Information Theory I

Fall Semester 2016

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

Model Answers to Exercise 6 of October 26, 2016

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

Problem 1 — *Strong Versus Weak Typicality*

By the definition of the strongly typical set (note that we do not follow the definition from the book *Elements of Information Theory*)

$$\mathcal{T}_\epsilon^{(n)}(P) \triangleq \left\{ \mathbf{x} \in \mathcal{X}^n : \left| \frac{1}{n} N(\xi|\mathbf{x}) - P(\xi) \right| \leq \epsilon \cdot P(\xi) \quad \forall \xi \in \mathcal{X} \right\},$$

a sequence $\mathbf{x} \in \mathcal{X}^{100}$ is strongly 0.01-typical if and only if the number of occurrences of "True" and "False" is between 49.5 and 50.5. The set of strongly typical sequences therefore consists of all sequences where "True" and "False" occur exactly 50 times. (This means that only a fraction of $\binom{100}{50}/2^{100} \approx 0.08$ of the sequences are strongly 0.01-typical.)

By the definition of the weakly typical set, a sequence $\mathbf{x} \in \mathcal{X}^{100}$ is weakly ϵ-typical if and only if its probability is not smaller than $2^{-100(H(X)+\epsilon)}$ and not larger than $2^{-100(H(X)-\epsilon)}$. Because "True" and "False" are equiprobable, $H(X) = 1$. Since the X_i are IID, every sequence $\mathbf{x} \in \mathcal{X}^{100}$ occurs with probability 2^{-100}. Consequently, every sequence is weakly typical, i.e., the set of weakly typical sequences is equal to $\mathcal{X}^{100}$ (irrespective of $\epsilon > 0$).

Problem 2 — *Random Box Size*

a) Because the X_i are IID, we have for all n

$$(\mathsf{E}[V_n])^{1/n} = \left(\prod_{i=1}^{n} \mathsf{E}[X_i] \right)^{1/n} = \left(\prod_{i=1}^{n} \frac{1}{2} \right)^{1/n} = \frac{1}{2},$$

and $\lim_{n\to\infty} (\mathsf{E}[V_n])^{1/n} = \frac{1}{2}$.

b) We have

$$\mathsf{E}[\ln X] = \int_0^1 \ln x \, \mathrm{d}x = (x \ln x - x)\Big|_{x=0}^{x=1} = -1,$$

and because the X_i (and therefore also $\ln X_i$) are IID, the claim follows by the weak law of large numbers (see Problem 4 of Exercise 1, which assumes that the variance is finite; more generally, the existence of the expectation is sufficient for the weak law of large numbers to hold, so we do not need to check whether the variance is finite or not).

c) The claim follows from Part b) because

$$L_n = \left(\prod_{i=1}^{n} X_i\right)^{1/n} = e^{\frac{1}{n}\ln\left(\prod_{i=1}^{n} X_i\right)} = e^{\frac{1}{n}\sum_{i=1}^{n}\ln X_i}$$

and because the exponential function is continuous. We now show the steps in more detail. Let $\epsilon_2 > 0$ be fixed and let $Z_n = \frac{1}{n}\sum_{i=1}^{n}\ln X_i$. By the continuity of the exponential function, there exists a $\delta > 0$ such that for all $z \in \mathbb{R}$, $|\, e^z - e^{-1}| < \epsilon_2$ whenever $|z-(-1)| < \delta$. Therefore,

$$\Pr\left[\left|e^{Z_n} - e^{-1}\right| \geq \epsilon_2\right] \leq \Pr[|Z_n - (-1)| \geq \delta]. \tag{1}$$

Using Part b) with $\epsilon_1 = \delta$ shows that $\Pr[|Z_n - (-1)| \geq \delta]$ tends to zero as n tends to infinity. Thus, the left-hand side of (1) also tends to zero as n tends to infinity, which proves the claim.

Remark 1. Comparing this result with the result from Part a) shows that the edge length L_n, which converges to e^{-1} in probability, does not capture the idea of the volume of the box.

Remark 2. Note the similarity between the concept of typical sets and the approach taken here. We considered only a subset of the outcomes, namely those which satisfy $|Z_n-(-1)| < \delta$; we have shown that $|\, e^{Z_n} - e^{-1}| < \epsilon_2$ is satisfied for those outcomes; and we have shown that the probability of this set of outcomes tends to one as n tends to infinity.

Problem 3 — *From AEP to Kraft's Inequality*

Let $\ell_1, \ldots, \ell_d$ be the codeword lengths of a uniquely decodable one-to-variable code $\mathscr{C}$ and assume that

$$\alpha = \sum_{i=1}^{d} 2^{-\ell_i} > 1. \tag{2}$$

We construct a memoryless source, i.e., a source that emits IID messages, with message probabilities

$$p_i = \frac{2^{-\ell_i}}{\alpha} \quad \forall i \in \{1, \ldots, d\}. \tag{3}$$

Using $\mathscr{C}$ to encode this source yields the expected codeword length

$$\mathsf{E}[\ell(X)] = \sum_{i=1}^{d} p_i \ell_i \overset{(3)}{=} \sum_{i=1}^{d} p_i \log\frac{1}{\alpha p_i} = H(X) - \log\alpha, \tag{4}$$

where $\log\alpha > 0$ because of (2).
(You do not need to show this, but setting $p_i \propto 2^{-\ell_i}$ in fact minimizes $\mathsf{E}[\ell(X)] - H(X)$ because

$$\begin{aligned}
\mathsf{E}[\ell(X)] - H(X) &= \sum_{i=1}^{d} p_i\ell_i - \sum_{i=1}^{d} p_i \log\frac{1}{p_i} \\
&= \sum_{i=1}^{d} p_i \log\frac{p_i}{2^{-\ell_i}} \\
&= \underbrace{\sum_{i=1}^{d} p_i \log\frac{p_i}{2^{-\ell_i}/\alpha}}_{= D(p||q) \geq 0} + \underbrace{\sum_{i=1}^{d} p_i \log(1/\alpha)}_{= -\log\alpha},
\end{aligned}$$

so the choice (3) corresponds to $D(p||q) = 0$, which minimizes the expression.)
For every $n \in \mathbb{Z}^+$, we construct a code that maps n source symbols to $\left\lceil n(H(X) - \frac{1}{2}\log\alpha)\right\rceil + 1$ bits as follows:

Step 1: Describe a source sequence $\mathbf{x} \in \mathcal{X}^n$ using the extension of $\mathscr{C}$.

Step 2: If the length of the description of $\mathbf{x}$ is less than or equal to $\lceil n(H(X) - \frac{1}{2}\log\alpha)\rceil$, append a single one to its end and then add zeros to its end until the length of the description is $\lceil n(H(X) - \frac{1}{2}\log\alpha)\rceil + 1$; use the resulting bitstring as the codeword for $\mathbf{x}$.

If the length of the description of $\mathbf{x}$ is larger than $\lceil n(H(X) - \frac{1}{2}\log\alpha)\rceil$, use the all-zero bitstring as the codeword for $\mathbf{x}$.

Observe that the rate of the described code satisfies

$$\lim_{n\to\infty} \rho_n = \lim_{n\to\infty} \frac{\lceil n(H(X) - \frac{1}{2}\log\alpha)\rceil + 1}{n} = H(X) - \frac{1}{2}\log\alpha < H(X),$$

where the last inequality follows from (2). Consequently, this code satisfies the condition for the converse of the general source coding theorem. We will show next that the probability of successful decoding of this code tends to one as n tends to infinity, which contradicts the theorem and establishes that no uniquely decodable one-to-variable code $\mathscr{C}$ satisfying (2) can exist.

To analyze the error probability of the code, observe that the source message can be uniquely determined if the length of the description in Step 1 is less than or equal to $\lceil n(H(X) - \frac{1}{2}\log\alpha)\rceil$: removing the trailing zeros and the single one from the codeword recovers the original description (in other words, the padding scheme is injective). We bound the probability that the description length exceeds $\lceil n(H(X) - \frac{1}{2}\log\alpha)\rceil$ as follows:

$$\begin{aligned}
&\Pr\left[\sum_{k=1}^{n} \ell(x_k) > \left\lceil n\biggl(H(X) - \frac{1}{2}\log\alpha\biggr)\right\rceil\right] \\
&\quad\le \Pr\left[\sum_{k=1}^{n} \ell(x_k) \ge n\biggl(H(X) - \frac{1}{2}\log\alpha\biggr)\right] \\
&\quad= \Pr\left[\frac{1}{n}\sum_{k=1}^{n} \ell(x_k) \ge H(X) - \frac{1}{2}\log\alpha\right] \\
&\quad\le \Pr\left[\left|\frac{1}{n}\sum_{k=1}^{n} \ell(x_k) - (H(X) - \log\alpha)\right| \ge \frac{1}{2}\log\alpha\right].
\end{aligned} \tag{5}$$

By the weak law of large numbers, the right-hand side of (5) tends to zero as n tends to infinity (we have an IID source; we computed the expected codeword length in (4); and $\log\alpha$ is positive). Hence, the probability that the description length is less than or equal to $\lceil n(H(X) - \frac{1}{2}\log\alpha)\rceil$, and also the probability of successful decoding, tend to one as n tends to infinity.