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

Lösungen zu Übungsblatt 10

10.1

Es gilt

$$\begin{aligned}
I(X^N;Y^N) &= H(Y^N) - H(Y^N|X^N) \\
&= H(Y^N) - \sum_{i=1}^{N} H(Y_i|Y_1,\dots,Y_{i-1},X^N) \\
&= H(Y^N) - \sum_{i=1}^{N} H(Y_i|X_i) \\
&\leq \sum_{i=1}^{N} H(Y_i) - \sum_{i=1}^{N} H(Y_i|X_i) \\
&= \sum_{i=1}^{N} I(X_i;Y_i) \\
&\leq NC.
\end{aligned}$$

Dabei wurde im dritten Schritt verwendet, dass Y_i bedingt auf X_i unabhängig vom Vektor der restlichen X_j, Y_j, $j \neq i$, ist (der Kanal ist gedächtnislos).

Wählt man für P_{X^N} die Produktverteilung $P_{X^N}(x_1 \dots x_N) = \prod_{i=1}^{N} P_X(x_i)$ für eine Verteilung P_X, die die Kapazität C, erreicht, dann werden die beide Ungleichungen oben zu Gleichungen (für das erste $\leq$ ist das der Fall, weil die X_i und damit auch die Y_i dann unabhängig sind) und es gilt $I(X^N;Y^N) = NC$. Also ist die Kapazität der N-ten Erweiterung gleich NC, bzw. C gemessen in bits pro Übertragung der ursprünglichen Kanals.

10.2

Gemäss Kanalcodierungstheorem können wir für den gegebenen Kanal und für gegebenes $\tilde{\varepsilon}$ ein N und einen (M,N)-Code finden, so dass die Rate $R = \frac{\log M}{N} \geq C - \tilde{\varepsilon}$ ist und gleichzeitig die maximale Blockfehlerwahrscheinlichkeit $\lambda_{\max} \leq \tilde{\varepsilon}$ ist. Der Code besteht aus einem Codierer $f : \{1,\dots,M\} \to \mathcal{X}^N$ und einem Decodierer $g : \mathcal{Y}^N \to \{1,\dots,M\}$. Wir definieren nun eine Familie von bedingten Verteilung auf $\mathcal{S} := \{1,\dots,M\}$ durch

$$P'(\hat{s}|s) = P(g(Y) = \hat{s}|X = f(s)); \tag{1}$$

auf der rechten Seite steht dabei die bedingten Verteilungen des ursprünglichen Kanals. Dies ist gerade die Wahrscheinlichkeit, dass beim Senden der Nachricht s die Nachricht $\hat{s}$ decodiert wird.

Wir betrachten nun den Kanal mit Ein- und Ausgabealphabet $\mathcal{S}$ und der durch (1) gegebenen Familie von bedingten Verteilungen. Für eine auf $\mathcal{S}$ gleichverteilte Zufallsvariable S erhalten wir durch Anwendung der Fano-Ungleichung

$$H(S|\hat{S}) = \frac{1}{M}\sum_{\hat{s}=1}^{M} H(S|\hat{S}=\hat{s}) \leq H(P_e, 1-P_e) + P_e \log M,$$

wobei $P_e = P(\hat{S} \neq S)$. Damit folgt

$$\begin{aligned} I(S;\hat{S}) &= H(S) - H(S|\hat{S}) \\ &\geq \log M - P_e \log M - H(P_e, 1-P_e) \\ &= (1-P_e)\log M - H(P_e, 1-P_e) \\ &\geq (1-\tilde{\varepsilon})\log M - H(\tilde{\varepsilon}, 1-\tilde{\varepsilon}), \end{aligned}$$

denn $P_e \leq \lambda_{\max} \leq \tilde{\varepsilon}$, und somit

$$\begin{aligned} \frac{I(S;\hat{S})}{N} &\geq (1-\tilde{\varepsilon})\frac{\log M}{N} - \frac{1}{N}H(\tilde{\varepsilon}, 1-\varepsilon) \\ &\geq (1-\tilde{\varepsilon})(C-\tilde{\varepsilon}) - \frac{1}{N}H(\tilde{\varepsilon}, 1-\tilde{\varepsilon}). \end{aligned}$$

Indem wir $\tilde{\varepsilon}$ ausreichend klein machen, können wir für jedes $\varepsilon > 0$ erreichen, dass

$$\frac{I(S;\hat{S})}{N} \geq C - \varepsilon$$

gilt. Wenn C' die Kapazität des neuen Kanals bezeichnet, folgt

$$\frac{C'}{N} \geq C - \varepsilon$$

für die Kapazität des neuen Kanals pro Übertragung des ursprünglichen Kanals. Es gilt ausserdem für alle $\hat{s}$

$$H(S|\hat{S}=\hat{s}) = -P(S=\hat{s}|\hat{S}=\hat{s})\log P(S=\hat{s}|\hat{S}=\hat{s}) - \sum_{s\neq\hat{s}} P(S=s|\hat{S}=\hat{s})\log P(S=s|\hat{S}=\hat{s});$$

wegen $P(S=\hat{s}|\hat{S}=\hat{s}) \geq 1-\lambda_{\max}$ und $P(S=s|\hat{S}=\hat{s}) \leq \lambda_{\max}$ für $s \neq \hat{s}$ und weil die Funktion $p \mapsto -p\log p$ in der Nähe von $p=0$ monoton wächst und in der Nähe von $p=1$ monoton fällt, können wir den ersten Summanden durch $-(1-\lambda_{\max})P(1-\lambda_{\max})$ und die restlichen Summanden jeweils durch $-\lambda_{\max}\log\lambda_{\max}$ abschätzen und erhalten damit

$$\begin{aligned} \max_{\hat{s}} H(S|\hat{S}=\hat{s}) &\leq -(1-\lambda_{\max})\log(1-\lambda_{\max}) - (M-1)\lambda_{\max}\log\lambda_{\max} \\ &\leq -(1-\tilde{\varepsilon})\log(1-\tilde{\varepsilon}) - (M-1)\tilde{\varepsilon}\log\tilde{\varepsilon}, \end{aligned}$$

was ebenfalls für $\tilde{\varepsilon} \to 0$ beliebig klein wird und damit den neuen Kanal annähernd rauschfrei macht.

10.3

Wir zeigen induktiv, dass $P(X_n \neq X_0) = \frac{1}{2}(1-(1-2\varepsilon)^n)$ gilt. Für $n = 1$ ist das wegen $\frac{1}{2}(1-(1-2\varepsilon)) = \varepsilon$ gerade die Voraussetzung, dass X_0 und X_1 Eingabe und Ausgabe eines $BSC(\varepsilon)$ sind. Für den Induktionsschritt berechnen wir

$$\begin{aligned} P(X_{n+1} \neq X_0) &= P(X_{n+1} \neq X_n)P(X_n = X_0) + P(X_{n+1} = X_n)P(X_n \neq X_0) \\ &= \varepsilon(1 - \frac{1}{2}(1-2\varepsilon)^n) + (1-\varepsilon)\frac{1}{2}(1-(1-2\varepsilon)^n) \\ &= \frac{1}{2}(1-(1-2\varepsilon)^{n+1}) \end{aligned}$$

unter Verwendung der Induktionsvoraussetzung im zweiten Schritt.