nach Küpfmüller (3)

Der Vollständigkeit halber seien hier noch Küpfmüllers Überlegungen dargelegt, die ihn zum Endergebnis $H = 1.51$ bit/Buchstabe führten. Da es für die Statistik von Wortgruppen oder ganzen Sätzen keine Unterlagen gab, schätzte er den Entropiewert der deutschen Sprache wie folgt ab:

- Ein beliebiger zusammenhängender deutscher Text wird hinter einem bestimmten Wort abgedeckt. Der vorhergehende Text wird gelesen, und der Leser soll versuchen, das folgende Wort aus dem Zusammenhang mit dem vorhergehenden Text zu ermitteln.
- Bei sehr vielen solcher Versuche ergibt die prozentuale Zahl der Treffer ein Maß für die Bindungen zwischen Wörtern und Sätzen. Es zeigt sich, dass bei ein und derselben Textart (Romane, wissenschaftliche Schriften, usw.) ein und desselben Autors relativ schnell (etwa 100 bis 200 Versuche) ein konstanter Endwert dieses Trefferverhältnisses erreicht wird.
- Das Trefferverhältnis hängt aber ziemlich stark von der Art des Textes ab. Für verschiedene Texte ergeben sich Werte zwischen 15% und 33%, mit dem Mittelwert bei 22%. Das heißt aber auch: Im Durchschnitt können 22% der Wörter in einem deutschen Text aus dem Zusammenhang heraus 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 signifikante Einbuße erfährt. Ausgehend vom Bezugswert $H_{5.5} = 2$ bit/Buchstabe (siehe Punkt **(5)**, letzte Seite) für ein mittellanges Wort ergibt sich somit die Entropie $H \approx 0.78 \cdot 2 = 1.56$ bit/Buchstabe.
- Küpfmüller überprüfte diesen Wert mit einer vergleichbaren empirischen Untersuchung der Silben und ermittelte den Reduktionsfaktor 0.54 hinsichtlich Silben. Als Endergebnis nennt Küpfmüller $H = 0.54 \cdot H_3 \approx 1.51$ bit/Buchstabe, wobei $H_3 \approx 2.8$ bit/Buchstabe der Entropie einer Silbe mittlerer Länge (≈ 3 Buchstaben, siehe Punkt **(3)**, vorletzte Seite) entspricht.

Die vielleicht als zu kritisch empfundenen Bemerkungen auf dieser Seite sollen die Bedeutung von Küpfmüllers Entropieabschätzung nicht herabsetzen, eben so wenig wie Shannon's Beiträge zur gleichen Thematik. Sie sollen nur auf die großen Schwierigkeiten hinweisen, die bei dieser Aufgabenstellung auftreten. Dies ist vielleicht auch der Grund dafür, dass sich seit den 1950er Jahren niemand mehr mit dieser Problematik intensiv beschäftigt hat.

Einige eigene Simulationsergebnisse (1)

Die Angaben von Karl Küpfmüller hinsichtlich der Entropie der deutschen Sprache sollen nun mit einigen Simulationsergebnissen verglichen werden, die vom Autor G. Söder dieses Kapitels am Lehrstuhl für Nachrichtentechnik der Technischen Universität München gewonnen wurden. Die Resultate basieren auf

- dem Programm **WDIT** (Wertdiskrete Informationstheorie) aus dem Praktikum **[Söd01];** der Link weist auf die Zip–Version des Programms,
- einer ASCII–Version der deutschen Bibel mit fast $N = 4.37$ Millionen Schriftzeichen, die auf den Symbolumfang $M = 33$ reduziert wurde:

  a, **b**, **c**, ... , **x**, **y**, **z**, **ä**, **ö**, **ü**, **ß**, **LZ**, **ZI**, **IP**.

Nicht unterschieden wurde bei unserer Analyse zwischen Groß– und Kleinbuchstaben. Gegenüber Küpfmüllers Analyse wurden hier noch zusätzlich berücksichtigt:

- die deutschen Umlaute „ä", „ö", „ü" und „ß", die etwa 1.2% des Bibeltextes ausmachen,
- die Klasse IP (Interpunktion) mit ca. 3%,
- die Klasse ZI (Ziffer) mit ca. 1.3% in Folge der Vers–Nummerierung,
- das Leerzeichen (LZ) als das häufigste Zeichen (17.8%), noch vor dem „e" (12.8%).

Die nachfolgende Tabelle fasst die Ergebnisse zusammen. N bezeichnet die jeweils analysierte Dateigröße in Schriftzeichen (Byte). Die Interpretation folgt auf der nächsten Seite.

Zeile	Symbolumfang	N	H_0	H_1	H_2	H_3
1	$M = 33$	4 368 593	5.044	4.176	3.657	3.216
2	$M = 32$ (ohne ZI)	4 312 675	5.000	4.130	3.641	3.212
3	$M = 31$ (ohne ZI, IP)	4 178 307	4.954	4.062	3.594	3.178
4	$M = 30$ (ohne ZI, IP, LZ)	3 422 356	4.907	4.132	3.755	3.414

Hinweis: Betrachten Sie diese Ergebnisse bitte nicht als Teil einer wissenschaftlichen Untersuchung, sondern nur als den Versuch, Studierenden die in Kapitel 1.3 behandelte Thematik in einem Praktikum näher zu bringen. Als Grundlage dieser Untersuchung wurde von der Bibel ausgegangen, da uns sowohl deren deutsche als auch die englische Fassung im geeigneten ASCII–Format zur Verfügung gestellt wurden.

Einige eigene Simulationsergebnisse (2)

Die in der **Tabelle** angegebenen Entropien H_0 (Entscheidungsgehalt), H_1, H_2 und H_3 wurden jeweils aus N Schriftzeichen ermittelt und sind jeweils in bit/Schriftzeichen angegeben. Die gesamte Datei „Bibel" (in deutscher Sprache) beinhaltet fast $N = 4.37$ Millionen Schriftzeichen, was bei 42 Zeilen pro Seite und 80 Zeichen pro Zeile etwa einem Buch mit 1300 Seiten entsprechen würde. Der Symbolumfang ist $M = 33$.

Die Ergebnisse lassen sich wie folgt zusammenfassen:

- In allen Zeilen nehmen die Entropienäherungen H_k mit wachsendem k monoton ab. Der Abfall verläuft konvex, das heißt, es ist $H_1 - H_2 > H_2 - H_3$. Die Extrapolation des Endwertes ($k \to \infty$) ist aus den jeweils ermittelten drei Entropienäherungen nicht (oder nur sehr vage) möglich.
- Verzichtet man auf die Auswertung der Ziffern (ZI, Zeile 2 $\Rightarrow$ $M = 32$) und zusätzlich auf die Auswertung der Interpunktionszeichen IP, Zeile 3 $\Rightarrow$ $M = 31$), so nehmen die Entropienäherungen H_1 (um 0.114), H_2 (um 0.063) und H_3 (um 0.038) ab. Auf den Endwert H als dem Grenzwert von H_k für $k \to \infty$ wirkt sich der Verzicht auf Ziffern (ZI) und Interpunktion (IP) voraussichtlich kaum aus.
- Lässt man bei der Auswertung noch das Leerzeichen (LZ, Zeile 4 $\to$ $M = 30$) außer Betracht, so ergibt sich nahezu die gleiche Konstellation wie von Küpfmüller ursprünglich betrachtet. Der einzige Unterschied sind die eher seltenen deutschen Sonderzeichen „ä", „ö", „ü" und „ß".
- Der in der letzten Zeile angegebene H_1–Wert 4.132 stimmt mit dem von Küpfmüller ermittelten Wert $H_1 \approx 4.1$ sehr gut überein. Hinsichtlich der H_3–Werte gibt es aber deutliche Unterschiede: Unsere Analyse ergibt $H_3 \approx 3.4$ gegenüber Küpfmüllers 2.8 (alle Angaben in bit/Buchstabe).
- Aus der Auftrittshäufigkeit des Leerzeichens (17.8%) ergibt sich hier eine mittlere Wortlänge von $1/0.178 - 1 \approx 4.6$, ein kleinerer Wert als von Küpfmüller (5.5) angegeben. Die Diskrepanz lässt sich mit unserer Analysedatei „Bibel" erklären (viele Leerzeichen aufgrund der Vers–Nummern).
- Interessant ist der Vergleich der Zeilen 3 und 4. Berücksichtigt man das Leerzeichen, so wird zwar H_0 von $\log_2(30)$ auf $\log_2(31)$ vergrößert, aber man verringert dadurch H_1 (um den Faktor 0.98), H_2 (um 0.96) und H_3 (um 0.93). Küpfmüller hat diesen Faktor intuitiv mit 85% berücksichtigt.

Obwohl wir unsere eigenen Recherchen als nicht so bedeutend ansehen, so glauben wir doch, dass für heutige Texte die von Shannon angegebenen 1.0 bit/Buchstabe für die englische Sprache und auch Küpfmüllers 1.3 bit/Buchstabe für Deutsch etwas zu niedrig sind, unter Anderem, weil

- der Symbolumfang deutlich größer ist, als von Shannon und Küpfmüller bei ihren Analysen berücksichtigt – beispielsweise gilt für den ASCII–Zeichensatz $M = 256$,
- die vielfachen Formatierungsmöglichkeiten (Unterstreichungen, Fett- und Kursivschrift, Einrückungen, Farben) den Informationsgehalt eines Dokuments erhöhen.

Synthetisch erzeugte Texte

In der Grafik sind künstlich erzeugte deutsche und englische Texte angegeben, die **[Küp54]** entnommen wurden. Der zugrundeliegende Symbolumfang ist $M = 27$, das heißt, berücksichtigt sind alle Buchstaben (ohne Umlaute und „ß") sowie das Leerzeichen.

- Die **Buchstabennäherung nullter Ordnung** geht von gleichwahrscheinlichen Zeichen aus. Hier ist kein Unterschied zwischen Deutsch (rot) und Englisch (blau) festzustellen.
- Bei der **ersten Buchstabennäherung** werden bereits die unterschiedlichen Häufigkeiten berücksichtigt, bei den Näherungen höherer Ordnung auch die vorangegangenen Zeichen.
- Bei einer **Synthese 4. Ordnung** ⇒ die Wahrscheinlichkeit für einen neuen Buchstaben hängt von den drei zuletzt ausgewählten Zeichen ab – erkennt man bereits sinnhafte Worte.
- Die **Wortnäherung 1. Ordnung** synthetisiert Sätze gemäß den Wortwahrscheinlichkeiten, die Näherung **2. Ordnung** berücksichtigt zusätzlich noch das vorherige Wort.

Buchstabennäherung 0. Ordnung	ITVWDGAKNAJTSQOSRMOIA QVFWTKHXD (Bei gleichem Symbolumfang für Deutsch und Englisch gleich)	
Buchstabennäherung 1. Ordnung	EME GKNEET ERS TITBL BTZENFNDBGD EAIE LASZ BETEATR IASMIRCH EGEOM	OCRO HLI RGWR NMIEL WIS EU LL NBNESEBYA TH EEI ALHENHHTTPA
Buchstabennäherung 2. Ordnung	AUSZ KEINU WONDINGLIN DUFRN IS AR 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 DJPFERENT THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITE THAT THE CHARACTER OF THIS
Wortnäherung 2. Ordnung	WEIL JEDER ANLAGE HATNACH DEM PFERDE NICHT ALLEIN DER HERR WILL ALS OB ICH FAST JEDES HAUS ZU SITZEN	

Weitere Information zur synthetischen Erzeugung von deutschen und englischen Texten finden Sie in **Aufgabe A1.8.**