Name: STAFF SOLUTIONS

DEPARTMENT OF EECS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.02 Fall 2011

Quiz I

October 18, 2011

"×" your section	Section	Time	Room	Recitation Instructor
☐	1	10-11	36-112	Paul Ampadu
☐	2	11-12	36-112	Sidhant Misra
☐	3	12-1	36-112	Sidhant Misra
☐	4	1-2	36-112	Paul Ampadu
☐	5	2-3	26-168	Karl Berggren
☐	6	3-4	26-168	Karl Berggren

"×" your TA
- ☐ Mukul Agarwal
- ☐ Jason Cloud
- ☐ Shuo Deng
- ☐ Lyla Fischer
- ☐ Rui Li
- ☐ Ruben Madrigal
- ☐ Surapap R.
- ☐ Xiawa Wang

There are **14 questions** (some with multiple parts) and **?? pages** in this quiz booklet. Answer each question according to the instructions given. You have **120 minutes** to answer the questions.

If you find a question ambiguous, please be sure to write down any assumptions you make. **Please be neat and legible.** *Please show your work for partial credit.* And please write your name above.

Use the empty sides of this booklet if you need scratch space. You may also use them for answers, although you shouldn't need to. *If you use the blank sides for answers, make sure to say so!*

One two-sided "crib sheet" and a calculator allowed. No other notes, books, computers, cell phones, PDAs, information appliances, carrier pigeons, etc.!

PLEASE NOTE: SEVERAL STUDENTS WILL TAKE THE MAKE-UP QUIZ TOMORROW AT 7.30 PM. UNTIL THEN, PLEASE DO NOT DISCUSS THIS QUIZ WITH ANYONE IN THE CLASS, UNLESS YOU ARE CERTAIN THEY HAVE TAKEN IT WITH YOU TODAY. THANKS!

Do not write in the boxes below

1-3 (x/11)	4-5 (x/14)	6-7 (x/11)	8-11 (x/16)	12-13 (x/28)	14 (x/20)	Total (x/100)

I Information, Entropy, and Source Coding

Suppose the temperature tomorrow can take on one of $N \geq 3$ possible distinct values, $V_1, V_2, \ldots, V_N$, whose respective probabilities are $p_1 \geq p_2 \ldots \geq p_N > 0$, with $\sum_{i=1}^{N} p_i = 1$. Call this distribution D.

1. [1 points]: Write the expression for the entropy of D in terms of the quantities defined above.

Solution: From the definition of entropy,

$$H(D) = -\sum_{1}^{N} p_i \log_2 p_i \ .$$

2. [4 points]: The weatherperson, who is never wrong, announces that the temperature tomorrow is guaranteed to be one of three distinct values.

A. What set of three values for tomorrow's temperature will convey the **most** information?

Solution: The less probable the event, the more informative it is to be told the event will occur. The event $\{V_i \text{ or } V_j \text{ or } V_k\}$ for distinct i, j, k has probability $p_i + p_j + p_k$, since the outcomes V_i, V_j and V_k are mutually exclusive. Hence the most informative event is $\{V_{N-2} \text{ or } V_{N-1} \text{ or } V_N\}$, i.e.,, the three temperatures are V_{N-2}, V_{N-1} and V_N.

B. How much information about tomorrow's temperature has the weatherperson given you if the three values are the most informative ones that you specified in Part A? Give **both** the **expression** and the **units** in which this information is expressed.

Solution: The corresponding information is

$$-\log_2(p_{N-2} + p_{N-1} + p_N) \ \text{bits.}$$

3. [6 points]: Suppose we take the distribution D and combine two values, V_1 and V_2, into a single "value", V_{12}, standing for the event "*the temperature can be either V_1 or V_2*", while keeping all the other $N-2$ values and probabilities the same (note that $N \geq 3$). Call the resulting distribution D'.

Which of the following statements is true? Select the best answer from the choices below (circle your choice). You **must** correctly explain your choice in the space below to receive full credit.

A. There exists some distribution D for which D has the same entropy as D'.

B. D always has strictly higher entropy than D'.

C. There exists some distribution D for which D has lower entropy than D'.

Solution: The correct answer is **B**. Intuitively, there is less uncertainty about the outcome of D', because there has been some "binning" of values — distinct items from D are now reported under the same label in D'. For a formal proof, we proceed as follows:

Keeping in mind that the probability of the event V_{12} is $p_1 + p_2$, we know that

$$\begin{aligned} H(D) - H(D') &= p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right) - (p_1 + p_2)\log_2\left(\frac{1}{p_1 + p_2}\right) \\ &= p_1\left[\log_2\left(\frac{1}{p_1}\right) - \log_2\left(\frac{1}{p_1 + p_2}\right)\right] + p_2\left[\log_2\left(\frac{1}{p_2}\right) - \log_2\left(\frac{1}{p_1 + p_2}\right)\right] . \end{aligned}$$

The first equality is because all other terms in $H(D)$ and $H(D')$ are identical. Note now that the terms in the square brackets are strictly positive, from which we conclude that $H(D) > H(D')$.

4. [6 points]: Ben Bitdiddle and Alyssa P. Hacker are taking the 6.02 quiz and encounter the following Huffman coding problem. They are given four symbols, A, B, C, D, and the associated symbol probabilities, $p_A \geq p_B \geq p_C \geq p_D$, and are asked to construct the Huffman code tree. They construct two different trees, both correct constructions. The code trees have the following properties:

Ben's tree: The length of the longest path from the root to a symbol is 3.

Alyssa's tree: The length of the longest path from the root to a symbol is 2.

Derive one constraint relating the symbol probabilities, which *ensures* that both Ben and Alyssa are correct. Write down the constraint, and show Ben's and Alyssa's trees below.

Solution: The constraint is

$$p_C + p_D = p_A.$$

The reason is that C and D combine first to form a super node CD. We are then left with CD, B, and A. In the next step, because $p_B \leq p_A$, B will need to combine with either CD or A. If B combines with CD, then we have Ben's tree; otherwise, we have Alyssa's tree. To obtain Ben's tree, we need the condition

$$p_B + p_c + p_D \leq p_A + p_B,$$

i.e., $p_A \leq p_C + p_D$. In contrast, to get Alyssa's tree, we need the condition

$$p_B + p_C + p_D \geq p_A + p_B,$$

i.e., $p_C + p_D \geq p_A$. The only consistent constraint between these two inequalities is

$$p_C + p_D = p_A.$$

Note that p_B **does not** have to equal p_A for both trees to be consistent. Of course, we are told that $p_B \leq p_A$, so one can't arbitrarily pick p_B. Here are some examples of probabilities that make both trees be correct:

$$(1/3, 1/3, 1/6, 1/6)(2/5, 1/5, 1/5, 1/5)(4/7, 3/7, 2/7, 2/7).$$

The two trees are shown below. Ben's tree is isomorphic to:

```
        /\
     /\   A
  /\   B
 C   D
```

Alyssa's tree is isomorphic to:

```
       /\
    /\   /\
   C   D B  A
```

5. [8 points]: Circle **True** or **False** for each of these statements about LZW. Briefly explain each answer to receive credit. Recall that a codeword in LZW is an index into the string table.

A. True / False Suppose the sender adds two strings with corresponding codewords c_1 and c_2 in that order to its string table. Then, it **may** transmit c_2 for the first time **before** it transmits c_1.

Solution: **True**. Whenever a new string is added to the table, the codeword corresponding to the string without the final symbol is transmitted, but the string that *gets added* may show up only much later in the text being compressed, or indeed may not show up at all! Consider, for example, the string $abbbbbbbbbbb\ldots$ The string ab is added to the table early in the process, but the corresponding codeword is *never* transmitted, while the codewords corresponding to strings added later, such a bb, bbb, etc. get transmitted earlier. Thus the order in which codewords are added to the table is not necessarily related to the order in which they are transmitted.

B. True / False Suppose the string table never gets full. If there is an entry for a string s in the string table, then the sender **must** have previously sent a distinct codeword for every non-null prefix of string s. (If $s \equiv p + s'$ where $+$ is the string concatenation operation and s' is some non-null string, then p is said to be a prefix of s.)

Solution: **True.** Suppose we add the string "str + s" to the table, where "s" is a single character. At this time, we know that "str" **must** be in the table, and the encoder must transmit the codeword corresponding to "str". This argument now recursively applies to "str" as well: when we added "str" to the table, its longest-prefix (assuming "str" is of length 2 or greater) must have already been in the table, and the corresponding codeword sent. This argument applies until we get to the first character of the string "str + s". Hence, the statement is true.

II Digital Signaling

Material not yet covered.

III Noise

Material not yet covered.

IV Linear Block Codes

See pset for solutions to a similar problem.

V Convolutional Codes

See pset for solutions.