Information & Communication Exercise Sheet #2

University of Amsterdam, Bachelor of Computer Science, January 2016
Lecturer: Christian Schaffner

Out: Wednesday, 6 January 2016
(due: Wednesday, 13 January 2016, 13:00)

To be solved in Class

1. ([MacKay], Example 2.15:) Three squares have average area $\overline{A} = 100\text{m}^2$. The average of the lengths of their sides is $\overline{\ell} = 10\text{m}$. What can be said about the size of the largest of the three squares? [Use Jensens inequality.]

2. ([Yeung]) Let X and Y be random variables over alphabets $\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4, 5\}$ and joint distribution P_{XY} given by the following matrix (where the entry in row i and column j is the probability $P_{XY}(i, j)$)

$$\frac{1}{25}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 0 & 0 \\ 2 & 0 & 1 & 1 & 1 \\ 0 & 3 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 & 3 \end{bmatrix}$$

   Calculate $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$, and $I(X;Y)$, and draw the entropy diagram.

3. ([MacKay], Example 2.13:) A source produces a character x from alphabet $\mathcal{A} = \{\mathtt{0}, \mathtt{1}, \mathtt{2}, \ldots, \mathtt{9}, \mathtt{a}, \mathtt{b}, \mathtt{c}, \ldots, \mathtt{z}\}$. With probability 1/3, x is a uniformly random numeral $\mathtt{0}, \mathtt{1}, \mathtt{2}, \ldots, \mathtt{9}$, with probability 1/3, x is a random vowel $\{\mathtt{a}, \mathtt{e}, \mathtt{i}, \mathtt{o}, \mathtt{u}\}$ and with probability 1/3, x is one of the 21 consonants. Estimate the entropy of X.

4. ([MacKay], Exercise 2.29) An unbiased coin is flipped until one head is thrown. What is the entropy of the random variable $X \in \{1, 2, 3, \ldots\}$, the number of flips? Repeat the calculation for the case of a biased coin with probability p of coming up heads.
   Hint: solve the problem both directly and by using the decomposability of the entropy, i.e. that for a probability distribution $\mathbf{p} = \{p_1, p_2, \ldots, p_n\}$, it holds that

$$H(\mathbf{p}) = H(p_1, 1 - p_1) + (1 - p_1) H\left(\frac{p_2}{1 - p_1}, \frac{p_3}{1 - p_1}, \ldots, \frac{p_n}{1 - p_1}\right).$$

5. *Maximal conditional entropy implies independence.* Let $n = \log(|\mathcal{X}|)$.
   (a) Prove that $H(X|Y) = n$ implies that X and Y are independent.
   (b) Give a joint distribution P_{XY} where $H(X) = n$, but X and Y are dependent.

6. For two distributions P and Q over $\mathcal{X}$, the *relative entropy* or *Kullback-Leibler divergence* is defined as

$$D(P||Q) := \sum_{\substack{x \in \mathcal{X} \\ P(x) > 0}} P(x) \log \frac{P(x)}{Q(x)}.$$

   Note that if $Q(x) = 0$ for some x, then $D(P||Q) = \infty$. Prove that $D(P||Q) \geq 0$, and that equality holds if and only if $P = Q$.
   Hint: Use Jensen's inequality.

7. [Cover-Thomas 5.18] Consider the code $C = \{0, 01\}$. Is it prefix-free? Is it uniquely decodable?

8. **Stirling's Approximation** Prove that

$$\ln(n!) \ \leq \ n \ln(n).$$

Then use the approximation $\ln(n!) \approx n \ln(n)$ to prove that

$$\frac{1}{n} \ln \binom{n}{np} \ \approx \ -p \ln(p) - (1-p) \ln(1-p) = h(p)\,,$$

where we assume that n is an integer, $p \in (0, 1)$, and np is an integer.

9. **The Weak Law of Large Numbers*** In this exercise, you will prove that averages converge to expectations in a certain precise sense. The proof is using a number of steps, each of which is interesting in its own right.

   (a) **Mean and Variance of Averages** Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed variables sampled from a distribution with mean $E[X]$ and variance $Var[X]$. Prove that

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \quad = \quad \mathrm{E}[X] \tag{1}$$

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \quad = \quad \frac{\mathrm{Var}[X]}{n} \tag{2}$$

   (b) **The Markov Bound** Suppose that S is a random variable which only takes on non-negative values (that is, $P(S \geq 0) = 1$). Prove that

$$P(S \geq s) \ \leq \ \frac{\mathrm{E}[S]}{s}$$

   (For instance, less than 1/5 of the population earns more than 5 times the average income.)

   (c) **Chebyshev's Inequality** Suppose X is a random variable with mean $\mathrm{E}[X]$ and variance $\mathrm{Var}[X]$. Prove that

$$P\left(|X - \mathrm{E}[X]| \geq \varepsilon\right) \ \leq \ \frac{\mathrm{Var}[X]}{\varepsilon^2}.$$

   (d) **The Weak Law of Large Numbers** Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with a shared mean $\mathrm{E}[X]$ and variance $\mathrm{Var}[X]$. Prove that

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathrm{E}[X]\right| \geq \varepsilon\right) \ \leq \ \frac{\mathrm{Var}[X]}{n\varepsilon^2}.$$

10. **Tail-heavy Distribution**** Give an example of a discrete random variable S for which the Markov bound holds with equality for every $s \in \{1, 2, 3, \ldots\}$.

Homework

1. **Entropy of functions of a random variable.** Let X be a discrete random variable. Show that the entropy of a function g of X is less than or equal to the entropy of X by justifying the following steps: 3 p.

$$\begin{aligned} H(X) &= H(X) + H(g(X)|X) && (3)\\ &= H(X, g(X)) && (4)\\ &= H(g(X)) + H(X|g(X)) && (5)\\ &\geq H(g(X)) && (6) \end{aligned}$$

2. **Sum Distribution** Let X and Y be independent binary random variables with 3 p.

$$P_X(1) \;=\; P_Y(1) \;=\; \frac{1}{2}.$$

   Compute $H(X+Y)$.

3. **Squares and Expectations** Use Jensen's inequality to derive an inequality between $\mathrm{E}[X^2]$ and $\mathrm{E}[X]^2$. Use this inequality as an alternative proof that $\mathrm{Var}[X] \geq 0$. 2 p.

4. **Mutual Information** The mutual information between two random variables X and Y is defined as $I(X;Y) := H(X) - H(X|Y)$

   (a) Show that the mutual information can be expressed in terms of the relative entropy, i.e. that $I(X;Y) = D(P_{XY}||P_X P_Y)$ 3 p.

   (b) Use (a) and Class exercise 6 to prove that $H(X|Y) \leq H(X)$. 1 p.

5. **Kraft's Inequality:** Below, six binary codes are shown for the source symbols $x_1, \ldots, x_4$.

	Code A	Code B	Code C	Code D	Code E	Code F
x_1	00	0	0	0	1	1
x_2	01	10	11	100	01	10
x_3	10	11	100	110	001	100
x_4	11	110	110	111	0001	1000

   (a) Which codes fulfill the Kraft inequality? 2 p.

   (b) Is a code that satisfies this inequality always uniquely decodable? 2 p.

   (c) Which codes are prefix-free codes? 2 p.

   (d) Which codes are uniquely decodable? 2 p.

6. **Optimal Huffman coding:** Consider a random variable X that takes on four values with probabilities $\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12}$. Show that there exist two different sets of optimal length for the (binary) Huffman codewords. 2 p.

7. **Huffman Coding:** Jane, a student, regularly sends a message to her parents via a binary channel. The binary channel is lossless (i.e. error-free), but the per-bit costs are quite high, so she wants to send as few bits as possible. Each time, she selects one message out of a finite set of possible messages and sends it over the channel. There are 7 possible messages:

   (a) "Everything is fine"
   (b) "I am short on money; please send me some"
   (c) "I'll come home this weekend"
   (d) "I am ill, please come and pick me up"
   (e) "My study is going well, I passed an exam (. . . and send me more money)"
   (f) "I have a new boyfriend"
   (g) "I have bought new shoes"

   Based on counting the types of 100 of her past messages, the empirical probabilities of the different messages are:

m	a	b	c	d	e	f	g
$P_M(m)$	19/100	40/100	12/100	2/100	16/100	4/100	7/100

   Jane wants to minimize the average number of bits needed to communicate to her parents (with respect to the empirical probability model above).

   (a) Design a Huffman code for Jane and draw the binary tree that belongs to it. 2 p.

   (b) For a binary source X with $P_X(0) = \frac{1}{8}$ and $P_X(1) = \frac{7}{8}$, design a Huffman code for blocks of $N = 1, 2$ and 3 bits. For each of the three codes, compute the average codeword length and divide it by N, in order to compare it to the optimal length, i.e. the entropy of the source. What do you observe? 4 p.

   (c) If you were asked at (b) to design a Huffman code for a block of $N = 100$ bits, what problem would you run into? 1 p.