Telecom ParisTech SOCOM205
Prof. Aslan Tchamkerten March 2017

ASSIGNMENT 2: SOLUTION

Exercise 1 (Guessing, Huffman)**.** There are 6 bottles of wine, one of which you know has gone bad. You do not know which bottle contains the bad wine, but you know that the probability of each bottle being bad is $(8/23, 6/23, 4/23, 2/23, 2/23, 1/23)$. The bad wine has a distinctive taste. To find the bad wine your friend suggests you to taste a little bit of each wine until you find the bad wine.

a. To have the least number of tastings on average, what should your strategy be? Which bottle should be tasted first?

b. What is the average number of tastings to find the bad wine?

c. Calculate the minimum average number of tastings if you are allowed to taste a mixture of different wines and detect a bad wine's taste inside (the distinctive taste is retained even when mixed with other good wines).

d. Is the strategy studied in (a) optimal if you are allowed to mix wines?

Solution. a. A guessing strategy for a random variable X can be written as a vector $G = (g_1, g_2, \ldots)$ with $g_i \in \mathcal{X}$ being the i-th guess of X. With this notation, the expected number of guesses is given by $\mathbb{E}(G) = \sum_i i\mathbb{P}(X = g_i)$. Now assume that for some $i < j$ we have $\mathbb{P}(X = g_j) > \mathbb{P}(X = g_i)$, and consider the new strategy G' where g_i and g_j are swapped. It then follows that $E(G) - E(G') = (j - i)(\mathbb{P}(X = g_j) - \mathbb{P}(X = g_i)) > 0$. It follows that the strategy that guesses the values of X in decreasing order of proabilities minimizes the expected number of guesses.

b. $56/23$

c. A sequence of questions is equivalent to a code. Indeed, any question depends on the sequence of answers to the questions before it. Since the sequence of answers uniquely determines a particular sample of X, if we represent the sequence of yes-no answers by 0 and 1, each sample of X is associated to a codeword. Conversely, from a binary code for each possible sample of X, we can find a sequence of questions that corresponds to the code. The i-th question is "Is the i-th bit bit equal to 1?" or, more specifically, "Does X belong to the set of samples whose codewords have the i-th bit equal to 1?"

Therefore, from the equivalence between guessing strategy and code, finding a guessing strategy that minimizes the number of questions is equivalent to finding a code whose average length is minimal. An optimal strategy to identify the bad bottle is thus obtained via the construction of the Huffman code of the bad bottle probability distribution. Note that we use here the fact that we are allowed to mix wines, hence we can ask, at each step, whether the bad wine belongs to some particular subset of bottles or not.

□

Exercise 2 (Entropy and Yes/No questions)**.** We are asked to determine an object by asking yes-no questions. The object is drawn randomly from a finite set according to a certain distribution. Playing optimally, we need 38.5 questions on the average to find the object. At least how many elements does the finite set have?

Solution. An optimal yes/no scheme corresponds to an optimal source code whose expected length is at most $H(X)+1$, where X is the hidden object. Hence $H(X)+1 \geq 38.5$. On the other hand we have $\log|\mathcal{X}| \geq H(X)$. These two yield that $\log|\mathcal{X}| \geq 37.5$, and so $n \geq \lceil 2^{37.5} \rceil$. □

Exercise 3 (Pure randomness from biased distributions)**.** Let $X_1, X_2, \ldots, X_n$ denote the outcomes of independent flips of a biased coin. Thus, for $i = 1, \ldots, n$ we have $\Pr(X_i = 1) = p, \Pr(X_i = 0) = 1 - p$, where p is unknown. We wish to obtain a sequence $Z_1, Z_2, \ldots, Z_K$ of fair coin flips from $X_1, X_2, \ldots, X_n$. To this end let $f : \mathcal{X}^n \to \{0,1\}^\star$ (where $\{0,1\}^\star = \{\Lambda, 0, 1, 00, 01, \ldots\}$ is the set of all finite length binary sequences including the null string Λ) be a mapping $f(X_1, X_2, \ldots, X_n) = (Z_1, Z_2, ..., Z_K)$, such that $Z_i \sim$ Bernoulli$(1/2)$ and where K possibly depends on $(X_1, ..., X_n)$. For the sequence $Z_1, Z_2, ..., Z_K$ to correspond to fair coin flips, the map f from biased coin flips to fair flips must have the property that all 2^k sequences $(z_1, z_2, \ldots, z_k)$ of a given length k have equal probability (possibly 0). For example, for $n = 2$, the map $f(01) = 0, f(10) = 1, f(00) = f(11) = \Lambda$ has the property that $\Pr(Z_1 = 1|K = 1) = \Pr(Z_1 = 0|K = 1) = 1/2$.

a. Justify the following (in)equalities

$$\begin{aligned} nH_b(p) &\overset{(a)}{=} H(X_1, ..., X_n) \\ &\overset{(b)}{\geq} H(Z_1, Z_2, \ldots, Z_K, K) \\ &\overset{(c)}{=} H(K) + H(Z_1, Z_2, ..., Z_K|K) \\ &\overset{(d)}{=} H(K) + E(K) \\ &\overset{(e)}{\geq} E(K) \end{aligned}$$

where $E(K)$ denotes the expectation of K. Thus, on average, no more than $nH_b(p)$ fair coin tosses can be derived from $(X_1, ..., X_n)$.

b. Exhibit a good map f on sequences of length $n = 4$.

Solution. a. (a.) the X_i's are i.i.d. Bernoulli(p) distributed; (b) (Z^K, K) is a function of X^n; (c) chain rule; (d) given $K = k$, $(Z_1, Z_2, \ldots, Z_k)$ is an i.i.d. Bernoulli$(1/2)$ sequence, hence $H(Z_1, Z_2, ..., Z_K|K = k) = k$, from which the result follows; (d) non-negativity of the entropy.

b. One possibility is as follows. Let T_k be the set of binary sequences of length 4 with exactly k ones ($k \in \{0, 1, 2, \ldots, 4\}$). Observe that T_1 and T_3 each have four elements, and each contains equiprobable elements (obviously, the elements in T_1 have a different probability than those in T_3). We map the 4 elements in T_1 in 00, 01, 10, and 11, and similarly for T_3. I follows that, given $K = 2$, (Z_1, Z_2) are purely random. To see this note that for any pair of bit (i, j)

$$\begin{aligned} \Pr((Z_1, Z_2) = (i,j)|K = 2) &= \Pr((Z_1, Z_2) = (i,j)|X^4 \in T_1 \cup T_3) \\ &= \Pr((Z_1, Z_2) = (i,j)|X^4 \in T_1)\Pr(X^4 \in T_1|X^4 \in T_1 \cup T_3) \\ &+ \Pr((Z_1, Z_2) = (i,j)|X^4 \in T_3)\Pr(X^4 \in T_3|X^4 \in T_1 \cup T_3) \\ &= \frac{1}{4}\Pr(X^4 \in T_1|X^4 \in T_1 \cup T_3) + \frac{1}{4}\Pr(X^4 \in T_3|X^4 \in T_1 \cup T_3) \\ &= \frac{1}{4}. \end{aligned}$$

All the elements in T_0, T_2, and T_4 are mapped into Λ.

□

Exercise 4 (Entropy bound)**.** Let $p(x)$ be a probability mass function of random variable X. Prove that

$$\log \frac{1}{d} Pr\{p(X) \leq d\} \leq H(X)$$

for any $d \geq 0$.

Solution. Let $Y = \log(\frac{1}{p(X)})$. Then since $Y \geq 0$, by Markov's inequality we have

$$Pr(Y \geq \log(\frac{1}{d})) \leq \frac{\mathbb{E}(Y)}{\log(\frac{1}{d})}$$

The result then follows by noticing that $\mathbb{E}(Y) = H(X)$ and that $Pr(Y \geq \log(\frac{1}{d})) = Pr(p(X) \leq d)$. □

Exercise 5 (Entropy and Mutual Information)**.** Prove the following inequalities:

a. $H(X, Y|Z) \geq H(X|Z)$,

b. $I((X, Y); Z) \geq I(X; Z)$,

c. $H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$.

Solution. a.

$$\begin{aligned} H(X, Y|Z) &\overset{(a)}{=} H(X|Z) + H(Y|X, Z) \\ &\overset{(b)}{\geq} H(X|Z) \end{aligned}$$

where (a) holds by the chain rule for entropy and where (b) follows by the non-negativity of entropy.

b.

$$\begin{aligned} I(X, Y|Z) &\overset{(a)}{=} I(X; Z) + I(Y; Z|X) \\ &\overset{(b)}{\geq} I(X; Z) \end{aligned}$$

where (a) holds by the chain rule for mutual information and where (b) holds the non-negativity of mutual information.

c.

$$\begin{aligned} H(X, Y, Z) - H(X, Y) &\overset{(a)}{=} (H(X, Z) + H(Y|X, Z)) - (H(X) + H(Y|X)) \\ &\overset{(b)}{\leq} H(X, Z) - H(X) \end{aligned}$$

where (a) is due to the chain rule for entropy and where (b) holds since conditioning cannot increase entropy.

□

Exercise 6 (Conditioning for mutual information)**.** Give examples of joint random variables X, Y, and Z such that

a. $I(X;Y|Z) < I(X;Y)$.

b. $I(X;Y|Z) > I(X;Y)$.

Solution. a. Let X be Bernoulli($\frac{1}{2}$) random variable and $Z = Y = X$. Then,

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|X) - H(X|X) = 0 - 0 = 0$$

$$I(X;Y) = H(X) - H(X|Y) = H(X) - H(X|X) = H(X) - 0 = 1.$$

b. Let X and Y be independent Bernoulli($\frac{1}{2}$) random variables and $Z = X + Y$. Then,

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X) - H(X|X,Y) = 1 - 0 = 1$$

$$I(X;Y) = H(X) - H(X|Y) = H(X) - H(X) = 0.$$

□