Information Theory I

Fall Semester 2017

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

Model Answers to Exercise 10 of November 22, 2017

http://www.isi.ee.ethz.ch/teaching/courses/it1.html

Problem 1 ***Channel Coding for a "Double"-Channel***

To prove the existence of such a codebook, we proceed along the same lines as for the achievability part of the channel coding theorem:

- Let Q and R be such that $\mathsf{R} < \min\{I(Q, W^{(1)}), I(Q, W^{(2)})\}$, and let $\epsilon' > 0$.
- For a fixed n, generate a random codebook $\mathcal{C}$ of block length n according to Q and reveal it to the encoder and both decoders.
- Use joint typicality decoding for both decoders, which fails if no codeword or more than one codeword is jointly typical with the received sequence. For a fixed message m and for n large enough, the performance analysis leads to

$$\Pr\big[\hat{M}^{(1)} \neq M \,\big|\, M = m\big] \leq \epsilon' + 2^{n\mathsf{R}} \cdot 2^{-n(I(Q,W^{(1)})-3\epsilon')},$$
$$\Pr\big[\hat{M}^{(2)} \neq M \,\big|\, M = m\big] \leq \epsilon' + 2^{n\mathsf{R}} \cdot 2^{-n(I(Q,W^{(2)})-3\epsilon')}.$$

  Averaging the error probabilities over all messages (which corresponds to uniform message probabilities) and applying the union bound leads to

$$\Pr\big[\hat{M}^{(1)} \neq M \text{ or } \hat{M}^{(2)} \neq M\big] \leq 2\epsilon' + 2^{n\mathsf{R}} \cdot 2^{-n(I(Q,W^{(1)})-3\epsilon')} + 2^{n\mathsf{R}} \cdot 2^{-n(I(Q,W^{(2)})-3\epsilon')}.$$

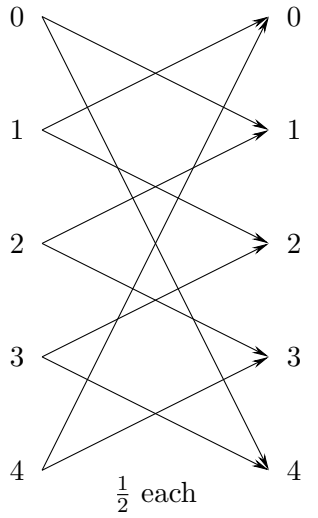
  For an appropriate choice of ϵ' and n large enough, the average probability of error satisfies

$$\Pr\big[\hat{M}^{(1)} \neq M \text{ or } \hat{M}^{(2)} \neq M\big] < \frac{\epsilon}{2} \tag{1}$$

  for any $\epsilon > 0$ because $\mathsf{R} < \min\{I(Q, W^{(1)}), I(Q, W^{(2)})\}$.
- Since the codebook is generated independently of the message, there exists a (deterministic) codebook $\mathcal{C}$ for which (1) holds.
- We obtain a maximal probability of error less than ϵ by throwing away half of the codewords.

Problem 2 ***Zero-Error Capacity***

The channel has the transition matrix $W(y|x)$ shown in Figure 1.



$$W(y|x) = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Figure 1: Transition matrix of the channel.

a) Since the channel is weakly symmetric, its capacity is

$$\mathsf{C} = \log|\mathcal{Y}| - H(\text{row of transition matrix}) = \log_2 5 - 1 = \log_2 2.5 \approx 1.322 \text{ bits}.$$

b) Let us construct a block code consisting of codewords of length 2. We want to achieve a rate greater than 1 bit per channel use, i.e., greater than 2 bits per two channel uses. Thus, we need more than 4 codewords. Let us pick 5 codewords with distinct first symbols: $\{0a, 1b, 2c, 3d, 4e\}$. We must choose a, b, c, d, and e so that the receiver will be able to determine which codeword was transmitted. A simple repetition code will not work, since if, say, 22 is transmitted, then 11 might be received, and the receiver could not tell whether the codeword was 00 or 22. Instead, we use the code $\{00, 13, 21, 34, 42\}$, i.e., the codewords can be described as $\mathbf{x}_i = (i, 3i \mod 5)$ for all $i = 0, \ldots, 4$. Then each codeword will be received as one of 4 possible 2-tuples which are all distinct:

$$\begin{aligned} 00 &\to \{44, 14, 41, 11\} \\ 13 &\to \{02, 22, 04, 24\} \\ 21 &\to \{10, 30, 12, 32\} \\ 34 &\to \{23, 43, 20, 40\} \\ 42 &\to \{31, 01, 33, 03\}. \end{aligned}$$

Since there are 5 possible error-free messages with 2 channel uses, the rate $\frac{1}{2}\log_2 5$ is achievable, and the zero-error capacity of this channel is at least $\frac{1}{2}\log_2 5 \approx 1.161$ bits.

The zero-error capacity of this channel is in fact exactly $\frac{1}{2}\log_2 5$, which was proved by László Lovász in the celebrated paper "On the Shannon capacity of a graph," *IEEE Transactions on Information Theory*, vol. IT-25, no. 1, pp. 1–7, January 1979.

Problem 3 ***An Elementary Converse for the Binary Erasure Channel***

a) Fix an erasure pattern $s^n \in \{0,1\}^n$, and define $\kappa \triangleq \frac{1}{n}\sum_{i=1}^n s_i$. Since exactly $n\kappa$ positions are erased, there are $n(1-\kappa)$ nonerased positions, and the number of different channel output sequences is $2^{n(1-\kappa)}$. The decoder can therefore only guess at most $2^{n(1-\kappa)}$ different messages, and at least $2^{n\mathsf{R}} - 2^{n(1-\kappa)}$ messages cannot be decoded correctly. The message distribution, conditional on $S^n = s^n$, is uniform because the erasures are independent of the channel input sequence. Thus, the probability of error conditional on $S^n = s^n$ satisfies

$$\Pr[E = 1 \,|\, S^n = s^n] \geq \frac{2^{n\mathsf{R}} - 2^{n(1-\kappa)}}{2^{n\mathsf{R}}}.$$

b) Observe that $S_1, \ldots, S_n$ are IID $\sim$ Bernoulli(ρ), thus $\mathsf{E}[S_1] = \rho$. Because $\delta > 0$,

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n S_i < \rho - \delta\right] \leq \Pr\left[\left|\frac{1}{n}\sum_{i=1}^n S_i - \rho\right| > \delta\right]. \tag{2}$$

By the weak law of large numbers, the RHS of (2) tends to zero as n tends to infinity, so the LHS of (2) must also tend to zero as n tends to infinity. Consequently,

$$\lim_{n\to\infty} \Pr\left[\frac{1}{n}\sum_{i=1}^n S_i \geq \rho - \delta\right] = 1.$$

c) Fix $\epsilon > 0$. Fix α so that $1 - \mathsf{R} < \alpha < \rho$ (this is possible since $\mathsf{R} > 1 - \rho$ implies $1 - \mathsf{R} < \rho$). Define the set $\mathcal{A} \triangleq \{s^n \in \{0,1\}^n : \frac{1}{n}\sum_{i=1}^n s_i \geq \alpha\}$. Observe that

$$\begin{aligned}
\Pr[E = 1] &\overset{\text{(i)}}{=} \sum_{s^n \in \{0,1\}^n} \Pr[S^n = s^n] \cdot \Pr[E = 1 \,|\, S^n = s^n] \\
&\geq \sum_{s^n \in \mathcal{A}} \Pr[S^n = s^n] \cdot \Pr[E = 1 \,|\, S^n = s^n] \\
&\overset{\text{(ii)}}{\geq} \sum_{s^n \in \mathcal{A}} \Pr[S^n = s^n] \cdot \frac{2^{n\mathsf{R}} - 2^{n(1-\frac{1}{n}\sum_{i=1}^n s_i)}}{2^{n\mathsf{R}}} \\
&\overset{\text{(iii)}}{\geq} \sum_{s^n \in \mathcal{A}} \Pr[S^n = s^n] \cdot \frac{2^{n\mathsf{R}} - 2^{n(1-\alpha)}}{2^{n\mathsf{R}}} \\
&= \Pr[S^n \in \mathcal{A}] \cdot \left(1 - 2^{n(1-\alpha-\mathsf{R})}\right),
\end{aligned} \tag{3}$$

where (i) follows from the law of total probability; (ii) follows from Part a); and (iii) holds because $\frac{1}{n}\sum_{i=1}^n s_i \geq \alpha$ for all $s^n \in \mathcal{A}$. By Part b), $\Pr[S^n \in \mathcal{A}]$ tends to one as n tends to infinity (define $\delta \triangleq \rho - \alpha > 0$, then $\rho - \delta = \alpha$). Because $1 - \mathsf{R} < \alpha$ implies $1 - \alpha - \mathsf{R} < 0$, $2^{n(1-\alpha-\mathsf{R})}$ tends to zero as n tends to infinity. Thus, the RHS of (3) tends to one as n tends to infinity, so for sufficiently large n,

$$\Pr[E = 1] \geq 1 - \epsilon.$$

Consequently, for sufficiently large n, the average probability of error satisfies

$$\lambda_{\text{avg}} = \frac{1}{2^{n\mathsf{R}}}\sum_{m=1}^{2^{n\mathsf{R}}} \Pr[E = 1 \,|\, M = m] \overset{\text{(i)}}{=} \Pr[E = 1] \geq 1 - \epsilon,$$

where (i) follows from the law of total probability since the message is chosen uniformly at random.

Problem 4 ***Average Bit-Error Probability***

a) The steps can be justified as follows:

(i) Define $M \triangleq U^k$. We have

$$\begin{aligned} I(U^k;\hat{U}^k) &= I(M;\hat{U}^k) \\ &\overset{\text{(A)}}{\leq} I(M;Y^n) \\ &\overset{\text{(B)}}{\leq} n\mathsf{C}, \end{aligned}$$

where (A) follows from the data processing inequality; and (B) was shown in the lecture.

(ii) This follows from the definition of the mutual information.

(iii) This holds because $U_1, \ldots, U_k$ are IID $\sim$ Bernoulli(1/2).

(iv) This follows from the chain rule for entropy.

(v) This holds because conditioning does not increase entropy.

(vi) This follows from Fano's inequality:

$$H(U_i|\hat{U}_i) \leq H_{\mathrm{b}}\big(\Pr[U_i \neq \hat{U}_i]\big) + \Pr[U_i \neq \hat{U}_i] \log(\underbrace{|\mathcal{U}_i|}_{=2} - 1) = H_{\mathrm{b}}\big(\Pr[U_i \neq \hat{U}_i]\big).$$

(vii) This follows from Jensen's inequality since entropy is a concave function.

b) The result from Part a) is equivalent to

$$H_{\mathrm{b}}\left(\frac{1}{k}\sum_{i=1}^{k} \Pr[U_i \neq \hat{U}_i]\right) \geq 1 - \frac{n}{k}\mathsf{C}.$$

Since $H_{\mathrm{b}}(\cdot)$ is invertible on [0, 1/2], we have

$$\frac{1}{k}\sum_{i=1}^{k} \Pr[U_i \neq \hat{U}_i] \geq H_{\mathrm{b}}^{-1}\Big(1 - \frac{n}{k}\mathsf{C}\Big) > 0,$$

where the last inequality holds because $\frac{k}{n} > \mathsf{C}$ by assumption.