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

Lösungen zu Übungsblatt 3

3.1

Z.B. sind die Codes mit den Codewörtern $\{0, 01, 11\}$ bzw. $\{01, 10, 11, 100\}$ nicht präfixfrei aber eindeutig decodierbar.

3.2

Wir modellieren das Problem als Wahrscheinlichkeitsraum $(\Omega = \{1^-, 1^+, \ldots, 12^-, 12^+\}, P)$, wobei die Elementarereignisse n^-, n^+ bedeuten, dass Ball n leichter bzw. schwerer als alle anderen ist, und mit der Gleichverteilung P. Für $I, J \subset \{1, \ldots, 12\}$ mit $I \cap J = \emptyset$ und $|I| = |J|$ können wir das Abwägen der Bälle $(B_i)_{i \in I}$ gegen die Bälle $(B_j)_{j \in J}$ als Zufallsvariable $X_{I,J} : \Omega \to \{-1, 0, 1\}$ betrachten, deren Werte den Ergebnissen "die $(B_i)_{i \in I}$ sind schwerer", "die $(B_j)_{j \in J}$ sind schwerer" und "beide sind gleich schwer" entsprechen.

Wir untersuchen zunächst, durch welche Wahl von I, J sich beim ersten Wägen die Unsicherheit darüber, welches Elementarereignis eingetreten ist, maximal reduzieren lässt. Die durchschnittliche Reduktion dieser Unsicherheit ist die Entropie $H(X_{I,J})$[1], und deswegen sollte die Verteilung von $X_{I,J}$ möglichst nah an einer Gleichverteilung sein. Mit $n := |I| = |J|$ gilt $P_{X_{I,J}}(-1) = \frac{n}{12} = P_{X_{I,J}}(1)$ und $P_{X_{I,J}}(0) = 1 - \frac{n}{6}$ (z.B. tritt $X_{I,J} = -1$ genau dann ein, wenn entweder einer der B_i schwerer oder einer der B_j leichter ist als alle anderen, also in $2n$ von 24 Fällen), was für $n = 4$ tatsächlich eine Gleichverteilung ist. Eine optimale Strategie für den ersten Schritt ist also das Abwägen von zwei Sets von vier Bällen gegeneinander. Die weiteren Schritte kann man sich analog überlegen. (Siehe [MacKay, Kapitel 4] für die komplette Beschreibung einer optimalen Strategie.)

Wir können eine Strategie, an deren Ende feststeht, welcher Ball schwerer oder leichter ist als alle anderen, als ternären Code $\Omega \to \{-1, 0, 1\}^*$ auffassen, der jedes Elementarereignis auf die Folge der Ergebnisse der jeweiligen Wägungen abbildet. Eine optimale Strategie entspricht einem Code mit minimaler erwarteter Länge.

3.3

Wir verwenden vollständige Induktion, wobei der Fall $n = 2$ schon aus der Vorlesung bekannt ist. Angenommen, wir haben die allgemeine Kettenregel für gegebenes $n - 1 \in \mathbb{N}$ bewiesen;

[1] Die Unsicherheit zu Beginn ist nämlich $H(P) = H(\mathrm{id}_\Omega)$, und damit ist die Reduktion der Unsicherheit gleich $H(\mathrm{id}_\Omega) - H(\mathrm{id}_\Omega | X_{I,J}) = H(\mathrm{id}_\Omega, X_{I,J}) - H(\mathrm{id}_\Omega | X_{I,J}) = H(X_{I,J})$.

indem wir $Y = (X_1, \ldots, X_{n-1})$ setzen, erhalten wir

$$\begin{aligned}
H(X_1, \ldots, X_n) &= H(Y, X_n) \\
&= H(Y) + H(X_n|Y) \\
&= H(X_1, \ldots, X_{n-1}) + H(X_n|X_1, \ldots, X_{n-1}) \\
&= H(X_1) + \sum_{i=2}^{n-1} H(X_i|X_1, \ldots, X_{i-1}) + H(X_n|X_1, \ldots, X_{n-1}) \\
&= H(X_1) + \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}),
\end{aligned}$$

unter Verwendung der schon bewiesenen Kettenregel für 2 und $n-1$ Zufallsvariablen.

Falls die Zufallsvariablen $X_1, \ldots, X_n$ *unabhängig* sind, sind die bedingten Entropien $H(X_i|X_1, \ldots, X_{i-1}) = H(X_i)$, und damit folgt $H(X_1, \ldots, X_n) = \sum_{i=1}^n H(X_i)$.

3.4

Indem wir die Definition $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$ mit Hilfe der Kettenregel umformen, erhalten wir

$$I(X;Y|Z) = H(X,Z) - H(Z) - H(X,Y,Z) + H(Y,Z),$$

was symmetrisch in X und Y ist.

Für gegebenes z mit $P_Z(z) > 0$ bezeichen wir mit X_z und Y_z die Zufallsvariablen, die wir durch Einschränkung von X und Y auf das Ereignis $\{Z = z\}$ erhalten (dieses ist mit der Verteilung $P(\cdot|Z = z)$ ein Wahrscheinlichkeitsraum). Damit gilt $H(X|Z) = E_z(X_z)$ und $H(X|Y,Z) = E_z(H(X_z|Y_z))$, und damit $I(X;Y|Z) = E_z(H(X_z) - H(X_z|Y_z)) = E_z(I(X_z;Y_z))$. Da für alle z die gegenseitige Information $I(X_z;Y_z) \geq 0$ ist (siehe Vorlesung), folgt $I(X;Y|Z) \geq 0$. (Alternativ kann man die Positivität natürlich auch durch Ausschreiben der bedingten Entropien und Anwendung der Jensen-Ungleichung zeigen.)

3.5

a) Ähnlich wie in der vorigen Aufgabe bezeichnen wir für gegebenes y mit $P_Y(y) > 0$ mit X_y bzw. Z_y die Zufallsvariablen, die wir durch Einschränken von X bzw. Z auf $\{Y = y\}$ erhalten. Es gilt $H(Z|Y) = E_y(H(Z_y))$ und $H(Z|X,Y) = E_y(H(Z_y|X_y))$, und somit

$$I(X;Z|Y) = E_y(H(Z_y) - H(Z_y|X_y)) = E_y(I(X_y;Z_y)).$$

Daran sieht man, dass $I(X;Z|Y)$ genau dann verschwindet (also $= 0$ ist), wenn $I(X_y;Z_y)$ für alle y mit $P_Y(y) > 0$ verschwindet; das wiederum ist genau dann der Fall, wenn X_y und Z_y für alle y mit $P_Y(y) > 0$ unabhängig sind, und das ist äquivalent zu $P(Z = z|X = x, Y = y) = P(Z = z|Y = y)$ für alle x, y, z mit $P_Y(y) > 0$, d.h. dazu, dass X, Y, Z eine Markovkette $X \to Y \to Z$ bilden.

b) Es gilt allgemein folgende *Kettenregel für die wechselseitige Information*:

$$\begin{aligned}
I(X;Y,Z) &= H(X) - H(X|Y,Z) \\
&= H(X) - H(X|Y) + H(X|Y) - H(X|Y,Z) \\
&= I(X;Y) + I(X;Z|Y);
\end{aligned}$$

durch Vertauschung der Rollen von Y und Z ergibt sich auf die gleiche Weise

$$I(X; Y, Z) = I(X; Z) + I(X; Y|Z).$$

Im Fall einer Markovkette $X \to Y \to Z$ gilt $I(X; Z|Y) = 0$, und damit ergibt sich aus der ersten Anwendung der Kettenregel $I(X; Y, Z) = I(X; Y)$. Zusammen mit der zweiten Anwendung der Kettenregel folgt

$$I(X; Y) = I(X; Z) + I(X; Y|Z) \geq I(X; Z),$$

da $I(X; Y|Z) \geq 0$.

c) Die Zufallsvariablen X, Y und $f(Y)$ bilden eine Markovkette $X \to Y \to f(Y)$, denn $P(f(Y) = z|X = x, Y = y) = P(f(Y) = z|Y = y)$ für alle x, y, z (beide Ausdrücke sind 1 falls $f(y) = z$ und 0 sonst). Daher folgt $I(X; f(Y)) \leq I(X; Y)$. Intuitiv ist klar, dass bei Anwendung einer Funktion (dem "Verarbeiten von Daten", vgl. die Bezeichnung "data processing inequality") höchstens Information verloren geht.