

Lösungsblatt Nr. 9

Aufgabe 30

a) Es sei $m < n$. Dann gilt wegen

$$H\left(\frac{1}{m}, \dots, \frac{1}{m}\right) = -\sum_{i=1}^m \frac{1}{m} \log \frac{1}{m} = \log m,$$

$$H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = -\sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} = \log n$$

und der strengen Monotonie der Logarithmusfunktion ($m < n \Leftrightarrow \log m < \log n$) auch

$$H\left(\frac{1}{m}, \dots, \frac{1}{m}\right) < H\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$$

b)

$$\begin{aligned} H\left(\frac{1}{m}, \dots, \frac{1}{m}\right) + H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) &\stackrel{(a)}{=} \log m + \log n \\ &= \log(m \cdot n) \stackrel{(a)}{=} H\left(\frac{1}{mn}, \dots, \frac{1}{mn}\right) \end{aligned}$$

c) Das Axiom

$$H(p_1, \dots, p_r, p_{r+1}, \dots, p_n) = H(a, b) + a \cdot H\left(\frac{p_1}{a}, \dots, \frac{p_r}{a}\right) + b \cdot H\left(\frac{p_{r+1}}{b}, \dots, \frac{p_n}{b}\right)$$

$$\text{mit} \quad a = \sum_{i=1}^r p_i \quad b = \sum_{i=r+1}^n p_i$$

läßt sich induktiv aus der einfacheren Form, quasi dem Basisfall¹

$$\begin{aligned} H(p_1, p_2, p_3) &= H(p_1, p_2+p_3) + p_1 \cdot \underbrace{H\left(\frac{p_1}{p_1}\right)}_{=-1 \log 1 = 0} + (p_2+p_3) \cdot H\left(\frac{p_2}{p_2+p_3}, \frac{p_3}{p_2+p_3}\right) \end{aligned}$$

gewinnen, dessen Gültigkeit wir im folgenden zeigen wollen:

$$H(p_1, p_2+p_3) + (p_2+p_3) \cdot H\left(\frac{p_2}{p_2+p_3}, \frac{p_3}{p_2+p_3}\right)$$

¹der echte, triviale Basisfall wäre natürlich

$$H(p_1, p_2) = H(p_1, p_2) + p_1 \underbrace{H\left(\frac{p_1}{p_1}\right)}_{=0} + p_2 \underbrace{H\left(\frac{p_2}{p_2}\right)}_{=0},$$

$$\begin{aligned}
&= -\left(p_1 \log p_1 + (p_2 + p_3) \log(p_2 + p_3)\right) \\
&\quad - (p_2 + p_3) \left(\frac{p_2}{p_2 + p_3} \log \frac{p_2}{p_2 + p_3} + \frac{p_3}{p_2 + p_3} \log \frac{p_3}{p_2 + p_3}\right) \\
&= -\left(p_1 \log p_1 + p_2 \log(p_2 + p_3) + p_3 \log(p_2 + p_3) + p_2 \log \frac{p_2}{p_2 + p_3} + p_3 \log \frac{p_3}{p_2 + p_3}\right) \\
&= -\left(p_1 \log p_1 + p_2 \underbrace{\log(p_2 + p_3)} + p_3 \underbrace{\log(p_2 + p_3)}\right) \\
&\quad + p_2(\log p_2 - \overbrace{\log(p_2 + p_3)}) + p_3(\log p_3 - \overbrace{\log(p_2 + p_3)}) \\
&= -(p_1 \log p_1 + p_2 \log p_2 + p_3 \log p_3) = H(p_1, p_2, p_3)
\end{aligned}$$

d) Daß

$$H(p, 1-p) = -(p \log p + (1-p) \log(1-p))$$

in $(0, 1) \ni p$ stetig ist, folgt aus der Stetigkeit des Logarithmierens im Intervall $(0, 1)$ sowie der Stetigkeit der Addition, Subtraktion und Multiplikation auf ganz \mathbb{R} .

q. e. d.

Aufgabe 31

a) Es ist nach Claude E. Shannon

$$H(p, 1-p) := -p \log p - (1-p) \log(1-p)$$

Im interessierenden Intervall $[0, 1]$ sind beide Summanden, $-p \log p$ und $-(1-p) \log(1-p)$, konkav, wie die zweite Ableitung² zeigt:

$$\begin{aligned}
\frac{d^2}{dp^2}(-p \log p) &= \frac{d}{dp} \left(-p \left[\frac{d}{dp} \log p \right] + \left[\frac{d}{dp} (-p) \right] \log p \right) \\
&= \frac{d}{dp} \left(-p \left[\frac{1}{p} \right] + [-1] \log p \right) = \frac{d}{dp} (-1 - \log p) \\
&= \frac{d}{dp} (-1) + \frac{d}{dp} (-\log p) = 0 + \left(-\frac{1}{p}\right) = -\frac{1}{p} < 0
\end{aligned}$$

(Die Argumentation für die Konkavität von $(1-p)$ läuft analog, aus Symmetriegründen im Intervall $[0, 1]$.) Da also die zweite Ableitung sowohl von $-p \log p$ als auch von $-(1-p) \log(1-p)$ im betrachteten Intervall negativ ist, gilt dies auch für die Summe der beiden; auch sie ist konkav.

Dies heißt jedoch, daß die Funktion $H(p, 1-p)$ selbst *höchstens ein* Extremum (das ein Maximum ist!) auf $[0, 1]$ hat, und zwar genau dort, wo die erste Ableitung Null wird:

$$\begin{aligned}
\frac{d}{dp} H(p, 1-p) &= \frac{d}{dp} (-p \log p - (1-p) \log(1-p)) \\
&= \frac{d}{dp} (-p \log p) + \frac{d}{dp} (-(1-p) \log(1-p))
\end{aligned}$$

²wir wenden im folgenden zunächst die Produktregel $(f \cdot g)' = f \cdot g' + f' \cdot g$ an

$$\begin{aligned}
&= \left[(-p) \frac{d}{dp} \log p + \frac{d}{dp} (-p) \log p \right] + \left[\underbrace{(- (1-p)) \frac{d}{dp} \log(1-p) + \frac{d}{dp} (- (1-p)) \log(1-p)}_{(f \circ g)' = g' \cdot (f' \circ g)} \right] \\
&= \left[(-p) \frac{1}{p} + (-1) \log p \right] + \left[\underbrace{(- (1-p)) \left(-1 \right) \frac{1}{1-p}}_{+1 \cdot \log(1-p)} + 1 \cdot \log(1-p) \right] \\
&= [-1 - \log p] + [1 + \log(1-p)] = \log(1-p) - \log p
\end{aligned}$$

Dieser Term, $f(1-p) - f(p)$, wird offenbar dann³ Null, wenn $1-p = p$ wird:

$$1-p = p \quad \Leftrightarrow \quad 2p = 1 \quad \Leftrightarrow \quad p = \frac{1}{2}$$

Hier liegt also ein bzw. das Maximum von $H(p, 1-p)$ vor.

q. e. d.

Bemerkung: Wir haben beim Ableiten vorausgesetzt, daß “log” hier der natürliche Logarithmus $\ln = \log_e$ ist,⁴ während doch in der Informationstheorie der Zweierlogarithmus $\text{ld} = \log_2$ gebräuchlich ist, weil wir üblicherweise in der Einheit “bit” rechnen. Man hätte die Rechnung auch mit “ld” führen können, hätte aber ein paar unschöne (konstante) Faktoren $\ln 2$ oder $\text{ld } e$ hineinbekommen, die sich aber schließlich wieder herausgekürzt hätten.

- b) Für den binären symmetrischen Kanal seien also die Übertragungswahrscheinlichkeiten wie folgt:

$$P(Y=0 | X=0) = P(Y=1 | X=1) = 1-\beta \quad (\text{kein Übertragungsfehler})$$

$$P(Y=1 | X=0) = P(Y=0 | X=1) = \beta \quad (\text{Übertragungsfehler})$$

Gegeben: diese bedingten Wahrscheinlichkeiten

Gesucht: die Kanalkapazität

Die Kanalkapazität ist die maximal erreichbare Transinformation; maximiert wird dabei über alle möglichen Quellenstatistiken, das sind die Wahrscheinlichkeitsverteilungen $P(X=x)$ an der Eingabe.

Aufgrund der Symmetrie des Kanals, d. h. der Symmetrie der Übertragungswahrscheinlichkeiten $P(Y|X)$, kann ausgeschlossen werden, daß das Maximum der Transinformation $H(X; Y)$ bei einer anderen Quellenstatistik als der symmetrischen angenommen wird. Setzen wir also diese symmetrische Quellenstatistik

$$P(X=0) = P(X=1) = \frac{1}{2}$$

voraus und berechnen die resultierende Transinformation $H(X; Y)$. Dies kann prinzipiell auf zwei verschiedene Arten geschehen: entweder als

$$H(X; Y) = H(X) - H(X|Y)$$

³und zwar sogar *genau dann*, da die Funktion $f = \log$ streng monoton und somit injektiv ist

⁴konkret: beim Schritt $(\ln p)' = 1/p$

oder umgekehrt als

$$H(X;Y) = H(Y) - H(Y|X)$$

Wir zeigen hier nur der Vollständigkeit halber beide Varianten:

Erste Methode:

$$\begin{aligned} H(X) &= H\left(\frac{1}{2}, \frac{1}{2}\right) \\ &= 1 \text{ bit} \end{aligned}$$

Dieser Wert (für die binäre Gleichverteilung) kann als bekannt vorausgesetzt werden. Der Subtrahend $H(X|Y)$ bestimmt sich zu

$$\begin{aligned} H(X|Y) &= \mathcal{E}_y \left[H(X | Y=y) \right] \quad (\text{Erwartungswert}) \\ &= P(Y=0) \cdot H(X | Y=0) + P(Y=1) \cdot H(X | Y=1) \\ &= P(Y=0) \cdot H(P(X=0 | Y=0), P(X=1 | Y=0)) \\ &\quad + P(Y=1) \cdot H(P(X=0 | Y=1), P(X=1 | Y=1)) \end{aligned}$$

Die bedingten Wahrscheinlichkeiten für die *Eingabe* in Abhängigkeit von der (gestörten) Ausgabe haben wir aber nicht. Mit einer einfachen, aber nicht völlig trivialen Hilfsrechnung⁵ können diese aber aufgrund der Symmetrie des Kanals als ebenso symmetrische $P(x|y) = P(y|x)$ ermittelt werden. Und weiter:

$$\begin{aligned} &= \frac{1}{2} \cdot H(1-\beta, \beta) + \frac{1}{2} \cdot H(\beta, 1-\beta) \\ &= H(\beta, 1-\beta) \\ &= -\beta \log \beta - (1-\beta) \log(1-\beta) \end{aligned}$$

Zweite Methode:

$$\begin{aligned} H(Y) &= H\left(\frac{1}{2}, \frac{1}{2}\right) \quad (\text{vgl. Fußnote}) \\ &= 1 \text{ bit} \end{aligned}$$

⁵Es muß "rückwärts" gerechnet werden, um über die bekannten $P(Y|X)$ die $P(X|Y)$ zu gewinnen:

$$\begin{aligned} P(X=0 | Y=0) &= \frac{P(X=0, Y=0)}{P(Y=0)} \\ &= \frac{P(X=0) \cdot P(Y=0 | X=0)}{P(X=0) \cdot P(Y=0 | X=0) + P(X=1) \cdot P(Y=0 | X=1)} \\ &= \frac{\frac{1}{2} \cdot (1-\beta)}{\frac{1}{2} \cdot (1-\beta) + \frac{1}{2} \cdot \beta} = 1-\beta \end{aligned}$$

Für die komplementäre Wahrscheinlichkeit gilt natürlich $P(X=1 | Y=0) = 1 - P(X=0 | Y=0) = \beta$; die Fälle für $Y=1$ laufen analog. Noch leichter ließe sich aufgrund der Symmetrie des Kanals für die Ausgabestatistik zeigen, daß

$$P(Y=0) = P(Y=1) = \frac{1}{2}$$

$$\begin{aligned}
H(Y|X) &= \mathcal{E}_x[H(Y|X=x)] \quad (\text{Erwartungswert}) \\
&= P(X=0) \cdot H(Y|X=0) + P(X=1) \cdot H(Y|X=1) \\
&= P(X=0) \cdot H(P(Y=0|X=0), P(Y=1|X=0)) \\
&\quad + P(X=1) \cdot H(P(Y=0|X=1), P(Y=1|X=1)) \\
&= \frac{1}{2} \cdot H(1-\beta, \beta) + \frac{1}{2} \cdot H(\beta, 1-\beta) \\
&= H(\beta, 1-\beta) \\
&= -\beta \log \beta - (1-\beta) \log(1-\beta)
\end{aligned}$$

Aufgabe 32

- a) Da dieser Informationskanal kein Gedächtnis (also keine verschiedenen Zustände) hat, geschieht die Übertragung der Bits unabhängig voneinander. Die Wahrscheinlichkeiten multiplizieren sich also:

$$\begin{aligned}
P(1001|1100) &= P(1|1) \cdot P(0|1) \cdot P(0|0) \cdot P(1|0) \\
&= 90\% \cdot 10\% \cdot 80\% \cdot 20\% = 1,44\%
\end{aligned}$$

- b) Die Quelleninformation erreicht (wie jede andere binäre Information auch) genau dann ihr Maximum von $H(X) = 1$ bit, wenn beide Zeichen, 0 und 1, gleich oft auftreten. Das heißt, wenn

$$p_0 := P(X=0) = \frac{1}{2} = P(X=1) =: p_1$$

ist. Dann gilt für den Empfang Y der Erwartungswert

$$P(Y=y) = \sum_x \underbrace{p_x}_{\text{Gewicht}} \cdot \underbrace{P(Y=y|X=x)}_{\text{Wert}}$$

Mit den konkreten Werten für die p_x erhalten wir:

$$P(Y=0) = \frac{1}{2} \cdot 80\% + \frac{1}{2} \cdot 10\% = \frac{1}{2} \cdot 90\% = 45\%$$

$$P(Y=1) = \frac{1}{2} \cdot 20\% + \frac{1}{2} \cdot 90\% = \frac{1}{2} \cdot 110\% = 55\%$$

Die empfangene Information $H(Y)$ ist also

$$H(Y) = H(45\%, 55\%) \approx 0,9928 \text{ bit}$$

- c)

$$H(Y) = 1 \text{ bit} \quad \Leftrightarrow \quad P(Y=0) = P(Y=1) = \frac{1}{2}$$

Um daraus nun die Quellenstatistik p_0, p_1 zu ermitteln, müssen wir "rückwärts" rechnen, was auf das folgende lineare Gleichungssystem führt:

$$\frac{1}{2} = P(Y=0) = 80\%p_0 + 10\%p_1$$

$$\frac{1}{2} = P(Y=1) = 20\%p_0 + 90\%p_1$$

Das System ist leicht nach p_1 aufzulösen, indem man das 4-fache der zweiten Gleichung von der ersten subtrahiert:

$$-\frac{3}{2} = 10\%p_1 - 360\%p_1 = -350\%p_1$$

$$p_1 = \frac{3}{2 \cdot 350\%} = \frac{3}{7} \approx 42,9\%$$

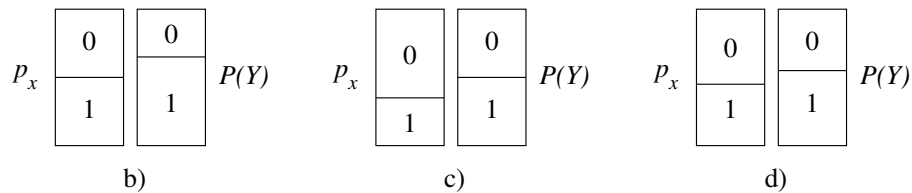
$$p_0 = 1 - p_1 = \frac{4}{7} \approx 57,1\%$$

Und so ist schließlich die Quelleninformation

$$H(X) = H\left(\frac{4}{7}, \frac{3}{7}\right) \approx 0,9852 \text{ bit}$$

d) Aus (b) wissen wir, daß für $p_0 = p_1$ gilt $P(Y=0) < P(Y=1)$.

Aus (c) wissen wir, daß für $P(Y=0) = P(Y=1)$ gilt $p_0 > p_1$.



Offensichtlich⁶ ist für den Kanal sowohl in Teil (b) als auch in Teil (c) $p_0 > P(Y=0)$ und $p_1 < P(Y=1)$.

Daher⁷ muß es eine Quellenstatistik p_0, p_1 geben, für die gilt

$$p_0 = P(Y=1) \quad \text{bzw.} \quad p_1 = P(Y=0)$$

Für diesen Betriebsmodus des Kanals wird dann $H(X) = H(Y)$.

Aus $p_0 = P(Y=1)$ erhält man:

$$\begin{aligned} p_0 &= P(Y=1) = 20\%p_0 + 90\% \overbrace{(1-p_0)}^{p_1} \\ &= 90\% - 70\%p_0 \\ p_0 &= \frac{90\%}{1+70\%} = \frac{9}{17} \approx 52,9\% \\ p_1 &= 1 - p_0 = \frac{8}{17} \approx 47,1\% \end{aligned}$$

Andere Lösung: Der Fall

$$p_0 = P(Y=0) \quad \text{bzw.} \quad p_1 = P(Y=1)$$

⁶weil nämlich die Übertragung der "0" stärker gestört ist (20%) als die Übertragung der "1" (10%)

⁷weil es hier nämlich stetig zugeht; wir können also interpolieren

kann auch eintreten. Aus $p_0 = P(Y=0)$ erhält man:

$$\begin{aligned}
 p_0 &= P(Y=0) = 80\%p_0 + 10\% \overbrace{(1-p_0)}^{p_1} \\
 &= 10\% + 70\%p_0 \\
 p_0 &= \frac{10\%}{1-70\%} = \frac{1}{3} \approx 33,3\% \\
 p_1 &= 1-p_0 = \frac{2}{3} \approx 66,7\%
 \end{aligned}$$

(In der Aufgabenstellung war allerdings nur eine der beiden Lösungen gefragt.)

- e) Die *Gesamtinformation* des Kanals berechnet sich aus den Wahrscheinlichkeiten aller kombinierten⁸ Basisfälle $p(X=x \wedge Y=y) = p(X=x, Y=y)$:

$$\begin{aligned}
 H(X, Y) &= H(p(0,0), p(0,1), p(1,0), p(1,1)) \\
 &= H(80\%p_0, 20\%p_0, 10\%p_1, 90\%p_1)
 \end{aligned}$$

Für die drei verschiedenen Betriebsmodi von b), c) und d) ergibt sich daraus durch Einsetzen:

	b)	c)	d)	d)
p_0	1/2	4/7	9/17	1/3
p_1	1/2	3/7	8/17	2/3
$H(X, Y) \approx$	1,5955 bit	1,5988 bit	1,6004 bit	1,4716 bit

- f) Aus der Gesamtinformation (vgl. Teil (e)) lassen sich leicht die bedingten Informationsgehalte, nämlich Fehlinformation $H(Y|X)$ und Äquivokation $H(X|Y)$, berechnen:

$$H(Y|X) = H(X, Y) - H(X)$$

$$H(X|Y) = H(X, Y) - H(Y)$$

$H(X)$ und $H(Y)$ haben wir bereits in Teil (b) und (c) berechnet; für Teil (d) ergibt sich zunächst:

$$H(X) = H(Y) = H\left(\frac{9}{17}, \frac{8}{17}\right) \approx 0,9975 \text{ bit}$$

bzw. für die alternative Lösung (vgl. (d))

$$H(X) = H(Y) = H\left(\frac{1}{3}, \frac{2}{3}\right) \approx 0,9183 \text{ bit}$$

Gemäß obigen Formeln für die Fehlinformation und die Äquivokation bekommen wir schließlich:

	b)	c)	d)	d)
$H(X)$	1 bit	$\approx 0,9852$ bit	$\approx 0,9975$ bit	$\approx 0,9183$ bit
$H(Y)$	$\approx 0,9928$ bit	1 bit	$\approx 0,9975$ bit	$\approx 0,9183$ bit
$H(X, Y)$	$\approx 1,5955$ bit	$\approx 1,5988$ bit	$\approx 1,6004$ bit	$\approx 1,4716$ bit
$H(Y X)$	$\approx 0,5955$ bit	$\approx 0,6136$ bit	$\approx 0,6029$ bit	$\approx 0,5533$ bit
$H(X Y)$	$\approx 0,6027$ bit	$\approx 0,5988$ bit	$\approx 0,6029$ bit	$\approx 0,5533$ bit

⁸nicht zu verwechseln mit $p(Y=y|X=x)$ oder $p(X=x|Y=y)$! Die Summe dieser *bedingten* Wahrscheinlichkeiten würde ja auch nicht Eins ergeben.

Hier ist deutlich zu sehen, daß trotz der relativ kleinen Fehlerraten des Kanals (10% bis 20%) der Großteil der Information “kaputt geht” im Zuge der – vermeintlichen – Übermittlung:

Die Äquivokation $H(X|Y)$ macht in allen Fällen über 50% der an der Quelle eingespeisten Information $H(X)$ aus; über 50% der Quelleninformation gehen also verloren.

Ebenfalls über 50% (also über die Hälfte!) der empfangenen Information $H(Y)$, nämlich genau die Fehlinformation $H(Y|X)$, ist störungsbedingt, also Unfug; reine Pseudoinformation.

- g) Die *Transinformation* $H(X;Y)$ bezeichnet denjenigen Anteil der Quelleninformation $H(X)$, der die Übertragung “überlebt” und beim Empfänger Y ankommt; also genau denjenigen Anteil der empfangenen Information $H(Y)$, der überhaupt von der Quelle X stammt (und nicht durch Störungen verursacht ist).

Sie läßt sich daher wahlweise durch eine der beiden Formeln berechnen:

$$H(X;Y) = H(X) - H(X|Y)$$

$$H(X;Y) = H(Y) - H(Y|X)$$

Mithilfe der schon in Teil (f) berechneten Werte erhalten wir:

	b)	c)	d)	d)
$H(X)$	1 bit	$\approx 0,9852$ bit	$\approx 0,9975$ bit	$\approx 0,9183$ bit
$H(X Y) \approx$	0,6027 bit	0,5988 bit	0,6029 bit	0,5533 bit
$H(Y)$	$\approx 0,9928$ bit	1 bit	$\approx 0,9975$ bit	$\approx 0,9183$ bit
$H(Y X) \approx$	0,5955 bit	0,6136 bit	0,6029 bit	0,5533 bit
$H(X;Y) \approx$	0,3973 bit	0,3864 bit	0,3946 bit	0,3650 bit

In allen vier Varianten (Betriebsmodi) werden also pro Bit Nenninformation nur weniger als 0,4 bit Realinformation übertragen.

- h) Experimentell⁹ läßt sich herausbekommen, daß die maximale Transinformation für Quellenstatistiken in der Nähe von

$$\begin{pmatrix} p_0 \\ p_1 \end{pmatrix} \approx \begin{pmatrix} 0,4825 \\ 0,5175 \end{pmatrix}$$

erreicht wird. Diese *maximale erreichbare Transinformation*

$$H(X;Y) \approx 0,39775 \text{ bit}$$

ist damit die Kapazität C des Übertragungskanals:

$$\begin{aligned} C &= \max_{p \in (0,1)} H(X;Y) \\ &= \max_{p \in (0,1)} [H(X) - H(X|Y)] \\ &= \max_{p \in (0,1)} [H(X) + H(Y) - H(X,Y)] \end{aligned}$$

⁹d. h. numerisch, also: wahlweise durch eine Schleife, die p_0 bzw. p_1 im Intervall $(0,1)$ fein genug abtastet, oder durch eine Intervallschachtelung (was eleganter ist, aber nicht so schnell zu programmieren)

$$= \max_{p \in (0,1)} \left[- \left(p \cdot \log_2 p + (1-p) \cdot \log_2(1-p) \right) - \right. \\ \left. \left((80\%p+10\%(1-p)) \cdot \log_2 (80\%p+10\%(1-p)) + (20\%p+90\%(1-p)) \cdot \log_2 (20\%p+90\%(1-p)) \right) + \right. \\ \left. \left(80\%p \cdot \log_2(80\%p) + 20\%p \cdot \log_2(20\%p) + 10\%(1-p) \cdot \log_2(10\%(1-p)) + 90\%(1-p) \cdot \log_2(90\%(1-p)) \right) \right] \text{ bit}$$

(Sobald wir den Logarithmus zu einer bestimmten Basis ziehen, haben wir uns damit auf eine Informationseinheit festgelegt. Da hier die Basis = 2 ist, wird in "bit" gemessen.)

Aufgabe 33

- a) Wir betrachten zunächst Wahrscheinlichkeiten für den Wert der Zufallsvariablen $X_t =: X$ und stellen fest:

$$P(X=0) = \frac{3}{4} \quad P(X=1) = \frac{1}{4}$$

Mit dieser stochastischen Bestimmung ist zugleich der Informationsgehalt der Quelle Q ("ohne Gedächtnis", d.h. ohne Wissen über X_{t-1}, X_{t-2}, \dots) gegeben:

$$H(X) = H(P(X=0), P(X=1)) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} \approx 0,811 \text{ bit}$$

Dies ist somit der durchschnittliche Informationswert pro emittiertem Zeichen X_t .

- b) Ist das vorhergehende Zeichen X_{t-1} (das wir nun der Einfachheit halber Y nennen) bekannt, so ändern sich – angesichts der Periodizität der Folge – die Wahrscheinlichkeiten für $X_t =: X$ (die ja nun bedingte sind):

$$P(X=0 | Y=0) = \frac{2}{3} \quad P(X=0 | Y=1) = 1$$

$$P(X=1 | Y=0) = \frac{1}{3} \quad P(X=1 | Y=1) = 0$$

Damit ändert sich aber auch der Informationsgehalt. (Bereits *vor* jeder Rechnung ist anzunehmen, daß mit dem Vorwissen über Y die wirklich neue Information, die X liefert, kleiner geworden sein muß.)

$$\underbrace{H(X|Y)}_{\text{Erwartungswert}} = \sum_y \underbrace{P(Y=y)}_{\text{Gewicht}} \cdot \underbrace{H(X|Y=y)}_{\text{Wert}} \\ = P(Y=0)H(X|Y=0) + P(Y=1)H(X|Y=1) \\ = \frac{3}{4}H(X|Y=0) + \frac{1}{4}H(X|Y=1) \\ = \frac{3}{4}H\left(\frac{2}{3}, \frac{1}{3}\right) + \frac{1}{4}\underbrace{H(1,0)}_{=0} \\ = \frac{3}{4}\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right) \approx 0,689 \text{ bit}$$

Dieser bedingte Informationsgehalt $H(X|Y) \approx 0,689$ bit ist, wie ja zu erwarten war, signifikant kleiner als $H(X) \approx 0,811$ bit. Das Wissen über Y hat sich also auszahlt.

- c) Die beiden vorausgehenden Zeichen (X_{t-2}, X_{t-1}) fassen wir als neue Zufallsvariable Y zusammen. Analog der Vorgehensweise in Teil (b) betrachten wir zunächst die bedingten Wahrscheinlichkeiten für alle möglichen¹⁰ Fälle:

$$P(X=0|Y=(0,0)) = \frac{1}{2} \quad P(X=0|Y=(0,1)) = 1 \quad P(X=0|Y=(1,0)) = 1$$

$$P(X=1|Y=(0,0)) = \frac{1}{2} \quad P(X=1|Y=(0,1)) = 0 \quad P(X=1|Y=(1,0)) = 0$$

Wir gehen analog wie in (b) vor und ermitteln den bedingten Informationsgehalt $H(X|Y)$ als Erwartungswert der konkret gegebenen Fälle für das Vorwissen Y :

$$\begin{aligned} H(X|Y) &= \sum_y \underbrace{P(Y=y)}_{\text{Gewicht}} \cdot \underbrace{H(X|Y=y)}_{\text{Wert}} \\ &= \frac{1}{2}H(X|Y=(0,0)) + \frac{1}{4}H(X|Y=(0,1)) + \frac{1}{4}H(X|Y=(1,0)) \\ &= \frac{1}{2}H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{4}H(1,0) + \frac{1}{4}H(1,0) \\ &= \frac{1}{2} \cdot 1 \text{ bit} + \frac{1}{4} \cdot 0 \text{ bit} + \frac{1}{4} \cdot 0 \text{ bit} = \frac{1}{2} \text{ bit} \end{aligned}$$

Durch weiteres Vorwissen über X_{t-2} ist die neue Information von X gegenüber (b) also – wie zu erwarten gewesen – weiter gefallen.

- d) Diesmal ist in der Zufallsvariablen Y das Wissen über die drei letzten Zustände $(X_{t-3}, X_{t-2}, X_{t-1})$ enthalten. Die bedingten Wahrscheinlichkeiten von X stellen sich nun wie folgt dar:¹¹

$$P(X=0|Y=(0,0,0)) = 0 \quad P(X=0|Y=(0,0,1)) = 1$$

$$P(X=1|Y=(0,0,0)) = 1 \quad P(X=1|Y=(0,0,1)) = 0$$

$$P(X=0|Y=(0,1,0)) = 1 \quad P(X=0|Y=(1,0,0)) = 1$$

$$P(X=1|Y=(0,1,0)) = 0 \quad P(X=1|Y=(1,0,0)) = 0$$

(Bereits in dieser Aufstellung ist abzulesen, daß nichts dem Zufall überlassen bleibt.) Die formale Berechnung der bedingten Information über X läuft folgendermaßen:

$$\begin{aligned} H(X|Y) &= \sum_y \underbrace{P(Y=y)}_{\text{Gewicht}} \cdot \underbrace{H(X|Y=y)}_{\text{Wert}} \\ &= \frac{1}{4}H(X|Y=(0,0,0)) + \frac{1}{4}H(X|Y=(0,0,1)) \\ &\quad + \frac{1}{4}H(X|Y=(0,1,0)) + \frac{1}{4}H(X|Y=(1,0,0)) \\ &= \frac{1}{4}H(0,1) + \frac{1}{4}H(1,0) + \frac{1}{4}H(1,0) + \frac{1}{4}H(1,0) \\ &= 0 \end{aligned}$$

Wie schon geahnt (und vom “gesunden” Menschenverstand nahegelegt), ist der Informationsgehalt dieser periodischen Quelle beim Wissen über genügend Vorgeschichte gleich Null.

¹⁰der nicht mögliche Fall $Y = (1,1)$ ist aus Gründen der Übersicht weggelassen; er hat eine Wahrscheinlichkeit = 0 und ist daher irrelevant

¹¹wieder sind hier die nicht möglichen Fälle (Wahrscheinlichkeit = 0) weggelassen

Noch einmal zur Übersicht der bedingte Informationsgehalt von Zeichen $X = X_t$ der Quelle Q in Abhängigkeit vom Vorwissen $Y = (X_{t-k}, \dots, X_{t-1})$:

k	$H(X Y)$
0	0,811 bit
1	0,689 bit
2	0,500 bit
3	0,000 bit