

Handout 5: Problem Set 3: (due Fri. Feb 27th by 4:30pm)

Prof: Jeff A. Bilmes <bilmes@ee.washington.edu>

Lecture 10, Feb 18, 2004

Book Problems Do problems 4.8,5.6,5.13,5.23,5.24,8.2,8.4,8.9 in C&T.

Problem 1:

- a Define the geometric mean.
- b Why is the geometric mean called the geometric mean?
- c What is similar about the geometric mean and entropy?

Problem 2:

- a Encode the string 00101001010101001010101010010101 using the basic Lempel-Ziv algorithm we described in class.
- b Encode the string 1100011011110000001001011100110000111 using the basic Lempel-Ziv algorithm we described in class.
- c Can you characterize the difference between these two strings? Did you notice this difference immediately when you first glanced at them? Why or why not?

Problem 3:

You've discovered a random number generator that generates random strings of bits $M_{1:n}$ ($M_i \in \{0, 1\}$) for any n . Each such bit string corresponds to sampled music recordings (one sample might be Miles Davis or John Coltrane, another might be Beethoven or Brahms, a third might be Eminem or the Sex pistols, etc.). Since this random source corresponds to natural signals, we know it can not have maximum entropy, meaning that either $H(M_i) < 1$ or $H(M_{i:j}) < \sum_{k=i}^j H(M_k)$ or (more likely) both. Why is this the case?

You give this bit string to your (best) friend who rather than listens to infinite free music (obviously not a Kazaa subscriber), does the following. She produces a new random bit string generator $B_{1:n}$ where $B_i = M_i \text{ xor } R_i$, and where R_i is a coin flip with entropy $H = 1$ (i.e., $R_{i:n}$ is an i.i.d. maximum entropy binary random process). xor is the xor function which produces the value 1 only if the two operands are not equal, and otherwise produces 0.

Q1: Is $B_{1:n}$ a maximum entropy binary random source in the same sense that $R_{1:n}$ is?

If your answer is "no", meaning that $B_{1:n}$ is not a maximum entropy source, prove that this is true (which means that you need to prove the $H(B_i) < 1$ and $H(B_{i:j}) < \sum_{k=i}^j H(B_k)$ conditions for all i and j).

If your answer is "yes", meaning that B is maximum entropy, then what is the entropy of the sequence $N_{1:n}$ constructed as $N_i = B_i \text{ xor } R_i$ where we assume that $N_{1:n}$ is constructed using the same instance of the random vector $R_{1:n}$ that $B_{1:n}$ used when $B_{1:n}$ was constructed. What is the sequence N ? What happened when we combined these two sources R and B which by themselves are entirely random, to get N , and why does this happen? Which of R or B contains the information? Explain.

Problem 4: Let X be a random variable with values in the set $A = \{1, \dots, M\}$, and let $P_X(m) = p_m$ for $m \in A$. Assume also that $p_1 \geq p_2 \geq \dots \geq p_m$.

- (a) Prove for any binary Huffman code that if $p_1 > 2/5$, then the corresponding symbol is assigned a codeword of length 1 bit.

(b) Prove for any binary Huffman code that if $p_1 < 1/3$, then the corresponding symbol is assigned a codeword of length ≥ 2 bits.

Problem 5: Using the Laplace model that we outlined in class, describe the Arithmetic coding algorithm for encoding a random bit string of length n with k ones, given n and k . For the case of the string of length $n = 4$, show in complete detail the intervals corresponding to all source substrings of length 1 – 4.