Informationstheorie, FS 2017 Dr. Luis Haug
ETH Zürich, Dept. Informatik Nico Gorbach

Lösungen zu Übungsblatt 8

8.1

Für jede Verteilung P_X gilt

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_{i=1}^{3} P_X(x_i) H(Y|X = x_i) \\ &= H(Y) - ((1 - P(x_2))H(\tfrac{1}{3}, \tfrac{2}{3}) + P(x_2)\log 3) \\ &= H(Y) - H(\tfrac{1}{3}, \tfrac{2}{3}) - P(x_2)(\log 3 - H(\tfrac{1}{3}, \tfrac{2}{3})) \\ &\le \log 3 - H(\tfrac{1}{3}, \tfrac{2}{3}). \end{aligned}$$

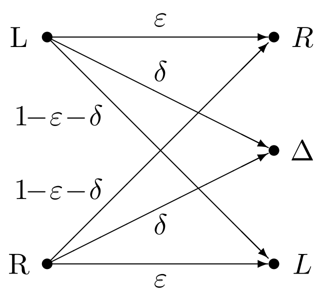
Wenn $P_X(x_2) = 0$ ist, verschwindet der letzte Term in der vorletzten Zeile; wenn zudem $P_X(x_1) = P_X(x_3) = \frac{1}{2}$ gilt, dann ist Y gleichverteilt und $H(Y) = \log 3$. Diese Verteilung maximiert also $I(X;Y)$, und damit ist die Kapazität des Kanals

$$C = \log 3 - H(\tfrac{1}{3}, \tfrac{2}{3}) = \tfrac{2}{3}.$$

Intuitiv leuchtet das ein, denn gegeben $X = x_2$ ist Y gleichverteilt, d.h. durch das Senden von x_2 wird keine Information übertragen.

8.2

a) Der Kanal zwischen Alice und Bob, den wir als *binären symmetrischen Auslöschungskanal* bezeichnen, kann folgendermassen dargestellt werden:



b) Um $I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x H(Y|X = x)P(x)$ zu maximieren, müssen wir hier nur $H(Y)$ maximieren, da $H(Y|X = x)$ unabhängig von x ist. Folgende Überlegungen zeigen, dass $H(Y)$ maximal ist, X gleichverteilt ist. Wir führen eine Zufallsvariable S ein:

$$S = \begin{cases} 0 & \text{falls } Y = \Delta \\ 1 & \text{falls } Y \neq \Delta. \end{cases}$$

Dann ist

$$H(Y) = H(YS) = H(Y|S) + H(S) = P_S(0)H(Y|S=0) + P_S(1)H(Y|S=1) + H(S)$$

Die erste Gleichung gilt, weil S durch Y vollständig bestimmt ist. Wegen $H(Y|S=0) = 0$ erhalten wir

$$H(Y) = P_S(1)H(Y|S=1) + H(S) = (1-\delta)H(Y|S=1) + H(S)$$

$H(Y)$ ist also maximal genau dann maximal, falls $H(Y|S=1)$ maximal ist (denn $(1-\delta)$ und $H(S)$ hängen nicht von der Verteilung von X ab). Dies wird erreicht, wenn $Y = L$ und $Y = R$ gleich wahrscheinlich sind. Dies wiederum ist dann der Fall, wenn $P_X(L) = \frac{1}{2}$ und $P_X(R) = \frac{1}{2}$.
Deshalb ist:

$$C = H\left(\tfrac{1}{2}(1-\delta), \delta, \tfrac{1}{2}(1-\delta)\right) - H\left(\varepsilon, \delta, 1-\varepsilon-\delta\right).$$

c) Die Rate eines Codes mit 4 Codewörtern mit einer Länge von je 15 bits ist $R = \frac{\log 4}{15} = \frac{2}{15}$. Für ausreichend kleine ε und δ gilt $R < C$. Gemäss dem Kanalcodierungstheorem kann Alice einen Code mit Codewortlänge $15N$ für die N Antworten finden, so dass die Wahrscheinlichkeit eines Decodierfehlers für Bob beliebig klein ist, falls N genügend gross gewählt wird.

8.3

Die Kanäle sind in Abbildung 1 illustriert. Man beachte, dass dies nicht dasselbe ist wie ein Kanal mit $|\mathcal{X}_1| + |\mathcal{X}_2|$ Input- und $|\mathcal{Y}_1| + |\mathcal{Y}_2|$ Outputsymbolen.

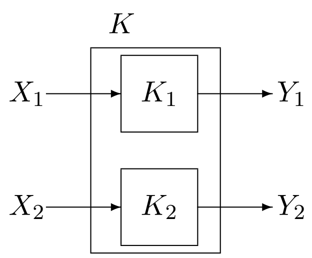


Abbildung 1: Zwei Kanäle in Parallelschaltung

a) Der Kanal K wird definiert durch die durch $P(y_1, y_2|x_1, x_2) = P(y_1|x_1)P(y_2|x_2)$ gegebene Familie von bedingten Verteilungen. Die gemeinsame Verteilung von X_1, X_2, Y_1, Y_2 erfüllt daher $P(x_1, x_2, y_1, y_2) = P(x_1, x_2)P(y_1|x_1)P(y_2|x_2)$. Daraus folgt insbesondere, dass $P(x_2, y_1|x_1) = P(x_2|x_1)P(y_1|x_1)$ ist, d.h. $X_2 \to X_1 \to Y_1$ ist eine Markovkette und somit $I(X_2; Y_1|X_1) = 0$, was äquivalent zu $I(X_1, X_2; Y_1) = I(X_1; Y_1)$ ist. Damit erhalten wir zunächst

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\ &= I(X_1; Y_1) + I(X_1, X_2; Y_2|Y_1), \end{aligned}$$

wobei im ersten Schritt die Kettenregel zum Einsatz kam. Zu zeigen ist also noch, dass $I(X_1, X_2; Y_2|Y_1) = I(X_2; Y_2) - I(Y_1; Y_2)$ gilt; das folgt aus

$$\begin{aligned} I(X_1, X_2; Y_2|Y_1) + I(Y_1; Y_2) &= I(X_1, X_2, Y_1; Y_2) \\ &= I(X_1, Y_1; Y_2|X_2) + I(X_2; Y_2) \\ &= I(X_2; Y_2), \end{aligned}$$

wobei wir in den ersten beiden Schritten die Kettenregel verwendet haben und im letzten Schritt die Tatsache, dass $(X_1, Y_1) \to X_2 \to Y_2$ eine Markovkette ist (das sieht man wiederum anhand der obigen Formel für die gemeinsame Verteilung).

b) Mit a) folgt

$$\begin{aligned} C &= \max_{P_{X_1 X_2}} I(X_1, X_2; Y_1, Y_2) \\ &= \max_{P_{X_1 X_2}} I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &\leq \max_{P_{X_1 X_2}} I(X_1; Y_1) + I(X_2; Y_2) \\ &\leq \max_{P_{X_1 X_2}} I(X_1; Y_1) + \max_{P_{X_1 X_2}} I(X_2; Y_2) \\ &= C_1 + C_2. \end{aligned}$$

Andererseits folgt mit a) auch

$$\begin{aligned} C &= \max_{P_{X_1 X_2}} I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &\geq \max_{P_{X_1} P_{X_2}} I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &= \max_{P_{X_1}} I(X_1; Y_1) + \max_{P_{X_2}} I(X_2; Y_2) \\ &= C_1 + C_2. \end{aligned}$$

Dabei haben wir ausgenutzt, dass das Maximum nur kleiner werden kann, wenn wir anstatt über alle Verteilungen $P_{X_1 X_2}$ nur über diejenigen Verteilungen maximieren, bei denen X_1 und X_2 unabhängig sind. In diesem Fall sind auch Y_1 und Y_2 unabhängig und damit ist $I(Y_1, Y_2) = 0$.

Insgesamt folgt also $C = C_1 + C_2$.