Musterlösung zur Aufgabe A1.1

a) Bei der Datei „Durchwachsen” sind die beiden Wahrscheinlichkeiten p_G und p_B gleich, jeweils 0.5. Damit ergibt sich für die Entropie:

$$H_\text{D} = 0.5 \cdot \log_2 \frac{1}{0.5} + 0.5 \cdot \log_2 \frac{1}{0.5} \underline{= 1\,\text{bit/Anfrage}}\,.$$

b) Mit p_B = 0.8 und p_G = 0.2 erhält man einen kleineren Entropiewert:

$$\begin{aligned} H_\text{R} &= 0.8 \cdot \log_2 \frac{5}{4} + 0.2 \cdot \log_2 \frac{5}{1} = 0.8 \cdot \log_2 5 - 0.8 \cdot \log_2 4 + 0.2 \cdot \log_2 5 = \\ &= \log_2 5 - 0.8 \cdot \log_2 4 = \frac{\lg 5}{\lg 2} - 0.8 \cdot 2 = \frac{0.699}{0.301} - 1.6 \underline{= 0.722\,\text{bit/Anfrage}}\,. \end{aligned}$$

c) In der Datei „Angenehm” sind die Wahrscheinlichkeiten gegenüber der Datei „Regenloch” genau vertauscht. Durch diese Vertauschung wird die Entropie nicht verändert:

$$H_\text{A} = H_\text{R} \underline{= 0.722\,\text{bit/Anfrage}}\,.$$

d) Mit p_B = 1/30 und p_G = 29/30 ergeben sich folgende Informationsgehalte:

$$\begin{aligned} I_\text{B} &= \log_2 30 = \frac{\lg 30}{\lg 2} = \frac{1.477}{0.301} \underline{= 4.907\,\text{bit/Anfrage}}\,, \\ I_\text{G} &= \log_2 \frac{30}{29} = \frac{\lg 1.034}{\lg 2} = \frac{1.477}{0.301} \underline{= 0.049\,\text{bit/Anfrage}}\,. \end{aligned}$$

e) Die Entropie H_P ist der mittlere Informationsgehalt der beiden Ereignisse „B” und „G”:

$$H_\text{P} = \frac{1}{30} \cdot 4.907 + \frac{29}{30} \cdot 0.049 = 0.164 + 0.047 \underline{= 0.211\,\text{bit/Anfrage}}\,.$$

Obwohl das Ereignis „B” seltener auftritt als „G”, ist sein Beitrag zur Entropie größer.

f) Die Ereignisse „B” und „G” sind bei der Datei „Unbekannt” tatsächlich gleichwahrscheinlich: Die 60 dargestellten Symbole teilen sich auf in 30 mal „G” und 30 mal „B”. Es bestehen nun aber starke statistische Bindungen innerhalb der zeitlichen Folge. Nach längeren Schönwetterperioden folgen meist viele schlechte Tage am Stück.

Aufgrund dieser statistischen Abhängigkeit innerhalb der **B**/**G**–Folge ist $H_\text{U} \approx 0.72$ bit/Anfrage kleiner als H_D = 1 bit/Anfrage. H_D ist gleichzeitig das Maximum für $M = 2$ ⇒ die letzte Aussage ist mit Sicherheit falsch. Richtig sind demnach die <u>Aussagen 1 und 3</u>.

Musterlösung zur Zusatzaufgabe Z1.1

Hinweis: Aus Platzgründen verwenden wir in der Musterlösung „ld” anstelle von „$\log_2$”.

a) Die Entropiefunktion $H'_{\rm bin}(p)$ lautet entsprechend der Angabe:

$$\begin{aligned} H'_{\rm bin}(p) &= p \cdot \ln \frac{1}{p} + (1-p) \cdot \ln \frac{1}{1-p} = \\ &= \ln 2 \cdot \left[p \cdot {\rm ld}\, \frac{1}{p} + (1-p) \cdot {\rm ld}\, \frac{1}{1-p} \right] \end{aligned}$$

$$\Rightarrow \quad H'_{\rm bin}(p) \ (\text{in nat}) = \ln 2 \cdot H_{\rm bin}(p) \ (\text{in bit}) = 0.693 \cdot H_{\rm bin}(p)\,.$$

Richtig ist also der <u>erste Lösungsvorschlag</u>. Die beiden weiteren Vorgaben machen keinen Sinn.

b) Die Optimierungsbedingung lautet $\mathrm{d}H_{\rm bin}(p)/\mathrm{d}p = 0$ bzw.

$$\frac{\mathrm{d}H'_{\rm bin}(p)}{\mathrm{d}p} \stackrel{!}{=} 0 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}p}\left[-p \cdot \ln p - (1-p) \cdot \ln(1-p) \right] \stackrel{!}{=} 0$$

$$\Rightarrow \quad -\ln p - p \cdot \frac{1}{p} + \ln(1-p) + (1-p) \cdot \frac{1}{1-p} \stackrel{!}{=} 0$$

$$\Rightarrow \quad \ln \frac{1-p}{p} = 0 \quad \Rightarrow \quad \frac{1-p}{p} = 1 \quad \Rightarrow \quad \underline{p = 0.5}\,.$$

Die Entropiewerte für $p = 0.5$ lauten somit:

$$\begin{aligned} H'_{\rm bin}(p = 0.5) &= -2 \cdot 0.5 \cdot \ln 0.5 = \ln 2 = 0.693\,\text{nat}\,, \\ H_{\rm bin}(p = 0.5) &= -2 \cdot 0.5 \cdot {\rm ld}\, 0.5 = {\rm ld}\, 2 \ \underline{= 1\,\text{bit}}\,. \end{aligned}$$

c) Für $p = 5\%$ erhält man:

$$\begin{aligned} H_{\rm bin}(p = 0.05) &= 0.05 \cdot {\rm ld}\, \frac{1}{0.05} + 0.95 \cdot {\rm ld}\, \frac{1}{0.95} = \\ &= \frac{1}{0.693} \cdot [0.05 \cdot \ln 20 + 0.95 \cdot \ln 1.053] = \\ &= \frac{1}{0.693} \cdot [0.05 \cdot 2.995 + 0.95 \cdot 0.051] \ \underline{\approx 0.286\,\text{bit}}\,. \end{aligned}$$

d) Diese Aufgabe lässt sich nicht in geschlossener Form lösen, sondern durch „Probieren”. Eine Lösung liefert das Ergebnis:

$$H_{\rm bin}(p = 0.10) = 0.469\,\text{bit}\,, \quad H_{\rm bin}(p = 0.12) = 0.529\,\text{bit}\,, \quad H_{\rm bin}(p = 0.11) \approx 0.5\,\text{bit}$$

$$\Rightarrow \quad p_1 \approx 0.11\,.$$

Die zweite (gesuchte) Lösung ergibt sich aus der Symmetrie von $H_{\rm bin}(p)$ zu $p_2 = 1 - p_1 \ \underline{= 0.89}$.

e) Mit $p = 0.45$ erhält man $H_{\rm bin}(p) = 0.993$ bit. Der relative Fehler bezüglich Entropie ist somit

$$\varepsilon_H = \frac{H_{\rm bin}(p = 0.45) - H_{\rm bin}(p = 0.5)}{H_{\rm bin}(p = 0.5)} = \frac{0.993 - 1}{1} \ \underline{= -0.7\,\%}\,.$$

Das Minuszeichen deutet darauf hin, dass der Entropiewert $H = 0.993$ zu klein ist. Hätte die Simulation den zu großen Wert $p = 0.55$ ergeben, so wäre H und auch der relative Fehler genau so groß.

f) Es gilt $H_{\text{bin}}(p = 0.045) = 0.265$ bit. Mit dem Ergebnis aus (c) ⇒ $H_{\text{bin}}(p = 0.05) = 0.286$ bit folgt daraus für den relativen Fehler bezüglich der Entropie:

$$\varepsilon_H = \frac{H_{\text{bin}}(p = 0.045) - H_{\text{bin}}(p = 0.05)}{H_{\text{bin}}(p = 0.05)} = \frac{0.265 - 0.286}{0.286} \underline{= -7.3\,\%}.$$

Eine falsche Bestimmung der Symbolwahrscheinlichkeiten um 10% macht sich für $p = 0.05$ aufgrund des steileren $H_{\text{bin}}(p)$–Verlaufs deutlich stärker bemerkbar als für $p = 0.5$. Eine zu große Wahrscheinlichkeit $p = 0.055$ hätte zu $H_{\text{bin}}(p = 0.055) = 0.307$ bit geführt und damit zu einer Verfälschung um $\varepsilon_H = +7.3\%$. In diesem Bereich verläuft die Entropiekurve also (mit guter Näherung) linear.

Musterlösung zur Aufgabe A1.2

a) Mit den Auftrittswahrscheinlichkeiten 1/2, 1/3 und 1/6 erhält man folgenden Entropiewert:

$$\begin{aligned} H &= 1/2 \cdot \log_2(2) + 1/3 \cdot \log_2(3) + 1/6 \cdot \log_2(6) = \\ &= (1/2 + 1/6) \cdot \log_2(2) + (1/3 + 1/6) \cdot \log_2(3) = \\ &= 2/3 \cdot 1\,\text{bit} + 1/2 \cdot 1.585\,\text{bit} \underline{\approx 1.46\,\text{bit}}\,. \end{aligned}$$

b) Richtig ist Lösungsvorschlag 2. Die Entropie hängt nur von den Auftrittswahrscheinlichkeiten ab. Es ist dabei egal, welche Zahlenwerte oder physikalische Größen man den einzelnen Symbolen zuordnet. Anders ist es bei Mittelwerten oder der AKF–Berechnung. Werden nur Symbole angegeben, so kann man hierfür keine Momente angeben. Außerdem hängen die Mittelwerte, Autokorrelation, usw. davon ab, ob man die Zuordnung bipolar (–1, 0, +1) oder unipolar (zum Beispiel: 0, 1, 2) vereinbart.

c) Die Entropie der Quelle Q_2 lässt sich wie folgt ausdrücken:

$$\begin{aligned} H &= p \cdot \log_2 \frac{1}{p} + 2 \cdot \frac{1-p}{2} \cdot \log_2 \frac{2}{1-p} = \\ &= p \cdot \log_2 \frac{1}{p} + (1-p) \cdot \log_2 \frac{1}{1-p} + (1-p) \cdot \log_2(2) = H_{\text{bin}}(p) + 1 - p\,. \end{aligned}$$

Für $p = 0.5 \Rightarrow H_{\text{bin}}(p) = 1$ ergibt sich $H \underline{= 1.5\ \text{bit}}$.

d) Die maximale Entropie einer gedächtnislosen Quelle mit dem Symbolumfang M ergibt sich, wenn alle M Symbole gleichwahrscheinlich sind. Für den Sonderfall $M = 3$ folgt daraus:

$$p_{\text{R}} + p_{\text{G}} + p_{\text{S}} = 1 \quad \Rightarrow \quad \underline{p = 1/3}\,.$$

Damit erhält man mit dem Ergebnis der Teilaufgabe (c) die folgende Entropie:

$$\begin{aligned} H &= H_{\text{bin}}(1/3) + 1 - 1/3 = 1/3 \cdot \log_2(3) + 2/3 \cdot \log_2(3/2) + 2/3 = \\ &= 1/3 \cdot \log_2(3) + 2/3 \cdot \log_2(3) - 2/3 \cdot \log_2(2) + 2/3 = \log_2(3) = 1.585\,\text{bit}\,. \end{aligned}$$

e) „Roulette 1” ist informationstheoretisch gleich der Konfiguration Q_2 mit $p = 1/37$:

$$p_{\text{G}} = p = \frac{1}{37}\,, \quad p_{\text{R}} = p_{\text{S}} = \frac{1-p}{2} = \frac{18}{37}\,.$$

Damit erhält man mit dem Ergebnis der Teilaufgabe (c):

$$\begin{aligned} H &= H_{\text{bin}}(1/37) + \frac{36}{37} = \frac{1}{37} \cdot \log_2(37) + \frac{36}{37} \cdot \log_2(37) - \frac{36}{37} \cdot \log_2 36 + \frac{36}{37} = \\ &= \log_2(37) + \frac{36}{37} \cdot (1 - \log_2(36)) = 5.209 - 4.057 \underline{= 1.152\,\text{bit}}\,. \end{aligned}$$

f) Setzt man bei Roulette auf einzelne Zahlen ⇒ Konfiguration „Roulette 2”, so sind alle Zahlen von **0** bis **36** gleichwahrscheinlich und man erhält:

$$H = \log_2(37) \underline{= 5.209\,\text{bit}}\,.$$

Musterlösung zur Aufgabe A1.3

a) Es handelt sich um die <u>Quelle Q3</u>, da die Symbole gleichwahrscheinlich sind ⇒ $H_1 = H_0$ und keine statistischen Bindungen zwischen den Symbolen bestehen ⇒ $H = ... = H_2 = H_1$.

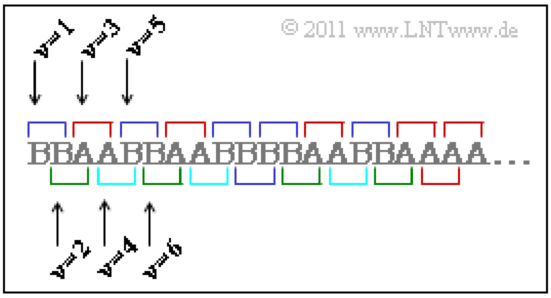
b) Man erkennt hier, dass **A** sehr viel häufiger auftritt als **B**, so dass $H_1 < H_0$ gelten muss. Entsprechend der Tabelle erfüllt nur die <u>Quelle Q1</u> diese Bedingung. Aus H_1 = 0.5 bit/Symbol kann man die Symbolwahrscheinlichkeiten $p_A \approx 0.89$ und $p_B \approx 0.11$ ermitteln.

c) Durch Ausschlussverfahren kommt man zum <u>Ergebnis Q2</u>: Die Quelle Q1 gehört zur schwarzen Folge, Q3 zur blauen und Q4 zum Wiederholungscode und damit offensichtlich zur grünen Symbolfolge. Die rote Symbolfolge weist folgende Eigenschaften auf:

- Wegen $H_1 = H_0$ sind die Symbole gleichwahrscheinlich: $p_A = p_B = 0.5$.
- Wegen $H < H_1$ bestehen statistische Bindungen innerhalb der Folge. Diese erkennt man daran, dass es mehr Übergänge zwischen **A** und **B** als bei statistischer Unabhängigkeit gibt.

d) Bei der grünen Symbolfolge (Quelle Q4) sind die Symbole **A** und **B** gleichwahrscheinlich:

$$p_A = p_B = 0.5 \quad \Rightarrow \quad H_1 = 1\,\text{bit/Symbol}\,.$$



Zur Berechnung von H_2 müssen Zweiertupel betrachtet und daraus die Verbundwahrscheinlichkeiten p_{AA}, p_{AB}, p_{BA} und p_{BB} berechnet werden. Aus der Skizze erkennt man:

- Die Kombinationen **AB** und **BA** sind nur dann möglich, wenn ein Tupel bei geradzahligem ν beginnt. Für die Verbundwahrscheinlichkeiten p_{AB} und p_{BA} gilt dann:

$$\begin{aligned} p_{AB} &= \Pr(\nu \text{ ist gerade}) \cdot \Pr(q_\nu = \mathbf{A}) \cdot \Pr(q_{\nu+1} = \mathbf{B}\,|q_\nu = \mathbf{A}) = \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = p_{BA}\,. \end{aligned}$$

- Dagegen gelten für die beiden weiteren Kombinationen **AA** und **BB**:

$$\begin{aligned} p_{AA} &= \Pr(\nu \text{ ist ungerade, zum Beispiel 1}) \cdot \Pr(q_1 = \mathbf{A}) \cdot \Pr(q_2 = \mathbf{A}\,|q_1 = \mathbf{A}) + \\ &+ \Pr(\nu \text{ ist gerade, zum Beispiel 2}) \cdot \Pr(q_2 = \mathbf{A}) \cdot \Pr(q_3 = \mathbf{A}\,|q_2 = \mathbf{A}) = \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{8} = p_{BB}\,. \end{aligned}$$

Damit ergibt sich für die Entropienäherung:

$$\begin{aligned} H_2 &= \frac{1}{2} \cdot \left[2 \cdot \frac{3}{8} \cdot \log_2 \frac{8}{3} + 2 \cdot \frac{1}{8} \cdot \log_2(8) \right] = \\ &= \frac{3}{8} \cdot \log_2(8) - \frac{3}{8} \cdot \log_2(3) + \frac{1}{8} \cdot \log_2(8) = \\ &= 1.5 - 0.375 \cdot 1.585 \; \underline{= 0.906\,\text{bit/Symbol}}\,. \end{aligned}$$

e) Nach ähnlichem Vorgehen kommt man bei Dreiertupeln zu den Verbundwahrscheinlichkeiten

$$\begin{aligned} p_{\rm AAA} &= p_{\rm BBB} = 1/4\,, \ \ p_{\rm ABA} = p_{\rm BAB} = 0\,, \\ p_{\rm AAB} &= p_{\rm ABB} = p_{\rm BBA} = p_{\rm BAA} = 1/8 \end{aligned}$$

und daraus zur Entropienäherung

$$H_3 = \frac{1}{3} \cdot \left[2 \cdot \frac{1}{4} \cdot \log_2(4) + 4 \cdot \frac{1}{8} \cdot \log_2(8) \right] = \frac{2.5}{3} \ \underline{= 0.833\,{\rm bit/Symbol}}\,.$$

Zur Berechnung von H_4 ergeben sich folgende 16 Wahrscheinlichkeiten:

$$\begin{aligned} p_{\rm AAAA} &= p_{\rm BBBB} = 3/16\,, \ \ p_{\rm AABB} = p_{\rm BBAA} = 2/16\,, \\ p_{\rm AAAB} &= p_{\rm ABBA} = p_{\rm ABBB} = p_{\rm BBBA} = p_{\rm BAAB} = p_{\rm BAAA} = 1/16\,, \\ p_{\rm AABA} &= p_{\rm ABAA} = p_{\rm ABAB} = p_{\rm BBAB} = p_{\rm BABB} = p_{\rm BABA} = 0\,. \end{aligned}$$

Daraus folgt:

$$\begin{aligned} H_4 &= \frac{1}{4} \cdot \left[2 \cdot \frac{3}{16} \cdot \log_2 \frac{16}{3} + 2 \cdot \frac{1}{8} \cdot \log_2(8) + 6 \cdot \frac{1}{16} \cdot \log_2(16) \right] = \\ &= [6 \cdot \log_2(16) - 6 \cdot \log_2(3) + 4 \cdot \log_2(8) + 6 \cdot \log_2(16)]\ /32 = 0.789\ {\rm bit/Symbol}\,. \end{aligned}$$

Man erkennt, dass auch die Näherung H_4 = 0.789 bit/Symbol noch weit vom tatsächlichen Entropiewert H = 0.5 bit/Symbol abweicht.

Hinweis: Der Wiederholungscode kann offensichtlich nicht durch eine Markovquelle modelliert werden. Wäre Q4 eine Markovquelle, so müsste nämlich gelten:

$$H = 2 \cdot H_2 - H_1 \quad \Rightarrow \quad H_2 = 1/2 \cdot (H + H_1) = 1/2 \cdot (0.5 + 1) = 0.75\,{\rm bit/Symbol}\,.$$

Musterlösung zur Aufgabe A1.4

a) Der Symbolumfang beträgt $M = 3$. Daraus ergibt sich der Entscheidungsgehalt mit dem *Logarithmus dualis* zur Basis 2 ($\log_2$ oder „ld"):

$$H_0 = \log_2 M = \log_2(3) \; \underline{= 1.585 \text{ bit/Symbol}}.$$

b) Die Entropienäherung erster Ordnung berücksichtigt nur die Symbolwahrscheinlichkeiten p_P, p_N und p_M und nicht die statistischen Bindungen innerhalb der Codefolge $\langle c_\nu \rangle$. Damit erhält man:

$$p_\text{N} = p_\text{L} = 1/2, \quad p_\text{P} = p_\text{M} = p_\text{H}/2 = 1/4$$

$$\Rightarrow \quad H_1 = \frac{1}{2} \cdot \log_2(2) + 2 \cdot \frac{1}{4} \cdot \log_2(4) \; \underline{= 1.5 \text{ bit/Symbol}}.$$

c) Zunächst müssen hier die $M^2 = 9$ Verbundwahrscheinlichkeiten von Zweiertupeln ermittelt werden, im Folgenden gekennzeichnet durch die beiden ersten Codesymbole c_1 und c_2:

- Da beim AMI–Code weder **P** auf **P** noch **M** auf **M** folgen kann, ist $p_\text{PP} = p_\text{MM} = 0$.
- Für die Verbundwahrscheinlichkeiten unter der Bedingung $c_2 = \textbf{N}$ gilt:

$$\begin{aligned} p_\text{NN} &= \Pr(c_1 = \textbf{N}) \cdot \Pr(c_2 = \textbf{N}\,|c_1 = \textbf{N}) = 1/2 \cdot 1/2 = 1/4, \\ p_\text{MN} &= \Pr(c_1 = \textbf{M}) \cdot \Pr(c_2 = \textbf{N}\,|c_1 = \textbf{M}) = 1/4 \cdot 1/2 = 1/8, \\ p_\text{PN} &= \Pr(c_1 = \textbf{P}) \cdot \Pr(c_2 = \textbf{N}\,|c_1 = \textbf{P}) = 1/4 \cdot 1/2 = 1/8. \end{aligned}$$

- Die Verbundwahrscheinlichkeiten der Zweiertupel „PM" und „MP" lauten:

$$\begin{aligned} p_\text{PM} &= \Pr(c_1 = \textbf{P}) \cdot \Pr(c_2 = \textbf{M}\,|c_1 = \textbf{P}) = 1/4 \cdot 1/2 = 1/8, \\ p_\text{MP} &= \Pr(c_1 = \textbf{M}) \cdot \Pr(c_2 = \textbf{P}\,|c_1 = \textbf{M}) = 1/4 \cdot 1/2 = 1/8. \end{aligned}$$

- Bei den restlichen Wahrscheinlichkeiten muss zusätzlich berücksichtigt werden, ob beim letzten Mal das Binärsymbol **H** mit **P** oder mit **M** codiert wurde $\Rightarrow$ weiterer Faktor 1/2:

$$\begin{aligned} p_\text{NM} &= \Pr(c_1 = \textbf{N}) \cdot \Pr(c_2 = \textbf{M}\,|c_1 = \textbf{N}) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8, \\ p_\text{NP} &= \Pr(c_1 = \textbf{N}) \cdot \Pr(c_2 = \textbf{P}\,|c_1 = \textbf{N}) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8. \end{aligned}$$

Damit ist die Entropie H_2' eines Zweiertupels bzw. dessen Entropie H_2 pro Codesymbol:

$$H_2' = \frac{1}{4} \cdot \log_2(4) + 6 \cdot \frac{1}{8} \cdot \log_2(8) \; = 2.75 \text{ bit/Zweiertupel}$$

$$\Rightarrow \quad H_2 = \frac{H_2'}{2} \; \underline{= 1.375 \text{ bit/Symbol}}.$$

d) Die Berechnung von H_3 erfolgt ähnlich wie bei der letzten Teilaufgabe für H_2, nur müssen nun $3^3 = 27$ Verbundwahrscheinlichkeiten ermittelt werden:

1. $p_\text{NNN} = 1/8 \quad (\text{nur einmal}),$
2. $p_\text{NMM} = p_\text{NPP} = p_\text{MNM} = \ldots = 0 \quad (\text{ingesamt } 12),$
3. $p_\text{NNM} = p_\text{NNP} = p_\text{PMP} = \ldots = 1/16 \quad (\text{ingesamt } 14)$

$$\Rightarrow \quad H_3 = \frac{1}{3} \cdot \left[\frac{1}{8} \cdot \log_2(8) + 14 \cdot \frac{1}{16} \cdot \log_2(16) \right] \underline{= 1.292\,\text{bit/Symbol}}\,.$$

e) Richtig sind die Lösungsvorschläge 1 und 2. Falsch ist dagegen die Aussage 3, da H_4 auf jeden Fall kleiner sein muss als H_3 = 1.292 bit/Symbol.

Musterlösung zur Zusatzaufgabe Z1.4

a) Da durch den AMI–Code weder neue Information hinzukommt noch Information verschwindet, ist die Entropie H_C der Codesymbolfolge $\langle c_\nu \rangle$ gleich der Quellenentropie H_Q. Bei gleichwahrscheinlichen und statistisch voneinander unabhängigen Quellensymbolen gilt deshalb:

$$H_{\rm Q} = 1\,{\rm bit/Binärsymbol} \quad \Rightarrow \quad H_{\rm C} \;\underline{= 1\,{\rm bit/Ternärsymbol}}\,.$$

b) Der Entscheidungsgehalt einer ternären Quelle beträgt $H_0 = \log_2(3) = 1.585$ bit/Symbol. Damit ergibt sich für die relative Redundanz

$$r_{\rm C} = 1 - H_{\rm C}/H_0 = 1 - 1/\log_2(3) \;\underline{= 36.9\,\%}\,.$$

c) Weiter gilt $H_C = H_Q$. Wegen den ungleichen Symbolwahrscheinlichkeiten ist nun H_Q kleiner:

$$H_{\rm Q} = \frac{1}{4}\cdot \log_2(4) + \frac{3}{4}\cdot \log_2(4/3) = 0.811\,{\rm bit/Binärsymbol}$$
$$\Rightarrow \quad H_{\rm C} = H_{\rm Q} \;\underline{= 0.811\,{\rm bit/Ternärsymbol}}\,.$$

d) In Analogie zur Teilaufgabe b) gilt

$$r_{\rm C} = 1 - 0.811/1.585 \;\underline{= 48.8\,\%}\,.$$

Man kann dieses Ergebnis verallgemeinern:

$$(1-0.488) = (1-0.189)\cdot(1-0.369)$$
$$\Rightarrow \quad (1 - r_{\rm Codefolge}) = (1 - r_{\rm Quelle})\cdot(1 - r_{\rm AMI-Code})\,.$$

e) Da jedes **L** auf **N** abgebildet wird und **H** alternierend auf **M** und **P**, gilt

$$p_{\rm N} = p_{\rm L} = 1/4\,,\quad p_{\rm P} = p_{\rm M} = p_{\rm H}/2 = 3/8$$
$$\Rightarrow \quad H_1 = 1/4\cdot\log_2(4) + 2\cdot 3/8\cdot\log_2(8/3) \;\underline{= 1.56\,{\rm bit/Ternärsymbol}}\,.$$

f) Nun ergeben sich die Symbolwahrscheinlichkeiten $p_N = 3/4$ sowie $p_P = p_M = 1/8$. Somit gilt:

$$H_1 = 3/4\cdot\log_2(4/3) + 2\cdot 1/8\cdot\log_2(8) \;\underline{= 1.06\,{\rm bit/Ternärsymbol}}\,.$$

Für $p_L = 1/4$, $p_H = 3/4$ ergibt sich $H_1 = 1.56$ bit/Symbol, bei $p_L = 3/4$, $p_H = 1/4$ dagegen ein deutlich kleinerer Wert: $H_1 = 1.06$ bit/Symbol. Für beide Parameterkombinationen gilt aber gleichermaßen:

$$H_0 = 1.585\,{\rm bit/Symbol}\,,\quad H_{\rm C} = \lim_{k\to\infty} H_k = 0.811\,{\rm bit/Symbol}\,.$$

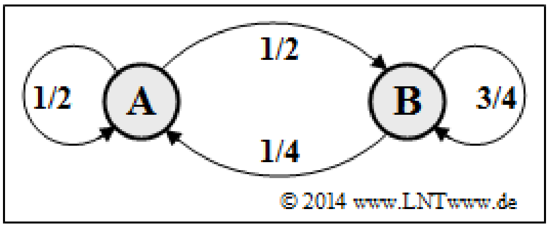
Daraus folgt: Betrachtet man zwei Nachrichtenquellen Q1 und Q2 mit gleichem Symbolumfang M ⇒ Entscheidungsgehalt H_0 = const., wobei bei der Quelle Q1 die Entropienäherung erster Ordnung deutlich größer ist als bei der Quelle Q2, so kann man daraus noch lange nicht schließen, dass die Entropie von Q1 tatsächlich größer ist als die Entropie von Q2. Vielmehr muss man für beide Quellen

- genügend viele Entropienäherungen H_1, H_2, H_3, ... , berechnen, und
- daraus (grafisch oder analytisch) den Grenzwert von H_k für $k \to \infty$ bestimmen.

Erst dann ist eine endgültige Aussage über die Entropieverhältnisse möglich.

Musterlösung zur Aufgabe A1.5

a) Hier gilt für die Übergangswahrscheinlichkeiten:



$$p_{\rm A|A} = 1 - p_{\rm B|A} = 1 - q \underline{= 0.5},$$
$$p_{\rm B|B} = 1 - p_{\rm A|B} = 1 - p \underline{= 0.75}.$$

Nach **A** sind **A** und **B** gleichwahrscheinlich. Nach **B** tritt **B** sehr viel häufiger als **A** auf.

b) Entsprechend den angegebenen Gleichungen gilt:

$$p_{\rm A} = \frac{p}{p+q} = \frac{0.25}{0.25 + 0.50} \underline{= 0.333},$$

$$p_{\rm B} = \frac{q}{p+q} = \frac{0.50}{0.25 + 0.50} \underline{= 0.667}.$$

c) Mit den unter (b) berechneten Wahrscheinlichkeiten gilt:

$$H_1 = H_{\rm bin}(p_{\rm A}) = 1/3 \cdot \log_2(3) + 2/3 \cdot \log_2(1.5) = 1.585 - 2/3 \underline{= 0.918 \text{ bit/Symbol}}.$$

d) Die Entropie der Markovquelle lautet entsprechend der Angabe

$$H = p_{\rm AA} \cdot \log_2 \frac{1}{p_{\rm A|A}} + p_{\rm AB} \cdot \log_2 \frac{1}{p_{\rm B|A}} + p_{\rm BA} \cdot \log_2 \frac{1}{p_{\rm A|B}} + p_{\rm BB} \cdot \log_2 \frac{1}{p_{\rm B|B}}.$$

Für die Verbundwahrscheinlichkeiten gilt:

$$p_{\rm AA} = p_{\rm A|A} \cdot p_{\rm A} = (1-q) \cdot \frac{p}{p+q} = \frac{1/2 \cdot 1/4}{3/4} = \frac{1}{6},$$
$$p_{\rm AB} = p_{\rm B|A} \cdot p_{\rm A} = q \cdot \frac{p}{p+q} = \frac{1/2 \cdot 1/4}{3/4} = \frac{1}{6},$$
$$p_{\rm BA} = p_{\rm A|B} \cdot p_{\rm B} = p \cdot \frac{q}{p+q} = p_{\rm AB} = \frac{1}{6},$$
$$p_{\rm BB} = p_{\rm B|B} \cdot p_{\rm B} = (1-p) \cdot \frac{q}{p+q} = \frac{3/4 \cdot 1/2}{3/4} = \frac{1}{2}$$
$$\Rightarrow \quad H = 1/6 \cdot \log_2(2) + 1/6 \cdot \log_2(2) + 1/6 \cdot \log_2(4) + 1/2 \cdot \log_2(4/3) =$$
$$= 1/6 + 1/6 + 2/6 + 1 - 1/2 \cdot \log_2(3) \underline{= 0.875 \text{ bit/Symbol}}.$$

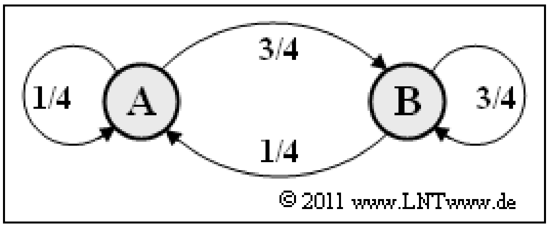
e) Allgemein gilt mit $H_{\rm M} = H$ für die k–Entropienäherung:

$$H_k = \frac{1}{k} \cdot [H_1 + (k-1) \cdot H_{\rm M}].$$

Daraus folgt:

$$H_2 = \frac{1}{2} \cdot [0.918 + 1 \cdot 0.875] \underline{= 0.897 \text{ bit/Symbol}},$$
$$H_3 = \frac{1}{3} \cdot [0.918 + 2 \cdot 0.875] \underline{= 0.889 \text{ bit/Symbol}},$$
$$H_4 = \frac{1}{4} \cdot [0.918 + 3 \cdot 0.875] \underline{= 0.886 \text{ bit/Symbol}}.$$

f) Mit dem neuen Parametersatz (p = 1/4, q = 3/4) erhält man für die Symbolwahrscheinlichkeiten: p_A = 1/4 und p_B = 3/4. Dieser Sonderfall führt demnach zu statistisch unabhängigen Symbolen:



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

Damit ist die Entropie H identisch mit der Entropienäherung H_1:

$$H = H_1 = 1/4 \cdot \log_2(4) + 3/4 \cdot \log_2(4/3) = 2 - 0.75 \cdot \log_2(3) \;\underline{= 0.811 \text{ bit/Symbol}}\,.$$

Die Entropienäherungen H_2, H_3, H_4, ... liefern hier ebenfalls das Ergebnis 0.811 bit/Symbol.

Musterlösung zur Zusatzaufgabe Z1.5

a) Bei einer **stationären binären Markovquelle** erster Ordnung gilt:

$$p_{\rm A} = p_{\rm A|A} \cdot p_{\rm A} + p_{\rm A|B} \cdot p_{\rm B} = (1-q) \cdot p_{\rm A} + q \cdot p_{\rm B}$$
$$q \cdot p_{\rm A} = q \cdot p_{\rm B} \quad \Rightarrow \quad p_{\rm A} = p_{\rm B} \underline{= 0.5}\,.$$

b) Zur Berechnung von H benötigt man alle vier Verbundwahrscheinlichkeiten:

$$\begin{aligned} p_{\rm AA} &= p_{\rm A} \cdot p_{\rm A|A} = 1/2 \cdot (1-q) = p_{\rm BB}\,, \\ p_{\rm AB} &= p_{\rm A} \cdot p_{\rm B|A} = 1/2 \cdot q = p_{\rm BA}\,. \end{aligned}$$

Setzt man diese Werte in die gegebene Entropie–Gleichung ein, so erhält man

$$\begin{aligned} H &= 2 \cdot \frac{1}{2} \cdot (1-q) \cdot \log_2 \frac{1}{1-q} + 2 \cdot \frac{1}{2} \cdot q \cdot \log_2 \frac{1}{q} = \\ &= q \cdot \log_2 \frac{1}{q} + (1-q) \cdot \log_2 \frac{1}{1-q} = H_{\rm bin}(q)\,. \end{aligned}$$

Der gesuchte Zahlenwert ist $H = H_{\rm bin}$ (0.25) $\underline{= 0.811 \text{ bit/Symbol}}$.

c) Bei gleichwahrscheinlichen Binärsymbolen ist $H_1 \underline{= 1 \text{ bit/Symbol}}$. Mit der für Markovquellen gültigen Gleichung gilt weiter:

$$\begin{aligned} H_2 &= \frac{1}{2} \cdot [H_1 + H] \underline{= 0.906 \,{\rm bit/Symbol}}\,, \\ H_3 &= \frac{1}{3} \cdot [H_1 + 2H] \underline{= 0.874 \,{\rm bit/Symbol}}\,. \end{aligned}$$

d) Das Maximum der binären Entropiefunktion ergibt sich für $q \underline{= 0.5}$. Damit beträgt die maximale Entropie H = 1 bit/Symbol. Man erkennt aus der Beziehung $H = H_1$ und aus dem vorne abgebildeten Übergangsdiagramm, dass q = 0.5 statistisch unabhängige Symbole zur Folge hat:

$$p_{\rm A} = p_{\rm A|A} = p_{\rm A|B} = 0.5\,, \quad p_{\rm B} = p_{\rm B|A} = p_{\rm B|B} = 0.5\,.$$

e) Richtig sind die <u>Lösungsvorschläge 1 und 2</u>. Die Symbolfolge ergibt sich entweder zu **AAAAAA**... oder zu **BBBBBB**..., je nachdem, welches Symbol als Startwert vorgegeben wurde. Die Entropie einer solchen Quelle ist $H = H_{\rm bin}(0) = 0$.

f) Nun kann weder **A** direkt auf **A** noch **B** direkt auf **B** folgen. Es ergibt sich eine alternierende Folge, je nach Startwert die Folge **ABABAB**... oder **BABABA**... ⇒ <u>Lösungsvorschlag 3</u>. Diese Quelle hat in beiden Fällen ebenfalls die Entropie $H = 0 = H_{\rm bin}(1)$.

Musterlösung zur Aufgabe A1.6

a) Die Symbolwahrscheinlichkeiten der ternären Markovquelle sind gegeben. Daraus lässt sich die Entropienäherung H_1 berechnen:

$$H_1 = 1/2 \cdot \log_2(2) + 2 \cdot 1/4 \cdot \log_2(4) \ \underline{= 1.5\,\text{bit/Symbol}}\,.$$

b) Die Verbundwahrscheinlichkeit ist $p_{\rm XY} = p_{\rm X} \cdot p_{\rm Y|X}$, wobei $p_{\rm X}$ die Symbolwahrscheinlichkeit von X angibt und $p_{\rm Y|X}$ die bedingte Wahrscheinlichkeit für **Y**, unter der Voraussetzung, dass vorher **X** aufgetreten ist. **X** und **Y** sind hier Platzhalter für die Symbole **N**, **P**, **M**. Dann gilt:

$$\begin{aligned}
p_{\rm NN} &= 1/2 \cdot 1/2 = 1/4\,, \ p_{\rm PP} = 1/4 \cdot 0 = 0\,, \ p_{\rm MM} = 1/4 \cdot 0 = 0\,,\\
p_{\rm NP} &= 1/2 \cdot 1/4 = 1/8\,, \ p_{\rm PM} = 1/4 \cdot 1/2 = 1/8\,, \ p_{\rm MN} = 1/4 \cdot 1/2 = 1/8\,,\\
p_{\rm NM} &= 1/2 \cdot 1/4 = 1/8\,, \ p_{\rm MP} = 1/4 \cdot 1/2 = 1/8\,, \ p_{\rm PN} = 1/4 \cdot 1/2 = 1/8
\end{aligned}$$

$$\Rightarrow \quad H_2 = \frac{1}{2} \cdot \left[1/4 \cdot \log_2(4) + 6 \cdot 1/8 \cdot \log_2(8)\right] \ \underline{= 1.375\,\text{bit/Symbol}}\,.$$

c) Da MQ3 Markoveigenschaften aufweist, können aus H_1 und H_2 alle Entropienäherungen H_k (k = 3, 4, ...) direkt angegeben werden und auch der Grenzwert $H = H_k$ für $k \to \infty$:

$$\begin{aligned}
H &= 2 \cdot H_2 - H_1 = 2 \cdot 1.375 - 1.5 && \underline{= 1.250\,\text{bit/Symbol}}\,,\\
H_3 &= (H_1 + 2 \cdot H)/3 = (1.5 + 2 \cdot 1.25)/3 && \underline{= 1.333\,\text{bit/Symbol}}\,,\\
H_4 &= (H_1 + 3 \cdot H)/4 = (1.5 + 3 \cdot 1.25)/4 && \underline{= 1.3125\,\text{bit/Symbol}}\,.
\end{aligned}$$

Die 10. Entropienäherung unterscheidet sich noch immer, wenn auch nur geringfügig (um 2%) vom Endwert H = 1.25 bit/Symbol:

$$H_{10} = (H_1 + 9 \cdot H)/10 = (1.5 + 9 \cdot 1.25)/10 = 1.275\,\text{bit/Symbol}\,.$$

d) Entsprechend der Angabe sind hier die Symbole gleichwahrscheinlich. Daraus folgt:

$$H_1 = H_0 = \log_2(4) \ \underline{= 2\,\text{bit/Symbol}}\,.$$

e) Von den M^2 = 16 möglichen Zweiertupeln sind acht Kombinationen nicht möglich: **NP**, **NO**, **PP**, **PO**, **OM**, **ON**, **MM**, **MN**. Die acht weiteren Kombinationen (Zweiertupel) ergeben jeweils den Verbundwahrscheinlichkeitswert 1/8, wie an zwei Beispielen gezeigt werden soll:

$$\begin{aligned}
p_{\rm NN} &= p_{\rm N} \cdot p_{\rm N|N} = 1/4 \cdot 1/2 = 1/8\,,\\
p_{\rm MP} &= p_{\rm M} \cdot p_{\rm P|M} = 1/4 \cdot 1/2 = 1/8\,.
\end{aligned}$$

$$\Rightarrow \quad H_2 = \frac{1}{2} \cdot \left[8 \cdot \frac{1}{8} \cdot \log_2(8)\right] \ \underline{= 1.5\,\text{bit/Symbol}}\,.$$

f) Aufgrund der Markoveigenschaft gilt hier:

$$\begin{aligned}
H &= 2 \cdot H_2 - H_1 = 2 \cdot 1.5 - 2 && \underline{= 1\,\text{bit/Symbol}}\,,\\
H_3 &= (H_1 + 2 \cdot H)/3 = (2 + 2 \cdot 1)/3 && \underline{= 1.333\,\text{bit/Symbol}}\,,\\
H_4 &= (H_1 + 3 \cdot H)/4 = (2 + 3 \cdot 1)/4 && \underline{= 1.250\,\text{bit/Symbol}}\,.
\end{aligned}$$

Auch hier unterscheidet sich die 10. Näherung noch deutlich vom Endwert, nämlich um 10%:

$$H_{10} = (H_1 + 9 \cdot H)/10 = (2 + 9 \cdot 1)/10 = 1.1\,\text{bit/Symbol}\,.$$

Eine Abweichung um 2% ergibt sich hier erst für k = 50. Zum Vergleich: Bei der Markovquelle MQ3 wurde diese Annäherung bereits mit k = 10 erreicht.

Musterlösung zur Zusatzaufgabe Z1.6

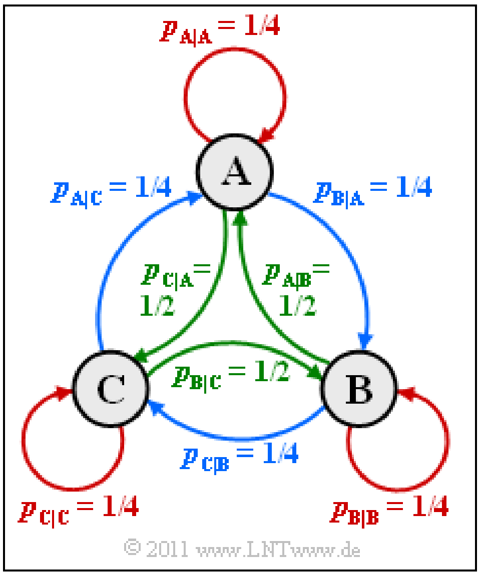
Hinweis: Aus Platzgründen verwenden wir in der Musterlösung „ld” anstelle von „$\log_2$”.

a) Die maximale Entropie ergibt sich dann, wenn die Symbole **A**, **B** und **C** gleichwahrscheinlich und die Symbole innerhalb der Folge statistisch voneinander unabhängig sind. Dann muss gelten:

- $p_A = p_{A|A} = p_{A|B} = p_{A|C} = 1/3$,
- $p_B = p_{B|A} = p_{B|B} = p_{B|C} = 1/3$,
- $p_C = p_{C|A} = p_{C|B} = p_{C|C} = 1/3$.

Beispielsweise erhält man aus $p_{C|C}$ = 1/3 der Wert $\underline{p = 1/3}$. Berücksichtigt man noch $p_{A|A} = q \cdot p$, so folgt $\underline{q = 1}$. Damit ergibt sich die maximale Entropie H_{max} = ld 3 $\underline{= 1.585 \text{ bit/Symbol}}$.

b) Mit den Parameterwerten p = 1/4 und q = 1 ergibt sich das nebenstehende Übergangsdiagramm, das folgende Symmetrien aufweist:



- $p_{A|A} = p_{B|B} = p_{C|C} = 1/4$ (rot markiert),
- $p_{A|B} = p_{B|C} = p_{C|A} = 1/2$ (grün markiert),
- $p_{A|C} = p_{B|A} = p_{C|B} = 1/4$ (blau markiert).

Es ist offensichtlich, dass die Symbolwahrscheinlichkeiten alle gleich sind:

$$p_A = p_B = p_C = 1/3$$
$$\Rightarrow \quad H_1 = \text{ld}\, 3 \; \underline{= 1.585 \text{ bit/Symbol}}\,.$$

c) Für die zweite Entropienäherung benötigt man die 3^2 = 9 Verbundwahrscheinlichkeiten. Mit dem Ergebnis der Teilaufgabe b) erhält man hierfür:

$$\begin{aligned} p_{AA} &= p_{BB} = p_{CC} = p_{AC} = p_{BA} = p_{CB} = 1/12\,,\\ p_{AB} &= p_{BC} = p_{CA} = 1/6 \\ \Rightarrow \quad H_2 &= \frac{1}{2} \cdot \left[6 \cdot \frac{1}{12} \cdot \text{ld}\, 12 + 3 \cdot \frac{1}{6} \cdot \text{ld}\, 6 \right] = \\ &= \frac{1}{4} \cdot \text{ld}\, 4 + \frac{1}{4} \cdot \text{ld}\, 3 + \frac{1}{4} \cdot \text{ld}\, 2 + \frac{1}{4} \cdot \text{ld}\, 3 = \frac{3}{4} + \frac{\text{ld}\, 3}{2} \; \underline{= 1.5425 \text{ bit/Symbol}}\,. \end{aligned}$$

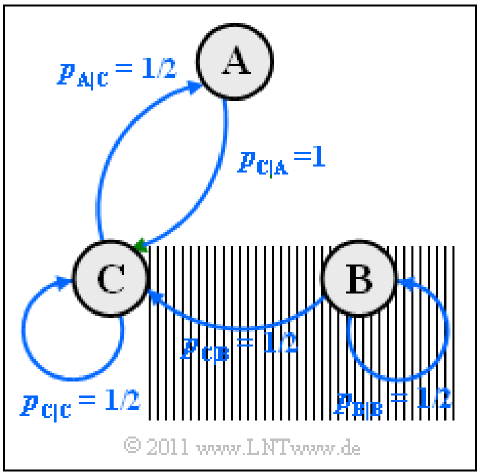
d) Aufgrund der Markoveigenschaft der Quelle gilt

$$H = 2 \cdot H_2 - H_1 = [3/2 + \text{ld}\, 3] - \text{ld}\, 3 \; \underline{= 1.5 \text{ bit/Symbol}}\,.$$

Zum gleichen Ergebnis würde man mit folgender Rechnung kommen:

$$\begin{aligned} H &= p_{AA} \cdot \text{ld}\, \frac{1}{p_{A|A}} + p_{AB} \cdot \text{ld}\, \frac{1}{p_{B|A}} + \ldots \\ &= 6 \cdot \frac{1}{12} \cdot \text{ld}\, 4 + 3 \cdot \frac{1}{16} \cdot \text{ld}\, 2 \; \underline{= 1.5 \text{ bit/Symbol}}\,. \end{aligned}$$

e) Aus nebenstehendem Übergangsdiagramm mit den aktuellen Parametern erkennt man, dass bei Stationarität $p_{\rm B} = 0$ gelten wird: **B** kann höchstens zum Starzeitpunkt einmal auftreten. Es liegt also eine binäre Markovkette mit den Symbolen **A** und **C** vor. Die Symbolwahrscheinlichkeiten ergeben sich zu:



$$p_{\rm A} = 0.5 \cdot p_{\rm C}\,, \quad p_{\rm A} + p_{\rm C} = 1$$
$$\Rightarrow \quad p_{\rm A} = 1/3\,, \quad p_{\rm C} = 2/3\,.$$

Damit erhält man folgende Wahrscheinlichkeiten:

$$\begin{aligned}
p_{\rm A|A} &= 0 &&\Rightarrow \quad p_{\rm AA} = 0\,,\\
p_{\rm A|C} &= 1/2 &&\Rightarrow \quad p_{\rm CA} = p_{\rm C} \cdot p_{\rm A|C} = 2/3 \cdot 1/2 = 1/3\,, \quad {\rm ld}\,(1/p_{\rm A\,|C}) = 1\,,\\
p_{\rm C|A} &= 1 &&\Rightarrow \quad p_{\rm AC} = p_{\rm A} \cdot p_{\rm C|A} = 1/3 \cdot 1 = 1/3\,, \quad {\rm ld}\,(1/p_{\rm C\,|A}) = 0\,,\\
p_{\rm C|C} &= 1/2 &&\Rightarrow \quad p_{\rm CC} = p_{\rm C} \cdot p_{\rm C|C} = 2/3 \cdot 1/2 = 1/3\,, \quad {\rm ld}\,(1/p_{\rm C\,|C}) = 1
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \quad H &= p_{\rm AA} \cdot {\rm ld}\,\frac{1}{p_{\rm A|A}} + p_{\rm CA} \cdot {\rm ld}\,\frac{1}{p_{\rm A|C}} + p_{\rm AC} \cdot {\rm ld}\,\frac{1}{p_{\rm C|A}} + p_{\rm CC} \cdot {\rm ld}\,\frac{1}{p_{\rm C|C}} = \\
&= 0 + 1/3 \cdot 1 + 1/3 \cdot 0 + 1/3 \cdot 1 \; \underline{= 0.667\,{\rm bit/Symbol}}\,.
\end{aligned}$$

Musterlösung zur Aufgabe A1.7

a) Der Symbolumfang bei Küpfmüllers Untersuchungen war $\underline{M = 26}$, da er im Gegensatz zu Shannon das Leerzeichen zunächst nicht berücksichtigte. Bei dem vorgegebenen deutschen Text dieser Aufgabe ist der Symbolumfang deutlich größer,

- da hier auch die typisch deutschen Zeichen „ä", „ö", „ü" und „ß" vorkommen,
- zwischen Klein– und Großschreibung unterschieden wird,
- und zudem noch Ziffern und Interpunktionszeichen hinzukommen.

b) Mit dem Entscheidungsgehalt $H_0 = \log_2 (31) \approx 4.7$ und der Entropie $H = 1.3$ (jeweils mit der Einheit bit/Zeichen) erhält man für die relative Redundanz:

$$r = \frac{H_0 - H}{H_0} = \frac{4.7 - 1.3}{4.7} \underline{\approx 72.3\,\%}\,.$$

c) Richtig ist nur der <u>erste Lösungsvorschlag</u>. Laut Küpfmüller benötigt man nur 1.3 Binärzeichen pro Quellenzeichen. Bei einer Datei der Länge N würden also $1.3 \cdot N$ Binärsymbole ausreichen, allerdings nur dann, wenn

- die Quellensymbolfolge unendlich lang ist ($N \to \infty$), und
- diese bestmöglich codiert ist.

Dagegen besagt Küpfmüllers Ergebnis und die in der Teilaufgabe (b) errechnete relative Redundanz von mehr als 70% nicht, dass ein Leser den Text noch verstehen kann, wenn 70% der Zeichen ausgelöscht sind. Ein solcher Text

- ist nie unendlich lang,
- noch wurde er vorher optimal codiert.

d) Richtig ist <u>Aussage 2</u>. Testen Sie es selbst: Der zweite Block der Grafik auf der Angabenseite ist leichter zu entschlüsseln als der letzte Block, weil man weiß, wo es Fehler gibt. Wenn Sie es weiter versuchen wollen: Für den unteren Block wurde genau die gleiche Zeichenfehlerfolge wie für Block 2 verwendet, das heißt, Fehler gibt es bei den Zeichen 6, 35, 37, usw..

Abschließend soll noch der Originaltext angegeben werden, der auf der **Angabenseite** nur durch Auslöschungen (*Erasures*) oder echte Zeichenfehler verfälscht wiedergegeben ist.

A Wie kam Küpfmüller zum Ergebnis H = 1.5 bit/Buchstabe? Da es für die Statistik von Wortgruppen oder ganzen Sätzen keine Veröffentlichungen gab, schätzte er die Entropie deutscher Texte wie folgt ab:

B Ein zusammenhängender, sonst beliebiger deutscher Text wird an einer bestimmten Stelle abgedeckt. Der Text vorher wird gelesen, und der Leser soll versuchen, das folgende Wort aus dem vorhergehenden Text und dem Zusammenhang zu ermitteln.

C Bei sehr vielen solcher Versuche ergibt die prozentuale Zahl der Treffer ein Maß für die Bindungen zwischen den Wörtern und Sätzen. Es zeigt sich, dass bei ein und derselben Textart desselben Autors verhältnismäßig schnell, ab etwa 100 bis 200 Versuchen, ein konstanter Endwert des Trefferverhältnisses erreicht wird.

D Das Trefferverhältnis hängt dabei stark von der Textart ab. Es ergeben sich für verschiedene Textarten Werte zwischen 15 und 33% mit dem Mittelwert bei 22%. Das heißt aber auch: Im Durchschnitt können 22% aller Wörter aus dem Zusammenhang ermittelt werden.

E Anders ausgedrückt: Die Zahl der Wörter eines langen Textes kann mit dem Faktor 0.78 reduziert werden, ohne dass dessen Nachrichtengehalt signifikant abnimmt. Damit wird definitionsgemäß die Entropie H = 0.78 mal 2.0 = 1.56 bit/Buchstabe.

F Küpfmüller hat sein Ergebnis mit einer vergleichbaren empirischen Untersuchung aller Silben überprüft und kam mit dem Reduktionsfaktor von 0.54 sowie H3 = 2.8 auf die Entropie 1.51 bit/Buchstabe.

Musterlösung zur Aufgabe A1.8

a) Richtig ist der Lösungsvorschlag 1. In der Datei 1 erkennt man viele englische Wörter, in der Datei 2 viele deutsche. Sinn ergibt keiner der beiden Texte.

b) Die Wahrscheinlichkeit eines Leerzeichens beträgt bei der Datei 1 (Englisch) 19.8%. Also ist im Mittel jedes 1/0.198 = 5.05–tes Zeichen ein Leerzeichen. Die mittlere Wortlänge ergibt sich daraus zu

$$L_{\rm M} = \frac{1}{0.198} - 1 \approx 4.05 \ \text{Zeichen}\,.$$

Entsprechend gilt für Datei 2 (Deutsch):

$$L_{\rm M} = \frac{1}{0.176} - 1 \approx 4.68 \ \text{Zeichen}\,.$$

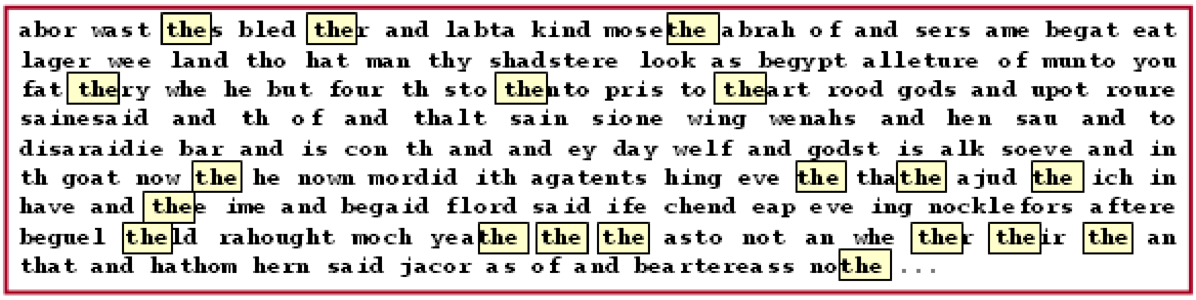
Richtig ist somit der Lösungsvorschlag 2. Die Abschätzungen von Shannon und Küpfmüller bestätigen unser Ergebnis.

c) Zur Bestimmung der Entropienäherung H_k müssen k–Tupel ausgewertet werden, zum Beispiel für k = 3 Tripel: „aaa”, „aab”, ... Nach der Generierungsvorschrift „Neues Zeichen hängt von den beiden Vorgängern ab” werden H_1, H_2 und H_3 von VORLAGE und SYNTHESE übereinstimmen, allerdings auf Grund der endlichen Dateilänge nur näherungsweise.

Dagegen unterscheiden sich die H_4–Näherungen stärker, da bei der Generierung der dritte Vorgänger unberücksichtigt bleibt. Bekannt ist nur, dass auch bezüglich SYNTHESE $H_4 < H_3$ gelten muss.

d) Richtig ist hier nur die Aussage 1. Nach einem Leerzeichen (Wortanfang) folgt „t” mit 17.8%, während am Wortende (vor einem Leerzeichen) „t” nur mit der Häufigkeit 8.3% auftritt. Insgesamt beträgt die Auftrittswahrscheinlichkeit von „t” über alle Positionen im Wort gemittelt 7.4%.

Als dritter Buchstaben nach Leerzeichen und „t” folgt „h” mit fast 82% und nach „th” ist „e” mit 62% am wahrscheinlichsten. Das lässt daraus schließen, dass „the” in einem englischen Text überdurchschnittlich oft vorkommt und damit auch in der synthetischen Datei 1, wie die folgende Grafik zeigt. Nicht bei allen Markierungen tritt „the” isoliert auf ⇒ direkt vorher und nachher ein Leerzeichen.


abor wast thes bled ther and labta kind mosethe abrah of and sers ame begat eat
lager wee land tho hat man thy shadstere look as begypt alleture of munto you
fat thery whe he but four th sto thento pris to theart rood gods and upot roure
sainesaid and th of and thalt sain sione wing wenahs and hen sau and to
disaraidie bar and is con th and and ey day welf and godst is alk soeve and in
th goat now the he nown mordid ith agatents hing eve the thathe ajud the ich in
have and thee ime and begaid flord said ife chend eap eve ing nocklefors aftere
beguel theld rahought moch yeathe the the asto not an whe ther their the an
that and hathom hern said jacor as of and beartereass nothe ...

e) Nach „de” ist tatsächlich „r” am wahrscheinlichsten (32.8%), gefolgt von „n” (28.5%), „s” (9.3%) und „m” (9.7%). Dafür verantwortlich könnten „der”, „den”, „des” und „dem” sein.

Weiterhin gilt:

- Nach „da” folgt „s” mit größter Wahrscheinlichkeit: 48.2%.
- Nach „di” folgt „e” mit größter Wahrscheinlichkeit: 78.7%.

Alle Aussagen treffen somit zu. Die Grafik zeigt die Datei 2 mit allen „der”, „die” und „das”.

aber scheinen zu ris und atter von kana akinestote ein wis der warbling dirkein
und sehr demt wer eiden zu ingerker angen die miten dasterkan ber auf desel den
dasua ungewistaren wohnet david de als land die eind das gen mittet sollem mach
und ihm zusameit nas der haberija kseiter west wachweinn und ma goll das aund
ken freuchlaufkottem de fam zurden ber ganigs spern tord jaher nit unde jern
den und eite meinern von unschens blend nachafte kinesusen zu ihr binerre ma
ewir kine auf deind seile meich imose hab willbs einund wich namm knes ine
dennem mung vonigewas begehm rudenn deinerzahwirdersat josall den nich zu zus
nachwegnichweger habeiner zogen ungegem nie dir dier dem dieber hon in ...