Informationstheorie, FS 2021
ETH Zürich, Dept. Informatik

Prof. Joachim M. Buhmann
Fabian Laumer

Lösungen zu Übungsblatt 9

9.1 (Fortsetzung von Aufgabe 8.2)

a) Zeigen Sie, dass für jede Zufallsvariable Z mit Werten in $\mathcal{Z}$ die Ungleichung

$$H(Z) \geq -\log P_Z(z^*)$$

gilt, wobei $z^* \in \mathcal{Z}$ ein Wert mit maximaler Wahrscheinlichkeit ist, d.h. $z^* \in \arg\max_{z \in \mathcal{Z}} P_Z(z)$. Folgern Sie, dass in der in Aufgabe 8.2 beschriebenen Situation für $\hat{X} = g_{MAP}(Y)$ und für jedes $y \in \mathcal{Y}$ die Ungleichung

$$H(X|Y = y) \geq -\log P(\hat{X} = X|Y = y)$$

gilt.

Es gilt

$$H(Z) = -\sum_z P_Z(z) \log P_Z(z) \geq -\sum_z P_Z(z) \log P_Z(z^*) = -\log P_Z(z^*),$$

da die Funktion $p \mapsto -\log p$ monoton fällt. Durch Anwendung auf die bedingte Verteilung $P_{X|Y=y}$ und Verwendung der Definition von g_{MAP} folgt

$$H(X|Y = y) \geq -\log P(\hat{X} = X|Y = y).$$

b) Zeigen Sie für $\hat{X} = g_{MAP}(Y)$ die Abschätzung

$$P(\hat{X} \neq X) \leq 1 - 2^{-H(X|Y)}.$$

Verwenden Sie dabei das Ergebnis aus (a) sowie die Ungleichung $\sum_z P_Z(z) 2^{-z} \geq 2^{-\sum_z P_Z(z) z}$ (eine Anwendung der Jensen-Ungleichung), die für jede reellwertige Zufallsvariable Z gilt.

Für $\hat{X} = g_{MAP}(Y)$ gilt

$$\begin{aligned}
P(\hat{X} \neq X) &= \sum_y P(y)(1 - P(g_{MAP}(y)|y)) \\
&= 1 - \sum_y P(y) P(g_{MAP}(y)|y) \\
&= 1 - \sum_y P(y) 2^{\log P(g_{MAP}(y)|y)} \\
&\leq 1 - \sum_y P(y) 2^{-H(X|Y=y)} \\
&\leq 1 - 2^{-\sum_y P(y) H(X|Y=y)} \\
&= 1 - 2^{H(X|Y)},
\end{aligned}$$

wobei die erste Ungleichung aus dem Ergebnis in (b) folgt und die zweite aus der in der Aufgabenstellung angegebenen Anwendung der Jensen-Ungleichung.

9.2

Gegeben sei ein Kanal $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ mit Kapazität C. Seine N-te Erweiterung ist der Kanal $(\mathcal{X}^N, \mathcal{Y}^N, P_{Y^N|X^N})$ mit $P_{Y^N|X^N}(\mathbf{y}, \mathbf{x}) = \prod_{i=1}^N P_{Y|X}(y_i, x_i)$. Zeigen Sie, dass für jede Verteilung P_{X^N} über $\mathcal{X}^N$ die Ungleichung

$$I(X^N; Y^N) \leq NC$$

gilt. Folgern Sie, dass die Kapazität der N-ten Erweiterung gemessen in Bits pro Übertragung des ursprünglichen Kanals gleich C ist.[1]

Es gilt

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

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 $P_{X^N}(x_1 \ldots x_N) = \prod_{i=1}^N P_X(x_i)$ 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.

9.3

Gegeben sei ein Kanal $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ mit Kapazität C. Konstruieren Sie davon ausgehend für gegebenes $\varepsilon > 0$ einen neuen Kanal $(\mathcal{X}', \mathcal{Y}', P_{Y'|X'})$, der annähernd rauschfrei ist (d.h. $H(Y'|X' = x') \approx 0$ für alle $x' \in \mathcal{X}'$) und dessen Kapazität in Bits pro Übertragung des ursprünglichen Kanals $\geq C - \varepsilon$ ist.

Hinweis: Verwenden Sie die Existenz eines Codes mit Rate $\approx C$ und maximaler Blockfehlerwahrscheinlichkeit ≈ 0 für den ursprünglichen Kanal.

[1] Das ist zu Vergleichszwecken das sinnvollere Mass ist als die Kapazität selbst.

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 $f : \{1, \ldots, M\} \to \mathcal{X}^N$ und einem Decodierer $g : \mathcal{Y}^N \to \{1, \ldots, M\}$. Wir definieren nun eine Familie von bedingten Verteilung auf $\mathcal{S} := \{1, \ldots, M\}$ durch

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

auf der rechten Seite steht dabei die bedingte Verteilung des ursprünglichen Kanals. (**??**) ist also 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 (**??**) gegebenen Familie von bedingten Verteilungen. Für Zufallsvariablen $S, \hat{S}$ mit Werten in $\mathcal{S}$, die die Eingabe bzw. Ausgabe modellieren, sagt die 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 $P_e = P(\hat{S} \neq S)$. Falls S gleichverteilt ist, gilt $H(S) = \log M$, und damit

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

denn $P_e \leq \lambda_{\max} \leq \tilde{\varepsilon}$. 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; damit ist der neue Kanal annähernd rauschfrei.