Information Theory Exercise Sheet #4

University of Amsterdam, Master of Logic, Fall 2014
Lecturer: Christian Schaffner
TA: Philip Schulz

Out: Wednesday, 19 November 2014 2014
(due: Wednesday, 26 November 2014, 13:00)

Previous Exam Questions

At this point in the course, you should be able to solve the following exercises which were exam questions in previous editions of the course:

Both entropy and variance are often used as measures of the "inherent uncertainty" in a distribution, so it is interesting to find out how similar they are. Consider sample space $\mathcal{X} = \{1, 2, \ldots, n\}$ for some $n \geq 2$.

1. What distribution $P_X^{\max}$ on $\mathcal{X}$ maximizes the entropy, and what is the entropy $H(P_X^{\max})$?
2. What distributions on $\mathcal{X}$ minimize the entropy?
3. What distribution $Q_X^{\max}$ on $\mathcal{X}$ maximizes the variance, and what is the variance $\mathrm{Var}[Q_X^{\max}]$?
4. What distributions on $\mathcal{X}$ minimize the variance?
5. Now let $\mathcal{X}$ be the positive natural numbers. Show that for every $\varepsilon > 0$, no matter how small, and for every finite C, no matter how large, there exists a distribution P_X on $\mathcal{X}$ that has entropy smaller than ε and variance greater than C.

   Hint: For some $\delta > 0$ and $n \geq 2$, consider the distribution $P_X(1) = 1 - \delta$, $P_X(n) = \delta$.

To be solved in Class

1. *Quantifiying the loss when using the wrong code.* Prove that when designing a code with length $\ell(X)$, believing that the distribution is Q_X when the true distribution is P_X incurs a penality of $D(P_X||Q_X)$ in the average description length. More formally, prove that

$$H(P_X) + D(P_X||Q_X) \leq \mathbb{E}_{P_X}[\ell(X)] \leq H(P_X) + D(P_X||Q_X) + 1$$

Homework

1. [4 points] *Size of binomial coefficient:* Let $h(p) = -p \log p - (1-p)\log(1-p)$ denote the binary entropy of a Bernoulli distribution when the probability of observing a zero is p, where log denotes the logarithm to base 2. Use Stirlings approximation

$$\ln(n!) = (n + \frac{1}{2}) \ln(n) - n + O(1)$$

to show that

$$\log \binom{n}{\gamma n} = nh(\gamma) - \frac{1}{2} \log n + O(1)$$

Reminder: $O(1)$ stands for any constant that does not depend on n.

2. *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) [2 points] Which codes fulfill the Kraft inequality?
   (b) [2 points] Is a code that satisfies this inequality always uniquely decodable?
   (c) [2 points] Which codes are prefix-free codes?
   (d) [2 points] Which codes are uniquely decodable?

3. *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) [2 points] Design a Huffman code for Jane and draw the binary tree that belongs to it.
   (b) [4 points] 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?

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

Consider the random variable Z with

z	1	2	3	4	5	6
$P_Z(z)$	1/10	3/10	2/10	2/10	1/10	1/10

(d) [2 points] Find an optimal *ternary* Huffman encoding for Z (i.e., using an alphabet with three symbols).

4. *Prefix-free arithmetic coding*

Definition 1 *A* binary interval *is defined as* $[\frac{s}{2^\ell}, \frac{s+1}{2^\ell})$ *for some natural numbers* s, ℓ*, where* $0 \leq s \leq 2^\ell - 1$*. Binary intervals have* names *consisting of the binary representation of* s *of length* ℓ*.*

Note that the intervals $[\frac{1}{3}, \frac{2}{3}), [\frac{1}{4}, \frac{3}{4})$ are *not* binary intervals.

Examples of binary intervals are $[\frac{2}{4}, \frac{3}{4}), [\frac{0}{8}, \frac{1}{8}), [\frac{1}{2}, \frac{2}{2})$ with corresponding names 10, 000, 1. The interval names can be thought of as describing the path in the binary tree that leads to that interval, see Mathias Madsen's slide for some nice illustrations: `http://informationtheory.weebly.com/uploads/2/4/6/2/24624554/arithmetic_coding.pdf`

(a) [1 point] What are the names of the binary intervals $[\frac{6}{8}, \frac{7}{8})$ and $[\frac{7}{16}, \frac{8}{16})$?

(b) [1 point] What are the binary intervals with the names 0110 and 011?

(c) [1 point] Prove that if the name of a binary interval I is the prefix of the name of another binary interval J, it must be that $J \subset I$.

Recall from the lecture that in arithmetic coding, we split the unit interval $[0, 1)$ into subintervals $I_1, I_2, \ldots I_m$ where the sizes of I_x correspond to the probabilities p_x. For any $x \in \mathcal{X}$, we now define the arithmetic code $AC^{pf}(x)$ to be the name of the largest binary interval that fits in I_x. For instance, for the source $P_X(a) = 0.3, P_X(b) = 0.3, P_X(c) = 0.4$, we obtain the intervals $I_a = [0, 0.3), I_b = [0.3, 0.6), I_c = [0.6, 1)$ and the maximal binary intervals are $[\frac{0}{4}, \frac{1}{4}) \subset I_a$, $[\frac{3}{8}, \frac{4}{8}) \subset I_b$, $[\frac{3}{4}, \frac{4}{4}) \subset I_c$ with names $AC^{pf}(a) = 00, AC^{pf}(b) = 011, AC^{pf}(c) = 11$.

(d) [2 points] Use (c) to prove that for any source, the resulting arithmetic code AC^{pf} is prefix-free.

(e) [4 points] Prove that for any source P_X, it holds that $\ell_{AC^{pf}}(X) \leq H(X) + 2$. **Hint:** Remember how we proved in class that $\ell_{AC}(X) \leq H(X) + 1$. Use a similar approach here, but set $\ell := \lceil \log(1/p) \rceil + 1$ instead.

(f) [2 points] Give the arithmetic code AC^{pf} for the source $P_X(a) = 1/6, P_X(b) = P_X(c) = 1/3, P_X(d) = 1/6$.