COMP2610/COMP6261 - Information Theory

Tutorial 3: Coding and Compression

Robert C. Williamson

Semester 2, 2018

1. *Probabilistic inequalities*

   Suppose a coin is tossed n times. The coin is known to land "heads" with probability p. The number of observed "heads" is recorded as a random variable X.

   (a) What is the *exact* probability of X being $n-1$ or more?

   (b) Using Markov's inequality, compute a *bound* on the same probability as the previous part.

   (c) Suppose $n = 2$. For what values of p will the bound from Markov's inequality be within 1% of the exact probability?

2. *AEP and source coding* (cf. Cover & Thomas, Problem 3.7)

   A sequence of bits its generated by i.i.d. draws from an ensemble with probabilities $p_0 = 0.995$ and $p_1 = 0.005$. Sequences are coded in 100-bit blocks. Every 100-bit block with at most three 1s is assigned a codeword. Those blocks with more than three 1s are not assigned codewords.

   (a) What is the minimum required length of the assigned codewords if they are all to be of the same length?

   (b) Calculate the probability of observing a 100-bit block that has no associated codeword.

   (c) (*Harder*) Use Chebyshev's inequality to bound the probability of observing a 100-bit block for which no codeword has been assigned. Compare the bound to the probability just calculated.

3. *Typical Sets and Smallest δ-Sufficient Subsets* (cf. Cover & Thomas, Problem 3.13)

   Let X^N be an extended ensemble for X with $\mathcal{A}_X = \{0, 1\}$ and $\mathcal{P}_X = \{0.4, 0.6\}$.

   (a) Calculate the entropy $H(X)$.

   (b) Let $N = 25$ and $\beta = 0.1$.

      i. Which sequences in X^N fall in the typical set $T_{N\beta}$? (You may find it helpful to refer to Table 1 below.)

      ii. Compute $P(\mathbf{x} \in T_{N\beta})$, the probability of a sequence from X^N falling in the typical set.

      iii. With reference to Table 1 below, how many elements are there in $T_{N\beta}$?

      iv. How many elements are in the smallest δ-sufficient subset S_δ for $\delta = 0.9$?

      v. What is the essential bit content $H_\delta(X^N)$ for $\delta = 0.9$?

4. *Source Coding Theorem*

   Recall that the source coding theorem (for uniform codes) says that for any ensemble X:

$$(\forall \epsilon > 0)\, (\forall \delta \in (0, 1))\, (\exists N_0)\, (\forall N > N_0)\, \left| \frac{1}{N} H_\delta(X^N) - H(X) \right| \leq \epsilon.$$

(a) Near, an enthusiastic software developer, has just learned about the source coding theorem. He exclaims: *"The theorem allows us to pick any $\epsilon > 0$. So, if I pick $\epsilon = H(X) - \epsilon'$, I get that for sufficiently large N,*

$$\frac{1}{N} H_\delta(X^N) \geq \epsilon'.$$

This means that by making ϵ' tiny, I can get away with using virtually zero bits per outcome. Great!".
Is Near's reasoning correct? Explain why or why not.

(b) Mello, a skeptical econometrician, has also just learned about the source coding theorem. He complains: *"The thereom is not really relevant to me. I am interested in coding blocks of outcomes where each outcome is dependent on the previous outcome, rather than them all being independent of each other. The source coding theorem is not useful in this case."*
Is Mello's reasoning correct? Explain why or why not.

5. *Coding*

A standard deck of cards contains 4 *suits* — $\heartsuit, \diamondsuit, \clubsuit, \spadesuit$ ("hearts", "diamonds", "clubs", "spades") — each with 13 *values* — $\mathrm{A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K}$ (The A, J, Q, K are called "Ace", "Jack", "Queen", "King").

Twenty cards are removed from a standard deck: All 13 $\spadesuit$s; the A$\heartsuit$, 2$\heartsuit$, 3$\heartsuit$, 4$\heartsuit$, and 5$\heartsuit$; and the A$\diamondsuit$ and K$\diamondsuit$. Assume the deck is thoroughly shuffled and that after each draw from the deck the card drawn is replaced and the deck reshuffled. That is, repeated card draws are i.i.d. from a uniform distribution.

We would like to code the *value* V of randomly drawn cards.

(a) How many bits are required to *uniformly code* V?

(b) Compute $p(V = v)$ for every possible value of v.

(c) Compute a *δ-sufficient subset* S_δ for V using $\delta = 0, 1/16, 1/2$ and their associated *essential bit content* H_δ.

(d) Compute a *typical set* $T_{N\beta}$ for $N = 1$ and $\beta = 0.3$.

(e) For $\delta = 1/16$, is it possible to construct a block code for large choices of N such that only 2 bits per outcome are used on average (i.e., $\frac{1}{N} H_\delta(V^N) = 2$)? Explain why or why not.

(f) Is it possible to construct a *prefix code* such that the face cards are given 3-bit codewords and non-face cards are given 4-bit codewords? If so, construct such a code.

6. *Prefix Codes*

Consider the codes $C_1 = \{0, 01, 1101, 10101\}$, $C_2 = \{00, 01, 100, 101\}$, and $C_3 = \{0, 1, 00, 11\}$

(a) Are C_1, C_2, and C_3 prefix codes? Are they uniquely decodable?

(b) Construct new prefix codes C_1', C_2', and C_3' that have the same lengths as C_1, C_2, and C_3, respectively. If this is not possible, explain why.

(c) Is it possible to construct a uniquely decodable code that has codeword lengths $\{1, 2, 3, 4, 4, 4\}$? If so, construct one.

(d) Is it possible to construct a uniquely decodable code with lengths $\{1, 3, 3, 4, 4, 4\}$? If so, construct one.

7. *Optimal Coding and Huffman Codes*

Consider the ensemble X with probabilities $\mathcal{P}_X = \mathbf{p} = \{\frac{1}{2}, \frac{1}{4}, \frac{31}{128}, \frac{1}{128}\}$ and the code $C = \{0, 11, 100, 101\}$.

(a) What is the entropy $H(X)$?

(b) What is the expected length $L(C, X)$? Is C an optimal code for X?

(c) What are the code lengths for X? Construct a prefix Shannon code C_S for X. Compute the expected code length $L(C_S, X)$.

(d) What are the probabilities $\mathbf{q} = \{q_1, q_2, q_3, q_4\}$ for the code lengths of C?

(e) Compute the relative entropy $D(\mathbf{p}\|\mathbf{q})$. What do you notice about $D(\mathbf{p}\|\mathbf{q})$, $H(X)$, $L(C, X)$, and $L(C_S, X)$?

(f) Construct a Huffman code C_H for X. How does its code lengths compare to C and C_S? How do their expected code lengths compare?

k	$\binom{N}{k}$	$\binom{N}{k} p_1^k p_0^{N-k}$	$-\frac{1}{N} \log_2 p(\mathbf{x})$
0	1	0.000000	1.321928
1	25	0.000000	1.298530
2	300	0.000000	1.275131
3	2300	0.000001	1.251733
4	12650	0.000007	1.228334
5	53130	0.000045	1.204936
6	177100	0.000227	1.181537
7	480700	0.000925	1.158139
8	1081575	0.003121	1.134740
9	2042975	0.008843	1.111342
10	3268760	0.021222	1.087943
11	4457400	0.043410	1.064545
12	5200300	0.075967	1.041146
13	5200300	0.113950	1.017748
14	4457400	0.146507	0.994349
15	3268760	0.161158	0.970951
16	2042975	0.151086	0.947552
17	1081575	0.119980	0.924154
18	480700	0.079986	0.900755
19	177100	0.044203	0.877357
20	53130	0.019891	0.853958
21	12650	0.007104	0.830560
22	2300	0.001937	0.807161
23	300	0.000379	0.783763
24	25	0.000047	0.760364
25	1	0.000003	0.736966

Table 1: Table for Question 3. Column 1 shows k, the number of 1s in a block of length $N = 25$. Column 2 shows the number of such blocks. Column 3 shows the probability $p(\mathbf{x}) = \binom{N}{k} p_1^k p_0^{N-k}$ of drawing such a block $\mathbf{x}$. Column 4 shows the Shannon information per symbol in $\mathbf{x}$.