# Information Theory I

Fall Semester 2017

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

## Model Answers to Exercise 11 of November 29, 2017

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

---

**Problem 1** ***Properties of*** $\mathsf{R}(\mathsf{D})$

a) For a fixed $P_X$ and a fixed $P_{\hat{X}|X}$, we have

$$\begin{aligned}\mathsf{E}\big[d'(X,\hat{X})\big] &= \mathsf{E}\big[d(X,\hat{X}) - w(X)\big] \\ &\stackrel{\text{(i)}}{=} \mathsf{E}\big[d(X,\hat{X})\big] - \mathsf{E}\big[w(X)\big] \\ &\stackrel{\text{(ii)}}{=} \mathsf{E}\big[d(X,\hat{X})\big] - \bar{w},\end{aligned}$$

where (i) follows from the linearity of expectation; and (ii) follows from the definition of $\bar{w}$. Thus, $\mathsf{E}\big[d'(X,\hat{X})\big] \le \mathsf{D}$ if and only if $\mathsf{E}\big[d(X,\hat{X})\big] \le \mathsf{D} + \bar{w}$, and

$$\mathsf{R}'(\mathsf{D}) = \min_{\substack{P_{\hat{X}|X}: \\ \mathsf{E}[d'(X,\hat{X})] \le \mathsf{D}}} I\big(X;\hat{X}\big) = \min_{\substack{P_{\hat{X}|X}: \\ \mathsf{E}[d(X,\hat{X})] \le \mathsf{D}+\bar{w}}} I\big(X;\hat{X}\big) = \mathsf{R}(\mathsf{D}+\bar{w}).$$

b) For every $x \in \mathcal{X}$, we choose

$$w(x) = \min_{\hat{x}\in\hat{\mathcal{X}}} d(x,\hat{x}).$$

Then, for every $x \in \mathcal{X}$,

$$\min_{\hat{x}\in\hat{\mathcal{X}}} d'(x,\hat{x}) = \min_{\hat{x}\in\hat{\mathcal{X}}} [d(x,\hat{x}) - w(x)] = \min_{\hat{x}\in\hat{\mathcal{X}}} d(x,\hat{x}) - w(x) = 0.$$

Thus, there is no essential loss of generality in assuming that for every $x \in \mathcal{X}$, there exists at least one $\hat{x} \in \hat{\mathcal{X}}$ that reproduces the source with zero distortion: if a distortion measure $d$ does not satisfy this, we can choose $w$ as above so that the modified distortion measure $d'$ satisfies the condition, and we can use Part a) to relate $\mathsf{R}(\mathsf{D})$ to $\mathsf{R}'(\mathsf{D})$.

c) Here, we choose $w(0) = 1$ and $w(1) = 3$. Then, $\bar{w} = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 3 = 2$ and

| $d'(x,\hat{x})$ | $\hat{x}=0$ | $\hat{x}=1$ |
|---|---|---|
| $x=0$ | 0 | 1 |
| $x=1$ | 0 | 1 |

The smallest expected distortion for $P_X$ and $d'$ is zero: $\mathsf{E}[d'(X,\hat{X})]$ cannot be smaller than zero, and $\mathsf{E}[d'(X,\hat{X})]$ is zero if $\hat{X}$ is deterministically zero. If $\hat{X}$ is deterministically zero, we do not need any rate, and we obtain that $\mathsf{R}'(\mathsf{D}) = 0$ for all $\mathsf{D} \ge 0$. Using Part a), we conclude that $\mathsf{R}(\mathsf{D}) = \mathsf{R}'(\mathsf{D} - \bar{w}) = \mathsf{R}'(\mathsf{D} - 2) = 0$ for all $\mathsf{D} \ge 2$. (An expected distortion smaller than 2 for $P_X$ and $d$ is not possible since $\mathsf{E}[d(X,\hat{X})]$ cannot be smaller than two.)

 

**Problem 2** ***Erasure Distortion***

a) We start with the computation of the expected distortion:

$$\begin{aligned}
\mathsf{E}\big[d(X,\hat{X})\big] &= \sum_{x\in\{0,1\}} \sum_{\hat{x}\in\{0,1,?\}} P(x)P(\hat{x}|x)d(x,\hat{x}) \\
&= \frac{1}{2}P_{\hat{X}|X}(0|0)\cdot 0 + \frac{1}{2}P_{\hat{X}|X}(1|0)\cdot\infty + \frac{1}{2}P_{\hat{X}|X}(?|0)\cdot 1 \\
&\quad + \frac{1}{2}P_{\hat{X}|X}(0|1)\cdot\infty + \frac{1}{2}P_{\hat{X}|X}(1|1)\cdot 0 + \frac{1}{2}P_{\hat{X}|X}(?|1)\cdot 1 \\
&= \frac{1}{2}P_{\hat{X}|X}(?|0) + \frac{1}{2}P_{\hat{X}|X}(?|1) \\
&= \Pr\big[\hat{X} = ?\big],
\end{aligned}$$

where we have to set $P_{\hat{X}|X}(1|0) = P_{\hat{X}|X}(0|1) = 0$ since otherwise the expected distortion is infinite. Hence, to find the rate distortion function we have to minimize $I(X;\hat{X})$ subject to the constraint

$$\Pr\big[\hat{X} = ?\big] \leq \mathsf{D}. \tag{1}$$

For $\mathsf{D} \geq 1$, choosing $\hat{X}$ to be deterministically ? satisfies this constraint: $\Pr\big[\hat{X} = ?\big] = 1 \leq \mathsf{D}$. Thus, $R(\mathsf{D}) = 0$ for $\mathsf{D} \geq 1$.

Since $P_{\hat{X}|X}(1|0) = P_{\hat{X}|X}(0|1) = 0$, we know that if $\hat{X} = 0$, then $X = 0$ with probability one, and if $\hat{X} = 1$, then $X = 1$ with probability one. Thus,

$$\begin{aligned}
I\big(X;\hat{X}\big) &= H(X) - H\big(X\big|\hat{X}\big) \\
&= 1 - \Big(\Pr\big[\hat{X}=0\big]\underbrace{H\big(X\big|\hat{X}=0\big)}_{=0} + \Pr\big[\hat{X}=1\big]\underbrace{H\big(X\big|\hat{X}=1\big)}_{=0} \\
&\qquad + \underbrace{\Pr[\hat{X}=?]}_{\leq \mathsf{D}\text{ by (1)}}\underbrace{H\big(X\big|\hat{X}=?\big)}_{\leq 1\text{ bit}}\Big) \\
&\geq 1 - \mathsf{D}.
\end{aligned}$$

For $\mathsf{D} \in [0,1]$, the inequality is met with equality for the choice

$$P(\hat{x}|x) = \begin{cases} 1-\mathsf{D} & \text{if } \hat{x} = x, \\ \mathsf{D} & \text{if } \hat{x} = ?, \\ 0 & \text{otherwise} \end{cases}$$

because then $\Pr\big[\hat{X} = ?\big] = \mathsf{D}$ and $\Pr\big[X = 0 \,\big|\, \hat{X} = ?\big] = \Pr\big[X = 1 \,\big|\, \hat{X} = ?\big] = \frac{1}{2}$.

Hence, the rate distortion function is given as follows (in bits):

$$\mathsf{R}(\mathsf{D}) = \begin{cases} 1-\mathsf{D} & \text{if } 0 \leq \mathsf{D} \leq 1, \\ 0 & \text{if } \mathsf{D} > 1. \end{cases}$$

The rate distortion region, i.e., the set of all achievable pairs $(\mathsf{R},\mathsf{D})$, is depicted in Figure 1.

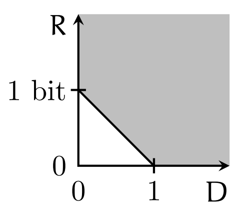


Figure 1: Rate distortion region (light gray).

b) We use the following scheme: for every source sequence of length $n$, choose a codeword that consists of the first $n(1-\mathsf{D})$ source symbols followed by $n\mathsf{D}$ question marks. This code needs $2^{n(1-\mathsf{D})}$ codewords and has therefore a rate of $(1-\mathsf{D})$ bits. The distortion achieved by this code is as follows: the first $n(1-\mathsf{D})$ digits have zero distortion, the rest has distortion 1. Hence, on average we get a distortion of $\mathsf{D}$.

(For $\mathsf{D} \geq 1$, we will only use question marks, i.e., we only use one codeword $\hat{\mathbf{x}} = (?, \ldots, ?)$. This code has zero rate and achieves a distortion of 1.)

## Problem 3 — *Rate Distortion Function with Infinite Distortion*

For a fixed $P_{\hat{X}|X}$, the expected distortion is

$$\begin{aligned}\mathsf{E}[d(X,\hat{X})] &= P_X(0)P_{\hat{X}|X}(0|0)d(0,0) + P_X(0)P_{\hat{X}|X}(1|0)d(0,1)\\ &\quad + P_X(1)P_{\hat{X}|X}(0|1)d(1,0) + P_X(1)P_{\hat{X}|X}(1|1)d(1,1)\\ &= \frac{1}{2}P_{\hat{X}|X}(1|0)\cdot\infty + \frac{1}{2}P_{\hat{X}|X}(0|1).\end{aligned}$$

We need $P_{\hat{X}|X}(1|0) = 0$ since otherwise the expected distortion is infinite. Let $P_{\hat{X}|X}(0|1) = p$ for some $p \in [0,1]$. Then, we have $\mathsf{E}[d(X,\hat{X})] \leq \mathsf{D}$ if and only if $p \leq 2\mathsf{D}$. For $\mathsf{D} \geq \frac{1}{2}$, we can take $\hat{X} = 0$ (which corresponds to $p = 1$); therefore $\mathsf{R}(\mathsf{D}) = 0$ for $\mathsf{D} \geq \frac{1}{2}$.
We now treat the case $\mathsf{D} \in \left[0, \frac{1}{2}\right]$. We have

$$P_{\hat{X}}(1) = P_X(0)\underbrace{P_{\hat{X}|X}(1|0)}_{=0} + \underbrace{P_X(1)}_{=1/2}\underbrace{P_{\hat{X}|X}(1|1)}_{=1-p} = \frac{1-p}{2},$$

so

$$\begin{aligned}I\big(X;\hat{X}\big) &= H\big(\hat{X}\big) - H\big(\hat{X}\big|X\big)\\ &= H_{\mathrm{b}}\left(\frac{1-p}{2}\right) - P_X(0)\underbrace{H\big(\hat{X}\big|X=0\big)}_{=0} - \underbrace{P_X(1)}_{=1/2}\underbrace{H\big(\hat{X}\big|X=1\big)}_{=H_{\mathrm{b}}(p)}\\ &= H_{\mathrm{b}}\left(\frac{1-p}{2}\right) - \frac{1}{2}H_{\mathrm{b}}(p).\end{aligned} \tag{2}$$

We need to minimize $I\big(X;\hat{X}\big)$ subject to $p \in [0, 2\mathsf{D}]$. We prove by contradiction that $p = 2\mathsf{D}$ achieves the minimum. To that end, let $p^* < 2\mathsf{D}$ be a minimizer. Denote by $W_1$ the transition probabilities $P_{\hat{X}|X}$ associated with $p^*$. Denote by $W_2$ the transition probabilities $P_{\hat{X}|X}$ associated with $p = 1$. Consider the mixture $\lambda W_1 + \bar{\lambda} W_2$. Denote by $D_1$ the expected distortion of $W_1$ and by $D_2$ the expected distortion of $W_2$. Since $p^* < 2\mathsf{D}$, $D_1 < \mathsf{D}$, and we have $D_2 = \frac{1}{2}$. The expected distortion of the mixture is $\lambda D_1 + \bar{\lambda} D_2$. Thus, there exists a $\lambda < 1$ such that the expected distortion of the mixture is less than or equal to $\mathsf{D}$. By the convexity of the mutual information in $W$,

$$I(P_X, \lambda W_1 + \bar{\lambda} W_2) \leq \underbrace{\lambda}_{<1} I(P_X, W_1) + \bar{\lambda}\underbrace{I(P_X, W_2)}_{=0} < I(P_X, W_1).$$

The mixture satisfies the distortion constraint and leads to a smaller mutual information, so $p^*$ cannot be a minimizer. Thus, (2) is minimized by $p = 2\mathsf{D}$. Combining these results, we obtain

$$\mathsf{R}(\mathsf{D}) = \begin{cases} H_{\mathrm{b}}\left(\frac{1-2\mathsf{D}}{2}\right) - \frac{1}{2}H_{\mathrm{b}}(2\mathsf{D}) & \text{if } 0 \leq \mathsf{D} \leq \frac{1}{2},\\ 0 & \text{if } \mathsf{D} > \frac{1}{2}.\end{cases}$$

**Problem 4** ***Rate Distortion for Uniform Source with Hamming Distortion***

First, we note that the expected distortion $\mathsf{E}\bigl[d(X,\hat{X})\bigr]$ is equal to $\Pr\bigl[X \neq \hat{X}\bigr]$:

$$\begin{aligned}
\mathsf{E}\bigl[d(X,\hat{X})\bigr] &= \sum_x p(x) \sum_{\hat{x}} p(\hat{x}|x) d(x,\hat{x}) \\
&= \sum_x p(x) \sum_{\hat{x}\neq x} p(\hat{x}|x) \\
&= \Pr\bigl[X \neq \hat{X}\bigr].
\end{aligned}$$

Hence, to find $R(\mathsf{D})$ we have to minimize $I(X;\hat{X})$ subject to the constraint that $\Pr\bigl[X \neq \hat{X}\bigr] \leq \mathsf{D}$. First we note that $R(\mathsf{D}) = 0$ for $\mathsf{D} \geq (m-1)/m$, because if $\hat{X}$ is independent of $X$, then, irrespective of $P_{\hat{X}}$, the expected distortion equals $(m-1)/m$ and the mutual information between $X$ and $\hat{X}$ equals 0. In the following we thus consider the case where $\mathsf{D} < (m-1)/m$. Let $P_{\mathrm{e}} = \Pr\bigl[X \neq \hat{X}\bigr]$ and observe that

$$\begin{aligned}
I\bigl(X;\hat{X}\bigr) &= H\bigl(X\bigr) - H\bigl(X|\hat{X}\bigr) \\
&\overset{\text{(i)}}{\geq} H\bigl(X\bigr) - H_{\mathrm{b}}(P_{\mathrm{e}}) - P_{\mathrm{e}} \log(m-1) \\
&\overset{\text{(ii)}}{\geq} H\bigl(X\bigr) - H_{\mathrm{b}}(\mathsf{D}) - \mathsf{D} \log(m-1) \\
&\overset{\text{(iii)}}{=} \log m - H_{\mathrm{b}}(\mathsf{D}) - \mathsf{D} \log(m-1),
\end{aligned}$$

where (i) follows from Fano's inequality; (ii) holds because $P_{\mathrm{e}} \leq \mathsf{D}$ and because the function $P_{\mathrm{e}} \mapsto H_{\mathrm{b}}(P_{\mathrm{e}}) + P_{\mathrm{e}} \log(m-1)$ is nondecreasing in $P_{\mathrm{e}}$ for $P_{\mathrm{e}} \in \bigl[0, \frac{m-1}{m}\bigr]$ (we will show this later); and (iii) holds because $X$ is uniformly distributed. Choosing $P_{\hat{X}|X}$ to be

$$P_{\hat{X}|X}(\hat{x}|x) = \begin{cases} 1-\mathsf{D} & \text{if } \hat{x} = x, \\ \frac{\mathsf{D}}{m-1} & \text{if } \hat{x} \neq x \end{cases}$$

satisfies $\Pr\bigl[X \neq \hat{X}\bigr] \leq \mathsf{D}$ and achieves equality because then

$$P_{\hat{X}}(\hat{x}) = \sum_x P_{X,\hat{X}}(x,\hat{x}) = \sum_x P_X(x) P_{\hat{X}|X}(\hat{x}|x) = \frac{1}{m}(1-\mathsf{D}) + (m-1)\frac{1}{m}\frac{\mathsf{D}}{m-1} = \frac{1}{m}$$

and

$$P_{X|\hat{X}}(x|\hat{x}) = \frac{P_{X,\hat{X}}(x,\hat{x})}{P_{\hat{X}}(\hat{x})} = \frac{P_{\hat{X}|X}(\hat{x}|x)P_X(x)}{P_{\hat{X}}(\hat{x})} = \frac{P_{\hat{X}|X}(\hat{x}|x)m^{-1}}{m^{-1}} = P_{\hat{X}|X}(\hat{x}|x).$$

Consequently,

$$\begin{aligned}
H\bigl(X|\hat{X}\bigr) &= \sum_{\hat{x}} P_{\hat{X}}(\hat{x}) H\bigl(X|\hat{X}=\hat{x}\bigr) \\
&= \sum_{\hat{x}} P_{\hat{X}}(\hat{x}) \biggl[(1-\mathsf{D}) \log\frac{1}{1-\mathsf{D}} + (m-1)\frac{\mathsf{D}}{m-1} \log\frac{m-1}{\mathsf{D}}\biggr] \\
&= (1-\mathsf{D}) \log\frac{1}{1-\mathsf{D}} + (m-1)\frac{\mathsf{D}}{m-1} \log\frac{m-1}{\mathsf{D}} \\
&= (1-\mathsf{D}) \log\frac{1}{1-\mathsf{D}} + \mathsf{D} \log\frac{1}{\mathsf{D}} + \mathsf{D} \log(m-1) \\
&= H_{\mathrm{b}}(\mathsf{D}) + \mathsf{D} \log(m-1).
\end{aligned}$$

Hence, the rate distortion function is

$$\mathsf{R}(\mathsf{D}) = \begin{cases} \log m - H_\mathrm{b}(\mathsf{D}) - \mathsf{D}\log(m-1) & 0 \le \mathsf{D} \le \frac{m-1}{m}, \\ 0 & \mathsf{D} > \frac{m-1}{m}. \end{cases}$$

It remains to show that the function $P_\mathrm{e} \mapsto H_\mathrm{b}(P_\mathrm{e}) + P_\mathrm{e}\log(m-1)$ is nondecreasing in $P_\mathrm{e}$ for $P_\mathrm{e} \in \bigl[0, \frac{m-1}{m}\bigr]$. This can be done for example by computing its derivative. An alternative way is the following concavity argument. Observe that

$$\begin{aligned} H_\mathrm{b}(P_\mathrm{e}) + P_\mathrm{e}\log(m-1) &= P_\mathrm{e}\log\frac{1}{P_\mathrm{e}} + (1-P_\mathrm{e})\log\frac{1}{1-P_\mathrm{e}} + P_\mathrm{e}\log(m-1) \\ &= P_\mathrm{e}\log\frac{m-1}{P_\mathrm{e}} + (1-P_\mathrm{e})\log\frac{1}{1-P_\mathrm{e}} \\ &= \log m + P_\mathrm{e}\log\frac{(m-1)/m}{P_\mathrm{e}} + (1-P_\mathrm{e})\log\frac{1/m}{1-P_\mathrm{e}} \\ &= \log m - D\Bigl(\bigl(P_\mathrm{e}, 1-P_\mathrm{e}\bigr)\big\|\bigl(\tfrac{m-1}{m}, \tfrac{1}{m}\bigr)\Bigr). \end{aligned}$$

Because relative entropy is nonnegative and equal to zero if and only if both arguments are equal, $P_\mathrm{e} \mapsto H_\mathrm{b}(P_\mathrm{e}) + P_\mathrm{e}\log(m-1)$ is maximized for $P_\mathrm{e} = \frac{m-1}{m}$. Since $P_\mathrm{e} \mapsto H_\mathrm{b}(P_\mathrm{e}) + P_\mathrm{e}\log(m-1)$ is concave in $P_\mathrm{e}$ (entropy and linear functions are concave; and the sum of concave functions is concave), it is nondecreasing in $P_\mathrm{e}$ for $P_\mathrm{e} \in \bigl[0, \frac{m-1}{m}\bigr]$.