Quiz 1 solutions COMP 423 Feb. 1, 2008

The class average was 11.5 / 15.

1. **(2 points)**

   (a)
   $$C(A_1) = 11, C(A_2) = 100, C(A_3) = 101, C(A_4) = 0$$

   (b)
   $$\overline{\lambda} - H = .35(2 + \log(.35)) + .1(3 + \log .1) + .15(3 + \log .15) + .4(1 - \log .4)$$

2. **(3 points)**

   (a)
   $$n = \sum_i n_i |S_i|$$
   $$totbits = \sum_i n_i |C(S_i)|$$

   (b) No. From the formula above, we see that the number of bits used to encode the string depends only on the length of the S_i codewords and on the frequency of the strings.

3. **(3 points)**

   (a)
   $$\lambda(i) = \lceil \frac{i}{2} \rceil + 1$$
   $$C(1) = 10, C(2) = 11, C(3) = 010, C(4) = 011, C(5) = 0010, C(6) = 0011, etc$$

   (b)
   $$p(i) = 2^{-\lambda(i)}$$
   and just plug in above formula for $\lambda(i)$.

   (c)
   $$C(1) = 10, C(2) = 110, C(3) = 111, C(4) = 010, C(5) = 0110, C(6) = 0111, etc$$
   The key idea here is that in each group of size $b = 3$, the first number has a codeword that is one bit shorter than the second and third numbers. i.e. Since $p(i)'s$ are non-increasing (see hint), the first number in each group should have a codeword length less than or equal to the second and third number.

   Marking scheme: I gave one point for each of the three parts. Several students gave all three numbers in each group the same codeword length, with one of the three having no sibling in the tree. I gave only half a point for this, since one can achieve better compression by using a code in which all leaves have a sibling.

4. **(3 points)**

(a)

$$a = \sum_{i=1}^{N} \frac{1}{i}.$$

(b)

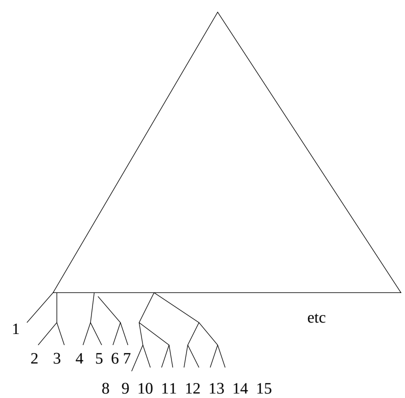
$$\lambda(i) = \log \frac{1}{p(i)} = \log \frac{i}{a} = \log i - \log a$$

The simplest way to think of this is to let

$$\lambda(i) = \lceil \log i \rceil + \lceil - \log a \rceil$$

so the codeword length should be a constant number of bits $\lceil - \log a \rceil$ plus a variable number of bits $\lceil \log i \rceil$.

(c)



5. **(2 points)**

$$\begin{aligned}
& \sum_{i=1}^{N} (i + \log(i+1)) \\
= \quad & \frac{N(N+1)}{2} + \sum_{i=1}^{N} (\log(i+1)) \\
\leq \quad & \frac{N(N+1)}{2} + N \log(\frac{1}{N} \sum_{i=1}^{N} (i+1)) \\
= \quad & \frac{N(N+1)}{2} + N \log(1 + \frac{1}{N} \sum_{i=1}^{N} i) \\
= \quad & \frac{N(N+1)}{2} + N \log(1 + \frac{N+1}{2})
\end{aligned}$$

6. **(2 points)**

The question says you should assume the list starts empty. Cranking through, you get:

$$C(1)C^*(M)C(2)C^*(a)C(3)C^*(t)C(1)C(4)C^*(\_)C(3)C(3)C(5)C^*(e)C(4)C(3)C(6)$$

$$C^*(h)C(4)C(4)C(6)C(6)C(6)$$

Here I am not distinguishing "M" and "m".